Christian Huemer
Pasquale Lops (Eds.)

LNBIP 123

E-Commerce and Web Technologies

13th International Conference, EC-Web 2012
Vienna, Austria, September 2012
Proceedings

Springer

Lecture Notes in Business Information Processing 123

Christian Huemer
Pasquale Lops (Eds.)

E-Commerce and Web Technologies

13th International Conference, EC-Web 2012
Vienna, Austria, September 4-5, 2012
Proceedings

Springer

Volume Editors

Christian Huemer
Vienna University of Technology
Institute of Software Technology and Interactive Systems
Business Informatics Group (BIG)
Vienna, Austria
E-mail: huemer@big.tuwien.ac.at

Pasquale Lops
Università di Bari
Dipartimento di Informatica
Bari, Italy
E-mail: pasquale.lops@uniba.it

ISSN 1865-1348 e-ISSN 1865-1356
ISBN 978-3-642-32272-3 e-ISBN 978-3-642-32273-0
DOI 10.1007/978-3-642-32273-0
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012943574

ACM Computing Classification (1998): J.1, K.4.4, I.2.11, H.3.5, H.4

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

We welcome you to the proceedings of the 13th International Conference on Electronic Commerce and Web Technologies—EC-Web 2012—which took place at Vienna University of Technology, Austria, during September 3–7, 2012.

The series of EC-Web conferences provides a platform for researchers and practitioners interested in the theory and practice of e-commerce and Web technologies. In 2012, EC-Web focused on the following topics:

Recommender systems. Recommender and business intelligence systems supporting both the customer and the provider in making better business decision is still a challenging issue.

e-Business architectures. E-business architectures leverage Web technologies to implement mission-critical e-business systems. Still there is a need for design principles, methods, and technologies for describing the structure of e-business systems, its composition of subsystems, and their relationship with the external environment.

Semantic representation of e-business and e-commerce information. Managing knowledge for the coordination of e-business processes through the systematic application of Semantic Web technologies is the focus of semantic e-business. It builds up on Semantic Web technologies, knowledge management, and e-business processes.

Agent-based e-commerce. Agents are computer systems situated in an environment and capable of autonomous action to meet their design objectives. Research on agent-based e-commerce has a vigorous tradition. However, new trends and concerns are emerging.

e-Business case studies. In constructive research, new prototypes to conduct e-business have emerged over the last couple of years. Although EC-Web focuses on new research ideas, we also welcome case studies that report on applying recent research results in real-world environments.

We were happy to see that our community was still active in contributing to the body of knowledge on future trends in e-commerce and Web technologies. Accordingly, we received 45 submissions from authors of 22 countries addressing the EC-Web topics mentioned above. Each submission received at least three review reports from Program Committee members, whereby the reviews were based on four criteria—originality, quality, relevance, and presentation—which resulted in a recommendation of each reviewer. Based on these recommendations we selected 15 full papers for publication and presentation at EC-Web 2012. Accordingly, the acceptance rate of EC-Web 2012 for full papers was about 33.3%.

In addition, these proceedings include four short papers that were presented at EC-Web 2012 as well.

These accepted papers were organized in six sessions:

- Recommender Systems I and II (2 sessions)
- Security and Trust
- Mining and Semantic Services
- Negotiation
- Agents and Business Services

When organizing a scientific conference, one always has to count on the efforts of many volunteers. We are grateful to the members of the Program Committee who devoted a considerable amount of their time in reviewing the submissions to EC-Web 2012. Not only did they deliver high-quality reviews that greatly facilitated our selection of papers, but they also performed these reviews in time.

We were privileged to work together with highly motivated people to arrange the conference and to publish these proceedings. We appreciate all the tireless support by the Publicity Chair Christian Pichler for announcing our conference on various lists. Special thanks go to Amin Anjomshoaa, who was always of great help in managing the conference submission system. Last, but not least, we want to express our thanks to Gabriela Wagner, who dedicated hours and hours in making EC-Web 2012 a success. Not only was she always of great help in solving organizational matters, but she also maintained the EC-Web 2012 website and was responsible for the compilation of all the papers in the proceedings.

We hope that you find these proceedings a valuable source of information on e-commerce and Web technologies.

September 2012

Christian Huemer
Pasquale Lops
Fernando Lopes

Organization

General Chairs

Christian Huemer	Vienna University of Technology, Austria
Pasquale Lops	University of Bari "Aldo Moro", Italy

Steering Chair

Fernando Lopes	National Research Institute, Portugal

Program Committee

Sarabjot Singh Anand	University of Warwick, UK
Liliana Ardissono	University of Turin, Italy
Giuliano Armano	University of Cagliari, Italy
Pierpaolo Basile	University of Bari Aldo Moro, Italy
Shlomo Berkovsky	CSIRO, Australia
Marko Boskovic	Research Studios Austria, Austria
Francesco Buccafurri	Università degli Studi Mediterranea di Reggio Calabria, Italy
Robin Burke	De Paul University, USA
Maria Laura Caliusco	Universidad Tecnológica Nacional, Argentina
Ivan Cantador	Universidad Autonoma de Madrid, Spain
Cinzia Cappiello	Politecnico di Milano, Italy
Sven Casteleyn	Universidad Politécnica de Valencia, Spain
Pablo Castells	Universidad Autónoma de Madrid, Spain
Federica Cena	University of Turin, Italy
Kuo-Ming Chao	University of Coventry, UK
Jen-Yao Chung	IBM Thomas J. Watson Research Center, USA
Marco Comuzzi	Eindhoven University of Technology, The Netherlands
Alfredo Cuzzocrea	ICAR-CNR & University of Calabria, Italy
Ernesto Damiani	Università degli Studi di Milano, Italy
Florian Daniel	University of Trento - DISI, Italy
Antonina Dattolo	University of Udine, Italy
Marco de Gemmis	University of Bari Aldo Moro, Italy
Ernesto William De Luca	DAI-Labor, Germany
Tommaso Di Noia	Polytechnic of Bari, Italy
Alexander Felfernig	University Klagenfurt, Austria
Alberto Fernández	University Rey Juan Carlos, Spain
Agata Filipowska	Poznan University of Economics, Poland
Fabio Gasparetti	Roma Tre University, Italy
Michele Gorgoglione	Polytechnic of Bari, Italy

Jingzhi Guo	University of Macau, P.R. China
Chang Heng	Huawei Technologies, Shenzhen, P.R. China
Birgit Hofreiter	TU Vienna, Austria
Leo Iaquinta	University of Bari Aldo Moro, Italy
Dietmar Jannach	TU Dortmund, Germany
Robert Jäschke	University of Kassel, Germany
Alípio Mário Jorge	University of Porto, Portugal
Alfred Kobsa	University of California, Irvine, USA
Christoph Lattemann	Jacobs University, Germany
Joerg Leukel	University of Hohenheim, Germany
Philipp Liegl	TU Vienna, Austria
Andre Ludwig	University of Leipzig, Germany
Heiko Ludwig	IBM T.J. Watson Research Center, USA
Andrea Maurino	Università di Milano Bicocca, Italy
Stuart E. Middleton	University of Southampton, UK
Cataldo Musto	University of Bari Aldo Moro, Italy
Fedelucio Narducci	University of Bari Aldo Moro, Italy
Alex Norta	University of Helsinki, Finland
Umberto Panniello	Polytechnic of Bari, Italy
Christian Pichler	TU Vienna, Austria
Francesco Ricci	Free University of Bozen-Bolzano, Italy
Juan A. Rodriguez-Aguilar	IIA, CSIC, Spain
Szabolcs Rozsnyai	IBM Thomas J. Watson Research Center, USA
Andreas Schönberger	University of Bamberg, Germany
Giovanni Semeraro	Univeristy of Bari "Aldo Moro", Italy
Jun Shen	University of Wollongong, Australia
Yan Tang	Free University Brusseles StarLab, Belgium
Eloisa Vargiu	Barcelona Digital, Spain
Laurent Vercouter	Graduate School of Engineering, Saint-Étienne, France
Andreas Wombacher	University Twente, The Netherlands
Jih-Shyr Yih	IBM Thomas J. Watson Research Center, USA
Markus Zanker	University Klagenfurt, Austria
Marco Zapletal	TU Vienna, Austria
Uwe Zdun	University of Vienna, Austria

External Reviewers

Carlos Rodriguez	University of Trento, Italy
Bart Knijnenburg	University of California, Irvine, USA
John O'Donnovan	University of California, Santa Barbara, USA
Roberto Mirizzi	Polytechnic of Bari, Italy
Marius Kaminskas	University of Bozen-Bolzano, Italy
Alejandro Bellogín	Universidad Autónoma de Madrid, Spain
Mehdi Elahi	University of Bozen-Bolzano, Italy
Soudip Roy Chowdhury	University of Trento, Italy

Table of Contents

Session 1

Robust Trust: Prior Knowledge, Time and Context 1
John Debenham and Carles Sierra

Validating the Relationship between Information Quality and Trust: The Moderating Effect from Customer Orientation 13
Shun Cai, Xina Yuan, and Kyounghee Chu

PAEAN – A Risk-Mitigation Framework for Business Transaction at Run-Time 25
Denada Cfarku, Yéhia Taher, Rafiqul Haque, Willem-Jan van den Heuvel, and Michael P. Papazoglou

Session 2

Mapping and Integration of Dimensional Attributes Using Clustering Techniques 38
Francesco Guerra, Marius-Octavian Olaru, and Maurizio Vincini

Service Offer Descriptions and Expressive Search Requests – Key Enablers of Late Service Binding 50
Maciej Zaremba, Tomas Vitvar, Sami Bhiri, Wassim Derguech, and Feng Gao

Performative-Based Mining of Workflow Organizational Structures 63
Chihab Hanachi, Walid Gaaloul, and Ravi Mondi

Session 3

Recommender Systems in Computer Science and Information Systems – A Landscape of Research 76
Dietmar Jannach, Markus Zanker, Mouzhi Ge, and Marian Gröning

Differential Context Relaxation for Context-Aware Travel Recommendation 88
Yong Zheng, Robin Burke, and Bamshad Mobasher

Multi-criteria Ratings for Recommender Systems: An Empirical Analysis in the Tourism Domain 100
Matthias Fuchs and Markus Zanker

Leveraging Social Media Sources to Generate Personalized Music Playlists 112
Cataldo Musto, Giovanni Semeraro, Pasquale Lops, Marco de Gemmis, and Fedelucio Narducci

Session 4

Computational Commerce: A Vision for the Future 124
Alegria Baquero and Richard N. Taylor

Usage Analysis of a Mobile Bargain Finder Application 137
Stephan Karpischek, Darshan Santani, and Florian Michahelles

Argumentation-Based Negotiation? Negotiation-Based Argumentation! 149
Jürgen Landes and Ricardo Buettner

Session 5

A Systematic Success Factor Analysis in the Context of Enterprise 2.0: Results of an Exploratory Analysis Comprising Digital Immigrants and Digital Natives 163
Dietmar Nedbal, Andreas Auinger, Alexander Hochmeier, and Andreas Holzinger

An Experimental Analysis of Online Unidirectional Conversion Problem 176
Iftikhar Ahmad and Günter Schmidt

Multichannel Sales Services for Enterprise Cloud Vendors (Short Paper) 188
Jih-Shyr Yih

Session 6

Applying Contextual Advertising to MultiModal Information Content (Short Paper) 195
Giuliano Armano, Alessandro Giuliani, Alberto Messina, Maurizio Montagnuolo, and Eloisa Vargiu

User Semantic Preferences for Collaborative Recommendations (Short Paper) 203
Sonia Ben Ticha, Azim Roussanaly, Anne Boyer, and Khaled Bsaies

A Multidimensional Model of Trust in Recommender Systems (Short Paper) 212
Martina Maida, Konradin Maier, Nikolaus Obwegeser, and Volker Stix

Author Index 221

Robust Trust: Prior Knowledge, Time and Context

John Debenham[1] and Carles Sierra[2]

[1] QCIS, University of Technology, Sydney, Australia
debenham@it.uts.edu.au
[2] Institut d'Investigació en Intel·ligència Artificial - IIIA,
Spanish Scientific Research Council, CSIC
08193 Bellaterra, Catalonia, Spain
sierra@iiia.csic.es

Abstract. *Trust* is an agent's expectation of the value it will observe when it evaluates the enactment of another agent's commitment. There are two steps involved in trust: first the action that another agent is expected to enact given that it has made a commitment, and second the expected valuation of that action when the result of that action is eventually consumed. A computational model of trust is presented that takes account of: *prior knowledge* of other agents, the evolution of trust estimates in *time*, and the evolution of trust estimates in response to changes in *context*. This model is founded on the principle of *information-based* agency that each and every utterance made contains valuable information. The computational basis for the model is substantially simpler and is more theoretically grounded than previously reported.

1 Introduction

The social concept of trust has received considerable attention. The seminal paper [1] describes two approaches to trust: first, as a belief that another agent will do what it says it will, or will reciprocate for common good, and second, as constraints on the behaviour of agents to conform to trustworthy behaviour. This paper is concerned with the first approach where trust is something that is learned and evolves. [2] presents a comprehensive categorisation of trust research: policy-based, reputation-based, general *and* trust in information resources. [3] presents an interesting taxonomy of trust models in terms of nine types of trust model. The scope described there fits well with this work with the possible exception of identity trust and security trust that are out of scope. [4] describes a powerful model that integrates interaction and role-based trust with witness and certified reputation that also relate closely to our model. *Reputation* is the opinion (more technically, a social evaluation) of a group about something — in a social environment — reputation feeds into trust [5].

The informal meaning of the statement "agent α trusts agent β" is that α expects β to act in a way that is somehow preferred by α. Human agents seldom trust another for *any* action that they may take — it is more usual to develop

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 1–12, 2012.

Action	Sign	Enact	Evaluate
Object	*(a,b)*	*(a',b')*	*b'*
Time	*t*	*t'*	*t''*

Fig. 1. Contract signing, execution and evaluation

a trusted expectation with respect to a particular set of actions. For example, "I trust John to deliver fresh vegetables" whilst the trustworthiness of John's advice on investments may be terrible. In this paper we discuss trust when the set of actions is restricted to negotiating, signing and enacting contracts that are expressed using some particular ontology. This then excludes morally founded trust as in: "Can I trust you to *do the right thing*?".

We assume that a multiagent system, $\{\alpha, \beta_1, \ldots, \beta_o, \xi, \theta_1, \ldots, \theta_t\}$, containing an agent α that interacts with negotiating agents, $\mathcal{X} = \{\beta_i\}$, information providing agents, $\mathcal{I} = \{\theta_j\}$, and an *institutional agent*, ξ, that represents the institution where we assume the interactions happen. Institutions give a normative context to interactions that simplify matters (e.g an agent can't make an offer, have it accepted, and then renege on it). The institutional agent ξ may form opinions on the actors and activities in the institution and may publish reputation estimates on behalf of the institution. The agent ξ also fulfils a vital role to compensate for any lack of sensory ability in the other agents by promptly and accurately reporting observations as events occur. For example, without such reporting an agent may have no way of knowing whether it is a fine day or not.

Our agents are information-based [6], they are endowed with machinery for valuing the information that they have, and that they receive. Information-based agency was inspired by the observation that "everything an agent says gives away information". They model how much they know about other agents, and how much they believe other agents know about them. Everything in their world, including their information, is uncertain; their only means of reducing uncertainty is acquiring fresh information. To model this uncertainty, their world model, $\mathcal{M}^t$, consists of random variables each representing a point of interest in the world. Distributions are then derived for these variables on the basis of information received. Over time agents acquire large amounts of information that are distilled into convenient measures including trust. By classifying private information into functional classes, and by drawing on the structure of the ontology, information-based agents develop other measures including a map of the 'intimacy' [7] of their relationships with other agents.

Section 2 discusses the notion of trust and develops a formal characterisation of it. The core mechanism for maintaining trust estimates is described in Section 3. Prior knowledge is then taken into account in Section 4 that includes a discussion of the *reliability* of an agent's utterances. Time is discussed in Section 5 and Context in Section 6. Section 7 concludes with a discussion of future work.

2 The Notion of 'Trust'

In this paper trust is concerned with valuing enactments made in fulfilment of commitments expressed in contracts. The scenario is: two agents α and β negotiate with the intention of leading to a *signed contract* that is a pair of commitments, (a, b), where a is α's and b is β's. A contract is signed by both agents at some particular time t. At some later time, t', both agents will have enacted their commitments[1] in some way, as say (a', b'). At some later time again, t'', α will consume b' and will then be in a position to evaluate the extent to which β's enactment of (a, b), b', was in α's interests. See Figure 1.

α's trust of agent β is expressed as α's expectation of its eventual valuation of β's future actions. We consider how α forms these expectations, how α will compare those expectations with observations, and how α then determines whether β's actions are preferred to α's expectations of them.

α forms expectations of β's future actions on the basis of all that it has: its full interaction history $H_\alpha \in \mathcal{H}_\alpha$ where $\mathcal{H}_\alpha$ is the set of all possible interaction histories that may be expressed in α's ontology[2]. H_α is a record of all interactions with each negotiating agent in $\mathcal{X}$ and with each information providing agent in $\mathcal{I}$. Let $\mathcal{B} = (b_1, b_2, \dots)$ denote that space of all enactments that β may make and $\mathcal{A}$ the space of α's enactments. Assuming that the space of contracts and enactments are the same, the space of all contracts and enactments is: $\mathcal{C} = \mathcal{A} \times \mathcal{B}$.

This raises the strategic question of given an expectation of some particular future requirements how should α strategically shape its interaction history to enable it to build a reliable expectation of β's future actions concerning the satisfaction of those particular requirements [8]. At time t'' α compares b' with α's expectations of β's actions, β having committed at time t to enact b at time t'. That is:

$$\text{compare}_\alpha^{t''}(\mathbb{E}_\alpha^t(\text{Enact}_\beta^{t'}(b)|\text{sign}_{\alpha,\beta}^t((a,b)), H_\alpha^t), b')$$

where $\text{sign}_{\alpha,\beta}^t((a,b))$ is a predicate meaning that the joint action by α and β of signing the contract (a, b) was performed at time t, and $\text{Enact}_\beta^{t'}(b)$ is a random variable over $\mathcal{B}$ representing α's expectations over β's enactment action at time t', $\mathbb{E}_\alpha^t(\cdot)$ is α's expectation, and $\text{compare}(\cdot, \cdot)$ somehow describes the result of the comparison.

Trust is the expectation of *the evaluation of* β's enactments made in fulfilment of its contractual commitments. Let $\mathcal{V} = (v_1, v_2, \dots, v_V)$ be the valuation space then α's expectation of the evaluation of a particular action that β may make is represented as a probability distribution over $\mathcal{V}$: $(f_1, f_2, \dots, f_V)$. We expect the set $\mathcal{V}$ to be smaller than the set $\mathcal{B}$, and so developing a sense of expectation for the value of β's actions should be easier than for the actions themselves. That is, we consider the expectation:

$$\mathbb{E}_\alpha^t(\text{Value}_\beta^{t''}(b)|\text{sign}_{\alpha,\beta}^t((a,b)), H_\alpha^t)$$

[1] For convenience we assume that both agents are presumed to have been completed their enactments by the same time, t'.

[2] The ontology is not made explicit to avoid overburdening the discussion.

where $\text{Value}^{t''}(b)$ is a random variable over $\mathcal{V}$ representing α's expectations of the value of β's enactment action given that he signed (a, b) and given H^t_α. At time t'' it then remains to compare expectation, $\mathbb{E}^t_\alpha(\text{Value}^{t''}_\beta(b)|\text{sign}^t_{\alpha,\beta}((a, b)), H^t_\alpha)$, with observation, $\text{val}_\alpha(b')$, where $\text{val}(\cdot)$ represents α's preferences — i.e. it is α's utility function[3].

We are now in a position to define 'trust'. *Trust*, $\tau_{\alpha\beta}(b)$, is a computable[4] [9] estimate of the distribution: $\mathbb{E}^t_\alpha(\text{Value}^{t''}_\beta(b)|\text{sign}^t_{\alpha,\beta}((a, b)), H^t_\alpha)$. τ is a summarising function that distils the trust-related aspects of the (probably very large) set H_α into a probability distribution that may be computed. $\tau_{\alpha\beta}(b)$ summarises the large set H_α. The set of contracts $\mathcal{C}$ is also large. It is practically unfeasible to estimate trust for each individual contract. To deal with this problem we appeal to the structure of the ontology, and aggregate estimates into suitable classes such as John's trustworthiness in supplying Australian red wine.

In real world situations the interaction history may not reliably predict future action, in which case the notion of trust is fragile. No mater how trust is defined we expect trusted relationships to develop slowly over time. On the other hand they can be destroyed quickly by an agent whose actions unexpectedly fall below expectation. This highlights the importance of being able to foreshadow the possibility of untrustworthy behaviour.

$\tau_{\alpha\beta}(b)$ is predicated on α's ability to form an expectation of the value of β's future actions. This is related to the famous question posed by Laplace "what is the probability that the sun will rise tomorrow?". Assuming that it has always previously been observed to do so and that there have been n prior observations then if the observer is in complete ignorance of the process he will assume that the probability distribution of a random variable representing the prior probability that the sun will rise tomorrow is the maximum entropy, uniform distribution on $[0, 1]$, and using Bayes' theorem will derive the posterior estimate $\frac{n+1}{n+2}$. The key assumption being that the observer is "in complete ignorance of the process". There may be many reasons why the sun may not rise such as the existence of a large comet on a collision trajectory with earth. These all important reasons are the *context* of the problem.

Laplace's naïve analysis above forms the basis of a very crude measure of trust. Suppose that the valuation space is: $\mathcal{V} = \{\text{bad}, \text{good}\}$, and that α is considering signing contract (a, b) with β. Let the random variable B denote the value of β's next action. Then assuming that we know nothing about the contract or about β except that this contract has been enacted by β on n prior occasions and that the valuation was "good" on s of those occasions. Using the maximum entropy prior distribution for B, $[0.5, 0.5]$, Bayes' theorem gives us a posterior distribution $[\frac{n-s+1}{n+2}, \frac{s+1}{n+2}]$. If at time t α signs the contract under consideration then the expected probability of a "good" valuation at time t'' is:

[3] It is arguably more correct to consider: $\text{Value}((a, b)) = \text{Value}(b) - \text{Value}(a)$, as β's actions may be influenced by his expectations of α's enactment of a — we choose to avoid this additional complication.

[4] *Computable* in the sense that it is easy to compute and not simply Turing computable.

$\frac{s+1}{n+2}$. This crude measure has little practical value although it readily extends to general discrete valuation spaces, and to continuous valuation spaces. The zero-information, maximum entropy distribution is the *trivial trust measure*. The crude Laplacian trust measure is in a sense the simplest non-trivial measure.

The weaknesses of the crude Laplacian trust measure above show the way to building a reliable measure of trust [10]. These are:

Prior knowledge. The use of the maximum entropy prior[5] is justified when there is absolutely no prior knowledge or belief of an agent's behaviour. In practical scenarios we expect prior observations, reputation measures or the opinions of other agents to be available to be reflected in the prior.

Time. There is no representation of time. In the crude trust measure all prior observations have the same significance, and so an agent that used to perform well and is deteriorating may have the same trust measure as one that used to perform badly and is now performing well.

Context. There is no model of general events in the world or of *how* those events may effect an agent's behaviour. This includes modelling causality, *why* an agent might behave as it does.

This section defines trust as a historic[6] estimator of the expected value of future enactments, and concluded with three features of a reliable measure of trust. This section also described the fundamental role that the structure of the ontology plays in the trust model. Following sections describe such a measure that uses new and improved computational methods of information-based agents [6] particularly their information evaluation, acquisition and revelation strategies that ideally suits them to this purpose. The core trust mechanism is detailed in Section 3 and subsequent sections then detail the incorporation of prior knowledge, time and context.

3 The Core Mechanism

Section 2 ends with three essential components of a reliable trust model. Those three components will be dealt with in due course. In this section we describe the core trust estimation mechanism. In subsequent sections we enhance the core with the three essential components. The final component, context, is incomplete as it relies on the solution to unsolved problems that are beyond the scope of this paper.

The general idea is that trust estimates are updated whenever α evaluates $\text{val}_\alpha^{t''}(b')$ for some previously signed contract (a, b). The contract space is typically very large and so estimates are not maintained for individual contracts; instead they are maintained for selected abstractions based on the ontology. Abstractions

[5] The maximum entropy prior expresses total uncertainty about what the prior distribution is.

[6] *Historic* in the sense that the estimation can be performed on the basis of the agent's interaction history.

are denoted by the 'hat' symbol: e.g. $\hat{a}$. For example, "red wine orders for more that 24 bottles" or "supply of locally produced cheese". As we will see when an evaluation $\text{val}_\alpha^{t''}(b')$ is performed, the trust estimates, $\tau_{\alpha\beta}(\hat{b})$, for certain selected nearby abstractions, $\hat{b}$, are updated.

In the absence of incoming information the integrity of an information-based agent's beliefs decays in time. In the case of the agent's beliefs concerning trust, incoming information is in the form of valuation observations $\text{val}_\alpha^{t''}(b')$ for each enacted contract. If there are no such observations in an area of the ontology then the integrity of the estimate for that area should decay.

In the absence of valuation observations in the region of $\hat{b}$, $\tau_{\alpha\beta}(\hat{b})$ decays to a *decay limit distribution* $\overline{\tau_{\alpha\beta}(\hat{b})}$ (denoted throughout this paper by 'overline'). The decay limit distribution is the zero-data distribution, but not the zero-information distribution because it takes account of reputation estimates and the opinions of other agents [11]. We assume that the decay limit distribution is known for each abstraction $\hat{b}$. At time s, given a distribution for random variable $\tau_{\alpha\beta}(\hat{b})^s$, and a decay limit distribution, $\overline{\tau_{\alpha\beta}(\hat{b})^s}$, $\tau_{\alpha\beta}(\hat{b})$ decays by:

$$\tau_{\alpha\beta}(\hat{b})^{s+1} = \Delta(\overline{\tau_{\alpha\beta}(\hat{b})^s}, \tau_{\alpha\beta}(\hat{b})^s) \tag{1}$$

where s is time and Δ is the *decay function* for the X satisfying the property that $\lim_{s\to\infty} \tau_{\alpha\beta}(\hat{b})^s = \overline{\tau_{\alpha\beta}(\hat{b})}$. For example, Δ could be linear:

$$\tau_{\alpha\beta}(\hat{b})^{s+1} = (1-\mu) \times \overline{\tau_{\alpha\beta}(\hat{b})^s} + \mu \times \tau_{\alpha\beta}(\hat{b})^s$$

where $0 < \mu < 1$ is the decay rate.

We now consider what happens when valuation observations are made. Suppose that at time s, α evaluates β's enactment b' of commitment b, $\text{val}_\alpha^s(b') = v_k \in \mathcal{V}$. The update procedure updates the probability distributions for $\tau_{\alpha\beta}(\hat{b})^s$ for each $\hat{b}$ that is "moderately close to" b. Given such a $\hat{b}$, let $\mathbb{P}^s(\tau_{\alpha\beta}(\hat{b}) = v_k)$ denote the prior probability that v_k would be observed. The update procedure is in two steps. First, we estimate the posterior probability that v_k would be observed, $\mathbb{P}^{s+1}(\tau_{\alpha\beta}(\hat{b}) = v_k)$ for the particular value v_k. Second, we update the entire posterior distribution for $\tau_{\alpha\beta}(\hat{b})$ to accommodate this revised value.

Given a $\hat{b}$, to revise the probability that v_k would be observed we work with three things: the observation: $\text{val}_\alpha^s(b')$, the prior: $\mathbb{P}^s(\tau_{\alpha\beta}(\hat{b}) = v_k)$, and the decay limit value: $\mathbb{P}^s(\overline{\tau_{\alpha\beta}(\hat{b})} = v_k)$. The observation $\text{val}_\alpha^s(b')$ may be represented as a probability distribution with a '1' in the k'th place and zero elsewhere, $\boldsymbol{u}_k$. To combine it with the prior we discount its significance for two reasons:

- b may not be semantically close to $\hat{b}$, and
- $\text{val}_\alpha^s(b') = v_k$ is a single observation whereas the prior distribution represents the accumulated history of previous observations.

To discount the significance of the observation $\text{val}_\alpha^s(b') = v_k$ we determine a value in the range between '1' and the zero-data, decay limit value $\mathbb{P}^s(\overline{\tau_{\alpha\beta}(\hat{b})} = v_k)$ by:

$$\delta = \text{Sim}(b, \hat{b}) \times \kappa + (1 - \text{Sim}(b, \hat{b}) \times \kappa) \times \mathbb{P}^s(\overline{\tau_{\alpha\beta}(\hat{b})} = v_k) \quad (2)$$

where $0 < \kappa < 1$ is the learning rate, and $\text{Sim}(\cdot, \cdot)$ is a semantic similarity function. Then the posterior estimate $\mathbb{P}^{s+1}(\tau_{\alpha\beta}(\hat{b}) = v_k)$ is given by:

$$\mathbb{P}^{s+1}(\tau_{\alpha\beta}(\hat{b}) = v_k) = \frac{\rho\delta(1-\omega)}{\rho\delta(1-\omega) + (1-\rho)(1-\delta)\omega} = \nu \quad (3)$$

where δ is given by Equation 2, $\rho = \mathbb{P}^s(\tau_{\alpha\beta}(\hat{b}) = v_k)$ is the prior value, and $\omega = \mathbb{P}^s(\overline{\tau_{\alpha\beta}(\hat{b})} = v_k)$ is the decay limit value. That is, we combine the two 'observed' probabilities ρ and δ in the context of the pre-observation value ω.

It remains to update the entire posterior distribution for $\tau_{\alpha\beta}(\hat{b})$ to accommodate the constraint $\mathbb{P}^{s+1}(\tau_{\alpha\beta}(\hat{b}) = v_k) = \nu$. Information-based agents [6] employ a standard procedure for updating distributions, $\mathbb{P}^t(X = x)$ subject to a set of linear constraints on X, $c(X)$, using:

$$\mathbb{P}^{t+1}(X = x | c(X)) = \text{MRE}(\mathbb{P}^t(X = x), c(X))$$

where the function MRE is defined by: $\text{MRE}(\boldsymbol{q}, \boldsymbol{g}) = \arg\min_{\boldsymbol{r}} \sum_j r_j \log \frac{r_j}{q_j}$ such that $\boldsymbol{r}$ satisfies $\boldsymbol{g}$, $\boldsymbol{q}$ is a probability distribution, and $\boldsymbol{g}$ is a set of n linear constraints $\boldsymbol{g} = \{g_j(\boldsymbol{p}) = \boldsymbol{a_j} \cdot \boldsymbol{p} - c_j = 0\}, j = 1, \ldots, n$ (including the constraint $\sum_i p_i - 1 = 0$). The resulting $\boldsymbol{r}$ is the *minimum relative entropy distribution*[7] [12]. Applying this procedure to $\tau_{\alpha\beta}(\hat{b})$:

$$\mathbb{P}^{s+1}(\tau_{\alpha\beta}(\hat{b}) = v) = \text{MRE}(\mathbb{P}^s(\tau_{\alpha\beta}(\hat{b}) = v), \mathbb{P}^{s+1}(\tau_{\alpha\beta}(\hat{b}) = v_k) = \nu)$$

where ν is the value given by Equation 3.

Whenever α evaluates an enactment $\text{val}^s_\alpha(b')$ of some commitment b, the above procedure is applied to update the distributions for $\mathbb{P}(\tau_{\alpha\beta}(\hat{b}) = v)$. It makes sense to limit the use of this procedure to those distributions for which $\text{Sim}(b, \hat{b}) > y$ for some threshold value y.

4 Prior Knowledge

The decay-limit distribution plays a key role in the estimation of trust. It is not directly based on any observations and in that sense it is a "zero data" trust estimate. It is however not "zero information" as it takes account of opinions and reputations communicated by other agents [11]. The starting point for constructing the decay-limit distribution is the maximum entropy (zero-data, zero-information) distribution. This gives a two layer structure to the estimation of trust: opinions and reputations shape the decay-limit distribution that in turn

[7] This may be calculated by introducing Lagrange multipliers $\boldsymbol{\lambda}$: $L(\boldsymbol{p}, \boldsymbol{\lambda}) = \sum_j p_j \log \frac{p_j}{q_j} + \boldsymbol{\lambda} \cdot \boldsymbol{g}$. Minimising L, $\{\frac{\partial L}{\partial \lambda_j} = g_j(\boldsymbol{p}) = 0\}, j = 1, \ldots, n$ is the set of given constraints $\boldsymbol{g}$, and a solution to $\frac{\partial L}{\partial p_i} = 0, i = 1, \ldots, I$ leads eventually to $\boldsymbol{p}$.

plays a role in forming the trust estimate that takes account of observed data [13]. Communications from other agents may not be reliable. α needs a means of estimating the reliability of other agents before they can be incorporated into the decay-limit distribution — reliability is discussed at the end of this section.

Reputation is the opinion (more technically, a social evaluation) of a group about something. So a group's reputation about a thing will be related *in some way* to the opinions that the individual group members hold towards that thing, or to shared evaluations that they may hold. An *opinion* is an assessment, judgement or evaluation of something. Opinions are represented in this paper as probability distributions on a suitable ontology that for convenience we identify with the *evaluation space* $\mathcal{V}$. That is, we assume that opinions communicated by β concerning another agent's trustworthiness are expressed as predicates using the same valuation space as $\mathcal{V}$ over which α represents its trust estimates.

An opinion is an evaluation of an *aspect* of a thing. A rainy day may be evaluated as being "bad" from the aspect of being suitable for a picnic, and "good" from the aspect of watering the plants in the garden. An aspect is the "point of view" that an agent has when forming his opinion. An opinion is evaluated in context. The *context* is everything that the thing is being, explicitly or implicitly, evaluated with or against. The set of valuations of all things in the context calibrates the valuation space. For example, "this is the best paper in the conference". The context can be vague: "of all the presents you could have given me, this is the best". If agents are to discuss opinions then they must have some understanding of each other's context.

Summarising the above, an *opinion* is an agent's evaluation of a particular aspect of a thing in context. A representation of an opinion will contain: the thing, its aspect, its context, and a distribution on $\mathcal{V}$ representing the evaluation of the thing. α acquires opinions and reputations through communication with other agents. α estimates the reliability of those communicating agents before incorporating that information into the decay-limit distributions. The basic process is the same for opinions and reputations; the following sub-section 4.1 describes the incorporation of opinions only.

4.1 The Decay-Limit Distribution and Reliability

Suppose agent β' informs agent α of his opinion of the trustworthiness of another agent β using an utterance of the form: $u = \texttt{inform}(\beta', \alpha, \tau_{\beta'\beta}(b))$, where conveniently b is in α's ontology. This information may not be useful to α for at least two reasons: β' may not be telling the truth, or β' may have a utility function that differs from α's. We will shortly estimate β''s "reliability", $R^t_\alpha(\beta')$ that measures the extent to which β' is telling the truth and that α and β' "are on the same page" or "think alike"[8]. Precisely, $0 < R^t_\alpha(\beta') < 1$; its value is used to moderate the effect of the utterance on α's decay-limit distributions. The estimation of $R^t_\alpha(\beta')$ is described below.

[8] The reliability estimate should perhaps also be a function of the commitment, $R^t_\alpha(\beta', b)$, but we choose to avoid that complication.

Suppose that α maintains the decay limit distribution $\overline{\tau_{\alpha\beta}(\hat{b})^s}$ for a chosen $\hat{b}$. In the absence of utterances informing opinions of trustworthiness, $\overline{\tau_{\alpha\beta}(\hat{b})^s}$ decays to the distribution with maximum entropy. As previously this decay could be linear:

$$\overline{\tau_{\alpha\beta}(\hat{b})^{s+1}} = (1-\mu) \times \text{MAX} + \mu \times \overline{\tau_{\alpha\beta}(\hat{b})^s}$$

where $\mu < 1$ is the decay rate, and MAX is the maximum entropy, uniform distribution.

When α receives an utterance of the form u above, the *decay limit distribution* is updated by:

$$\overline{\tau_{\alpha\beta}(\hat{b})^{s+1}} \mid \texttt{inform}(\beta', \alpha, \tau_{\beta'\beta}(b)) = \\ \Big(1 - \kappa \times \text{Sim}(\hat{b}, b) \times R^s_\alpha(\beta')\Big) \times \overline{\tau_{\alpha\beta}(\hat{b})^s} + \kappa \times \text{Sim}(\hat{b}, b) \times R^s_\alpha(\beta') \times \tau_{\beta'\beta}(b)$$

where $0 < \kappa < 1$ is the learning rate and $R^s_\alpha(\beta')$ is α estimate of β''s reliability. It remains to estimate $R^s_\alpha(\beta')$.

Estimating $R^s_\alpha(\beta')$ is complicated by its time dependency. First, in the absence of input of the form described following, $R^s_\alpha(\beta')$ decays to zero by: $R^{s+1}_\alpha(\beta') = \mu \times R^s_\alpha(\beta')$. Second, we describe how $R^s_\alpha(\beta')$ is increased by comparing the efficacy of $\overline{\tau_{\alpha\beta}(\hat{b})^s}$ and $\tau_{\beta'\beta}(b)^s$ in the following interaction scenario. Suppose at a time s, α is considering signing the contract (a, b) with β. α requests β''s opinion of β with respect to b, to which β may respond $\texttt{inform}(\beta', \alpha, \tau_{\beta'\beta}(b))$. α now has two estimates of β's trustworthiness: $\overline{\tau_{\alpha\beta}(\hat{b})^s}$ and $\tau_{\beta'\beta}(b)^s$. If α then signs the contract (a, b) at time t, and at some later time t'' evaluates β's enactment $\text{val}^{t''}_\alpha(b') = v_k$, say. $\overline{\tau_{\alpha\beta}(\hat{b})^s}$ and $\tau_{\beta'\beta}(b)^s$ are both probability distributions that each provide an estimate of $\mathbb{P}^s(\text{Value}_\beta(b) = v_k)$. If:

$$\mathbb{P}(\tau_{\beta'\beta}(b)^s = v_k) > \mathbb{P}(\overline{\tau_{\alpha\beta}(\hat{b})^s} = v_k)$$

then β''s estimate is better than α's and α increases $R^s_\alpha(\beta')$ using:

$$R^{s+1}_\alpha(\beta') = \kappa + (1-\kappa) \times R^s_\alpha(\beta')$$

where $0 < \kappa < 1$ is the learning rate.

5 Time

The core trust mechanism in Section 3 and the prior knowledge in Section 4 both give greater weight to recent observations than to historic. This may be a reasonable default assumption but has no general validity. Trust, $\tau_{\alpha\beta}(\hat{b})^s$, estimates *how* we expect β to act. If an agent is considering repeated interaction with β then he may also be interested in how β's actions are expected to *change* in time.

The way in which the trust estimate is evolving is significant in understanding which agents to interact with. For example, and agent for whom $\tau^s_{\alpha\beta}(\hat{b})$ is fairly constant in time may be of less interest than an agent who is slightly less trustworthy but whose trust is consistently improving. To capture this information we need something like the finite derivative: $\frac{\delta}{\delta s}\tau^s_{\alpha\beta}(\hat{b})$. The sum of the elements in such a vector will be zero, and in the absence of any data it will decay to the zero vector.

Estimating the rate of change of $\tau^s_{\alpha\beta}(\hat{b})$ is complicated by the way it evolves that combines continual integrity decay with periodic updates. Evolution due to decay tells us nothing about the rate of change of an agent's behaviour. Evolution caused by an update is performed following a period of prior decay, and may result in compensating for it. Further, update effects will be very slight in the case that the commitment b is semantically distant from $\hat{b}$. In other words, the evolution of $\tau^s_{\alpha\beta}(\hat{b})$ itself is not directly suited to capturing the rate of change of agent behaviour.

The idea for an indirect way to estimate how β's actions are evolving comes from the observation that $\tau_{\alpha\beta}(\hat{b})^s$ is influenced more strongly by more recent observations, and the extent to which this is so depends on the decay rate. For example, if the decay rate is zero then $\tau_{\alpha\beta}(\hat{b})^s$ is a time-weighted "average" of prior observations. Suppose that $\tau_{\alpha\beta}(\hat{b})^s$ has been evaluated. We perform a parallel evaluation using a lower decay rate to obtain $\tau^-_{\alpha\beta}(\hat{b})^s$, then the vector difference, $\tau_{\alpha\beta}(\hat{b})^s - \tau^-_{\alpha\beta}(\hat{b})^s$, is a vector the sum of whose elements is zero, and in which a positive element indicates a value that is presently "on the increase" compared to the historic average.

The preceding method for estimating change effectively does so by calculating a first difference. If we calculate another first difference using an even lower decay rate then we can calculate a second difference to estimate the *rate of* change. This may be stretching the idea too far!

6 Trust in Context

The informal meaning of context is information concerning everything in the environment that could effect decision making *together with* rules that link that information to the deliberative process. That is, *context* consists of facts about the environment *and* rules that link those facts to the agent's reasoning. Those rules may rely on common sense reasoning.

Human and artificial agents have rather different practical problems in dealing with context. One practical difficulty for human agents is assimilating new information in an information-overloaded environment. Humans then rely on common sense and experience to learn how to key contextual information to their deliberation. Storage permitting, artificial agents can assimilate real-time data flows with ease, and can manage the integrity decay of old information. After that things become tricky for artificial agents; identifying and dealing with inconsistency is a hard problem, and so is keying context to deliberation.

To make matters worse, both human and artificial agents can reasonably assume that their knowledge of their context is substantially incomplete. Dealing with context is arguably *the* major impediment to delivering trustworthy negotiation by artificial agents in the real world. After this grim observation we consider the context of trust.

Following the procedure described in Section 3 an agent builds up a sense of trust on the basis of its own past experience and statements of opinion and reputation from other agents. In a sense those statements of opinions and reputation are contextual information for the business of estimating trust.

Suppose that an agent has built up a sense of trust in another agent based on their prior interaction, before relying on that trust estimate as an indicator of future performance the agent will consider whether there are any perceivable changes in the context that cause it to distrust its previous observations as an indicator of future behaviour. As a simple example, if John has an impeccable history of delivering goods on time then the contextual information that John has sprained his ankle, or that he is overseas, may cause us to distrust our experience as an indicator of John's timeliness in the near future.

In this paper 'trust in context' is concerned with just one issue: is there any reason to distrust our trust estimate due to a change in context. Supposing that α is considering signing a contract (a, b) at time t, to address this issue we require:

1. knowledge of the context of previous observations of behaviour. Their *context* is the state of each of the observables in the environment and of the states of the other agents when those previous observations of behaviour were made — given the way that observations are aggregated in Section 3 the more recent the observation the greater its significance.
2. founded beliefs concerning the context that will pertain at the future time of the evaluation of the presumed future behaviour — i.e. at time t'' in Figure 1.
3. some reasoning apparatus that enables us to decide whether differences between the believed future context and the observed previous contexts cause us to modify our experience-based trust estimate.

Taken together these three points are the *context* of the trust estimate that α has for the act of signing (a, b) with β. As stated the context of an observation of behaviour is the state of *all* observables at the time the observation is made. This is a potentially massive exercise. A causal model that identified only those observables that could be seen to cause or affect the behaviour would simplify things but is a major issue in its own right and is beyond the scope of this discussion.

The information-based architecture makes a modest contribution to the maintenance of trust estimates through the persistent decay of information integrity by Equation 1. Beyond that we offer no 'magic bullet' solutions to the contextual problems described above and leave the discussion as a pointer to the work that is required to increase the reliability of trust estimation in dynamic environments.

7 Future Work

Current work is focussed heavily on the issues of context identified in Section 6. In particular we are exploring the application of the minimum message length principle to reduce the complexity of models of context — unfortunately this comes with a very high computational overhead.

References

1. Ramchurn, S., Huynh, T., Jennings, N.: Trust in multi-agent systems. The Knowledge Engineering Review 19(1), 1–25 (2004)
2. Artz, D., Gil, Y.: A survey of trust in computer science and the semantic web. Web Semantics: Science, Services and Agents on the World Wide Web 5(2), 58–71 (2007)
3. Viljanen, L.: Towards an Ontology of Trust. In: Katsikas, S.K., López, J., Pernul, G. (eds.) TrustBus 2005. LNCS, vol. 3592, pp. 175–184. Springer, Heidelberg (2005)
4. Huynh, T., Jennings, N., Shadbolt, N.: An integrated trust and reputation model for open multi-agent systems. Autonomous Agents and Multi-Agent Systems 13(2), 119–154 (2006)
5. Sabater, J., Sierra, C.: Review on computational trust and reputation models. Artificial Intelligence Review 24(1), 33–60 (2005)
6. Sierra, C., Debenham, J.: Information-based agency. In: Proceedings of Twentieth International Joint Conference on Artificial Intelligence IJCAI 2007, Hyderabad, India, pp. 1513–1518 (January 2007)
7. Sierra, C., Debenham, J.: The LOGIC Negotiation Model. In: Proceedings Sixth International Conference on Autonomous Agents and Multi Agent Systems AAMAS 2007, Honolulu, Hawai'i, pp. 1026–1033 (May 2007)
8. Debenham, J., Sierra, C.: When Trust Is Not Enough. In: Huemer, C., Setzer, T., Aalst, W., Mylopoulos, J., Rosemann, M., Shaw, M.J., Szyperski, C. (eds.) EC-Web 2011. LNBIP, vol. 85, pp. 246–257. Springer, Heidelberg (2011)
9. Matt, P.A., Morge, M., Toni, F.: Combining statistics and arguments to compute trust. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, Richland, SC, pp. 209–216. International Foundation for Autonomous Agents and Multiagent Systems (2010)
10. Burnett, C., Norman, T.J., Sycara, K.: Bootstrapping trust evaluations through stereotypes. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, Richland, SC, pp. 241–248. International Foundation for Autonomous Agents and Multiagent Systems (2010)
11. Sierra, C., Debenham, J.: Information-based reputation. In: Paolucci, M. (ed.) First International Conference on Reputation: Theory and Technology (ICORE 2009), Gargonza, Italy, pp. 5–19 (2009)
12. MacKay, D.: Information Theory, Inference and Learning Algorithms. Cambridge University Press (2003)
13. Buechner, J., Tavani, H.: Trust and multi-agent systems: applying the "diffuse, default model" of trust to experiments involving artificial agents. Ethics and Information Technology 13(1), 39–51 (2011)

Validating the Relationship between Information Quality and Trust: The Moderating Effect from Customer Orientation

Shun Cai[1], Xina Yuan[1], and Kyounghee Chu[2]

[1] School of Management, Xiamen University,
Xiamen, Fujian Province, 361005, China
{caishun,zinayuan}@xmu.edu.cn
[2] Business School, Korea University,
Seoul 135-701 South Korea
750202@korea.ac.kr

Abstract. The Internet has grown extremely fast and given birth to E-taling, a new and popular way to sell goods online. Developing a trustworthy environment to generate customer trust is a powerful tool to sustain customer loyalty for E-tailers. This research proposes and validates the importance of information quality and customer orientation in website design on trust-building. More importantly, our research model suggests that customer orientation not only has a direct effect on online trust, but also moderates the relationship between information quality and trust. An laboratory experiment was setup and a large-scale survey was conducted to test our research model and proposed hypotheses. Results support our hypotheses. Managerial and theoretical implications are discussed.

Keywords: Information Quality, Customer Orientation, Trust, E-tailing.

1 Introduction

The Internet has grown extremely fast and given birth to E-taling, a new and popular way to sell goods online [1-3]. In order to attract customers to visit and revisit E-tailers' sites, they attempt to build web sites that meet consumer needs [1]. However, an electronic hypermedia environment faces new challenges to establish and maintain long-term relationship with the customers, because the customer's perceived risk and insecurity on Internet shopping arising from the physical separation between buyers and sellers provide these challenges to E-tailers [4-6]. Therefore, developing a trustworthy environment to generate customer trust is a powerful tool to sustain customer loyalty for E-tailers [7, 8].

Evidence has confirmed that trust is a significant predictor of the customer's online purchase intention, personal information disclosure, and use of services offered by the vendor [1, 9, 10]. It is the lack of trust between consumers and E-tailers that delay accomplishing its e-commerce potential scale [11, 12].

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 13–24, 2012.

The extant research has recognized that trust building is a psychological process. In this process, evidence about a firm, the trustee, is interpreted with respect to the goodwill of the firm, probability of opportunistic behavior by the firm, and its ability to fulfill obligations [13, 14]. If such reasoning favors the conclusion that a firm can and will fulfill its obligations, the customer's trust is formed. But where does the evidence come from? One of the key evidences is information quality commercial websites. In a computer-mediated environment, customers' purchasing decisions on a firm's products and services can be determined by their perceived quality of information [1, 15, 16]. However, it has been reported that current websites still contain numerous usability problems [7, 17]. Lack of interaction and reliability, lack of professionalism (e.g. low-quality images), irrelevant content, difficulty in reading text (e.g. different fonts, typefaces) and learning functionality (e.g. inconsistent design structures) are often cited as usability problems. Becker and Mottay state that "online business failures are increasing as consumers are turning away from unusable sites. The 'build it and they will come' attitude has led to the demise of several e-commerce sites which are too slow, too buggy, or too complex for ease of use" (p. 54) [18]. Moreover, from an online consumer's perspective, the evidences leading for trust-building might be highly context-dependent [19]. Even though, the effect of website's information quality on customer trust had been proposed in the literature, some studies reported that the effect of information quality on customer trust is unexpectedly weak or non-significant. Therefore, we expect that the causal relationship between information quality and trust could be contingent on certain consumer orientations.

This study attempts to investigate how perceived information quality (IQ) affects online consumer trust under different level of customer orientation (CO). These two constructs are expected to have direct and synergistic effects on customer trust during an online shopping. A laboratory experiment manipulating information quality (IQ) and customer orientation (CO) were conducted to test our hypotheses (Study 1). Further, a survey was employed to test our research model (Study 2). By combining two research methods, this research exploits the moderating role CO in the effect of IQ on customer trust, in addition to the direct effects of IQ and CO. Results show the substantial direct effect of IQ and CO, and the interaction effect of IQ and CO. The moderating effect of CO means that the strength of the relationship between IQ and customer trust in E-tailers decreases as CO increases.

2 Theory and Hypotheses

This study focuses on identifying the independent impact of IQ and synergistic effect with CO on consumer's trust in E-tailers. The conceptual model underlying our research is shown as Figure 1.

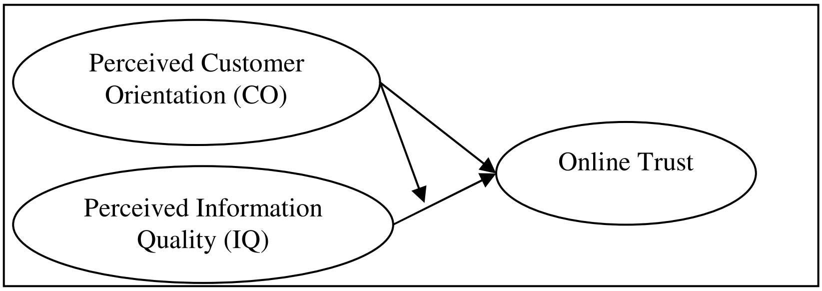


Fig. 1. Proposed Research Model and Hypotheses

2.1 Perceived Customer Orientation (CO) and Trust

The importance of customer orientation design in e-commerce has been widely recognized by academia and practitioners. In practice, senior management reports emphasize that customer satisfaction, positive word of mouth, and repeat business are the outcome of a firm's strategy of customer orientation. In a similar vein, Luo and Seyedian [20] argue that firms in the virtual market space should be committed to meet customer needs to make them feel satisfied and become loyal to the sites. In other words, customers' perception of customer orientation strategy firms has influence over customer satisfaction of internet storefronts.

Previous research has investigated the influence of CO on behavioral intentions (purchase intention, repurchase intention, revisit intention) in the wholesaler–retailer context[21]. For example, Deshpande et al. [22] investigate the relationship between customer evaluation of the firm's customer orientation strategy (perceived customer orientation) and business performance. They emphasize that the evaluation of a customer oriented organization should come from its customers rather than merely from the company itself. In other words, the effect of customer orientation should be measured from the customers' perspective. Day and Nedungadi [23] indicate that customer satisfaction and loyalty are important indicators to measure the performance of customer-oriented business.

Equally important, a Website could be evaluated based on the effectiveness of product organization and the degree of customer focusing. However, most of the customer orientation studies focus on antecedents and performance outcome of customer orientation from the organization-centric view (i.e., profitability, market share, and revenue, etc.), rather than consider the perspective from customer such as CO and its outcome (i.e., customer response and behaviors, etc.). DeLone and McLean [24] asserted that the information system's success (i.e., website success) should together consider the independent variables such as customer focus (i.e., customer orientation). Moreover, Chang et al. [25] found that companies which exhibit customer orientation, by being vigilant regarding the need of customers, would achieve better performance in the e-commerce. A consumer is likely to trust a customer-oriented e-retailer over a less customer-oriented one. Thus, we hypothesize that:

HYPOTHESIS 1: The perceived customer orientation (CO) of e-tailer will positively affect consumer's trust towards an e-tailer.

2.2 Perceived Information Quality (IO) and Trust

With the dramatic growth of the Internet, customers and lodging companies rely on the Internet as a major information resource and marketing tool. The information quality of commercial websites is becoming a focal point in the current consumer behavior and e-commerce literature [15, 26-29]. In an invented computer-mediated environment, customer purchasing decisions are influenced by their IQ. The information quality is top-ranked in all types of products as a key success factor by exploring web success factors. In addition, the information quality of a company's offerings is also believed to be crucial to create a positive image of the company and build an ongoing relationship with the customer [30-32].

Website information is a critical factor for website design because customers' online purchasing decision is directly affected by the quality of the website [5, 15, 26-29, 33]. Poor quality websites can result in loss of customers to competitors, escalation of costs, and reduction in profits. Moreover, trust will evaporate when people perceive that the information quality on the website is not good [34]. In other words, if the information provided by the E-tailer is useless, insufficient, inaccurate, or incomplete, customer trust will be diminished, since they may consider that E-tailers engage in detrimental opportunistic behavior. In contrast, providing high information quality on the website will increase trust. Specifically, we hypothesize that:

HYPOTHESIS 2: The perceived information quality (IQ) of an e-retailer's website will positively affect the consumer's trust towards an e-retailer.

2.3 The Moderating Effects of Perceived Customer Orientation

The existing evidences for the relationship between IQ and trust have been mixed. Some studies have reported the positive linkages between IQ and trust [10, 35]. However, some research also mentions that there is no significant effect of information quality on customer response such as customer satisfaction and trust [36]. The inconsistent results show that there might be some moderators that affect the strength and/or direction of the relationship between IO and trust.

The mixed results could be due to the fact that trust and the need for information are alternative mechanisms to cast away uncertainty and that there is an inverse relationship between trust and the need for information [37, 38]. We can infer this that the more trust is given, the less information is required. That is, the more one trusts in an e-retailer, the less information one needs to make a decision. In contrast, if there is little trust, one will need a more accurate and complete information in order to decrease uncertainty and build more trust.

According to this logic, the influence of CO may be conjectured as the following expectations: in the high level of CO, IQ has a strong positive impact on the consumer trust in E-retailers; consumers might have to some extent form trust in E-retailers, indicating that they have less need for information to cope with uncertainty [38]. Therefore it is relatively less important to provide high quality information (e.g., more accurate and complete information) to instill customer trust in e-retailer, meaning the relatively low effect of good information on trust in e-retailers. On the other hand, in the low level of CO, which is likely to lower consumer trust in e-retailer, it creates further doubts and perceived risk thus discouraging purchasing decision. Therefore it is relatively important to provide high quality information concerning products and services to reduce the uncertainty and perceived risk, hence, encouraging the customer trust in E-retailer.

Based on the previous reasoning, we might expect that there is a synergistic effect between IQ and CO: First, when IO is good, high CO would help in conveying the same and hence improve the consumer's trust in E-tailers. Second, even when IQ is good, low CO will lower the possibility of consumers building high trust. Third, when IQ is poor, low perceived CO would only lead to the further decline of trust in E-tailers. Fourth, even when IQ is poor, high CO can prevent it from degrading trust in E-tailers. We therefore suggest that IQ is a significant and positive determinant to build trust in e-retailers. However, we also argue that the higher consumers perceive an e-retailer as customer-oriented, the less consumers rely on PIQ to build trust, thus the effect of IQ weakens as CO increases. Therefore, we hypothesize:

HYPOTHESIS 3: The effect of IQ on consumer's trust in E-retailers would be moderated by the CO.

3 Research Methodology

We designed 2 empirical studies to test our research model and hypotheses.

3.1 Study 1: A Laboratory Experiment

In study 1, a laboratory experiment was used to test our hypotheses. The empirical design of this study comprised two phases: a pre-test and a main experiment.

The pre-test served as a process for validating the manipulations of information quality (IQ) and customer orientation (CO). In particular, we first selected a set of actual web stores, as will be explained shortly. A proxy program was developed to manipulate these web stores. When subjects visited these web stores through the proxy server, they saw a manipulated version of the actual web store. IQ and CO were manipulated through configurations of the proxy server. Measures for IQ and CO were developed for manipulation checks (please refer to Table 1). In the pre-test, we recruited student subjects to evaluate various versions of websites and rate those websites based on IQ and CO scales. Based on results from pre-test, we selected the most appropriate website designs for our main experiment (due to the page limit, the details of the pre-test results are not reported here but available on request).

The main experiment is a 2X2 factorial design, manipulating IQ and CO. A sample of 164 students was recruited for the 2x2 experiment with manipulations of the information quality and customer orientation of the website. The 161 subjects were randomly assigned into four experimental groups. After they browsed the website, they were asked to complete a questionnaire. 3 questionnaires were discarded due to incompleteness. Finally, we have 41 valid questionnaires for high IQ and high CO group, 41 for high IQ and low CO group, 39 for low IQ and high CO group, and 40 for low IQ and low CO group. Among the 161 subjects, 92 of them are male (57%) and 69 are female (43%). Regarding daily Internet usage, 129 out of the 161 subjects spend 1 to 3 hours on Internet (80%), 30 of them spend 4 to 7 hours (19%), and 2 of them spend more than 7 hours on Internet (1%).

For control check, we performed statistical tests on subjects' gender, age, Internet usage experience to check the results of random assignments. A Chi-square test for gender was performed. There was no significant difference in terms of gender across different treatments. ANOVA were conducted for other variables. No significant difference was found. Therefore, the randomization was considered adequate.

To check the construct validity of multiple-item scales used in our experiment, we conducted confirmative factor analysis (CFA) with a LISREL path diagram [39]. A series of tests suggest that our measurement model is adequate. Therefore, we averaged the scores for each construct for further test.

Table 1. Constructs and Measures

Constructs	Measures	Reference
Customer Orientation	1. It is likely that this company's strategy for competitive advantage is based on the understanding of customers' needs	[40, 41]
	2. It is likely that this company's business strategies are driven by beliefs about how it can create greater value for customers.	
	3. It is likely that this company's business objectives are driven by customer-satisfaction.	
	4. It is likely that this company measures customer satisfaction systematically and frequently.	
	5. It is likely that this company gives close attention to After-sales service.	
Information Quality	1.This company provides useful information to the customers	[24,42-44]
	2. This company provides complete products description information to the customers	
	3. This company provides timely information to the customers	
	4. This company provides relevant information to the customers	
Trust	1. This company is trustworthy	[13,27, 34]
	2. I trust this company keeps my best interest in mind	
	3. This company wants to be known as one who keeps promises and commitments	
	4. I believe that my transaction with this store is likely to be safe	
	5. My transaction with this store is likely to be reliable	

For hypothesis testing, we conducted a series of ANOVA tests. The results of the hypotheses analysis in our proposed model are presented in Table 2, Table 3, Table 4, and Figure 2. Hypothesis 1 predicted that the perceived customer orientation for E-tailers will positively affect the consumer's trust towards E-tailers. Table 2 reveals a significant positive effect in the hypothesized direction ($F = 66.654$, $p = .000$, $\eta 2 = .265$), supporting Hypothesis 1. In Hypothesis 2, we proposed that the information quality of a website will positively affect the consumer's trust towards E-tailers. Table 2 reveals a significant positive effect in the hypothesized direction ($F = 23.992$, $p = .000$, $\eta 2 = .133$), supporting Hypothesis 2. More importantly, Hypothesis 3 predicted that the positive causal effect of information quality on trust would be moderated by CO. Table 2 reveals a significant interaction effect in the hypothesized direction ($F =7.991$, $p = .010$, $\eta 2 = .042$), supporting Hypothesis 3. The source of this interaction was examined further by testing the simple effect at high and low level of CO (see Table 3). In the case of which CO is low, the effect of IQ on trust is significant and positive ($F = 28.35$, $p= .000$, $\eta 2 = .14$). However, in the case of which CO is high, the effect of PIQ on trust is only marginally significant ($F = 2.86$, $p=.092$, $\eta 2 = .01$). That is, the simple effect of IQ was more pronounced in the low CO than in the high CO ($F (1,157) = 28.35$ and 3.56, $p=.000$ and .092, $\eta 2 = .14$ and .01, respectively for low CO and high CO.

Table 2. Results for ANOVA Tests

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
IQ	28.213	1	28.213	23.992	.000
CO	66.654	1	66.654	56.743	.000
IQ x CO	7.991	1	7.991	6.803	.010
Error	184.421	157	1.175		

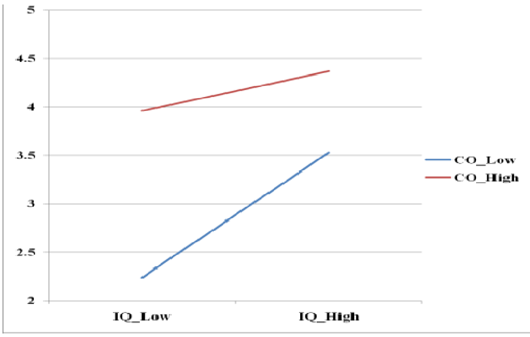


Fig. 2. Estimated Marginal Means of Trust in Different Conditions

Although all hypotheses proposed in our research are supported, we note the limitations of our experiment test. Two of the major limitations of our experiment research is that 1) the manipulations of IQ and CO could lead to changes in our aspects of the website, 2) the extent to which the conclusions drawn from such an experiment design could be applied in real shopping context is limited, i.e., the external validity is relatively weak. Therefore, we conducted Study 2, a survey research, to further validate our research model

Table 3. The Effects of IQ and CO on Trust

PCO	PIQ		
	High PIQ	Low PIQ	Total
High PCO	4.37 (1.01, 41)	3.96 (0.97, 39)	4.17 (1.07, 80)
Low PCO	3.53 (1.04, 41)	2.25 (0.82, 40)	2.89 (1.08, 81)
Total	3.95 (1.09, 82)	3.10 (0.95, 79)	

*(a, b), a: Std. Deviation, b: cell sample size

3.2 Study 2: A Survey Research

The survey method was used to test the model because it provides a basis for establishing generalizability, allows replicability, and has statistical power[45]. First, a literature search was carried out within the domain of the constructs to generate sample items. Short interviews with 15 Internet shoppers were next conducted to assess their face validity followed by a process of conceptual validation. Table 2 presents the instruments used in our survey. All constructs were measured in a 5-point likert scale. A survey was conducted in a large university in Korea. 260 completed questionnaires were collected.

We examined the response bias issue with the procedure suggested by Armstrong and Overton [46]. We compared early and late respondents on key demographic variables, namely, their work experience in the current organization and the current industry. The results of the Mann-Whitney tests show no significant differences, suggesting that response bias would not likely affect our findings [46].

As with all self-reported data, there is the potential for the occurrence of common method variance (CMV), i.e., variance that is attributable to the measurement method rather than to the constructs the measures represent [47]. To address this issue, we performed a Harman’s one-factor test via CFA by specifying a hypothesized method factor as an underlying driver of all of the indicators. The results revealed that the fit of the single-factor model was extremely unsatisfactory, indicating the common method variance is not a major source of the variations in the items [48]. While the results of this analysis do not explicitly preclude the possibility of common method variance, they do suggest that common method variance is not of great concern in this study.

We then conducted data analysis in accordance with a two-stage methodology [39]. The first step in the data analysis is to establish the convergent and discriminant validity of the constructs. We test the measurement model using Principal Components Analysis (PCA) in SPSS (SPSS, Chicago, USA) and Confirmatory Factor Analysis (CFA) in LISREL [49]. In the second step, the structural models are examined using Structural Equation Modelling (SEM) based on the cleansed measurement model. The results from PCA and SEM reveal that our measurement model has appropriate level of construct validity (due to the page limit, the detailed results are not presented here).

Table 4 presents our hypotheses testing results. IQ and CO are found to be statistically significant to online consumer trust (H1 and H2). Our analysis also reveals significant moderating effects from CO on the relationship between IQ and Trust. Hence, all the hypotheses in this study are supported (see Table 4).

Table 4. Hypothesis Testing Results

Paths (Hypotheses)	Beta	T value	p value	Support Hypothesis?
IQ - Trust(H1)	.448	6.840	.000	Yes
CO – Trust (H2)	.413	7.298	.000	Yes
IQ*CO-Trust (H3)	-.054	-2.492	.013	Yes

4 Conclusions

In conclusion, our research proposes and validates the importance of information quality and customer orientation in website design on trust-building. More importantly, our results reveal that customer orientation not only has a direct effect on online trust, but also moderates the relationship between information quality and trust.

The importance of the two independent variables, namly, information quality and customer orientation, would be of substantial worth. For E-tailer managers, understanding about whether the information quality of the E-tailer's website or the continuous customer-oriented efforts, has a greater effect on improving customer trust in e-retailers would be of substantial importance. Investigation of the moderating role of perceived customer orientation is of considerable value to E-tailer marketers. For the purpose of optimum resource allocation, our empirical results suggest that E-tailer managers should devote their resources to developing some cues (e.g., services, slogans, or systems), which make consumers perceive customer-orientated company, that directly create trust in E-tailers.

Acknowledgments. This research is supported by the National Research Foundation for Doctoral Program of Higher Education of China, Grant No.20100121120013, the National Ministry of Education of China (Humanities and Social Science), Grant No. 10YJC630010, and the Fundamental Research Funds for the Central Universities, Grant No. 2011221016. Dr Xina Yuan is the corresponding author of this paper.

References

1. Mukherjee, A., Nath, P.: Role of electronic trust in online retailing: A re-examination of the commitment-trust theory. European Journal of Marketing 41, 1173–1202 (2007)
2. Cai, S., Xu, Y.: Effects of outcome, process and shopping enjoyment on online consumer behavior. Electronic Commerce Research and Application 5, 272–281 (2006)

3. Cai, S., Xu, Y.: Designing not just for pleasure: Effects of Web site aesthetics on consumer shopping value. International Journal of Electronic Commerce 15, 159–187 (2011)
4. Mauldin, E., Arunachalam, V.: An Experimental Examination of Alternative Forms of Web Assurance for Business-to-Consumer e-Commerce. Journal of Information Systems 16, 33–54 (2002)
5. Pavlou, P.A., Gefen, D.: Building Effective Online Marketplaces with Institution-Based Trust. Information Systems Research 15, 37–59 (2004)
6. Teo, T.S.H., Yeong, Y.D.: Assessing the consumer decision process in the digital marketplace. OMEGA 31, 349 (2003)
7. Tung, L.L., Xu, Y., Tan, F.B.: Attributes of Web Site Usability: A Study of Web Users with the Repertory Grid Technique. International Journal of Electronic Commerce 13, 97–126 (2009)
8. Kim, H.-W., Xu, Y., Koh, J.: A comparison of online trust building factors between potential customers and repeat customers. Journal of the Association for Information Systems 5, 392–420 (2004)
9. Gefen, D., Straub, D.W.: Managing user trust in B2C e-services. e-Service Journal 2, 7–24 (2003)
10. McKnight, D.H., Choudhury, V., Kacmar, C.: Developing and validating trust measures for e-commerce: An integrative typology. Information Systems Research 13, 334–359 (2002)
11. Dinev, T., Hart, P.: Internet privacy concerns and their antecedents - measurement validity and a regression model. Behaviour & Information Technology 23, 413–422 (2004)
12. McCole, P., Ramsey, E., Williams, J.: Trust considerations on attitudes towards online purchasing: The moderating effect of privacy and security concerns. Journal of Business Research 63, 1018–1024 (2010)
13. Doney, P., Cannon, J.P.: An examination of the nature of trust in buyer-seller relationships. Journal of Marketing 61, 35–51 (1997)
14. Doney, P., Cannon, J.P.: Understanding the influence of national culture on the development of trust. Academy of Management Review 23, 601–620 (1998)
15. Agarwal, R., Venkatesh, V.: Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability. Information Systems Research 13, 168–186 (2002)
16. Kumar, A., Grisaffe, D.B.: Effects of Extrinsic Attributes on Perceived Quality, Customer Value, and Behavioral Intentions in B2B Settings: A Comparison Across Goods and Service Industries. Journal of Business-to-Business Marketing 11, 43–74 (2004)
17. Treiblmaier, H., Pinterits, A.: Developing metrics for web sites. Journal of Computer Information Systems 50, 1–10 (2010)
18. Becker, S.A., Mottay, F.E.: A global perspective on web site usability. IEEE Software, 54–61 (2001)
19. Bigley, G.A., Pearce, J.L.: Straining for Shared Meaning in Organization Science: Problems of Trust and Distrust. The Academy of Management Review 23, 405–421 (1998)
20. Luo, X., Seyedian, M.: Contextual marketing and customer-orientation strategy for e-commerce: an empirical analysis. International Journal of Electronic Commerce 8, 95–118 (2004)
21. Cano, C., Carrillat, F., Jaramillo, F.: A meta-analysis of the relationship between market orientation and business performance: evidence from five continents. International Journal of Research in Marketing 21, 179–200 (2004)

22. Deshpandé, R., Farley, J.U., Webster, F.E.: Corporate culture customer orientation, and innovativeness in Japanese firms: A quadrad analysis. Journal of Marketing 57, 23–37 (1993)
23. Day, G.S., Nedungad, P.: Managerial representations of competitive advantage. Journal of Marketing 58, 31–44 (1994)
24. DeLone, W.H., McLean, E.R.: Measuring e-Commerce Success: Applying the DeLone & McLean Information Systems Success Model. International Journal of Electronic Commerce 9, 31–47 (2004)
25. Chang, K., Jacksona, J., Grover, V.: E-commerce and corporate strategy: an executive perspective. Information & Management 40, 663–675 (2003)
26. Torkzadeh, G., Dhillon, G.: Measuring Factors that Influence the Success of Internet Commerce. Information Systems Research 13, 187–204 (2002)
27. Koufaris, M.: Applying the technology acceptance model and flow theory to online consumer behavior. Information Systems Research 13, 205–223 (2002)
28. Delone, W.H., McLean, E.R.: Information Systems Success: The Quest for the Dependent Variable. Information Systems Research 3, 60–95 (1992)
29. McKinney, V., Kanghyun, Y., Zahedi, F.M.: The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach. Information Systems Research 13, 296–315 (2002)
30. Hoffman, D.L.: The Revolution Will Not Be Televised: Introduction to the Special Issue on Marketing Science and the Internet. Marketing Science 19, 1–3 (2000)
31. Hoffman, D.L., Novak, T.P.: Marketing in hypermedia computer-mediated environments: Conceptual foundations. Journal of Marketing 60, 50–68 (1996)
32. Straub, D.W., Hoffman, D.L., Weber, B.W., Steinfield, C.: Toward New Metrics for Net-Enhanced Organizations. Information Systems Research 13, 227–238 (2002)
33. Hoffman, D.: Building consumer trust online. Communications of the ACM 42, 80–85 (1999)
34. Morgan, R.M., Hunt, S.D.: The commitment-trust theory of relationship marketing. Journal of Marketing 58, 20–38 (1994)
35. McKnight, D., Cummings, L., Chervany, N.: Initial Trust Formation in new Organizational Relationships. Academy of Management Review 23, 473–490 (1998)
36. Jang, H., Olfman, L.I., Ko, J., Koh, J., Kim, K.: The influence of online brand community characteristics on community commitment and brand loyalty. International Journal of Electronic Commerce 12, 57–80 (2008)
37. Grabner-Krauter, S., Kaluscha, E.A.: Empirical research in on-line trust: a review and critical assessment. International Journal of Human - Computer Studies 58, 783–812 (2003)
38. Tomkins, C.: Interdependencies, Trust and Information in Relationships, Alliances and Networks. Accounting, Organizations and Society 26, 161–191 (2001)
39. Anderson, J.C., Gerbing, D.W.: Structural equation modelling in practice: A review and recommended two-step approach. Psychological Bulletin 103, 411–423 (1988)
40. Narver, J.C., Slater, S.F.: The Effect of a Market Orientation on Business Profitability. Journal of Marketing 54, 20–35 (1990)
41. Greenley, G.E., Hooley, G.J., Rudd, J.M.: Market orientation in a multiple stakeholder orientation context: implications for marketing capabilities and assets. Journal of Business Research 58, 1483–1494 (2005)
42. Aladwani, A.M., Palvia, P.C.: Developing and validating an instrument for measuring user-perceived web quality. Information & Management 39, 467–476 (2002)

43. Delone, W.H., McLean, E.R.: The DeLone and McLean model of information systems success: A ten-year update. Journal of Management Information Systems 19, 9–30 (2003)
44. Liu, C., Arnett, K.: Exploring the factors Associated with Website Success in the context of Electronic Commerce. Information & Management 38, 23–33 (2000)
45. Teo, H.H., Wei, K.K., Benbasat, I.: Predicting Intention to Adopt Interorganizational Linkages: An Institutional Perspective. MIS Quarterly 27, 19–49 (2003)
46. Armstrong, J.S., Overton, T.S.: Estimating Nonresponse Bias in Mail Surveys. Journal of Marketing Research 14, 396–402 (1977)
47. Podsakoff, P.M., MacKensie, S.B., Lee, J.-Y., Podsakoff, N.P.: Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology 88, 879–903 (2003)
48. Malhotra, N.K., Kim, S.S., Patil, A.: Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research. Management Science 52, 1865–1883 (2006)
49. Jöreskog, K.G., Sörbom, D.: LISREL for Windows. Scientific Software International, Inc., Lincolnwood (2006)

PAEAN – A Risk-Mitigation Framework for Business Transaction at Run-Time

Denada Cfarku[1], Yéhia Taher[1], Rafiqul Haque[2], Willem-Jan van den Heuvel[1], and Michael P. Papazoglou[1]

[1] European Reseach Institute for Service Science,
Tilburg University, The Netherlands
d.cfarku@gmail.com, y.taher@TilburgUniversity.edu,
{W.J.A.M.vdnHeuvel,mikep}@uvt.nl
[2] Lero - The Irish Software Engineering Research Centre,
University of Limerick, Ireland
rafiqul.haque@lero.ie

Abstract. The increasing complexity of the business transaction results in a higher potential risk in terms of SLA violation. The fulfillment of the QoS constraints specified in global SLA can be threatened at any time by different unexpected events that could occur during the execution of a business transaction. Unfortunately, there is no solution found that can efficiently mitigate this risk. Some of the current business transaction monitoring and managing solutions can monitor business transaction and report fault *only* after it happens. The framework proposed in this paper, monitors business transactions, computes potential risks and performs *proactively* adaptation actions in order to avoid global SLA violation that could causes transaction abortion.

Keywords: Service Composition, Non-functional QoS, SLA constraints, Business transaction Monitoring, CEP.

1 Introduction

Business transaction is a series of collaborative activities that are performed in a flexible manner by transaction participants in order to accomplish the agreed-upon business objectives [3]. For example, in an *order management* business transaction, *order placement*, *making payment*, and *shipping goods* (products) are the collaborative activities carried out by the participants, buyer, seller, and shipper complying with the agreement (better known as *service level agreement*). Typically, service level agreement (SLA) is the outcome of negotiation among participants, it happens before designing and deploying a business transaction. SLA contains the quality of services measured by *metrics,* such as the *processing time* of an 'order management business transaction'. To each metrics it is associated a *value* or *range* comprises of *lower* and *upper* threshold. For example, processing time could be *1* to *5 days* that must be satisfied by the while a business transaction is executing. Any otherwise case in particular, processing time of business transaction is more than 5

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 25–37, 2012.

days, will be treated as a *violation of agreement.* Such cases are serious faults that can result severe consequences, particularly abortion of the business transaction. In modern day's business transaction, such non-functional fault in particular, *violation of agreement* is critical since it fetches detrimental effect for running business transaction. Thus, a solution that efficiently mitigates the potential risks of SLA violation is indispensible. We consider two parameters in relation to the efficiency of the solution: (a) potential risk computation ability and (b) pro-activeness to avoid violation of SLA.

We emphasize on pro-activeness because, to the best of our understanding, reactive method is not adequately efficient to mitigate risks (SLA violation) owing to its (a) lack of ability to foresee potential risk, and (b) action-pattern. A solution that has built on a reactive method acts *only* after the failure or fault happens; it is not concerned with fault or failure that *will or may* occur at some point of the business transaction lifetime. Thus, resisting potential failure is out of the scope of the reactive method.

In this research, we offer a risk-mitigation framework named [1]*PAEAN* that facilitates foreseeing the risks involved in business transaction and carry out action proactively in order to avoid potential SLA violation that may cause transaction abortion. Proactive method is the base of the PAEAN risk management framework.

PAEAN integrates an automatic risk-computation mechanism that enables the framework to calculate the risk based on the information of the business transaction gathered at runtime. Additionally, the framework integrates a *monitor*, the most essential component, which piles up real-time information of the business transaction. It is not possible to compute and forecast future violation without knowing the current information of business transaction. Thus, the monitor is an important component of the framework. The automated risk-computation mechanism works in tandem the monitor to enable PAEAN to prognosticate the potential risks of SLA violation without any manual intervention. The most significant task that our framework is able to do is *repairing* the fault without any manual intervention. PAEAN integrates adaptation mechanism that acts proactively to repair the fault.

The paper is structured as follows: in Section 2, we present a motivating example, which shows the potential risk associating a business transaction at run-time. Section 3 then describes PAEAN, the risk-mitigation framework, our approach for solving the problem. Section 4 includes an experimental demonstration by showing the result of applying PAEAN. Section 5 presents other related work in terms of business transaction monitoring technologies and finally Section 6 presents the conclusions.

2 Motivation

In this section, we describe an example (Fig. 1) that highlights the motivation of this research paper. In the example, a buyer and a seller build through negotiation an SLA, which is a structured document containing the quality of services. To simplify the example, our framework will consider only two quality metrics: *processing time* and *acknowledgement time* that associate respectively the value of *5 days* and *5 minutes*. We assume that the agreed QoSs are annotated at design-time, at the same phase

[1] *PAEAN* is a Latin word means *the healer.*

when the business process, composed of activities (*aka* services, we will use them interchangeably in the paper), is designed. Our example does not explicitly show the annotation because the '*business process design phase'* is not within the focus of this research. However, after annotating QoS, the process model turns into a transactional process model that later is deployed on the process execution engine. Once it has deployed, it is the responsibility of the engine to manage risks while the business transaction is running. It is worth noting that transaction processing engine integrates off-the-shelf management software to perform the management tasks.

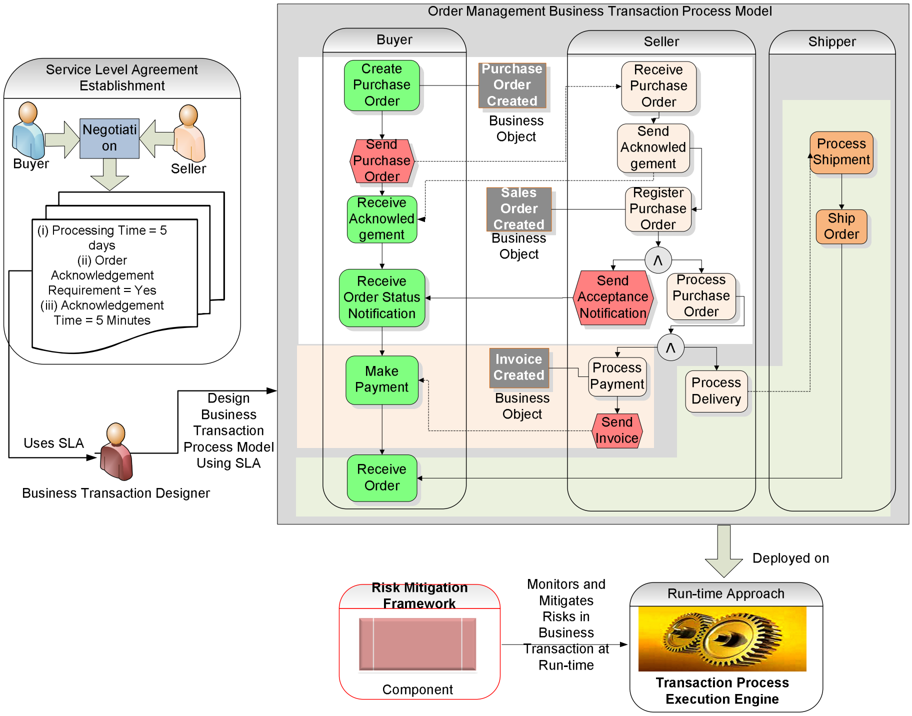


Fig. 1. An example of an order management business transaction at design-time and run-time

While the business transaction is running, at any time, the QoSs are venerable to threat to various events that could happen during the execution of the business transaction. For example, the supplier could fail to deliver the product within the deadline, and consequently the seller could not deliver the goods to the buyer within the agreed *processing time* (5 days). This would results in a SLA violation and the buyer could cancel the order, which would ultimately cause a business transaction abortion. This is a business-oriented fault. There could also be technical fault, such as a component simply stops working, 'Halting failure', or a service becomes unavailable because the server where the service is running could stop working for different reasons; it needs to be removed, substituted or has to restart.

All these failure scenarios demonstrate the risks involved in a business transaction at run-time. In addition, these scenarios create the critical need of running a transaction monitoring activity in parallel to a service execution. The monitoring activity would be responsible of monitoring not only the service functionality, but also especially the QoS levels and trigger proactive adaptation actions when the QoS constraints are violated [1]. In other words, there must be a framework associating the transaction processing engine that will trace and mitigate the risks in a business transaction execution. Such a framework is missing in the state of the art. Through this framework, there are a number of interesting objectives to aim at:

- Checking the availability of each service of the business transaction;
- Estimating the value of the QoS constraints that are in scope of this work;
- Checking regularly QoS constraints for any case of SLA violation;
- Performing adequate adaptation action if a SLA violation is identified.

We develop PAEAN Risk-Mitigation Framework to achieve these objectives. We in the following section describe the framework.

3 PAEAN – The Risk-Mitigation Framework

In this section, we give detail explanation of the basic building blocks the PAEAN framework. In addition, we discuss PAEAN's theoretical model, which automates risk computation and fault repairing.

3.1 Overview

Since the proposed framework deals with business transaction, it was essential to decide what behavioral principle of transactions would be suitable for the framework. Our intention was not to define any new principle but to find a business transaction model, which relies on flexible principle. The classical atomicity [8] is not suitable since its semantic is very strict at the point that it cannot tolerate even trivial faults. Thus, PAEAN would not be effective or even applicable in such a strict environment. We found the extension of classical atomicity in [3], called *eventual failure atomicity* which informally described as follows: "*for every failure if occurs in a transaction, ignore the failure if the failed operation is not vital or try all possible options to resist the abortion*". We found this principle compatible with the key notion of our risk-mitigation technique - substituting service in case of failure or unavailability of a service at runtime. Thus, we develop the PAEAN framework adopting eventual failure atomicity principle instead of classical atomicity. It was important to ratify the existence of such a flexible (behavioral) principle of the business transaction; otherwise, the applicability of the proposed framework would be in question.

The aim of PAEAN framework is to guarantee the consistency of QoS metrics relevant in the context of the provider-client contract and their acceptable values specified in the SLA. Inconsistent business transaction – global QoS constraints not fulfilled at runtime – causes violation of SLA that might result in transaction abortion. Note that, global QoS is contained in *global SLA* is a type of service level agreement involves the major business participants include buyer and the seller. The other type is *local SLA* that engages the seller with another participant type such as supplier, shipper etc. PAEAN framework is concentrated on mitigating the potential risks of global SLA violation.

To the best of our understanding, business transaction is possible to be continued (using efficient techniques) if any violation happens in local SLA, but violation of global SLA jeopardizes the entire transaction. We develop PAEAN based on this understanding. The framework reacts to local SLA violation and pro-acts to repairing every possible fault to mitigate risks of global SLA violation. The *adaptation mechanism* PAEAN uses to repair the transaction faults.

The pivotal constituents of PAEAN framework are *QoS metrics* and *composite flow patterns*. In this research, our approach with respect to QoS can be defined as multi-dimensional because it focuses in more than one QoS metric. QoS metrics can be divided into two groups: quantitative properties, such as time and cost, and qualitative properties, such as reliability and availability. The composition flow patterns used by the PAEAN framework include sequential, switch and loop. We discuss QoS metrics and service composition patterns in detail in next section.

3.2 The Fundamental Constituents

Qos metrics and *composite flow patterns* are the fundamentals constituents of our approach. In this section, we describe these essentials in detail. The fundamental constituents are determined based on our definition of the QoS constraints as relations that represent values of QoS metrics based on the composite flow patterns.

At current stage, the PAEAN framework uses a limited number of QoS metrics that are briefly described in the followings:

- **Time:** In the case of a composite service, time is an aggregation of the execution time of each component service S_i. In the workflow system, time can be defined as the total time needed for a workflow instance to be processed by a task. Time (T) can be seen as composed of two major components. The first component is *Invoke time* (IT). This is the time that the instance needs to be ready to be processed by the task. The second dimension is the *Process time* (PT). This is the time where the workflow instance is being processed by the task. Therefore, *Time* for a task *t* can be computed as follows: $T(t) = IT(t) + PT(t)$
- **Cost:** It represents the cost associated with the composite service defined as a total value of the cost of each component service S_i. In the workflow system, cost represents the total cost of executing all workflow tasks.
- [2]**Reliability:** This metric corresponds to the probability that a component will perform correctly within the expected time. The reliability of a composite service can be expressed as the production of the reliability of each individual component service S_i.
- **Availability:** Availability of a composite service is the probability that each component service is available. Therefore, it can be expressed as a product of each individual component service S_i.

[2] `http://www.s-cube-network.eu/km/terms/r/reliability`

The other constituent of our approach is the composite flow patterns (Fig. 2). There are different ways that services can be composed into a composite service. In this research, we focus in three types of composition relationships: sequential, switch and loop associations. Figure 2 shows a combination of the basic composite flow patterns into a composite service.

The composite service shown above is composed of seven services and it includes different composite flow patterns. Services 'S1', 'S6', and 'S7' follow the sequential model and they are not part of any loop or switch. For example, when the buyer puts an order in the system, first the personal data should be entered. This process is executed only once. Nevertheless, there could be the case, as it often happens, that for the same process instance, the buyer puts in the system more than one order. The data for each order need to be entered. This service follows the loop model and it corresponds to service 'S2'. In this case the time metric is expressed as a product of the constant K, K*T(i). K is a *counter* and represents the number of iteration of the service execution. Other services, such as *checking goods availability* in warehouse or *checking the data validity,* follow a switch model ('S3' in Fig. 2) because the output of this services would result in the execution of different services depending on the condition result (true or false).

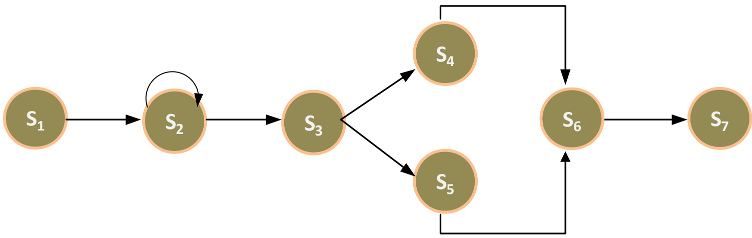


Fig. 2. Composite flow patterns in a composite service

Based on the QoS metrics we have described and the composite flow patterns that are applied to a business process, is feasible to build QoS constraints that can be stated in global SLA. Our framework will monitor these QoS constraints continuously during the business transaction execution. For example, the business transaction described in Section 2 could be designed to meet the following time and cost constraint:

T < 180 time units and C < 100 price units

The PAEAN's monitor will be policing whether business transactions satisfy the time and cost constraint.

3.3 The Theoretical Model

The constituents depicted in Section 3.2 are the *building blocks* of the PAEAN framework. In this section, we describe how the framework uses these constituents to mitigate the run-time risks of business transactions in a Service-Based Application (SBAs). The services contained in SBAs are usually composite. A composite service can formally be defined as follows.

Definition 1: *A composite service* S $\stackrel{\text{def}}{=}$ $\{s_1, s_2,....,s_n\}$ *where 's' represents an atomic service.* By 'atomic', we meant a non-decomposable service, which is indeed a *task* (aka *operation*) that is performed at runtime.
We assume that the design-time specification of SBA describes the composite services and the QoS that are associating those services. The risk-computer, a component of the PAEAN framework, computes risks using as input the information supplied by the monitor. The risk-calculator relies on a mathematical model that we have developed for the purpose of this research. At current stage, PAEAN can compute only temporal risk. The mathematical model for risk computation is described in the following.

- ***Risk-computation Model***

Let consider that T_i is the processing time for the service S_j. The total processing time for a composite service is expressed as $T = \sum T_i$ where i = 1…k. For every S_j, there is T_i such that i = j. T_i has an expected value (specified as a QoS constraint) that has to be satisfied after the completion of the service execution. On the other hand, the monitor provides actual value of T_i, which represents the total execution time taken by the service S_j. The expected values of all services in an SBA are temporal constraints that we mathematically define as follows:

- C $(A_{SBA}) = (\sum_{i=1}^{k} T(EV)) \leq T(AV)$ such that $T_1 (AV) \leq T_1 (EV)$, $T_2 (AV) \leq T_2 (EV)$, ……., $T_k (AV) \leq T_k (EV)$ where C, ASBA, AV , and EV denotes respectively constraint, Service-Based Application, actual value and expected value.

The constraint model, used by the PAEAN framework during the execution of each service, infers whether the business transaction is potentially in risk. The risk-calculator of the framework uses the following equation.

- R $(A_{SBA}) = (\sum_{i=1}^{k} T(AV)) > T_i(EV)$ where, R denotes the risk. From this equation, we can say that if the actual execution time is more than expected execution time then the business transaction is potentially in risk of violating the global SLA.

If the risk-calculator alarms a potential risk, which could be caused by a failure or a delay of a service execution, the framework exploits an adaptation mechanism to mitigate the risk. The adaptation mechanism is discussed in the following.

- ***Adaptation Mechanism***

The adaptation mechanism is another main component of our framework. It performs the tasks, as in (i-iv):

(i) ***Estimation of the Remaining Time of Execution***: the component computes the remaining time after the occurrence of a failure or a delay. The component uses the equation below:

$$E(T_R) = T - T_{elapsed} - (E(T_x) + E(T_y) + E(T_z))$$

T presents the total processing time for a composite service, $T_{elapsed}$ is the time already elapsed for failed or delayed services, and $E(T_x)$, $E(T_y)$, $E(T_z)$ denotes respectively the estimated time for service discovery, service rescheduling and service replacement. $T_{elapsed}$ is computed by using the following formula: $\sum_{i=1}^{k} T(AV)$, which is the time consumed by the services that have completed their execution.

(ii) ***Service Discovery***: a new service(s) needs to in two cases: (i) a running service fails to execute or (ii) a running service take more time to complete. The framework finds a new service that is functionally equivalent to the current failed one.

(iii) ***Service Rescheduling***: the component reschedules newly discovered services.

(iv) ***Service Replacement***: the adaptation component replaces the failed service by the newly discovered one.

The exact point from where the adaption should start (nodes it should include) depends on the position of the node-point where the deviation or failure is found. We described the adaptation using an example shown in Fig 3.

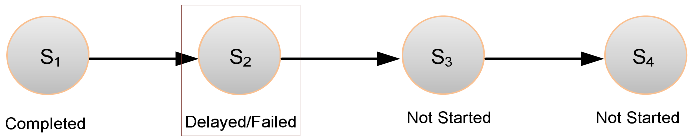


Fig. 3. A workflow model composed of services and their status of execution

Fig. 3 shows a running workflow model composed of services that have different states, entailing: completed, delayed, failed, and not started. If a service in the workflow model fails or delays, the adaptation mechanism initiates the adaptation action. In Fig. 3, the state of service S_2 has shown delayed or failed. We use two states of the same service to explain two cases. If the service has failed then the adaptation starting point is S_2 itself and it includes S_3 and S_4 within adaptation scope. All these services should be rescheduled and replaced by the newly discovered services. If the service has completed with delay then the adaptation start from *nearest next point* (NNP) which is S_3 and includes the subsequent service node-point S_4 within the adaptation scope.

In our framework, the adaptation process does not begin unless a failure or a delay occurs. This demonstrates the reactive nature of the framework. Conversely, adapting new services in advance is the proactive behavior the framework.

However, to automate the risk-mitigation process, we develop an algorithm. Eliminating the manual intervention in risk-mitigation is our concern in this research. The PAEAN framework underlies the algorithm shown in Fig. 4 that automate the mitigation of risks in business transaction in SBAs.

Algorithm:

Input:

(i) composite services S expressed as: S= {s1, s2,….,sn} partitioned into S1, S2 and S3 such that: $S = S1 \cup S2 \cup S3 \neq Ø$ *and* $S1 \cap S2 \cap S3 = Ø$

(ii) $C\,(ASBA) = \sum_{i=1}^{k} T(EV) \geq T(AV))$

Output: *All services executed within global SLA, shown by the condition S3= Ø*

Global variable:

R ← repository of services
S1 ← services that have been completed – at the beginning this set is empty
S2 ← services that are currently under execution
S3 ←services that have not started yet – at the beginning this set contains all services
NNP ← nearest next point
T (EV) ← QoS constraint on processing time metric
If $S_3 \neq Ø$
For *i=1* ***to*** *n* ***do***
S_i *← call (R,*S_i*); invoke service*
If S_i *'available'* ***then***
$S_2(i)$ *← put* S_i *services to executing*
$S_2(i) \leftarrow S_2+S_i$ *;*
$S_1(i)$ *← move* S_i *to* S_1
$S_1 \leftarrow S1 + S_i$
If $T_i\,(AV) \leq T_i\,(EV)$
No adaptation needed
S_{i+1}*← call (R,*S_{i+1}*);*
invoke *next service*
Else
{
Adaptation starting point
NNP ←S3
NNP ← needs to be adapted
Estimate *the remaining time of execution*
$E(T_R) = T - T(AV)_{elapsed} - (E(T_x) + E(T_y) + E(T_z))$
For each *Sj Є NNP* ***do***
{
Sj ← find call (R,S_j*)*
$E(T_x)$ *← time for service discovery*
$E(T_y)$ *←time for service* ***rescheduling***
$E(T_z)$ *← time for service* ***replacement***
If $(E(T_x) + E(T_y) + E(T_z)) \leq T_j\,(EV)$
Sj ← ***invoke*** *service*
Else
Sj ← ***reschedule*** *again*
} // end of for loop for adapting

If S_i *'executed' with success* ***then***
{
$T_i\,(AV)$ *←* ***compute*** *the actual processing time*
NNP
} // end of if condition of Qos constraint
S_{i+1}*← continue with then next service*
}// end of first branch of checking service execution status
Else if *Si 'not available' or 'failed'* ***then***
{
NNP ←Si $\cup$ *S3*
NNP← ***needs to be adapted***
estimate the remaining time of execution
$E(T_R) = T - T(AV)_{elapsed} - (E(T_x) + E(T_y) + E(T_z))$
For each *pj Є Ssj* ***do***
{
Sj← ***reschedule*** *service*
$E(T_x)$ *← time for service discovery*
$E(T_y)$ *←time for service rescheduling*
$E(T_z)$ *← time for service replacement*
If $(E(T_x) + E(T_y) + E(T_z)) \leq T_j\,(EV)$
Sj ← ***invoke*** *service*
Else
Sj ← ***reschedule*** *again*
} // end of loop for adapting NNP
} // end of else branch when service is not available
}// end of composite service execution

Fig. 4. Algorithm for risk-mitigation in business transaction

3.4 The Architecture of PAEAN

This section describes the architecture of our PAEAN framework. PAEAN has four main components: a real-time QoS monitoring component - CEP engine, an automatic risk-computation mechanism, an adaptation mechanism (a java component), and a Service Repository. The coordinated interaction among these components enables PAEAN to carry out action proactively in order to avoid SLA violation that causes business transaction abortion. As it can be noticed in Fig. 5, the composite service is the input of the engine that will monitor the execution process. Every service execution is triggered by the event that signals the start point of the service. The execution process is then traced systematically by the real-time QoS Monitoring component, which keeps information about the start and end point of execution.

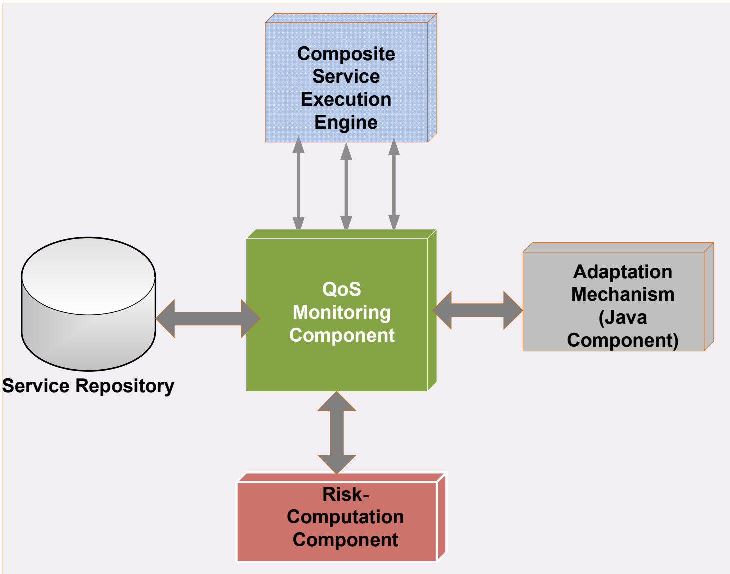


Fig. 5. PAEAN Architecture

Moreover, it keeps information about the QoS metrics of the service. On the base of this information, it is possible to be computed by the *automatic risk-computation mechanism,* the risk related to the QoS constraints. The local SLA violations inferred are then treated by the *Adaptation mechanism*, which is a java component that performs adequate adaptation action. The adaptation action could consist of service substitution, service re-execution, service compensation, or remain execution plan re-orchestration. In all these cases, the *Service Repository* deposit is used to filter and select new services based on the business transaction composition and QoS constraints.

4 Experimental Evaluations

An experiment was conducted with the purpose to test the approach presented in this research. We chose CEP – Complex Event Processing - as an implementation technology. The CEP engine was running in a PC with a configuration of 2 processors AMD Opteron(tm) Processor 6128 2,15GHz and 4.0 GB RAM.

The engine runs continuous computational queries (CCQ) on a stream of data, which contain information about the service execution. Thus, *suspicious* events are detected as they occur. This characteristic provides a real-time environment, which complies with the requirement of monitoring the global QoS constraints of a business transaction at run-time. The CEP mechanism used for the detection of the events is the event-pattern-detection. After an event is detected, the engine takes different actions, such as generating other events in order to make adequate adaptations of the workflow and assure the successful completion of the workflow within the agreed-upon global SLA.

For example, in the context of our approach, some event patterns that are useful to be identify would be:

- The arrival of a request for the 'Payment' service while the 'Shipment' service has not ended;
- The arrival of a message that the 'Place Order' service has completed with a higher cost than originally agreed;
- The arrival of a message that the 'Shipment' took more time to complete;
- The arrival of a message showing that the 'Payment' service is not available.

In order to give a comprehensive understanding of the tool we developed, we found it more appropriate and meaningful to show the result by a video demonstration rather than by a textual description of the tool. The video shows better all the details, characteristics, and behaviors of the tool during the whole process, from the start until the end of the business transaction execution. It shows how the events threatening the global SLA are detected for later triggering proactive adaptation actions.
A demonstration of our experimental results can be seen at:
(http://dl.dropbox.com/u/13869335/demo.avi).

5 Related Works

The increasing complexity of the business transactions results in a higher potential risk (SLA violation) associating the business transaction execution. Monitoring business transaction enables a system to avoid such risks [10][11][12]. We deeply review the current solutions that should not only be able to monitor business transaction but also to repair them if necessary. Such review was of a critical importance while trying to build our own framework. Framework such as [16] monitors and analyzes performance metric of a process and its QoS metric is not adequate to serve the purpose of this research. These frameworks only monitor and generate report that is the outcome of analysis. Some of the current business transaction monitoring and managing solutions, such as [5], [15] can monitor business transaction and report faults (e.g., operation failure). A solution named *OpTier* reported in [6] can repair faults *only* after it happens. The shortcomings of this solution are as follows: (i) it relies on reactive method and (ii) repairs only system fault (e.g., software crash, server outage, network failure).

During our study, we focused on different composition technologies used for constructing SBAs that carry out complex business transactions. From the research emerged that the current composition technologies lack the ability to mitigate the risks involved in a business. Some of these technologies, such as C-BPEL[7], and SELF-SERV platform [13] do consider local SLA, and other technologies, such as SCMP [4], Optimization problem [2], and QoS broker [17] consider global SLA but they do not monitor business transaction. For example, BPEL [9] focus on message and control flow rather than on business object and QoS constraints.

The lack of current solutions in efficient monitoring and repairing business transaction has inspired this research. It is important to mention that this research has influenced from [14].

6 Conclusions

The approach presented in this paper offers a structure for managing and controlling the QoS of a business transaction (carried out by a Service-Based Application) at run-time. There are different unexpected events that could occur during the execution of a business transaction that could threaten the fulfillment of the QoS metrics specified in SLA. This sets out the need for a solution to deal with this issue. The framework proposed in this paper monitors business transactions, computes potential risks, and performs proactive adaptation actions in order to prevent the possible risks of violating global SLA.

In terms of results, we identify the actual possible cases of SLA violation during run-time and provide an approach for mitigating them by substituting services that could have failed or triggering changes of the composite services in terms of its compounding components.

The limitation of PAEAN framework is that it in its current version cannot resist the violation of local SLAs. Extending the functionality of the framework in terms of local SLA violation prevention is our future work. Additionally, the proposed framework has evaluated for a small process but not for a process for instance, the value chain process that may pose problems. Mitigating the risks efficiently in value chain process would be challenging for the framework. Our plan is to use the framework for end-to-end business processes.

References

[1] Ardagna, D., Comuzzi, M., Mussi, E., Pernici, B., Plebani, P.: PAWS: A framework for executing adaptive web-service processes. IEEE Software 24(6), 39–46 (2007)

[2] Benatallah, B., Dumas, M., Sheng, Q.Z.: The self-serv environment for Web services composition. IEEE Computer 7(1), 40–48 (2003)

[3] Carroll, N., Haque, R., Whelan, E., Richardson, I.: Modelling Business Transactions across Service Supply Chain Networks. In: 20th International Conference on Information Systems Development, ISD 2011 (2011)
[4] Chi, Y., Lee, H.: A formal modeling platform for composing web services. Expert Systems with Applications 34, 1500–1507 (2008)
[5] DynaTrace, http://www.dynatrace.com/en/
[6] Garbini, J.P.: Competitive Analysis: Application Performance Management and Business Transaction Management, pp. 18–25 (September 2010)
[7] Ghedira, C., Mezni, H.: Through Personalized Web Service Composition Specification: From BPEL to C-BPEL. Electronic Notes in Theoretical Computer Science 146, 117–132 (2006)
[8] Harder, T., Reuter, A.: Principle of transaction oriented database recovery. ACM Computing Survey 15(4), 287–317 (1983)
[9] OASIS. Ws-bpel 2.0 (2006), http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html
[10] Papazoglou, P.M.: Making Business Processes Compliant to Standards and Regulations. In: International Enterprise Distributed Object Computing Conference(EDOC), pp. 141–150 (2009)
[11] Papazoglou, M.P., Kratz, B.: Web services technology in support of business transactions. Service Oriented Computing and Applications 1(1), 51–63 (2007)
[12] Papazoglou, M.P., Traverso, P., Dustdar, S., Leymann, F.: Service-Oriented Computing: State of the Art and Research Challenges. IEEE Computer 40(11), 38–45 (2007)
[13] Sheng, Q.Z., Benatallah, B., Dumas, M., Mak, E.: SELF-SERV: A Platform for RapidComposition of Web services in a Peer-to-Peer Environment. In: Proc. of the 28th Very LargeDataBase Conference (VLDB 2002), Hong Kong, China (2002)
[14] Taher, Y., Parkin, M., Papazoglou, M.P., van den Heuvel, W.-J.: Adaptation of Web Service Interactions Using Complex Event Processing Patterns. In: Kappel, G., Maamar, Z., Motahari-Nezhad, H.R. (eds.) ICSOC 2011. LNCS, vol. 7084, pp. 601–609. Springer, Heidelberg (2011)
[15] WebMthods. Business Activity Monitoring- the Face of Monitoring (June 2006), http://www.business.unr.edu/faculty/kuechler/788/BAM-The_New_Face_of_BPM.pdf
[16] Wetzstein, B., Leitner, P., Rosenberg, F., Dustdar, S., Leymann, F.: Monitoring and Analyzing influential factors of business process performance. In: International Enterprise Distributed Object Computing Conference(EDOC), pp. 141–150 (2009)
[17] Yu, T., Lin, K.: Service selection algorithms for Web services with end-to-end QoS constraints. ISeB 3, 103–126 (2005), doi:10.1007/s10257-005-0052-z

Mapping and Integration of Dimensional Attributes Using Clustering Techniques

Francesco Guerra, Marius-Octavian Olaru, and Maurizio Vincini

Department of Information Engineering
University of Modena and Reggio Emilia
Via Vignolese 905, 41025 Modena, Italy
firstname.lastname@unimore.it
http://www.dii.unimore.it

Abstract. Following recent trends in Data Warehousing, companies realized that there is a great potential in combining their information repositories to obtain a broader view of the economical market. Unfortunately, even though Data Warehouse (DW) integration has been defined from a theoretical point of view, until now no complete, widely used methodology has been proposed to support the integration of the information coming from heterogeneous DWs. This paper deals with the *automatic* integration of dimensional attributes from heterogeneous DWs. A method relying on topological properties that similar dimensions maintain is proposed for discovering mappings of dimensions, and a technique based on clustering algorithms is introduced for integrating the data associated to the dimensions.

Keywords: Multidimensional Schema Mapping, Dimension integration.

1 Introduction

During the last two decades, Data Warehousing has been the main Business Intelligence (BI) instrument for the analysis of large banks of operational data. It allows managers to take strategic decisions based on highly valuable aggregated information synthesized from the operational data. The classic approach has been to create the enterprise DW from the company's operational repository through a process called *Extract-Transform-Load* (ETL). The approach allows managers to have access only to the information of their own company.

In recent years, however, companies realized that a higher potential can be obtained from the DW, by combining information coming from one or more companies. For example, it is now common for two or more companies to decide to merge, or to work together in a *federation-like* environment. In both cases, the DWs of the independent companies have to be combined in order to allow participants to share not only their own local information repository, bot also the entire available information. The widest used approach, is to extract the data from the repositories of all the participants through complex ETL processes, and then to rebuild a new DW from the unified data repository. However, this

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 38–49, 2012.

solution requires an enormous amount of work and it usually has high costs and long development times. This approach may be considered a *low-end* solution, as the actual *integration* is being made at the early stages of the DW building procedure. A more effective solution would be to make a *high-end* integration of the DWs, which means integrating the multidimensional information contained in the final DWs.

The method we propose is able to combine information from dimensions of two or more different DWs by (1)generating a set of consistent mappings that express the semantic similarity between various dimension levels, and (2) importing remote instance information, compatible with the local multidimensional schema. This allows the interrogation of all the compatible DWs in the same, uniform way.

The rest of this paper is organized as follows: Section 2 provides an overview on related work; Section 3 presents the schema mapping generation step of our method, while Section 4 describes the schema importation procedure. Finally, Section 5 draws the conclusions of our work.

2 Related Work

One of the first attempts for DW integration is mentioned in [1], where the authors propose a design methodology for *conformed dimensions* (dimensions that share consistent dimension keys, consistent column names, consistent attribute definitions and consistent attribute values), although such solution may apply only to cases where the intent to integrate heterogeneous DW information (or to perform *drill-across* queries) is clear from the beginning.

A systematic solution is proposed in [2], where the possible conflicts are identified and a strategy for the design of DW integration techniques, divided in 4 steps, is proposed: (1)identification of the facts that can be integrated and the dimensions of these facts that can be combined to perform the integration, (2)resolution of conflicts between common dimensions, (3)resolution of conflicts between facts to be integrated, (4)reconciliation and integration of dimensions and facts according to the desired level of interoperability.

In [3], the authors define an automatic mapping generation technique based on *similarity* functions and on earlier work in data integration[4,5,6,7]. The solution lacks, however, a data integration strategy and it fails to exploit the discovered mappings for integration purposes.

In [8], the authors introduce a *Dimension Algebra* (DA) that describes operations such as *selection*, *projection* and *aggregation* that can be computed on a dimension. The DA is then used to define *compatible dimensions* and to express similarities between two heterogeneous dimensions. After formalizing properties that mappings can have, the authors define the *intersection* between two dimensions, as a dimension that can be computed from either one of two dimensions using a DA expression. The paper provides a solid theoretical ground on which to build dimension integration methodologies, however it does not provide a method to automatically compute the DA expressions required for the integration purposes.

3 Mapping Dimensions of Different DWs

3.1 Motivating Scenario

Consider, for example, the dimensions depicted in Fig. 1. Suppose the first schema (S_1) contains the *REVENUE* fact table of a DW (DW_1), and that the second schema (S_2) contains the *SALE* fact schema from another DW (DW_2), while the third schema (S_3) contains the *warehouse* information. Combining the information from the three repositories offers users the possibility of executing *drill-accross* queries. Nevertheless, this possibility depends on the compatibility of the schemas and instances. For example, a user may require the total revenue from the DWs, divided by *city* and *month*, or the total revenue obtained for a specific item that was available in the warehouse, divided by *month* and *region*. Note that the first query provides an exact answer, as the information is available at the required granularity level, while the second is impossible, due to the incompatibility of the schemas (in particular, of the dimensions). These inconsistencies can be removed by uniforming the dimensions of the given schemas. For this purpose, we propose a method that combines a *cardinality*-based mapping generation technique and an instance integration method based on *clustering* and *relevant* values generation[9].

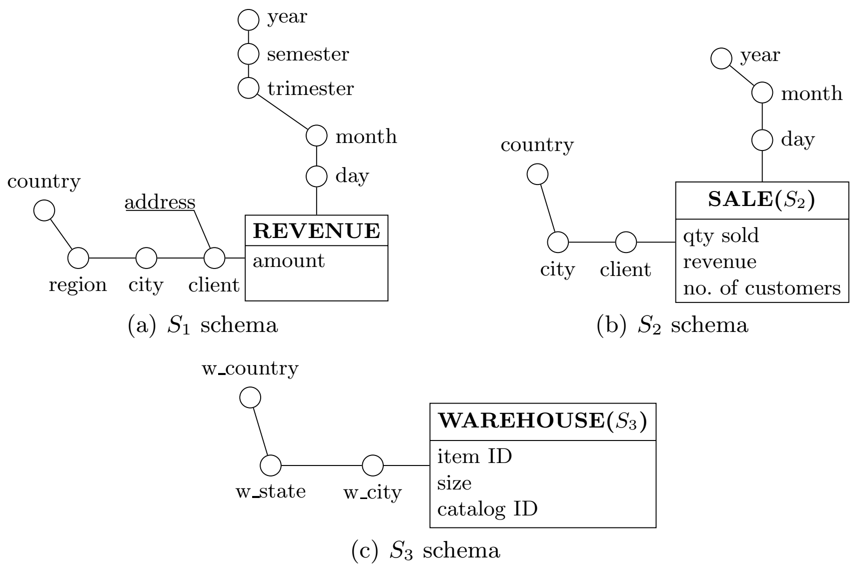


Fig. 1. Example

3.2 Detection of Semantically Similar Nodes

The first step of our method consists in identifying similar dimension levels inside different dimension hierarchies. For this purpose, classical data-integration approaches, like the ones based on semantics and similarity measure functions (see [10,11] for an example), may be used. However, even if this way it is possible to find semantically similar elements, we believe that such approach would not be able to provide the required accuracy in a multidimensional environment. The main reason is that two similar, but different dimensions, may contain related information (like a *time* hierarchy), but structured differently. This may lead to inconsistent analysis capabilities.

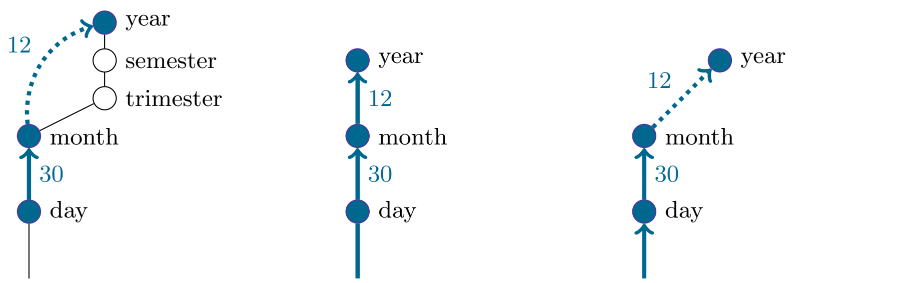
In order to automatically detect these *equivalences*, the method we propose makes use of *cardinality*-based properties. In fact, dimensions usually maintain a common characteristics, which is a *tree-like* structure (also called *quasi-tree* in [12]) imposed by the partial order relationship on the dimensional attributes set. This property is usually maintained when dimensions represent a concept of the real world with a common, well-known, structure. Consider, for example, that the time dimension of schema S_1 contains all the *days* from January 1^{st} 2007 to December 31^{st} 2010 (4 complete years), and that the time dimension of the schema S_2 contains all the *days* from January 1^{st} 2009 to December 31^{st} 2011 (3 complete years). Although the sets of the values of the attributes are only partially overlapped, this information may not be sufficient to discover that a certain dimension level from the first dimension hierarchy is semantically equivalent to another dimension level in the second hierarchy.

We rely, instead, on another property of dimension hierarchies, the *cardinality-ratio*, which is simply the ratio among the number of different elements between two dimension hierarchy levels. For example, in the *time* dimension in schema S_1, every element of the *month* level is an aggregation of roughly 30 different elements of the *day* level. Although it covers a different time period, the same property can be observed in the second *time* dimension. This information is maintained not only between directly connected dimension hierarchies. For example, in the schema S_2 a *year* is an aggregation of 12 different *months*. In schema S_1, a *year* is composed of 2 *semesters*, every *semester* is composed of 2 *trimesters*, where each *trimester* is composed of 3 different *months*. This means that a *year* is an aggregation of $2 \times 2 \times 3$ different months, which is the same information that is directly available in the other hierarchy.

This property may be used not only on time dimensions, but on all dimensions that represent a concept of the real world with a fixed structure. For example, the geographical distribution inside one country is similar among all DWs: an address contains a city, cities are organized in comunes, grouped into regions, and so on.

By using this property, we can consider the dimension hierarchies as directed labeled graphs, where the label of each edge is the cardinality-ratio among different elements. Figure 2(a) is a directed labeled graph that represents the *time* dimension of the first schema (S_1), while Fig. 2(b) represents the dimension of the second schema (S_2). Starting from these two graphs, it is possible to compute a common subgraph that can be used for identifying pairs of equivalent nodes in the initial graphs (Fig. 2(c)).

For the sake of simplicity, we do not develop the first step of the method in it's finest details, as we believe that this is trivial, for example by making use of connectivity matrices and matrix theory.



(a) First Graph (b) Second Graph (c) Common Subgraph

Fig. 2. Dimension Graphs

The sub-graph is obtained from one of the initial graphs, by eliminating one or more nodes (and the incoming and outgoing edges), but maintaining all the paths of which the removed edges were part of. For example, the common sub-graph may be obtained from the first graph by eliminating the nodes *trimester* and *semester*, and by adding the directed labeled edge $(month, year)$, with the appropriate label, computed as the product of the labels of the edges that the path was composed of.

The common sub-graph can be used to map elements of the initial graphs. For example, the node *day* of the common sub-graph is obtained from the node *day* of the first graph, or from the node with the same name of the second dimension. We can thus state that nodes $S_1.day$ (for identification purposes, we will refer to it by using the name of the dimensional attribute it represents) is *equivalent* to node $S_2.day$. Following the same approach, $S_1.month$ is *equivalent* to $S_2.month$ and $S_1.year$ is *equivalent* to $S_2.year$[1].

3.3 Mapping Set Generation

In order to express the complex relationships between various dimension levels, we will make use of a subset of the mapping predicates proposed in [13], in particular, we used the following mapping predicates:

– **equi-level** predicate: used to state that two attributes in two different md-schemas have the same granularity and meaning;

[1] For simplicity reasons, we made use of similar or identical names for the labels of the attributes, although in real cases the attribute labels are different or are likely to be contain or to be composed of abbreviations and/or acronyms.

- **roll-up** predicate: used to indicate that an attribute (or set of attributes) of the first md-schema aggregates an attribute (or set of attributes) of the second md-schema;
- **drill-down** predicate: used to indicate that an attribute (or set of attributes) of the first md-schema disaggregates an attribute (or set of attributes) of the second md-schema;
- **related** predicate: indicates that between two attributes there is a $many - to - many$ relation;

The *cardinality-ratio* concept expressed earlier is useful not only for finding pairs of equivalent nodes, but also to express other types of relationships among dimension levels. For example, we computed that $S_1.month$ is equivalent to $S_2.month$, which means that they represent the same concept of the real world, at the same aggregation level. In the first dimension, S_1, a *trimester* is an aggregation (or *roll-up*) of *months*. Being $S_1.month$ equivalent to $S_2.month$, we can infer that $S_1.trimester$ is also an aggregation of $S_2.month$. Similar rules may be used to discover other types of relationships among attributes.

To generate the complete mapping set, we make use of the following inference rules:

Let P_x and P_y be two nodes of the first dimension such that there is a path from P_x to P_y, and P_h and P_k two nodes of the second dimension such that there is a path from P_h and P_k

1. **Rule 1**: If P_x and P_h are equivalent, add the mappings:
   * $P_x\ (equi - level)\ P_h$
2. **Rule 2**: if $P_x\ (equi - level)\ P_h$, add the mappings (see Figure 3(a)):
   * $P_y\ (roll - up)\ P_h$
   * $P_h\ (drill - down)\ P_y$
3. **Rule 3**: if $P_y\ (equi - level)\ P_h$, add the mappings (see Figure 3(b)):
   * $P_x\ (drill - down)\ P_h$
   * $P_h\ (roll - down)\ P_x$
4. **Rule 4**: if $P_y\ (equi - level)\ P_h$, add the mappings (see Figure 3(c)):
   * $P_x\ (drill - down)\ P_k$
   * $P_k\ (roll - up)\ P_x$
5. **Rule 5**: for every nodes P_x and P_h of the two graphs for which there has not been found any mapping rule, add the mapping:
   * $P_x\ (related)\ P_h$

The rules allow the generation of the complete mapping list (for the sake of simplicity, for the Rules 2&3 we enumerate only the first generated mapping; Rules 4&5 produced redundant mappings):

- **Rule 1**:
  - ω_1 : $S_1.day$ *<equi-level>* $S_2.day$
  - ω_2 : $S_1.month$ *<equi-level>* $S_2.month$
  - ω_3 : $S_1.year$ *<equi-level>* $S_2.year$

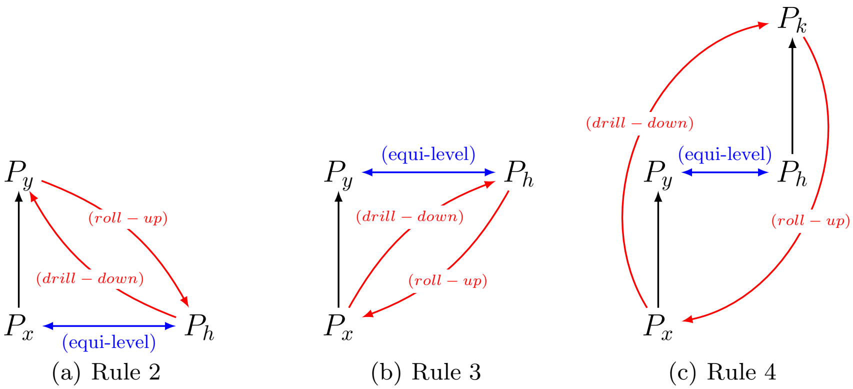


Fig. 3. Graphical representation of Rules 2, 3 and 4

- **Rule 2**:
  - $\omega_4 : S_1.month <roll\text{-}up> S_2.day$
  - $\omega_5 : S_1.trimester <roll\text{-}up> S_2.day$
  - $\omega_6 : S_1.trimester <roll\text{-}up> S_2.month$
  - $\omega_7 : S_1.semester <roll\text{-}up> S_2.day$
  - $\omega_8 : S_1.semester <roll\text{-}up> S_2.month$
  - $\omega_9 : S_1.year <roll\text{-}up> S_2.day$
  - $\omega_{10} : S_1.year <roll\text{-}up> S_2.month$
- **Rule 3**:
  - $\omega_{11} : S_1.day <drill\text{-}down> S_2.month$
  - $\omega_{12} : S_1.day <drill\text{-}down> S_2.year$
  - $\omega_{13} : S_1.month <drill\text{-}down> S_2.year$
  - $\omega_{14} : S_1.trimester <drill\text{-}down> S_2.year$
  - $\omega_{15} : S_1.semester <drill\text{-}down> S_2.year$

4 Exploiting Mappings for Querying Different DWs

The discovered mappings can be exploited for rewriting queries, formulated according to the dimensions of a specific DW, into queries executable in other DWs having different dimensions. Only few approaches have been developed in the literature for dealing with this issue. Among them, in [14], a Peer-to-Peer network of DWs, where the user can formulate a query on a local peer and the query is automatically rewritten on remote peers, is proposed. The approach integrates information contained in the remote peers with the one contained in the local peer. This way, the user receives an answer that includes data from all the peers in the network. Obviously, the accuracy of the approach in rewriting queries depends on the compatibility of the schemas that compose the network, i.e. the fact that the DWs contain information related to the same topic and described with dimensions having similar level of granularity.

The situation complicates even more when the network is composed of a larger number of peers, as for each query, in some cases a user may obtain only information coming from the local peer, in other cases information coming from all (or a subset of) the network. To address this kind of problem, a collaborative approach not only may be useless, but may also become harmful and misleading.

The method proposed in this paper deals with this issue by presenting a two-step integration process: in the first step, mappings are exploited for comparing the granularities of the dimensions and identifying and importing the (possible) missing levels, at schema level; in the second step, mappings and heuristic rules are exploited for populating the values of the new introduced dimensions.

4.1 Importing Dimensions

Mappings between the dimensions are exploited for discovering and, in case, importing, when possible, "compatible parts" of remote DW dimensions. In particular, the presence of at least one <*equi-level*> mapping implies that the involved dimensions share common information, and, consequently, their schemas are (partially) overlapping. Such overlapping schema is used as a starting point for importing other dimensional attributes. In particular, the attributes are firstly inserted as optional attributes, and then, modified to mandatory attributes, according to the actual data in the dimensions.

To formalize this step, we propose to use the *Dimensional Fact Model* (DFM)[12], which descrbes a fact schema as a sextuple $f = (M, A, N, R, O, S)$, where:

- M is a set of *measures* defined by a numeric or Boolean value
- A is a set of *dimensional attributes*
- N is a set of *non-dimensional attributes*
- R is a set of ordered couples that define the $quasi - tree$ representing the dimension hierarchy
- $O \subseteq R$ is a set of *optional* relationships.
- S is a set of aggregation statements

The first step of the schema importation procedure is based on the following rule:

Rule 1. *Given two fact schemas* $f_1 = (M', A', N', R', O', S')$ *and* $f_2 = (M'', A'', N'', R'', O'', S'')$, *and* $\mathcal{M}$ *the set of mappings generated by the first step, let* $a_i, a_j \in A'$ *such that* $(a_i, a_j) \in R'$, *and* $b_i \in A''$. *If* $\{(a_i < equi-level > b_i)\} \subseteq \mathcal{M}$, *then:*

$$A'' := A'' \cup \{a_j\} \tag{1}$$
$$O'' := O'' \cup \{(b_i, a_j)\}$$

$\forall\, b_k \in A'', b_k \neq b_i$ *such that* $\{b_k < roll{-}up > a_j\} \in \mathcal{M}$ *(or* $\{b_k < drill{-}down > a_j\} \in \mathcal{M}$*), then:*

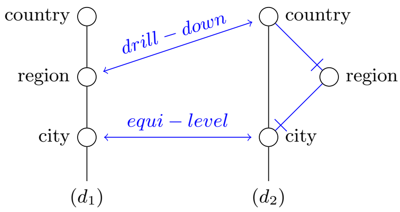


Fig. 4. Graphical example of the importation rule

$$O'' := O'' \cup \{(a_j, b_k)\} \quad or \qquad (2)$$
$$O'' := O'' \cup \{(a_j, b_k)\}$$

Example. Let us obtain the following mappings by exploiting the method proposed in Section 3:

– $\omega_1 : S_1.city$ <equi-level>$S_2.city$
– $\omega_2 : S_1.country$ <equi-level>$S_2.country$
– $\omega_3 : S_1.region$ <roll-up>$S_2.city$
– $\omega_4 : S_1.region$ <drill-down>$S_2.country$

Figure 4 contains a graphical example of the importation Rule[2]. Dimensions d_1 and d_2 are the corresponding time dimensions of DW_1 and DW_2 as defined in Fig. 1. As $region, city \in A'$ and $(city, region) \in R'$ and $city \in A''$, and $\{(S_1.city < equi - level > S_2.city)\} \in \mathcal{M}$, then, according to (1), the attribute *region* is inserted among the attributes of S_2 (more precisely, in the set A'') and the ordered tuple $(city, region)$ is inserted in O'' (because it is an optional attribute). Subsequently, as $\{(S_1.region < drill - down > S_2.country) \in \mathcal{M}$, according to (2), the order couple $(region, country)$ is inserted into O''.

4.2 Importing Values with RELEVANT

One important step in the DW integration procedure is the integration and reconciliation of common information. This implies that: (1) the newly imported attributes have to be populated with consistent values and (2) the possible inconsistencies among attribute values must be resolved. For this purpose, we propose an extension of RELEVANT (RELEvant VAlues geNeraTor)[9] that is specially conceived for the integration of multidimensional information. RELEVANT performs clustering of attribute values and, for each cluster, identifies one value, the "relevant value", that is the representative value for the whole cluster. By applying the RELEVANT techniques to the values of the dimensions, we obtain clusters of related dimensions. The "relevant value" provided by each cluster is then used for populating the missing values. In this way, RELEVANT is able to provide "approximate values" to the new dimension attributes.

[2] Only the relevant mappings have been drawn.

Table 1. Dimension values

client	city	region	country	dimension
M.ROSSI	FIRENZE	TOSCANA	ITALY	d1
P.BIANCHI	ROMA	LAZIO	ITALY	d1
A.RENZO	BOLOGNA	E.ROMAGNA	ITALY	d1
A.MANCINO	MODENA	E.ROMAGNA	ITALY	d1
S.RUSSO	MILANO	LOMBARDIA	ITALY	d1
P.ESPOSITO	MILAN	*NULL*	ITALY	d2
S.ROMANO	TURIN	*NULL*	ITALY	d2
A.COLOMBO	ROME	*NULL*	ITALY	d2
B.RICCI	PARMA	*NULL*	ITALY	d2
F.MARINO	PALERMO	*NULL*	ITALY	d2
F.GRECO	ROME	LAZIO	ITALY	d3
M.GALLO	FLORENCE	TUSCANY	ITALY	d3
D.LOMBARDI	MODENA	EM.ROMAGNA	ITALY	d3
R.MORETTI	PALERMO	SICILY	ITALY	d3
T.CONTI	TURIN	PIEDMONT	ITALY	d3

Clusters of related elements are computed by using some similarity measures. In this extension of RELEVANT, clusters are created by means of two similarity measures: 1. syntactic similarity, which compares the alphabets used for describing the attribute values; 2. containment, which measures the closeness of attributes belonging to different dimension. In particular, the containment is based on the $< roll - up >$, $< drill - down >$, and $< equi - level >$ mappings holding among the sources. Some further semantic measures based on lexical similarity and external ontologies can be exploited for dimensions belonging to specific domains, but this is outside the scope of the paper.

Example. Let us consider that the geographical dimensions shown in Figure 1 are populated with the values in Table 1. The application of RELEVANT generates clusters of related values. For each cluster, the relevant value is the approximate values that we can exploit for populating missing values. Note

Table 2. Main Clusters

city	region	assigned value
FIRENZE\FLORENCE	TOSCANA\TUSCANY	TOSCANA
ROMA\ROME	LAZIO*NULL*	LAZIO
BOLOGNA	E.ROMAGNA	E.ROMAGNA
MODENA\MODENA	E.ROMAGNA\EM.ROMAGNA	E.ROMAGNA
MILANO\MILAN	LOMBARDIA*NULL*	LOMBARDIA
TURIN\TURIN	*NULL*\PIEDMONT	PIEDMONT
PARMA	*NULL*	NULL
PALERMO	*NULL*\SICILY	SICILY

that in some cases the relevant values can be exploited for identifying similar values in different dimensions, e.g., "Milan" and "Milano" that are two different names of the same city. (see Table 2).

5 Conclusion

The work presented in this paper describes a method for the mapping and integration of heterogeneous DW dimensions. We showed how it is possible to use dimension topological properties to find sets of mappings among heterogeneous Data Warehouse dimensions, that we then used to insert remote dimensional attributes and to populate them with the available data. Overall, the proposed method reiterates the idea that although Data Warehouse integration may be seen as a special case of classical data integration, specific techniques are required for effective results.

We believe that the proposed method could fully show its usefulness in large environments, characterized by a dynamic set of DWs, where the continuous changes in the composition of the network makes the classical ETL approach impossible to apply. However, we believe that the method can be used in any case of DW integration, as it could imply a reduction of the costs and the development times.

References

1. Kimball, R., Ross, M.: The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, vol. 32. John Wiley & Sons, Inc., New York (2002)
2. Torlone, R.: Interoperability in Data Warehouses. In: Liu, L., Özsu, T. (eds.) Encyclopedia of Database Systems, pp. 1560–1564. Springer, Berlin (2009)
3. Banek, M., Vrdoljak, B., Tjoa, A.M., Skocir, Z.: Automated Integration of Heterogeneous Data Warehouse Schemas. IJDWM 4(4), 1–21 (2008)
4. Cabri, G., Guerra, F., Vincini, M., Bergamaschi, S., Leonardi, L., Zambonelli, F.: Momis: Exploiting Agents to Support Information Integration. Int. J. Cooperative Inf. Syst. 11(3), 293–314 (2002)
5. Bergamaschi, S., Gelati, G., Guerra, F., Vincini, M.: An Intelligent Data Integration Approach for Collaborative Project Management in Virtual Enterprises. World Wide Web 9(1), 35–61 (2005)
6. Madhavan, J., Bernstein, P.A., Rahm, E.: Generic Schema Matching with Cupid. In: Apers, P.M.G., Atzeni, P., Ceri, S., Paraboschi, S., Ramamohanarao, K., Snodgrass, R.T. (eds.) VLDB, pp. 49–58. Morgan Kaufmann (2001)
7. Melnik, S., Garcia-Molina, H., Rahm, E.: Similarity Flooding: A Versatile Graph Matching Algorithm and Its Application to Schema Matching. In: ICDE, pp. 117–128 (2002)
8. Cabibbo, L., Torlone, R.: On the Integration of Autonomous Data Marts. In: SSDBM, pp. 223–231. IEEE Computer Society (2004)
9. Bergamaschi, S., Sartori, C., Guerra, F., Orsini, M.: Extracting Relevant Attribute Values for Improved Search. IEEE Internet Computing 11(5), 26–35 (2007)
10. Bergamaschi, S., Bouquet, P., Giacomuzzi, D., Guerra, F., Po, L., Vincini, M.: An Incremental Method for the Lexical Annotation of Domain Ontologies. Int. J. Semantic Web Inf. Syst. 3(3), 57–80 (2007)

11. Beneventano, D., Bergamaschi, S., Guerra, F., Vincini, M.: The SEWASIE Network of Mediator Agents for Semantic Search. J. UCS 13(12), 1936–1969 (2007)
12. Golfarelli, M., Maio, D., Rizzi, S.: The Dimensional Fact Model: A Conceptual Model for Data Warehouses. Int. J. Cooperative Inf. Syst. 7(2-3), 215–247 (1998)
13. Golfarelli, M., Mandreoli, F., Penzo, W., Rizzi, S., Turricchia, E.: Towards OLAP query reformulation in Peer-to-Peer Data Warehousing. In: Song, I.Y., Ordonez, C. (eds.) DOLAP, pp. 37–44. ACM (2010)
14. Golfarelli, M., Mandreoli, F., Penzo, W., Rizzi, S., Turricchia, E.: OLAP Query Reformulation in Peer-to-Peer Data Warehousing. Information Systems (2011)

Service Offer Descriptions and Expressive Search Requests – Key Enablers of Late Service Binding

Maciej Zaremba[1], Tomas Vitvar[2], Sami Bhiri[1], Wassim Derguech[1], and Feng Gao[1]

[1] Digital Enterprise Research Institute
National University of Ireland in Galway, Ireland
firstname.lastname@deri.org
[2] Web Engineering Group
Faculty of Information Technologies
Czech Technical University in Prague
tomas@vitvar.com

Abstract. Late service binding is a long-term goal for Service Computing. It fully enables service loose coupling by allowing service consumers to dynamically identify services at runtime. We present models of expressive search requests and service descriptions enabling matchmaking of highly configurable services whose properties are request-dependent and dynamic. Both models are grounded in lightweight semantic formalisms of RDF and SPARQL, and use Linked Data. Our hierarchical service model provides a foundation for runtime generation of service offer descriptions, while the majority of service models and service matchmaking approaches do not sufficiently address service dynamicity aspects and operate on static service descriptions. We apply our approach to a shipping domain in order to show its feasibility for solving real world problems and its benefits for the late service binding.

Keywords: Service Computing, service modelling, late service binding, semantic services, service discovery, highly configurable services.

1 Introduction

Late service binding supports forming short-term business interactions between service consumers and provisioned services where references to services are not hard-coded in applications and business processes of service consumers [8]. In late service binding scenarios, details of the tasks to be performed by a set of services become available at runtime. Services discovered during late service binding are often available for a limited period of time. Late service binding is comprised of service matchmaking and service invocation. In this paper, we primarily focus on models of service descriptions and search requests enabling matchmaking in the late service binding. We understand services in business terms as intangible artefacts with valuable outcomes where technologies like Web services and RESTful services are examples of service access mechanisms. Business services are often individually tailored (configured) to the service requests by allowing service consumers to select service options and values of service properties.

Existing service descriptions are often incomplete and underspecified [9] due to: (1) dependency on service requests (e.g., service availability and throughput depends on

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 50–62, 2012.

a customer location), (2) changing business conditions (e.g., service price depends on resource availability, currency exchange rates, third-party factors), (3) privacy concerns (e.g., a health insurance price is only available on an individual request). Therefore, we distinguish between *service variants* and *service offers*. **Service variant** is a service reference model that can be customised for a particular service consumer and usage scenario and does not have to reflect an invocable service provided by a specific service provider. Service variants foster reuse where new service variants are created by reusing and modifying existing service variants. **Service offer** is the most specific service variant which cannot be further customised and often has been generated for an individual search request. An offer has a "well-defined" business meaning and it should include at least the following four conditions [12]: delivery date, price, terms of payment and fair description of an item or service. Service offers are quantifiable artifacts suitable for objective ranking and making services truly comparable. **Highly configurable service** [6] is a service variant highly dependent on a service request that provide a large number of distinctive service offers. Concrete values of service properties are available only on the service offer level and play a key role in service matchmaking. Our distinction between service variants and service offers is similar to: abstract and concrete services [9], Web services and services offers [5], and services and service configurations [6].

In this paper, we present models of service variant descriptions and search requests used in the matchmaking process. We propose Service Variant Hierarchy where service variant descriptions differ by the level of their specialisation. Service Variant Hierarchy promotes reuse and facilitates creation of new service descriptions. Our service model is suited for modelling highly configurable services whose properties are request-dependent and dynamic. Our service model uses RDF[1] as a lightweight language for describing and linking various entities, and SPARQL[2] as a language for expressing queries, rules and constraints. In our previous work on service offers [14,15], we proposed genetic algorithms in the matchmaking process that take advantages of the search request model flexibility for finding the best available services. In this paper, we contribute with a formalisation of the Service Variants Hierarchy and we elaborate on our service and search request models for the matchmaking phase of late service binding.

We apply our discovery approach to a shipping scenario where we match expressive search requests with real online services of shipping companies like FedEx, TNT, An-Post, USPS, and others. From a service consumer perspective, it is much more valuable to operate on concrete shipping offers rather than to operate on abstract service descriptions. Offers of shipping companies vary significantly and late service binding to the best of them, according to the shipping task at hand, offered price and delivery time, provides a significant business value.

In Section 2, we discuss early and late service binding. In Section 3, we present a shipping scenario and show the benefits of late service binding in this scenario. We introduce our models of search requests and service descriptions in Section 4. In Section 5, we shortly discuss service matchmaking and implementation. In Section 6, we describe related work. We conclude the paper and discuss future work in Section 7.

---

1 http://www.w3.org/RDF

2 http://www.w3.org/TR/rdf-sparql-query/

2 Service Binding

Services can be bound to applications and business processes activities either at design time (early binding) or at runtime (late binding [4]). As shown in Figure 1 we distinguish three kinds of stakeholders in our service binding process: service consumer, broker and service providers. Provided services are described in *service descriptions* and service consumers formulate their service requirements in *search requests*. Both service descriptions and search requests are communicated to the broker whose role is to provide the best match between a *search request* and available services. Service broker generates service offer descriptions (the most specific, quantifiable, often request-dependent service variants), and then filters and ranks them in the process of service matchmaking.

Currently, early binding is a prevalent style of service interactions. Services used in the early binding are prescribed at design time what leads to tight coupling. Typically, a long-term business relationship has been previously established and it is not required to switch between different implementations of functionally similar services.

A late binding service interaction style allows service consumers to dynamically determine and invoke services at runtime [4]. A consumer's application or business process can be specified as a set of activities to be performed by various services determined at runtime. Late binding facilitates switching between services offered by different service providers. Late service binding provides also a solution for self-healing service-based systems where alternative services are identified on-the-fly and used as replacements of original services which no longer satisfy consumer's requirements (e.g., incur higher cost or provide unreliable service). In order to realise late service binding, service matchmaking has to operate on quantifiable artifacts suitable for objective service ranking.

Our conceptual models of search requests and service descriptions in Service Variant Hierarchy presented in Section 4 provide a foundation for generating service offers. The service model is suited for modelling highly configurable services whose properties are

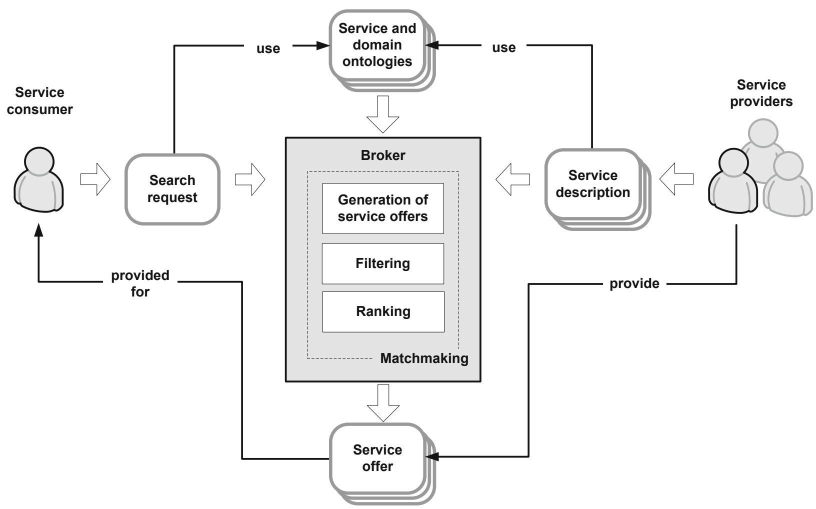


Fig. 1. Service bindings framework

request-dependent and dynamic. Generating service offers from highly configurable services is more challenging and relevant for late service binding scenarios. Service offers are often generated for a specific search request in the matchmaking phase and are often valid for a limited period of time. In order to use the best service offers determined in the matchmaking phase, service invocation phase should proceed shortly.

3 Application Example

We have chosen a shipping scenario as an application area to illustrate our models of service descriptions and search requests. Services offers of shipping companies have a highly dynamic and configurable character which depends on service requests. Early binding of shipping services is a limited business solution since required data (e.g., the package size, the source and target destinations, etc.) and service consumer preferences become available only prior to the execution of the shipping task. For example, a company selling its product to overseas customers, might want to dynamically switch between shippers such as: FedEx, USPS, AnPost, GLS, EcoParcel and TNT depending on the shipping destinations, requested insurance, etc. There is no *“one size fits all”* shipping company and the ability to easily switch to the optimal (cheapest, fastest, most reliable - a combination of these) shipping service provides a significant business value. Figure 2 illustrates late service binding scenario in a shipping domain.

Shipping services are highly configurable and their service offers (comprised of: availability, price, delivery time, etc.) depend on the number of required properties (the source and target destinations, package size and weight, required insurance, etc.). Shipping companies offer a great variability in their services. A service might offer an optimal solution for a given shipping task, while for another task it might be unavailable (e.g., due to the package size or unsupported target country) or offer an uncompetitive shipping option (more expensive, slow, unreliable). In our shipping scenario, there is a number of shipping details (e.g., source and target address, freight size, required delivery time, etc,.) that strongly influence the optimal choice of the shipping company. The capabilities of shippers vary significantly and the ability to late bind to their offerings provides a significant business value. In the remainder of this paper, we use the shipping scenario for illustrating our conceptual model.

Fig. 2. Late service binding in the shipping scenario

4 Conceptual Model

Our conceptual model provides the foundation for the matchmaking of highly configurable services [6] which is a core part of the late service binding process. Search request $\mathcal{R}$ expresses service consumers' needs. A service description $\mathcal{S}$ can be defined with a varying levels of specialisation. We assume that ontology $\mathcal{O}$ defines shared elements (e.g., domain specific concepts) between $\mathcal{R}$ and $\mathcal{S}$. We do assume that $\mathcal{R}$ and $\mathcal{S}$ refer to the same $\mathcal{O}$ – it is a valid assumption when well-established, standardised domain and service ontologies are used. $\mathcal{O}$, $\mathcal{R}$ and $\mathcal{S}$ are specified using lightweight semantics as a combination of RDF and SPARQL as a language for expressing queries, rules and constraints. For the brevity reasons, we do not present complete listings of domain ontology, search requests and service descriptions. In our examples (Listings 1.1, 1.2, and 1.3) we use RDF in a human-readable N3 notation and SPARQL focusing on the most relevant parts and omit variable bindings (e.g., *?srcAddress a so:SourceAddress; vcard:country ?srcCountry*).

4.1 Domain Ontology

We use domain ontology $\mathcal{O}$ as a shared conceptualisation between search requests and service descriptions. Information expressed in $\mathcal{R}$ and $\mathcal{S}$ such as properties and exchanged messages map to elements defined in the domain ontology. Examples of domain ontologies include open vocabularies defining data schema like SKOS[3] and vcard[4] and large data sets expressed using these open vocabularies like *dbpedia*[5]. We refer to *dbpedia* entities for example to identify geographical locations (e.g., Ireland in Listing 1.2 line 6).

Listing 1.1 shows a fragment of the shipping domain ontology example. In here, the shipping domain contains entities such as *SourceAddress, TargetAddress, Package, ShippingPrice*, and *ShippingDelivery*. We use *vcard* vocabulary for expressing addresses (line 7).

```
1  @prefix so: <http://usoa.deri.ie/ontology/shipping_domain#>.
2  @prefix vcard: <http://www.w3.org/2006/vcard/ns#>.
3
4  so:SourceAddress a rdfs:Class.
5  so:TargetAddress a rdfs:Class.
6  so:hasAddress a rdf:Property;
7   rdfs:domain so:SourceAddress, so:TargetAddress; rdfs:range vcard:VCard.
8
9  so:Package a rdfs:Class.
10 so:hasWeight a rdf:Property;
11  rdfs:domain so:Package; rdfs:range so:Weight.
12
13 so:ShippingPrice a rdfs:Class.
14 so:ShippingDelivery a rdfs:Class.
```

Listing 1.1. Shipping domain ontology snippet

3 http://www.w3.org/2008/05/skos
4 http://www.w3.org/Submission/vcard-rdf/
5 http://dbpedia.org

4.2 Service Description

Service descriptions and search requests are the core artefacts used in the matchmaking process. In order to promote reuse and facilitate modelling of service descriptions, we propose Service Variant Hierarchy (SVH) – a tree structure capturing relationships between service variant descriptions. Service variant descriptions $\mathcal{S}$ do not have to reflect capabilities, constraints and properties of services provided by a specific provider and available via specific endpoint. $\mathcal{S}$ can also model an abstract service capturing commonalities of service categories (e.g., shipping services, car insurance services, ISP services and others). $\mathcal{S}$ in SVH differ by their level of specialisation as shown in Figure 3. $\mathcal{S}$ which are higher in the hierarchy are less specialised than $\mathcal{S}$ which are lower in the hierarchy. Service offer $\mathcal{S}_{Offer}$ is the most specific $\mathcal{S}$ that cannot be further specialised. $\mathcal{S}_{Offer}$ contains concrete values of service properties and can be readily consumed by a service consumer.

SVH follows prototype-based inheritance [1] instead of a classical class-based inheritance. In a class-based inheritance objects are generated from classes and their properties and constrains are defined at the design time. In prototype-based inheritance objects are derived from existing objects by cloning and refining. Service descriptions are inherently incomplete [9] and their specifics cannot often be prescribed at the design time. For example, a shipping service $\mathcal{S}$ may provide a number of individually generated alternative service offers such as: (1) express shipping, (2) regular shipping, and (3) economy shipping. These service offers and their specific properties (or lack of some properties) become available at runtime. Prototype-based inheritance provides less restrictive model facilitating dynamic generation of new service variants. Traversing SVH and determining specifics of service variant descriptions and their availability depends on constrains and properties specified in search request $\mathcal{R}$. Prototype-based inheritance is more suitable for dynamic creation of service variant descriptions where it is often impossible to prescribe properties and constraints of service variants (including service offers) at design time.

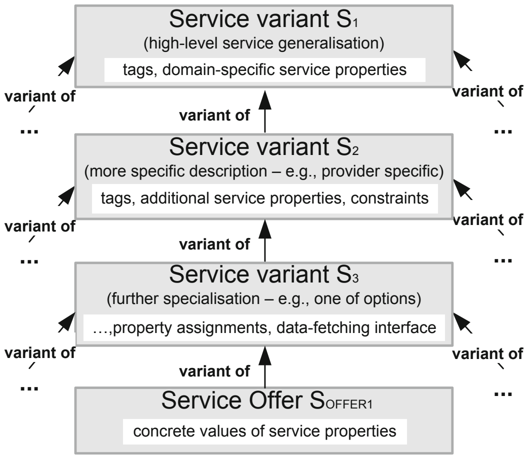


Fig. 3. Service Variants Hierarchy

Definition 1 (Service variant description). *Service variant description $\mathcal{S}$ is composed of a number of optional elements whose occurrences depend on the level of variant's concreteness. More formally: $\mathcal{S} = (\mathcal{O}, \mathcal{S}_{\mathcal{T}}, \mathcal{S}_{\mathcal{F}}, \mathcal{S}_{\mathcal{C}}, \mathcal{S}_{\mathcal{D}}, \mathcal{S}_{\mathcal{X}})$, where $\mathcal{O}$ are referenced ontologies. $\mathcal{S}_{\mathcal{T}}$ is a set of tags used for coarse-grained service matchmaking. $\mathcal{S}_{\mathcal{F}}$ is a set of properties required $P^R = \{p_1^R, ..., p_n^R\}$, $P^R \in C$ (expected by a $\mathcal{S}$ from $\mathcal{R}$) and provided $P^P = \{p_1^P, ..., p_m^P\}$, $P^P \in C$ (provided by $\mathcal{S}$). P^P and P^R map to concepts C defined in $\mathcal{O}$. Depending on the level of specialisation of a given service variant, $\mathcal{S}_{\mathcal{F}}$ may contain data type definitions of P^R and P^P properties or concrete values of P^R and P^P. $\mathcal{S}_{\mathcal{C}}$ are hard constraints imposed on P^P and P^R. $\mathcal{S}_{\mathcal{D}}$ is a set of property assignments defining how to create P^P. $\mathcal{S}_{\mathcal{X}}$ is a data-fetching interface used for handling dynamic P^P which cannot be a part of a service variant description. In order to formalise prototype-based inheritance relationships between service variants, we represent $\mathcal{S}$ as a set of tuples (E, V) where E uniquely identifies entity in $\mathcal{S}$ by URI and V is the entity value specified at entity URI. E and V map to entities and values which are defined as part of $\mathcal{O}, \mathcal{S}_{\mathcal{T}}, \mathcal{S}_{\mathcal{F}}, \mathcal{S}_{\mathcal{C}}, \mathcal{S}_{\mathcal{D}}$, and $\mathcal{S}_{\mathcal{X}}$.*

Definition 2 (Service offer description). *Service offer description $\mathcal{S}_{Offer}$ is a special case of a service variant description $\mathcal{S}$. Service offer description contains concrete values of service variant properties and cannot be further specialised. More formally: $\mathcal{S}_{Offer} = (\mathcal{S}_F, \mathcal{S}_{\mathcal{C}})$ where service properties $\mathcal{S}_F$ and service constraints $\mathcal{S}_C$ should be very specific and detailed in order to constitute fair description of a service offer in legal terms [12].*

Definition 3 (is-variant-of relationship). *Is-variant-of relationship $\overrightarrow{ivo}$ captures hierarchical prototype-based inheritance between parent $\mathcal{S}^{Parent}$ and child $\mathcal{S}^{Child}$ (including $\mathcal{S}^{Child}$ as a $\mathcal{S}_{Offer}$). $\overrightarrow{ivo}$ implies cloning and optionally overriding ($\overrightarrow{ovr}$) and extending entities coming from a $\mathcal{S}^{Parent}$. Overriding means replacing entities overlapping between a $\mathcal{S}^{Parent}$ and a $\mathcal{S}^{Child}$ with entities and values coming from the $\mathcal{S}^{Child}$. Extending means adding new entities and values to the ones that are defined in $\mathcal{S}^{Parent}$ and do not overlap with entities defined in $\mathcal{S}^{Child}$. More formally:*

$$\overrightarrow{ovr}, \overrightarrow{ivo} := \mathcal{S} \times \mathcal{S} \to \mathcal{S}$$
$$\overrightarrow{ovr}(\mathcal{S}^{Child}, \mathcal{S}^{Parent}) = \mathcal{S}^{Parent}.E \backslash (\mathcal{S}^{Parent}.E \cap \mathcal{S}^{Child}.E), \quad \mathcal{S}^{Parent}.V \backslash (\mathcal{S}^{Parent}.V \cap \mathcal{S}^{Child}.V)$$
$$\overrightarrow{ivo}(\mathcal{S}^{Child}, \mathcal{S}^{Parent}) = (\mathcal{S}^{Parent} \backslash \overrightarrow{ovr}(\mathcal{S}^{Child}, \mathcal{S}^{Parent})) \cup \mathcal{S}^{Child}$$

Table 1 shows an example of SVH. It uses required properties P^R like *SourceAddress* and *TargetAddress* and provided properties P^P like *ShippingPrice*, *ShippingDelivery* provided by a property assignment $\mathcal{S}_{\mathcal{D}}$ or data-fetching interface $\mathcal{S}_{\mathcal{X}}$ in response to P^R. Our example starts with a high-level service variant $\mathcal{S}^{Shipping}$ encompassing common characteristics for all shipping services (e.g., common tags - e.g., *shipping*, *parcel*, *delivery*) and common properties (e.g., source and target address, package dimension and weight). Property assignments and data-fetching interfaces are used for creating and determining the service offers in late service binding process.

Property assignments define dependencies between required and provided service properties. Property assignments are utilised during the matchmaking process to provide

Table 1. Service Variant Hierarchy example

Service variant	Content
$\mathcal{S}^{Shipping}$	$\mathcal{O} = ref(\mathcal{O}_{Shipping}, \mathcal{O}_{Service})$ $\mathcal{S}_{\mathcal{T}} = \{shipping, parcel, delivery\}$ $\mathcal{S}_{\mathcal{F}} = \{P_1^P = so : Availability, P_2^P = so : ShippingPrice,$ $P_3^P = so : ShippingDelivery, P_1^R = so : SourceAddress,$ $P_2^R = so : TargetAddress, P_3^R = so : Package\}$ $\mathcal{S}_{\mathcal{C}} = \{P_1^R \neq P_2^R\}$
$\overrightarrow{ivo}(\mathcal{S}^{FedEx}, \mathcal{S}^{Shipping})$	$\mathcal{S}_{\mathcal{T}} = \{FedEx, premium\}$ $\mathcal{S}_{\mathcal{F}} = \{P_4^P = \{so : Insurance\}\}$ $\mathcal{S}_{\mathcal{C}} = \{P_3^R.weight \leq 200kg\}$ $\mathcal{S}_{\mathcal{X}} = \{http// : localhost : 8080/../fedex/quote, ...\}$
$\overrightarrow{ivo}(\mathcal{S}^{FedExExpr}, \mathcal{S}^{FedEx})$	$\mathcal{S}_{\mathcal{C}} = \{P_3^R.weight \leq 50kg\}$ $\mathcal{S}_{\mathcal{D}} = \{\{P_3^P = 1day$ $if \quad P_1^R.country \in EU \wedge P_2^R.country \in EU\}\}$
$\overrightarrow{ivo}(\mathcal{S}^{FedExExpr}_{Offer}, \mathcal{S}^{FedExExpr})$	$\mathcal{S}_{\mathcal{F}} = \{P_1^P = true, P_2^P = 200EUR, P_3^P = 1day,$ $P_1^R = AddrInIreland, P_2^R = AddrInGermany,$ $P_3^R.weight = 20kg, P_3^R.length = 30cm,$ $P_3^R.height = 40cm, P_3^R.width = 30cm, P_4^P = 500EUR\}$

concrete values for provided properties P^P. Dependencies between properties that do not change frequently can be formalised in property assignments. Each property assignment defines constraints (SPARQL FILTER) and assignment statements (SPARQL BIND).

Listing 1.2 in lines 3–8 shows an example of property assignment. It creates a value of the *ShippingDelivery* provided property if source and destination constraints are satisfied (i.e., both source and destination countries are in Europe).

Data-fetching interface is an interface that represents a request-response interaction with the service endpoint (e.g., HTTP endpoint). Exchanged messages map to some of the required P^R and some of the provided P^P properties. In many cases not all dependencies between required and provided service properties can be explicitly specified using *property assignments*. It can be due to the dynamicity of service properties, complexity of dependencies between properties, or their business sensitivity. In such cases, values of provided properties P^P must be obtained on-the-fly from service provider's back-end system that is accessed via a referred endpoint. A data-fetching interface is a public service interface that is used during the discovery process to dynamically fetch provided properties P^P.

Listing 1.2 in lines 11–18 shows an example of the data-fetching interface. This data fetching interface is executed only if the source country is in Europe and the destination country is in Asia (lines 14–15). Instances of *ShippingPrice* and *ShippingDelivery* concepts will be obtained by invoking the service endpoint (line 16). Lowering in line 17 refers to the mapping from P^R in RDF to outgoing XML message, while lifting in line 18 refers to the mapping from incoming XML message to P^P in RDF.

```
1  @prefix s: <http://usoa.deri.ie/ontology/soffd#>.
2  /*Property assignment – delivery for addresses in EU*/
3  :deliveryEU a s:PropertyAssignment;
4   s:hasReqProperty so:SourceAddress,so:TargetAddress,so:Package;
5   s:hasProvProperty so:ShippingPrice, so:ShippingDelivery;
6   s:hasValue ”?srcCountry skos:subject dbpedia:European_countries.
7               ?trgCountry skos:subject dbpedia:European_countries.
8               BIND(3, ?delivery). BIND(d:Business_day, ?deliveryUnit).”^^s:SPARQL.
9
10 /*Data–fetching (dynamic price) – shipping between Europe and Asia*/
11 :priceDeliveryEurope a ser:DataFetching;
12  s:hasReqProperty so:SourceAddress,so:TargetAddress,so:Package;
13  s:hasProvProperty so:ShippingPrice, so:ShippingDelivery;
14  s:hasConstr ”?srcCountry skos:subject d:European_countries.
15               ?trgCountry skos:subject d:Asian_countries.”^^s:SPARQL;
16  s:hasEndpoint ”http://.../FedEx/endpoint”;
17  s:hasLowering ”http://.../Lowering.rq”;
18  s:hasLifting ”http://.../Lifting.rq”.
19
20 /* Service hard constraints – max. package size */
21 :supportedMaxPckgSize a ser:HardConstraint;
22  s:hasReqProperty so:Package;
23  s:hasValue ”FILTER (?length + ?width + ?height < 105 && ?lUnit=d:Centimetre && ...)”^^s:SPARQL.
```

Listing 1.2. Service description snippet

Hard constraints specify constraints on a service usage. Service hard constraints can be specified on: (1) required properties (e.g., $P^R_{length} + P^R_{girth} < 108cm$), (2) provided properties (e.g., $P^P_{deliveryTime} < 8$), and (3) combination of provided and required properties (e.g., $P^P_{price}/P^R_{weight} < 10$). Listing 1.2 in lines 21–23 shows an example of a hard constraint imposed on the package size.

4.3 Search Request

Our model of search requests facilitates expressing complex matching criteria. Simple search requests define fixed values of required properties P^R, whereas more complex search requests can define required properties as ranges (e.g., package weight: 8–12 kg) and enumerations (e.g., target address: home address or office address). More formally, $\mathcal{R} = (\mathcal{O}, \mathcal{R}_\mathcal{T}, \mathcal{R}_\mathcal{C}, \mathcal{R}_\mathcal{P})$, where $\mathcal{O}$ is a reference to the domain ontologies, $\mathcal{R}_\mathcal{C}$ are hard constraints that capture conditions which must be satisfied including flexible specification of required properties P^R in terms of ranges and enumerations, and $\mathcal{R}_\mathcal{P}$ are preferences that express the consumer's ranking criteria. Listing 1.3 shows an example that we will use in this section to illustrate the search request elements.

Hard constraints express service consumer's hard constraints defined over P^R and P^P. Hard constraints are defined using logical expressions on: (1) required properties only, (2) provided properties only, (3) combination of required and provided properties. Hard constraints defined in the case (1) are used to project required properties P^R. Service offers are generated with required properties P^R provided in $\mathcal{R}$. However, a service consumer might specify a flexible search request where P^R are defined in terms of ranges (e.g., package weight: 8–12 kg), enumerations (e.g., target address: home address or office address). Fixed values (e.g., package weight: 10 kg, target address: home address) are required by $\mathcal{S}$ in order to generate service offers that include details such as price and delivery time. In order to evaluate property assignments and data-fetching interfaces as well as search request hard constraints and preferences, we need to *project* (map) required properties P^R to *fixed* values. Hard constraints in the case (2) and (3) are evaluated on $\mathcal{S}$ and on the generated service offers.

```
1  /* Search Request Hard Constraints */
2  :hardConstraints a s:HardConstraint;
3   s:hasValue ":fromAddresses :hasValue ?fromAddr.
4               FILTER(?from = :fromAddresses). FILTER(?to = :addrGalway).
5               FILTER(?weight>=12.0 && ?weight<=15 && ?wUnit=d:Kilogram).
6               FILTER(?price < 120 && ?priceUnit = d:Euro).
7               FILTER(?deliveryDays<5 && ?deliveryUnit=d:Business_day)."^^s:SPARQL.
8  :fromAddresses s:hasValue :addrBerlin,:addrMunich,:addrVienna,:addrEindhoven.
9
10 /* Search Request Preferences */
11 :utilityFunction a s:UtilityFunction;
12  s:hasValue "BIND(−1*?price−10*?deliveryDays+5*?weight+0.01*?insurance, ?ranking)."^^s:SPARQL
       .
```

Listing 1.3. Shipping search request snippet

Listing 1.3 (lines 2–7) shows an example of hard constraints defined by a service consumer who wants a package to be shipped to a specific location under the condition that the shipping price should be below 120 EUR (line 6) and that the delivery has to take less than 5 business days (line 7). In here, the hard constraints use concepts from the domain ontology: *ShippingPrice* and *ShippingDelivery* defined in the domain ontology in Listing 1.1. Required properties P^R like *SourceAddress*, *TargetAddress* and *Package* are provided in $\mathcal{R}$ and P^P provided properties like *ShippingPrice* and *ShippingDelivery* are provided by the service during the discovery time.

Preferences express service consumer's ranking criteria. In our search request model, preferences can be specified using utility functions [**?**], weighted rules, and TCP-nets [2]. Our approach can support different types of preferences and we define preferences as a generic fitness function that indicates service offer ranking.

Utility functions are suitable for expressing min/max criteria over service properties, whereas weighted rules are more suitable for expressing various conditional utilities (similarly like TCP-nets) over service properties. Listing 1.3 (lines 11–12) shows a utility function that expresses a preference to minimize the shipping price (*-1 * ?price*), minimize the delivery time at the cost of 10 EUR per each day less (*-10 * ?deliveryDays*), maximize the package weight at the cost of 5 EUR per each kg more (*+5 * ?weight*), and maximize the package insurance coverage at negligible cost (*+0.01 * ?insurance*). This utility function is applied to generated service offers containing concrete values of provided and required properties. Calculated value of the utility function is used for service offer ranking.

5 Service Matchmaking and Implementation

In our search request model, search request properties P^R can be specified as fixed values or in a flexible way using enumerations and ranges. In order to evaluate constraints, execute property assignments, perform data-fetching and calculate rankings, we need to *project* (map) flexible search request properties P^R to *fixed* properties. The aim of our service matchmaking algorithm is to generate, filter and rank service offers (Figure 1). The matchmaking algorithm consists of the following steps: (1) calculation of valid property projections considering search request and service variant hard constraints, (2) random creation of an initial set of property projections, (3) dynamic generation of service offers, (4) filtering and ranking of the generated service offers, (5) selection of new projections (e.g., based on genetic algorithms). The algorithm continues until N projections have been tried. More detailed descriptions of our matchmaking algorithm is provided in [15].

We applied our approach to the shipping domain to address dynamic creation of a large number of possible service offers. Shipping companies provide shipping offers per individual request only. Due to the lack of shipping Web services, we have built service wrappers that perform on-the-fly interactions and advanced screen scraping with the shipping web sites leveraging Selenium2[6] Web application testing tool as an advanced, stateful screen scraper. Generating service offers can be a time-consuming operation when using *data-fetching interfaces* due to the complex processing (e.g., multi step browser-based advanced screen scraping) taking place behind a service endpoint.

6 Related Work

We identified related work in relation to service matchmaking, models of service descriptions and search requests which are the core ingredients to realise late service binding of highly configurable services. QoS-based service discovery [13] and QoS-based service composition aim to make applications adaptive, fault-tolerant and optimized in terms of discovered services. In our work, we do not address the modelling and monitoring of QoS characteristics (e.g., service responsiveness, latency, etc.), but we focus on the service properties which are request-dependent and dynamic.

GoodRelations service description model [3] identifies *more advanced price modeling* and *multi-dependent properties* as one of its core research challenges. Both challenges are of the primary focus of our service description model. Static service descriptions (e.g., expressed using GoodRelations) do not capture search request dependencies (e.g., in terms of complex pricing models and multi-property dependencies). In our matchmaking approach we promote dynamic service offer generation process in order to address these challenges.

Our service and search request models are related to Semantic Web services (SWS) frameworks like OWL-S [7] and WSMO [10] which are based on heavyweight logics foundations. SWS frameworks require complex modelling and strong background in logics – what is one of the main obstacles to their wider adoption. In contrast, our approach is based on lightweight semantic technologies (RDF, SPARQL and Linked Data) that more and more Web developers are familiar with and has been successfully used in various areas.

The approach proposed in [11] is also related to ours since it leverages Linked Data for describing services. However, the authors focus mainly on data services and

[6] `http://seleniumhq.org`

perceive services as atomic API operations that can be annotated with Linked Data. In our case, we perceive services more from a business perspective with special emphasis on different concreteness levels of service descriptions.

The notion of service offer is currently not addressed in business service description languages such as Universal Service Description Language (USDL[7]). Service descriptions have a varying level of concreteness depending on the stage of the matchmaking and configuration. This notion is currently missing from USDL. Introducing different levels of concreteness in USDL Functional Module will make USDL applicable for describing highly configurable services which are of particular interest for late service binding process.

7 Conclusion and Outlook on Future Work

In this paper we have addressed matchmaking of service offers in the late service binding. We contribute Service Variant Hierarchy model which promotes reuse and facilitates creation of new service descriptions where service variant descriptions differ by the level of their specialisation. Our approach is applicable to dynamic, highly configurable services whose properties are volatile and request-dependent. Only descriptions of service offers can satisfy the concrete needs of service consumers, while higher level service descriptions require manual examination.

We presented our work on service offer discovery as part of the late service binding mechanism that allows service consumers to dynamically determine and bind to the best available service offers. Late binding introduces more flexibility into applications and business processes, thus easing the switching between services offered by different providers. We applied our late service binding approach to the shipping domain. In this domain, service matchmaking cannot be achieved using high level service descriptions, as shipping details (e.g., price, delivery time, insurance) dynamically change and depend on the search request (e.g., target address and package weight).

As a future work we plan to apply our service offer discovery approach in the areas of Cloud Computing and Internet of Things (IoT) scenarios. Similar to shipping services addressed as an application example in this paper, Amazon Elastic Cloud[8] pricing model is also request-dependent (e.g., price of a computing instance depends on the requested usage) and dynamic (e.g., price of spot instances fluctuate depending on supply and demand). Services and resources on various layers of Cloud Computing stack (IaaS, PasS, SaaS) and demands of Cloud Computing consumers are highly dynamic. Our approach can be employed for matching supply and demand across different Cloud Computing vendors. On the other hand, in IoT scenarios we deal with feature-rich services which are user-driven, context-specific, dependent on the physical world sources of data where services must dynamically adjust their content and behaviour, depending on the specifics of the user requests and real-world data. In IoT scenarios our approach will provide consolidated view on IoT services by determining service properties at the matchmaking phase in contrast with the current practices where services are represented by static artefacts.

[7] http://www.internet-of-services.de
[8] http://aws.amazon.com/ec2/pricing/

Acknowledgments. This work is supported by the Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2).

References

1. Abadi, M., Cardelli, L.: A Theory of Objects. Springer (1996)
2. Brafman, R.I., Domshlak, C., Shimony, S.: On Graphical Modeling of Preference and Importance. Journal of Artificial Intelligence Research (JAIR) 25, 389–424 (2006)
3. Hepp, M.: GoodRelations: An Ontology for Describing Products and Services Offers on the Web. In: Gangemi, A., Euzenat, J. (eds.) EKAW 2008. LNCS (LNAI), vol. 5268, pp. 329–346. Springer, Heidelberg (2008)
4. van den Heuvel, W.-J., et al.: Software Service Engineering: Tenets and Challenges. In: ICSE PESOS Workshop (2009)
5. Kopecký, J.J., Simperl, E., Fensel, D.: Semantic Web Service Offer Discovery. In: SMR2 Workshop at International Semantic Web Conference, ISWC (2007)
6. Lamparter, S., Ankolekar, A., Studer, R., Grimm, S.: Preference-based Selection of Highly Configurable Web Services. In: World Wide Web, WWW (2007)
7. Martin, D.L., et al.: Bringing Semantics to Web Services with OWL-S. World Wide Web Journal, 243–277 (2007)
8. Papazoglou, M.P., Traverso, P., Dustdar, S., Leymann, F.: Service-Oriented Computing: State of the Art and Research Challenges. IEEE Computer 40, 38–45 (2007)
9. C. Preist. A Conceptual Architecture for Semantic Web Services (Extended Version), HP Laboratories. In *International Semantic Web Conference (ISWC)*. 2004.
10. Roman, D., et al.: Web Service Modeling Ontology. Applied Ontologies 1(1) (2005)
11. Speiser, S., Harth, A.: Integrating Linked Data and Services with Linked Data Services. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) ESWC 2011, Part I. LNCS, vol. 6643, pp. 170–184. Springer, Heidelberg (2011)
12. Treitel, G.H.: The Law of Contract. Sweet & Maxwell (2003)
13. Vu, L.-H., Porto, F., Aberer, K., Hauswirth, M.: An Extensible and Personalized Approach to QoS-enabled Service Discovery. In: IDEA (2007)
14. Zaremba, M., Vitvar, T., Bhiri, S., Hauswirth, M.: Preference-based Discovery of Dynamically Generated Service Offers. In: SCC (2011)
15. Zaremba, M., Vitvar, T., Bhiri, S., Hauswirth, M.: Service Offer Discovery Using Genetic Algorithms. In: ECOWS (2011)

Performative-Based Mining of Workflow Organizational Structures

Chihab Hanachi[1], Walid Gaaloul[2], and Ravi Mondi[2]

[1] IRIT, University Toulouse 1, France
[2] Institut TELECOM; TELECOM & Management SudParis, Évry, France

Abstract. Continuous and unforeseeable evolution of business rules, processing logics, and organizational structures within enterprises, require from business process management systems to integrate continuous design. Supporting business process rediscovery based on workflow logs analysis, workflow mining gathers retroactive (re)design techniques necessary to understand business process execution reality. Most of the work in this area focus on the control-flow perspective, while very few of them address the organizational aspect. In this paper, we propose to enrich workflow logs using a performative-based agent communication language in order to integrate conversations among workflow performers. Thereafter, we present innovative mining algorithms in order to discover organizational workflow structures based on the enriched logs. Those algorithms have been implemented as new plug-ins in ProM framework.

Keywords: Process mining, Organizational mining, Performatives.

1 Introduction

A great diversification of company services and products lead to a continuous process evolution. New requirements emerge and existing processes change. Consequently, the alignment of a process in regard to its observed evolution requires a permanent attention and reaction during the process life cycle. To maintain this alignment it is important to detect changes, i.e. the deviations of the described or prescribed behavior. Analyzing interactions of those complex systems will enable them to be well understood, controlled, and redesigned. It is obvious that the discovery, and the analysis of workflow interactions at runtime, would enable the designer to be alerted of design gaps, to better understand the model. Indeed, this kind of analysis is very useful in showing cause effect relationships and to analyze the discrepancies between the discovered model and the initially designed model. These discrepancies can be used to detect initial design gaps which may be used in a re-engineering process. *Process mining* provides an important bridge between data mining and business process modeling and analysis [19,17,18]. Process mining techniques are able to *extract knowledge from event logs* commonly available in today's information systems. These techniques provide new means to *discover, monitor, and improve processes* in a variety of application domains.

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 63–75, 2012.

Nowadays different approaches and techniques promote the use of process mining so that decision makers are in a better position to know who did what, when was it done, how did it end-up, and where did it happen. However, most of them have been focusing on control-flow discovery (e.g. How are the processes actually been executed?), whereas a few researches are interested in organizational mining (e.g. Who is handing over work to whom?). Indeed, both perspectives are critical as they permit to detect what is going on in a business in terms of organizational structure and flow. Nevertheless, very few work deal with the discovery of the workflow organizational models [11,10,13]. Furthermore, studies in [11,10,15] have proven the limitations of logs when it comes to identify a workflow's informational and organizational models.

The organizational models describe how the workflow users communicate for sharing their skills and knowledge. They help managers to understand clearly how the processes had been actually executed in order to improve their business process management. The aim of this paper is focused on mining organizational models derived from workflow logs. By organizational model, we mean here either a static relationship between performers or a dynamic one. The static relationship is described by a graph where nodes represent performers and arcs represent the relationships between them, e.g. delegation, cooperation, etc. The dynamic structure is represented by an interaction protocol, e.g. contract-net, negotiation, auction, etc., between performers that each of them plays a well-defined role such as bidder, vendor, etc.

Logs that keep track of all executed activities are the primary source of process mining. These activities are usually part of workflows upon which different applications like Customer Relationship Management (CRM) and Enterprise Resource Planning (ERP) are being built. However, several studies like [11,10,15] have proven the limitations of these logs when time comes to identify a workflow's informational and organizational models.

In order to understand the problem, let us refer to the Table 1 which represents an example of a common workflow log used in recent work and tools. Existing process mining tools could discover a workflow diagram as in figure 1.a and a social network diagram as in figure 1.b using Table 1. In the first case, activity A was performed by John. Then, Julie completed task B and gave duty to William to finish that case by realizing activity D. However, this task assignment among the performers (Julie, William, John, Ines) can be done in different ways (arbitrary assignment, call for participation, delegation, etc.) depending on the related organizational models which are defined through the semantic of performers' interactions.

Table 1. Log sample

Case Id.	Task id.	Perfor.
1	A	John
1	B	Julie
1	D	William
2	A	John
2	C	Julie
2	D	Ines
3	A	Ines
3	F	Ines

The semantics of interaction among performers is very ambiguous in common logs. Therefore, a comprehensible model of log structure is an essential

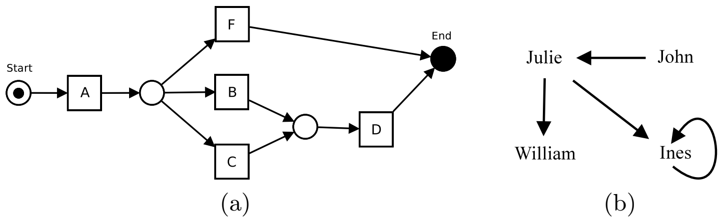


Fig. 1. Mined process and organizational models

factor. The Agent technology, suitable for the design and development of co-operative systems, offers interaction languages and high-level protocols allowing the representation of fine-grained organizational structures such as hierarchy, federation etc. So, we propose enriching logs with interaction details based on agent technology interaction languages. There are practically no approaches in workflow mining that tackle this issue (mining more fine-grained organizational structures using more meaningful log description). In the *process mining manifesto* [19] published recently, extracting and using the most meaningful logs for process mining techniques was cited as a guiding principle.

Our work is based on the use of agent technology that provides a new view on workflow mining. This technology handles and provides representation for organizational concepts, such as groups, roles, commitments, and also organizational structures, such as federation, hierarchy or market, inherited from "Organizational Theory". All these concepts are useful to structure rules or at least understand at a macro level the coordination patterns of the different partners involved in a workflow or an Inter-Organizational Workflow. In this paper, we show how to integrate conversations among the performers of processes using a performative-based agent communication language (FIPA-ACL) and discover workflow organizational structures. The organizational structures are represented by graphs where the nodes are the performers and the arcs are relations among them.

The remainder of this article is structured as follows. In section 2, we present the structure of workflow event logs. The organizational mining algorithms and the interaction protocol checking are detailed respectively in sections 4 and 3. In section 5, we illustrate the implementation efforts done to validate our approach. Section 6 discusses the related work and concludes this paper.

2 Structure of Workflow Event Logs

Information systems might not be concerned with the details of the internal processing of their process activity executions. However, most process-aware management systems, such as ERP, CRM, SCM, log the external events that capture the activities life cycle. Extensible event stream (XES) is a generic XML-based format suitable for representing and storing workflow event log data [15]. The logs used for our work is of XES format.

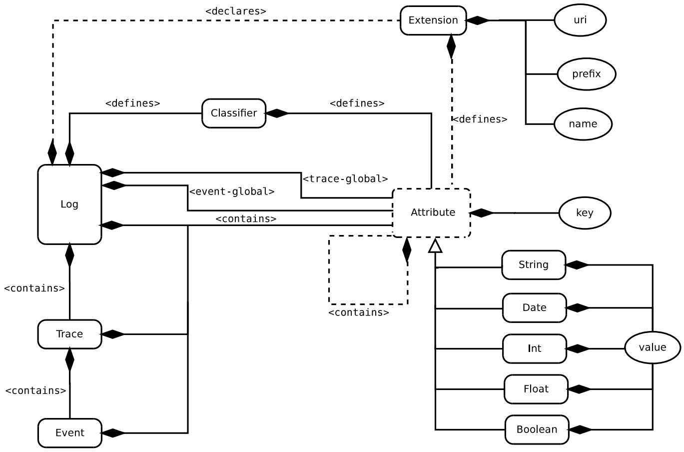


Fig. 2. XES meta model [15]

Figure 2 shows the data meta model of XES. An XES file contains one log containing multiple number of traces. Each trace contains a sequential list of events that were executed within one workflow execution instance. XES does not prescribe any mandatory set of attributes for the log, trace and event elements. To generalize the use of the same attributes across various logs, XES uses the so-called 'extensions'. They provide the semantics to the attributes. There are five standard extensions: concept, lifecycle, organizational, time and semantic. New extensions can be added as and when necessary. XES also has the concepts of 'classifiers', which is a list of attributes to differentiate between two events in a trace.

In order to capture the interaction in its entirety, we specify three attributes related to the event: *Sender*, *Receiver* and *Performative*. A sender is a performer who desires to communicate with a receiver using performatives. A receiver is a performer who waits for active communications with the sender. A performative describes the communication intention expressed using an expressive communication language. In fact, these interactions enable the determination of the real task allocation policy between performers, and indirectly the underlying protocols or organizational structures. Agent Communication Language (ACL), which is based on speech-act theory, is applied in the message's content language to express the action [2]. ACL's approach [5], based on the speech-act theory, gives the communication facility among performers by exchanging the message in various protocols. Speech acts are articulated by means of standard vocabulary, also known as 'performatives'. Some of the performatives we use in our "augmented" log structure are: *Delegate*, *Inform*, *Execute*, *Query Request*, *Answer*, *Call for proposal*, *Propose*, *Accept proposal*, *Reject proposal*, *Failure*, etc.. Listing 1.1 shows a list of events in a given workflow trace enriched with our performatives.

Listing 1.1. Excerpt of a workflow log

```
...
<event>
    <string key="Activity" value="FileFine"/>
    <string key="Performative" value="Cfp"/>
    <string key="Sender" value="A4"/>
    <string key="Receiver" value="A3"/>
</event>
<event>
    <string key="Activity" value="FileFine"/>
    <string key="Performative" value="Cfp"/>
    <string key="Sender" value="A4"/>
    <string key="Receiver" value="A5"/>
</event>
<event>
    <string key="Activity" value="FileFine"/>
    <string key="Performative" value="Propose"/>
    <string key="Sender" value="A5"/>
    <string key="Receiver" value="A4"/>
</event>
<event>
    <string key="Activity" value="FileFine"/>
    <string key="Performative" value="Propose"/>
    <string key="Sender" value="A3"/>
    <string key="Receiver" value="A4"/>
</event>
<event>
    <string key="Activity" value="FileFine"/>
    <string key="Performative" value="Reject"/>
    <string key="Sender" value="A4"/>
    <string key="Receiver" value="A3"/>
</event>
<event>
    <string key="Activity" value="FileFine"/>
    <string key="Performative" value="Accept"/>
    <string key="Sender" value="A4"/>
    <string key="Receiver" value="A5"/>
</event>
...
```

3 Organizational Structures Mining Techniques

In this section, we present two mining techniques to discover organizational structures (hierarchy, federation) using only workflow logs as input. An organizational structure can be represented by a graph where nodes correspond to members and labeled arcs to their links with several possible semantics: communication, control, etc. [7] classified numerous models of organizational structures which differ in the structure of subordination, the form of the organizational structure, and the presence/absence of a common objective.

3.1 Graph Toolkit

Based on workflow logs, we extract an interaction graph. By definition, this graph is denoted as $G(V, E)$ where $|V|$ (respectively $|E|$) represents the nodes' set (respectively the edges' set) which is the performers' set (respectively the performatives' set). Indeed, there is an edge p between two vertices x, y in V *iff* $\exists l : Event \in L : Processlog| \ (l.sender = x \wedge l.receiver = y \wedge l.performative = p)$. Some high level operations that are used in our mining algorithms are explained below.

```
procedure HIERARCHY-DISCOVERY(L, HList)
Input: L: Log
Output: HList: G
Local variables: S: nodeList        ▷ list of sources (without vertices father)
G(V, E): graph        ▷ partial graph obtained from L restricted to delegation relations
    G(V, E) ← partialGraphP(L, "delegate"),
    if Tree(G) then
        HList ← G
    else
        S ← Sources(G)
        for each vertex in S do
            CC ← ConnectedGraph(G, s)
            if Tree(CC) then
                HList ← HList.add(CC)
            end if
        end for
    end if
    return HList;
end procedure
```

Algorithm 1. Hierarchy mining algorithm

1. $partialGraphP(L, P)$: Returns a graph $G(V, E)$ where vertices are restricted to performers of L:log whose performative is equal to P:performative. There is an edge between two vertices x, y *iff* $\exists l : Event \in L : Log|\ (l.Sender = x) \wedge (l.Receiver = y) \wedge (l.Performative = P)$.
2. $Connected(G)$: Returns true if the graph G is connected, false if not. A connected graph is a graph such that there is at least a path between all pairs of vertices.
3. $Tree(G)$: Returns true if G is a tree false if not. In a directed graph, a tree will be called a polytree. A polytree is a weakly connected graph with no circuit.
4. $Symmetric(G(V, E))$: Returns true if the graph G is symmetric, false if not. A graph G is symmetric (or arc-transitive), if each pair of vertices connected in a sense is also in the other.
5. $ElemenPath(a, b, G)$ Returns true if there is an elementary path between vertices a and b of G. A path is called elementary if no vertices appear more than once in it;
6. $ConnectedGraph(G, s)$: a function that returns a graph CC such as $CC \subset G \wedge s \in CC \wedge Connected(CC) \wedge \nexists CCi \subset G|CC \subset CCi \wedge Connected(CCi)$
7. $Parent(a, b, G)$: Returns true if it exists an edge from a to b in G;
8. $Sources(G(V, E))$: Returns a list of vertices such as $s \in Source(G) \Leftrightarrow \nexists x \in V|\ Parent(x, s, G))$. A source is a vertex that has only outgoing edges;
9. $ChildList(s, G)$: Returns a list of vertices such as $x \in ChildList(s, G) \Leftrightarrow Parent(x, s, G)$;
10. $artVertGraph(G)$: Returns the list of articulation points in G. An articulation point is a node whose removal will disconnect the graph.

3.2 Hierarchy Discovery

In a hierarchy structure, performers are conceptually structured as a polytree where a higher-level agent has more global view than their lower-level agents. A hierarchy is the only organizational structure which has mono-directional relationship that has as type "superior/subordinate". Hence, we can conclude that each performer has at most one superior. In a polytree, nodes are agents and edges express a superior/subordinate control structure that has two major roles which are (i) simplification and (ii) control of information transmission. Let $G(V, E)$ be a directed graph restricted to the relation of delegation. There is a hierarchical link between two vertices a and b, where a is the superior of b, and noted $HierarchicalLink(a, b, G)$ iff $(Parent(a, b, G) \wedge \nexists ElemenPath(a, b, G))$. Algorithm 1 is our proposed hierarchy mining algorithm. It proceeds in three main steps. The first step involves extracting partial graph G restricted to the performative 'delegate' (line 6). The next step is to check whether the obtained graph G is a tree or a polytree. If it is a tree, G represents the complete hierarchy (lines 7;8). If G is a polytree, the next step involves extracting the connected components CCi from G, and return the set of partial hierarchy(ies) (lines 9-17).

3.3 Federation Discovery

A federation is a structure representing many performer groups. In fact, each group is formed by several performers that communicate via an intermediate performer. This specific performer (a federation representative) is represented as an articulation point and other performers are the leaves. A federation must obey four properties which are:

- P1: The interaction graph G is connected and symmetric;
- P2: There are at least two articulation points and each point has at least one child who is not an articulation point;
- P3: Each vertex of G which is not an articulation point (federated performer) has a single parent who is an articulation point;
- P4: The partial graph of G restricted to delegations is not a tree (which excludes hierarchies).

To check if a workflow organizational structure is a federation or not, we propose a second mining algorithm (see Algorithm 2) which takes as input the log L and returns a Boolean output. The algorithm begins with the extraction of interaction graph G, from the log L (line 10). Next, we check whether the graph G is connected and symmetric (line 11). If it is not (i.e. violates P1), there is no federation. Otherwise, we extract sub-graph Gd, limited to the performative 'delegate' (checking P4) (line 12). If Gd is not a polytree, we build the set of representatives ER (articulation points) (line 14) and Federated points EF (line 14) (the leaves) and make sure the properties P2 (lines 18-23) and P3 (lines 25-30) hold true.

4 Interaction Protocol Compatibility

In this section, we provide an algorithm for checking the compatibility of performers' interactions to an interaction protocol during runtime. In contrast to the described mining techniques in Section 3, we use two inputs: workflow logs and

```
procedure FEDERATION(L, FederationG)
Input: L: Processlog
Output: FederationG: Boolean
Variables: Fed, Rep : Vertices    ▷ Fed and REP are resp. the federated and representative vertices from the federated structure
ER: Vertex List    ▷ list of all federation representatives
EF: Vertex List    ▷ List all Federated
G(N, A): graph    ▷ resulting interactions graph of L
Gd(Nd, Ad)    ▷ interaction graph
Prop2, Pro3: Boolean
    G(N, A) ← partialGraphP(L, all)
    if (Connected(G) ∧ Symmetric(G)) then
        Gd(Nd, Ad) ← GraphP(L, "delegate")
        if ¬Tree(Gd) then
            ER ← artVertGraph(G)
            EF ← N − ER
            if ER ≠ ∅ then
                Prop2 ← True
                while (ER ≠ ∅) ∧ (Prop2) do
                    Rep ←ER.next()    ▷ each vertex has at least one federated
                    if (ChildList(Rep, G) ∩ EF ≠ ∅ then
                        Prop2 ← false
                    end if
                end while
                Prop3 ← True
                while EF ≠ ∅ ∧ Prop3 do
                    Fed ← EF.next()    ▷ each federate has a single representative father
                    if (ParentList(Fed, G) ⊂ ER)∧ |ParentList(Fed, G)| = 1) then
                        Prop3 ← False
                    end if
                end while
                if (Prop2) ∧ (Prop3) then FederationG ← True
                end if
            end if
        end if
    end if
    return FederationG;
end procedure
```

Algorithm 2. Federation mining algorithm

an interaction protocol. In the literature, a protocol is a package of regulation which is used by computers to communicate with each other across a network. A protocol is a convention or standard that controls or enables the connection, communication and data transfer between computing endpoints which defines the concept of conversation. The concept of conversation provides each interaction with a context that determines its purpose and it conversely organizes the whole set of atomic interactions among agents. Therefore many protocols have been designed and maintained depending on enterprise structure.

Real process executions can be different from the initial design. In order to detect such deviation, we propose an algorithm to check that each execution instance is compatible to a specific protocol. A protocol may be described by a notation such as sequence diagrams or by a rigorous formalism such as Petri Nets. As input we assume that we have already the protocol graph and the process instance log to check. In this paper, we propose an algorithm to check, for instance, the compatibility of the Call For Proposal (CFP) protocol Call For Proposal. (CFP) protocol consists of different transitions (send invitation, rejection, propose, etc.) which communicate one to another for exchanging message using edge. Message which is transferred from a transition to another is stored in a place. A performer can send a CFP to different performers. The possible interactions among two performers a and b in CFP protocol are:

1. $cfp(a,b).refuse(b,a)$ a proposes a duty to b but b rejects.
2. $cfp(a,b).propose(b,a).reject(a,b)$ b accepts the proposal but a rejects.
3. $cfp(a,b).propose(b,a).accept(a,b).inform(b,a,\text{“}fail\text{”})$ a accepts the proposal of b but b fails in the task execution.
4. $cfp(a,b).propose(b,a).accept(a,b).inform(b,a,\text{“}done\text{”})$ b executes successfully the task after accepting the proposal.

```
1:  procedure CompatibleProtocol(C,GM(R,M0),Comp)
2:  Input: C: ProcessInstance; GM(R, M0): markingGraph
3:  Output: Comp: Boolean
4:  Local variable: l: ATE
5:  S: PerformativeList                  ▷ list of performatives extracted from C
6:  CM: Vertex                           ▷ current Vertex in one GM path
7:  A: Edge                             ▷ an edge starting from the current marking vertex
8:  p: Performative
9:                                                                    ▷ Step 1
10:     S ← ∅
11:     for each l in C do
12:         S ← S ∪ (l.performative, l.sender, l.receiver)
13:     end for                                           ▷ Parallel Path is S et R
14:     CM ← M0
15:     p ← First(S)
16:     Comp ← True
17:                                                                   ▷ Step 2
18:     while ( Comp ∧ p ≠ ∅ ∧ ¬(Leaf(CM))) do
19:                                   ▷ Leaf(CM) returns true if CM is a leaf in GM
20:         if ∃ a:edge | a.vertex() = CM ∧ a == p then
21:             p ← S.next()                                    ▷ go forward in S
22:             CM ← R.next()                                   ▷ go forward in M
23:         else
24:             Comp ← False
25:         end if
26:     end while
27:                                                                   ▷ Step 3
28:     if ( ¬Leaf(CM) ∨ p ≠ ∅) then
29:         Comp ← False
30:     end if
31: end procedure
```

Algorithm 3. Compatible Protocol Algorithm

Listing 1.1 shows a list of events in a given trace and it is enriched with the performatives of the Call For Proposal (CFP) protocol. ‘Activity’ is a classifier for the log, which means all the events mentioned in the excerpt belong to one activity (“FileFine”). The interaction sequence “Cfp.Cfp.Propose.Propose.Reject.Accept” can be interpreted as the execution of a call-for-proposal protocol launched by A4, which is submitted to A3 and A5. The proposal from A3 is rejected while that from A5 is accepted. So, A5 executes the task.

The main steps of our CFP compatibility algorithm (Algorithm 3) are as follows: (i) we capture the execution sequence of performatives S extracted from the instance log C (lines 10-16); (ii) we walk in parallel through the protocol graph GM and S as long as the performative considered in S is equal to the label of the edge in GM (lines 18-26) (iii) if we reach a terminal node in GM at the same time we reach the end of S, then C is considered to be conform to the protocol P, if not C is considered to be incompatible with P (lines 28-30). Following the same principles, we can specify other compatibility algorithms for other communication protocols.

5 Implementation

Our implementation is composed of two steps. First, we collect logs based on an enhanced XES format based on an agent interaction protocol. Then, we use these logs to validate our purposed organizational mining techniques using plug-ins developed in ProM framework.

We propose different means to validate the workflow log collecting issue. We have used two complementary tools: collecting real execution logs and generating simulated logs. The first objective was achieved by implementing an API which was grafted into a WfMS to collect log instances from already designed and executed workflow processes. To satisfy the second objective, we have used CPN Petri nets simulation tools to generate simulated logs. Indeed, getting real logs from big size workflow examples turns out to be a difficult task. The advantage in using simulated logs is that it is easier to fix and vary external factors ensuring a better diversity of the examples and more accurate validation. The scaling issue of our validation test is consequently better dealt with simulated logs that enable us to cover a qualitatively and quantitatively varying set.

In order to do this, we use a log simulating tool [9] which creates random XML logs by simulating already designed workflow processes based on CPN tools[1]. The program enables us to create simulated logs. Modifications were brought to the program to call predefined functions that create "enriched" XES logs with respect to our multi-agent organizational log events. This stage implies modifications in the modeling level of CPN workflow declarations, particularly in the actions and the transition input/output levels. Thereafter, we used ProMimport[2] to group or gather the simulated workflow logs in one log.

(a) (b)

Fig. 3. Screenshots of the mined structures

[1] http://cpntools.org/
[2] www.promimport.sourceforge.net

Our approach has been validated within the ProM framework [14], as plug-ins. The ProM framework is flexible with respect to the input and output format, and is also open enough to allow for the easy reuse of code during the implementation of new process mining ideas. In this paper, the three proposed algorithms were implemented: hierarchy and federation mining algorithms, and CFP compatibility protocol. Figure 3.a and 3.b show the respective screen shots of the output of the plug-ins. The hierarchy in figure 3.a shows actors $A2$ and $A5$ form the top of the hierarchy, followed by $A4$, $A8$ and $A10$, $A6$. The federation structure in figure 3.b shows that actors $A4$, $A5$ and $A8$ represent the articulation points and others interact through one of them.

6 Discussion

Process mining techniques have proven to be a valuable instrument to gain insight into how business processes are compromised within organizations [14]. They aid the decision maker on re-designing the workflows [20,6,16]. The idea of applying process mining in the context of process management was first introduced in [1]. This work proposes methods for automatically deriving a formal model of a process from a log of events related to its executions and is based on workflow graphs. To discover and analyze social networks, [13] combines many business process concepts, management and sociometry. [4] proposes to use a less rich log structure to discover process models from event logs.

However, the approaches do not allow to discover complex structures as the organizational ones. To the best of our knowledge there are practically no approaches in workflow mining that use performative-based log to mine organizational structures. This article may be seen as a first step in this area. In our approach, we have tried to have a smart trade-off between the complexity of the mining techniques input (log structure) and the output (discovered structures). Our approach proposed to enrich workflow logs with performatives describing interactions between process performers. Each interaction is recorded with the performative and the performers who are involved (sender, receiver). This allowed us to remove many ambiguities about the relationships between the performers themselves, and learn about the workflow organizational structures using the original field of Multi-Agent Systems.

In this paper, we presented organizational mining techniques to discover static (hierarchy, federation) relationships between performers, and check dynamic (CFP protocol) relationships between performers. Currently, we are working with industrial partners in order to assess the effectiveness of our approach in a concrete use case. Besides this, we have tested our tool qualitatively and quantitatively against various set of organizational structures and protocols. A second prospect for us is to have a semantically annotated performatives, as organizations can use different communication languages. The semantic enrichment of logs should help to identify the organizational and informational models of a workflow [21].

References

1. Agrawal, R., Gunopulos, D., Leymann, F.: Mining Process Models from Workflow Logs. In: Schek, H.-J., Saltor, F., Ramos, I., Alonso, G. (eds.) EDBT 1998. LNCS, vol. 1377, pp. 469–498. Springer, Heidelberg (1998)
2. Andonoff, E., Bouaziz, W., Hanachi, C., Bouzguenda, L.: An agent-based model for autonomic coordination of inter-organizational business processes. Informatica, Lith. Acad. Sci. 20(3), 323–342 (2009)
3. Jagadeesh Chandra Bose, R.P., van der Aalst, W.: Trace Alignment in Process Mining: Opportunities for Process Diagnostics. In: Hull, R., Mendling, J., Tai, S. (eds.) BPM 2010. LNCS, vol. 6336, pp. 227–242. Springer, Heidelberg (2010)
4. Ferreira, D.R., Gillblad, D.: Discovering Process Models from Unlabelled Event Logs. In: Dayal, U., Eder, J., Koehler, J., Reijers, H.A. (eds.) BPM 2009. LNCS, vol. 5701, pp. 143–158. Springer, Heidelberg (2009)
5. FIPA. FIPA ACL message structure specification. FIPA agent communication language specifications (2002)
6. Gaaloul, W., Bhiri, S., Haller, A.: Mining and Re-engineering Transactional Workflows for Reliable Executions. In: Parent, C., Schewe, K.-D., Storey, V.C., Thalheim, B. (eds.) ER 2007. LNCS, vol. 4801, pp. 485–501. Springer, Heidelberg (2007)
7. Horling, B., Lesser, V.: A survey of multi-agent organizational paradigms. Knowl. Eng. Rev. 19, 281–316 (2004)
8. Mans, R.S., Russell, N.C., van der Aalst, W.M.P., Bakker, P.J.M., Moleman, A.J., Jaspers, M.W.M.: Proclets in healthcare. Journal of Biomedical Informatics 43(4), 632–649 (2010)
9. Medeiros, A.K.A.D., Gnther, C.W.: Process mining: Using cpn tools to create test logs for mining algorithms. In: Proceedings of the Sixth Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools, pp. 177–190 (2005)
10. Reijers, H.A., Song, M., Jeong, B.: Analysis of a collaborative workflow process with distributed actors. Information Systems Frontiers 11(3), 307–322 (2009)
11. Song, M., van der Aalst, W.M.P.: Towards comprehensive support for organizational mining. Decision Support Systems 46(1), 300–317 (2008)
12. van der Aalst, W.M.P., Pesic, M., Song, M.: Beyond Process Mining: From the Past to Present and Future. In: Pernici, B. (ed.) CAiSE 2010. LNCS, vol. 6051, pp. 38–52. Springer, Heidelberg (2010)
13. van der Aalst, W.M.P., Reijers, H.A., Song, M.: Discovering social networks from event logs. Computer Supported Cooperative Work 14(6), 549–593 (2005)
14. van der Aalst, W.M.P., van Dongen, B.F., Günther, C.W., Rozinat, A., Verbeek, E., Weijters, T.: Prom: The process mining toolkit. In: de Medeiros, A.K.A., Weber, B. (eds.) BPM (Demos). CEUR Workshop Proceedings, vol. 489, CEUR-WS.org (2009)
15. van der Aalst, W.M.P., van Dongen, B.F., Herbst, J., Maruster, L., Schimm, G., Weijters, A.J.M.M.: Workflow mining: A survey of issues and approaches. Data Knowl. Eng. 47(2), 237–267 (2003)
16. van der Aalst, W.M.P., van Hee, K.M., van der Werf, J.M.E.M., Verdonk, M.: Auditing 2.0: Using process mining to support tomorrow's auditor. IEEE Computer 43(3), 90–93 (2010)
17. Gaaloul, W., Gaaloul, K., Bhiri, S., Haller, A., Hauswirth, M.: Log-based transactional workflow mining. Distributed and Parallel Databases 25(3), 193–240 (2009)

18. Gaaloul, W., Bhiri, S., Haller, A.: Mining and Re-engineering Transactional Workflows for Reliable Executions. In: Parent, C., Schewe, K.-D., Storey, V.C., Thalheim, B. (eds.) ER 2007. LNCS, vol. 4801, pp. 485–501. Springer, Heidelberg (2007)
19. van der Aalst, W., Adriansyah, A., de Medeiros, A.K.A., Arcieri, F., Baier, T., Blickle, T., Bose, J.C., van den Brand, P., Brandtjen, R., Buijs, J., Burattin, A., Carmona, J., Castellanos, M., Claes, J., Cook, J., Costantini, N., Curbera, F., Damiani, E., de Leoni, M., Delias, P., van Dongen, B.F., Dumas, M., Dustdar, S., Fahland, D., Ferreira, D.R., Gaaloul, W., van Geffen, F., Goel, S., Günther, C., Guzzo, A., Harmon, P., ter Hofstede, A., Hoogland, J., Ingvaldsen, J.E., Kato, K., Kuhn, R., Kumar, A., La Rosa, M., Maggi, F., Malerba, D., Mans, R.S., Manuel, A., McCreesh, M., Mello, P., Mendling, J., Montali, M., Motahari-Nezhad, H.R., zur Muehlen, M., Munoz-Gama, J., Pontieri, L., Ribeiro, J., Rozinat, A., Seguel Pérez, H., Seguel Pérez, R., Sepúlveda, M., Sinur, J., Soffer, P., Song, M., Sperduti, A., Stilo, G., Stoel, C., Swenson, K., Talamo, M., Tan, W., Turner, C., Vanthienen, J., Varvaressos, G., Verbeek, E., Verdonk, M., Vigo, R., Wang, J., Weber, B., Weidlich, M., Weijters, T., Wen, L., Westergaard, M., Wynn, M.: Process Mining Manifesto. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) BPM Workshops 2011, Part I. LNBIP, vol. 99, pp. 169–194. Springer, Heidelberg (2012)
20. Gaaloul, W., Bhiri, S., Godart, C.: Discovering Workflow Transactional Behavior from Event-Based Log. In: Meersman, R., Tari, Z. (eds.) CoopIS/DOA/ODBASE 2004. LNCS, vol. 3290, pp. 3–18. Springer, Heidelberg (2004)
21. Belecheanu, R., Cabral, L., Domingue, J., Gaaloul, W., Hepp, M., Filipowska, A., Kaczmarek, M., Kaczmarek, T., Nitzsche, J., Norton, B., Pedrinaci, C., Roman, D., Stollberg, M., Stein, S.: Business process ontology framework. In: SUPER Deliverable (2007)

Recommender Systems in Computer Science and Information Systems – A Landscape of Research

Dietmar Jannach[1], Markus Zanker[2], Mouzhi Ge[3], and Marian Gröning[1]

[1] TU Dortmund, 44227 Dortmund, Germany
{Dietmar.Jannach,Marian.Groening}@tu-dortmund.de
[2] Alpen-Adria-Universität Klagenfurt, 9020 Klagenfurt, Austria
Markus.Zanker@aau.at
[3] Bundeswehr University Munich, 85577 Neubiberg, Germany
Mouzhi.Ge@unibw.de

Abstract. The paper reviews and classifies recent research in recommender systems both in the field of Computer Science and Information Systems. The goal of this work is to identify existing trends, open issues and possible directions for future research. Our analysis is based on a review of 330 papers on recommender systems, which were published in high-impact conferences and journals during the past five years (2006-2011). We provide a state-of-the-art review on recommender systems, propose future research opportunities for recommender systems in both computer science and information system community, and indicate how the research avenues of both communities might partly converge.

1 Introduction

Rooted in the fields of information retrieval (IR), machine learning (ML) and decision support systems (DSS), recommender systems (RS) have emerged as a research field of their own during the last twenty years. Recommender systems propose ranked lists of items (that are subsets of a larger collection) according to their presumed relevance to individual users. Relevance is determined from explicit and implicit user feedback such as ratings on items, commercial transactions or explicitly stated requirements [1]. With the rapid growth of electronic commerce, the ubiquity of mobile information access and the advent of the Social Web, the interest in RS research has grown enormously during the past years. This is for example documented by the rapidly growing ACM Recommender Systems conference series as well as by the publication of various focused journal special issues and books. The reasons for this high attractiveness of the field are manifold and include highly visible competitions such as the Netflix prize, increased industrial interest or the new application opportunities for recommendation techniques in the Mobile and Social Web. Based on what is the dominant goal of a recommender system, research can be considered from a variety of different perspectives:

– The IR perspective addresses the problem of information overload. The purpose of a RS is therefore to identify the items relevant to a user's information need.

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 76–87, 2012.

– The ML perspective on recommender systems is to learn a model that predicts the user feedback on a specific item as accurately as possible. The goal is therefore to reduce the error between the predicted and the actual feedback (i.e. typically a rating value) of a given user.
– From the point of view of DSS, recommender systems can be understood as tools supporting consumers in their decision making process. Therefore, the ultimate goal of a RS is to increase the quality of decisions made. Measuring the achievement of this goal is less straightforward compared to the goals of the two previous perspectives. A variety of factors actually influence the users' decision making such as their appreciation of the system, trust in the information and service provider, experience and domain expertise, word-of-mouth as well as their real preferences. Researchers in marketing, e-commerce and information systems research might be interested in proxies such as the impact of such systems on customer behavior, loyalty and sales figures.

Given this diversity of research perspectives, the goal of our work is to review and classify recent research in recommender systems in order to quantify the research interests and identify opportunities for future research. In addition, this analysis should serve as a basis to understand limitations of current research practice in this field. As RS are IT applications we naturally limit our analysis to publications in the neighboring fields of Computer Science (CS) and Information Systems (IS). In the next section, we will describe the methodology of our literature analysis. In Section 3 we present detailed findings and conclude with a discussion on under-researched areas.

2 Methodology

We systematically evaluated all publications of a pre-defined set of high-impact journals and conferences in the fields of Computer Science (CS) and Information Systems (IS) during the period from January 2006 to July 2011. We included both journal articles as well as full papers appearing in conference proceedings. In particular, we considered those journals, where special issues on recommender systems have appeared. Table 1 lists the publication outlets, their type, i.e. journal (jrnl) or conference proceedings (proc), their presumed belonging to either CS or IS and the respective number of publications considered for further analysis. In total 330 publications have been identified, out of which 73 appeared in journals and 257 in conference proceedings. 65 publications (∼20% from total) appeared in outlets that belong to the IS community, if such an attribution is permitted. Not astoundingly, the newly established ACM conference series on RS is the single most important publication venue in this field. We classify publications according to the following scheme presented in Table 2. The classification task has been performed by the authors. We will discuss the possible evaluations for each class attribute in conjunction with the results in the next section.

Table 1. Considered publication outlets

Name	Type	Field	Nbr.
ACM Conf. on Human Factors in Comp. Syst. (CHI)	proc	CS	13
ACM Conf. on Recommender Syst. (RecSys)	proc	CS	86
Int. Conf. on Int. User Interfaces (IUI)	proc	CS	17
Int. Conf. on Knowl. Disc. and DM (SIGKDD)	proc	CS	22
Int. Conf. on Res. and Dev. in IR (SIGIR)	proc	CS	33
Int. Conf. on World Wide Web (WWW)	proc	CS	21
Int. Joint Conf. on AI (IJCAI)	proc	CS	13
AAAI Conf. on AI (AAAI)	proc	CS	10
Int. Conf. on Data Mining (ICDM)	proc	CS	5
Americas Conf. on Information Systems (AMCIS)	proc	IS	8
European Conf. on Information Systems (ECIS)	proc	IS	6
Int. Conf. on Information Systems (ICIS)	proc	IS	7
Med. Conf. on Information Systems (MCIS)	proc	IS	5
Pac. Asia Conf. on Information Systems (PACIS)	proc	IS	11
ACM Trans. on Intell. Syst. and Techn. (TOIST)	jrnl	CS	6
ACM Trans. on the Web (TWeb)	jrnl	CS	5
AI Comm.	jrnl	CS	12
IEEE Intelligent Systems	jrnl	CS	14
Int. Jrnl. of Human Computer Studies (IJHCS)	jrnl	CS	5
World Wide Web (WWW)	jrnl	CS	3
Dec. Supp. Syst. Jrnl. (DSS)	jrnl	IS	9
Inf. Syst. Res. (ISR)	jrnl	IS	3
Int. Jrnl. of Electronic Comm. (IJEC)	jrnl	IS	7
Jrnl. of Mgt. Information Systems (JMIS)	jrnl	IS	7
Mgt. Information Systems Quarterly (MISQ)	jrnl	IS	2

Table 2. Classification scheme

Research contribution	What is the main contribution of the paper? For instance proposing a novel algorithm.
Recommended items	For instance, media and entertainment resources, people or diverse e-commerce products
Recommendation paradigm	Collaborative filtering, content-based filtering or knowledge-based recommendation techniques.
Research method	E.g. experimental research on datasets, studies involving real users in lab or field conditions, formal proof.
Data sets	Which data sets are used in the paper?
Evaluation measures	Employed metrics and choice of a baseline.

3 Results

3.1 Research Contribution

We limit our taxonomy of research contributions first to *constructive* contributions that developed a novel technical artifact, notably an algorithm or recommendation technique, and *empirical* research that advanced the body of theory by, for instance, hypotheses driven user-involved studies in the lab or the real-world (experimental as well as observational research approaches). Not surprisingly empirical research plays compared to the other categories a relatively more important role in the IS field while in contrast CS publications focus heavily on algorithmic improvements.

Table 3. Research contributions

Type of contribution	IS outlets	CS outlets
Technical artifacts	24 (36.9%)	189 (71.3%)
Empirical research	21 (32.3%)	18 (6.8%)
Both	9 (13.8%)	43 (16.2%)
Other	11 (16.9%)	15 (5.7%)
Total	65 (100%)	265 (100%)

Furthermore, the following observations can be made based on an additional content analysis.

- More than 25% of CS papers were related to recommendation in the context of the Social and Semantic Web, while only 6% of IS papers considered this context.
- The CS community also addressed questions on non-functional requirements such as scalability or privacy (15% of all CS papers). These areas have been mostly ignored in IS research.
- Cold-start recommendations, i.e. proposing items to users that newly entered the system or recommending novel items, is an issue in both fields. About 10% of the CS papers and about 5% of the IS papers referred to this issue.
- Questions of user interface design (CS 5.8%, IS 12.3%) and transparency (CS 6.8%, IS 10.8%) are relatively more relevant to IS research.
- Topics such as group recommendations, context-awareness, diversity of recommendations, multi-criteria and knowledge-based recommendations, as well as methodological questions still play a very small role in both communities.

Summarizing, according to our analysis the IS community focuses more on the user perspective and the interplay of computerized systems and users whereas research in CS more often takes an algorithmic perspective. Even though questions regarding human computer interaction in RS have been addressed in the early CS literature, see for example Swearingen and Sinha [2], there is still more work required in this area. Topics such as user-centric evaluation, human decision-making and user interaction, have only recently gained more attention also in the CS field. This trend can be observed from recent publications such as Pu et al. [3] and Knijnenburg et al. [4], or from recent journal special issues, for example, a special issue on measuring the impact of recommendation of personalization on user behavior [5]. This can be considered as an emerging trend of CS and IS in RS research.

3.2 Recommended Items

The next aspect we consider in our analysis is related to the application domain and the recommended items. In the CS field, the main application areas are media and entertainment (>45% from CS total), social networks (>25%), as well as general e-commerce and browsing and search (each 10%). In the IS field, recommendations in e-commerce play a dominant role (>50% from IS total). Beside media and entertainment (12%) also digital libraries (11%) and the use of RS within the organization (12%), e.g. for team collaboration or staffing, were relevant. We subsumed the most popular item category of movie recommendations under media and entertainment even though this problem could be also subsumed under the e-commerce umbrella. Figure 1 depicts this distribution of IS and CS publications over these item categories.

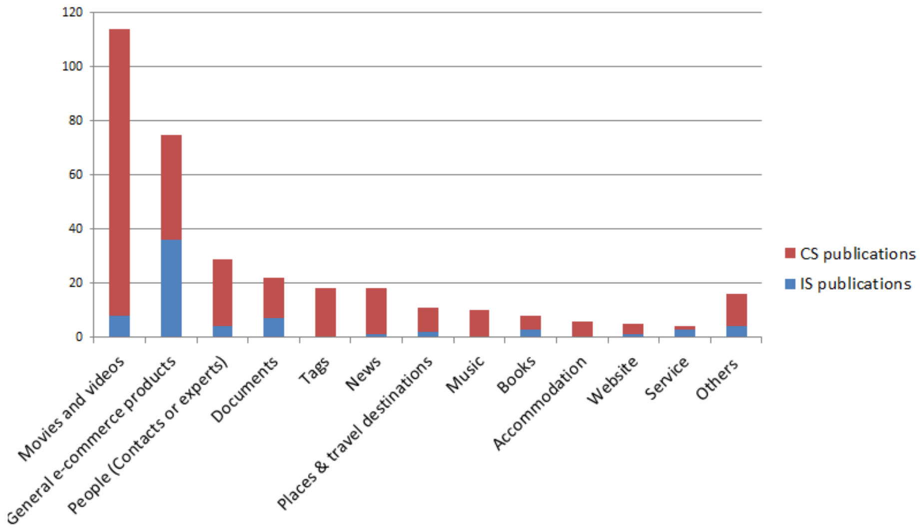


Fig. 1. Recommended items

The availability of public datasets for evaluation such as MovieLens, Netflix or data from Social Web platforms seems to strongly bias the choice of researchers on which application domains to work on. This is particularly the case for the CS field, while IS researchers focus on the recommendation of shopping goods and documents. With respect to avenues for future research, it would be of interest to see if and how the algorithm models of movie recommenders can be transferred to other commercial and business domains. The fact that over 40% of CS papers are focusing on the movie domain shows that other benchmark datasets are badly required. Such datasets would help to stimulate new research directions in particular for CS research and prevent the community from further optimizing the predictive accuracy for the media & entertainment domain. While general e-commerce and intra-organizational document retrieval is at the core of

IS researchers' interest, the advent of the Social Web is not yet reflected in their work. However, we can assume that once the intra-organizational use of social networks becomes more popular, RS in social networks will also move into the focus of IS researchers.

3.3 Recommendation Paradigms

The recommendation paradigm determines the principal underlying mechanism how a RS computes recommendations [6,7]. Collaborative filtering, for instance, reasons that users who had similar opinions in the past will also more likely agree in the future, while content-based filtering determines recommendable items based on their similarity with items the user has liked in the past. While these two mechanisms are most popular, the analysis of our sample reveals significant differences in the research practice of CS and IS communities (see Figure 2). In particular knowledge-based recommendation systems that exploit explicitly codified domain expertise are comparably popular in IS research but play a minor role in CS research. In IS research, RS are often referred to as Recommendation Agent [8] or Digital Advisor [9]. IS papers on decision support or expert systems are usually based on constraints and interactive preference elicitation because business decisions are more often guided by strict rules rather than user experience. In our review, we found that for both IS and CS research, two prominent types of knowledge-based systems are constraint-based and critiquing approaches [10,11]. Hybridizations of different paradigms play only a minor role in both fields. Considering the importance of collaborative filtering technique in RS, we looked into more details of the used algorithms. Figure 3 shows a more detailed picture of the algorithms used in collaborative filtering in both IS and CS fields. Some papers fall into several categories at the same time.

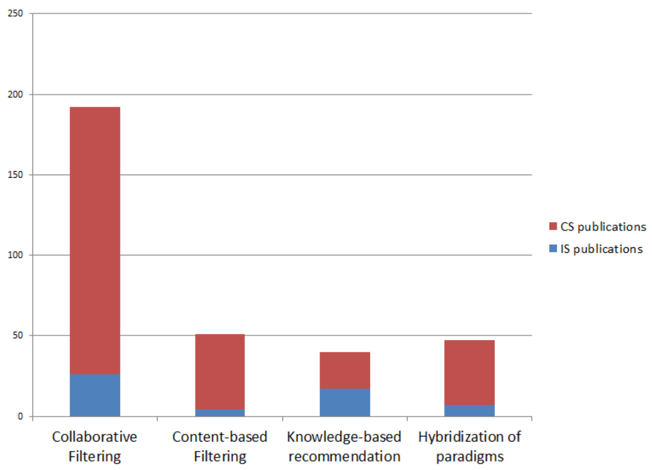


Fig. 2. Recommendation paradigms

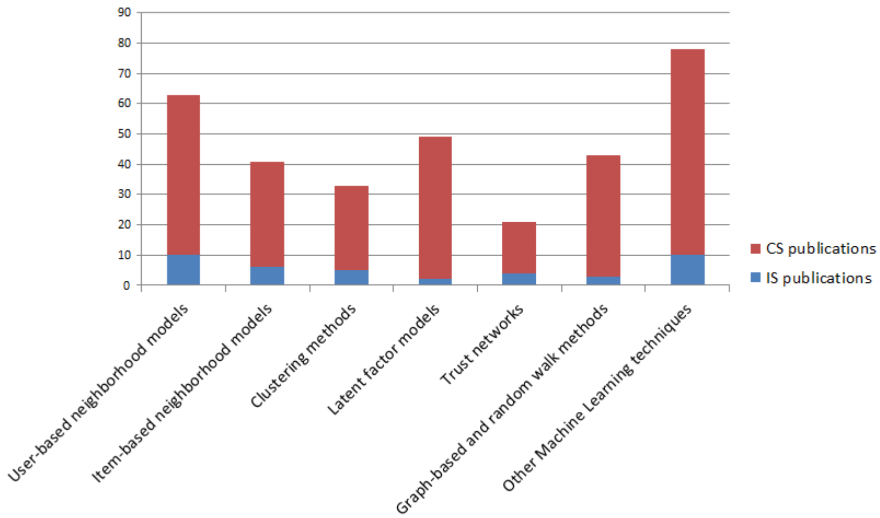


Fig. 3. Collaborative filtering techniques

The analysis shows that classical RS techniques such as nearest-neighbor and clustering are still relevant in today's RS research. During the past ten years and particularly boosted by the Netflix prize, various types of latent factor models including Singular Value Decomposition (SVD) and Latent Dirichlet Allocation (LDA) have become popular and are nowadays often used as a baseline algorithm for comparative evaluations in CS [7]. In IS research, however, these models only play a minor role. In addition, we found that various machine learning methods such as probabilistic or regression models are relevant in both fields. As for the future research, a better understanding of different aspects of latent factor approaches that go beyond predictive accuracy is required. It is for example unclear how recommendations based on these models can be explained to the users in order to increase the user's trust. Also, we know little about how these methods perform with respect to the diversity, serendipity or novelty of the generated recommendations. Furthermore, as early work such as Balabanovic & Shoham [12] has showed that combining different techniques or sources of knowledge can advance the system performance, more research is required in the selection of recommender algorithms to construct high performance RS. Also, recommender strategies may perform differently in different situations [13]. For example, the study of Jannach & Hegelich [14] revealed that the choice of the most effective recommender depends on the specific situation and goal of the user. In the future it is interesting to further understand how to select different recommender strategies in line with various situations.

3.4 Research Method

In this section, we will discuss the research design of the selected papers, which includes research methods, underlying theories, data sets, evaluation metrics and data analysis methods. Figure 4 classifies the sampled publications according to the employed research methods. However, note that we did not use a comprehensive taxonomy of research methods, but denoted the most commonly used terminology in RS research.

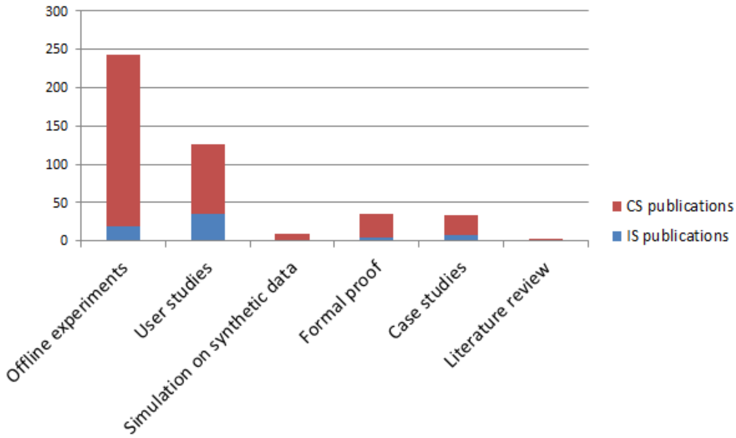


Fig. 4. Research methods

Over the last five years, two third of the selected papers in CS are based on offline experiments using historical user data. In IS research, in contrast, user studies and user-involved experiments in conjunction with questionnaires or interviews are more popular. Simulation and formal proofs play a minor role in both fields. While most CS papers typically improve the RS performance, IS papers often try to explain a phenomenon or theories related to RS. With respect to hypothesis development and social theories, IS research is more based on theoretical considerations than CS. In about half of the IS papers, theories such as the Technology Acceptance Model, Social Agency, Social Presence or the Theory of Reasoned Action are mentioned as underlying theories.

As offline experimentation on data sets is the most important method, we provide an overview on the most popular types of data (Figure 5). MovieLens, Netflix, Eachmovie, IMDB as well as a major share of Yahoo! data denote data sets from the movie domain. This documents the dominating role of benchmark data from media and entertainment that was employed by at least 50% of all papers that are based on offline evaluations. Besides movies, recommendations in the context of Web 2.0 (included in *Others*) also played an important role in the past five years. For instance, the corresponding datasets can be derived from Epinions, Delicious, YouTube, CiteULike, Bibsonomy, Flickr, Orkut, Twitter, Digg, or Wikipedia. In total, about 25% of CS papers were evaluated based on Web 2.0 data. To some smaller extent, news recommendation is also a relevant topic in CS research. In addition, a small number of IS experiments were done in the e-commerce domain, based for example on the data obtained from Amazon.com.

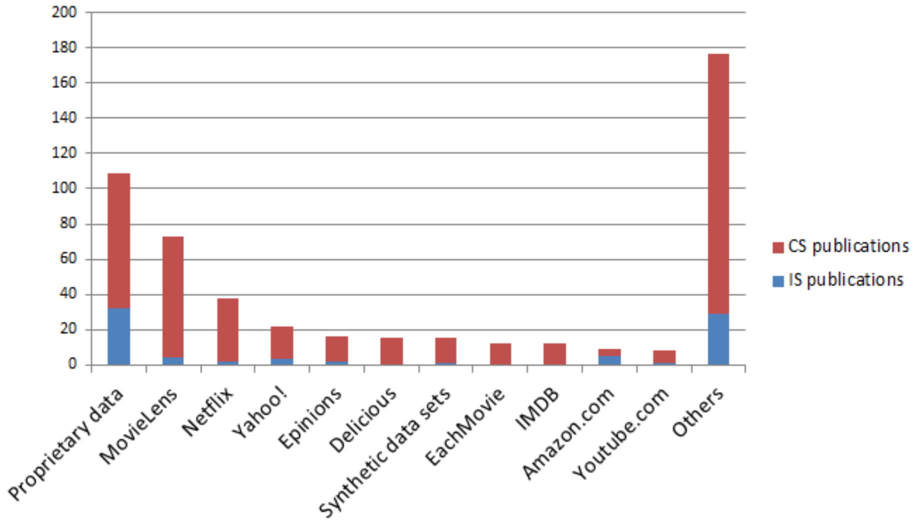


Fig. 5. Datasets

Table 4. Evaluation measures

	IS outlets	CS outlets
IR measures		
Precision and Recall	12	115
F1	2	20
Rank measures (e.g. NDCG)	9	27
ROC curve	1	11
Area under ROC	0	8
ML measures		
Mean absolute Error	6	57
Root Mean Squared Error	0	49
Application quality		
Computation time	2	28
Coverage metrics	2	28
Decision support quality		
Perceived utility or user satisfaction	11	7
Online conversion	3	12
Diversity metrics	0	10

Table 4 quantifies which metrics are used in the literature to assess the quality of recommendations such as the whole quality of RS or factors used to test research hypotheses. It is structured according to the perspectives denoted in the introduction.

– Classic Information Retrieval (IR) metrics measuring classification accuracy such as Precision, Recall and their harmonic mean F1 as well as ROC curves

and the area below the curve are particularly popular in the CS community. Rank measures that consider the position of items in recommendation lists such as the normalized discounted cumulated gain (NDCG) or half-life utility also fall into this category, but they are only used to a minor extent in RS research.
- In Machine Learning (ML), traditionally aggregate deviations between actual and predicted rating values are measured. The Mean Absolute Error (MAE) metric averages the deviations between predicted and actual rating values, while the Root Mean Squared Error (RMSE) puts more weight to larger deviations. Although several authors have critiqued these measures that, for instance, combine prediction errors for items the user hates as well as for highly recommended items, a considerable share of work still argues its contribution by improvements in terms of these error measures.
- The third section in Table 4 groups measures describing aspects of the technical *application quality.* Coverage measures determine the share of the user population that can actually receive recommendations, while computation time is an important aspect for the system's responsiveness.
- Finally, we identified measures that focus on quality aspects of a system's decision support capabilities. For instance, perceived utility is a well known concept from technology acceptance research. Online conversion rates do measure short-term commercial success and the persuasive traits of a recommendation system, but do not measure medium and long-term satisfaction with the bought item. Furthermore, diversity metrics measure if a system provides a broad view on the offered choices.

Most CS papers are more homogenous with respect to the applied metrics and focus only on the accuracy of recommendations. This can again be explained by the existence of standardized benchmark problems. For the metrics that go beyond accuracy, only diversity measures are applied in recent literature. Even though the problems of using only accuracy metrics in RS have been discussed in the last few years, for example in [15], more evaluation metrics such as novelty, popularity and serendipity are not largely applied so far. Future RS research can therefore focus on evaluating and improving RS by considering a variety of metrics to achieve an overall high quality. In IS research, most papers are focused on measuring the perceived quality such as utility or user satisfaction and explaining the latent relationships and effects based on user-centric evaluation. As typically practiced in social science research, papers are often organized by proposing specific research hypotheses and validated via a corresponding experimental design. The appropriate evaluation method therefore depends on the design of the proposed models or hypotheses. This may lead to the effect that a multitude of different measures are proposed but barely reused. We observed that individual metrics which were used in about 30% of the papers were barely used in other papers. Therefore future research can focus on proposing validated and widely accepted measuring instruments for more rigorous research.

4 Discussion

The large amount of papers which aim at optimizing the accuracy of predictions based on historical data in the movie domain underlines the strong need for the CS research community to focus both on other domains as well as on other types of experimental designs to understand the real impact of recommender systems on users. Even for the supposedly well-understood movie domain, it is not definitely clear in which situations and to which extent lower RMSE values finally lead to higher user satisfaction or sales. Consider for example a RS that recommends the fifth sequel of a movie to a user who has liked all other movies in this series. Such a prediction might be accurate but not valuable. The context of movie consumption (Am I watching a movie alone or with friends? Am I looking for entertainment or for intellectual challenge?) is largely not taken into account in the majority of papers. In these areas, only the time of the day or week was considered as a context factor to some extent in recent works. The same holds for the development of the user's taste over time and long-term user models. In future work, more focus should therefore be put on context-aware RS, see for example [16]. The rapid development of the Social and Semantic Web has the potential to further boost the field of RS. On the one hand, new application areas arise, for example the recommendation of tags, pictures, friends or links. At the same time, more and more data is available for exploitation by recommendation algorithms, as users are increasingly willing to contribute, participate, or interact, for example, on social networks, review platforms or blogging sites. One final observation in the context of RS evaluation is that theoretical considerations about the computational complexity or questions of scalability of algorithms are covered only by a small fraction of research papers in both fields. Most of the recent algorithmic approaches are model-based, rely on an offline learning phase and support the efficient generation of predictions. However, in particular in the context of the Social Web and the massive amounts and different types of data that have to be processed, appropriate strategies to efficiently compute models and predictions might become more relevant in the future.

5 Conclusions

Our literature review indicated the importance of recommender systems in the fields of Information Systems and Computer Science. Given the different roots of the fields, CS researchers focus more on algorithms whereas IS researchers are more interested in the systems-perspective and the effects of RS on the users. Correspondingly, different research designs and methods dominate in the two communities as has been documented by this work. As an outlook, we see evidence that increased mutual exchange of results from the two communities can help further advance the research of recommender systems. In the survey of Xiao and Benbasat [8] the authors discuss the role of recommender systems for example in the context of consumer research and marketing, human decision making, electronic commerce or human computer interaction. Only few works in

CS focus on these topics today. Thus we see the development of techniques that exploit the insights from these different areas as a field of future RS research for the CS community. In parallel, the IS community can benefit from incorporating recent algorithmic results from the CS community.

References

1. Zanker, M., Jessenitschnig, M.: Case-studies on exploiting explicit customer requirements in recommender systems. UMUAI 19(1-2), 133–166 (2009)
2. Swearingen, K., Sinha, R.: Beyond algorithms: An HCI perspective on recommender systems. In: ACM SIGIR 2001 Workshop on Recommender Systems (2001)
3. Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: Proc. ACM RecSys 2011, pp. 157–164 (2011)
4. Knijnenburg, B., Willemsen, M., Gantner, Z., Soncu, H., Newell, C.: Explaining the user experience of recommender systems. UMUAI 22(4), 441–504 (2012)
5. Zanker, M., Ricci, F., Jannach, D., Terveen, L.G.: Measuring the impact of personalization and recommendation on user behaviour. IJHCS 68(8), 469–471 (2010)
6. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: Recommender Systems - An Introduction. Cambridge University Press (2011)
7. Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.): Recommender Systems Handbook. Springer (2011)
8. Xiao, B., Benbasat, I.: E-commerce product recommendation agents: Use, characteristics, and impact. MIS Quarterly 31(1), 137–209 (2007)
9. Felfernig, A., Friedrich, G., Jannach, D., Zanker, M.: An Integrated Environment for the Development of Knowledge-Based Recommender Applications. International Journal of Electronic Commerce 11(2), 11–34 (2006-2007)
10. Zanker, M., Jessenitschnig, M., Schmid, W.: Preference reasoning with soft constraints in constraint-based recommender systems. Constraints 15(4), 574–595 (2010)
11. Felfernig, A., Friedrich, G., Jannach, D., Zanker, M.: Developing constraint-based recommenders. In: [7], pp.187–215
12. Balabanović, M., Shoham, Y.: Fab: content-based, collaborative recommendation. Communications of the ACM 40(3), 66–72 (1997)
13. Pilászy, I., Tikk, D.: Recommending new movies: even a few ratings are more valuable than metadata. In: Proc. ACM RecSys 2009, pp. 93–100 (2009)
14. Jannach, D., Hegelich, K.: A case study on the effectiveness of recommendations in the mobile internet. In: Proc. ACM RecSys 2009, New York, pp. 41–50 (2009)
15. McNee, S., Konstan, J.R.J.: Being accurate is not enough: How accuracy metrics have hurt recommender systems. In: EA ACM CHI 2006, pp. 997–1001 (2006)
16. Adomavicius, G., Tuzhilin, A.: Context-aware recommender systems. In: [7], pp. 217–253

Differential Context Relaxation for Context-Aware Travel Recommendation

Yong Zheng, Robin Burke, and Bamshad Mobasher

Center for Web Intelligence
School of Computing, DePaul University
Chicago, Illinois, USA
{yzheng8,rburke,mobasher}@cs.depaul.edu

Abstract. Context-aware recommendation (CARS) has been shown to be an effective approach to recommendation in a number of domains. However, the problem of identifying appropriate contextual variables remains: using too many contextual variables risks a drastic increase in dimensionality and a loss of accuracy in recommendation. In this paper, we propose a novel treatment of context – identifying influential contexts for different algorithm components instead of for the whole algorithm. Based on this idea, we take traditional user-based collaborative filtering (CF) as an example, decompose it into three context-sensitive components, and propose a hybrid contextual approach. We then identify appropriate relaxations of contextual constraints for each algorithm component. The effectiveness of context relaxation is demonstrated by comparison of three algorithms using a travel data set: a contenxt-ignorant approach, contextual pre-filtering, and our hybrid contextual algorithm. The experiments show that choosing an appropriate relaxation of the contextual constraints for each component of an algorithm outperforms strict application of the context.

Keywords: Context-aware Recommender System, Collaborative Filtering, Contextual Pre-Filtering, Context Relaxation.

1 Introduction

The application of context to recommender systems has a strong intuitive appeal. It makes sense that a user's preferences will change from situation to situation: the set of restaurants desirable for a dinner date is not the same as for a quick business lunch. This insight is the motivation behind the expansion of recent research in context-aware recommender systems (CARS) [3,4,2]. A number of researchers have shown that appropriate application of context can create effective recommendation solutions in a variety of domains [5,7,10,15].

One of the key problems in CARS research is the identification of contextual variables. It is clear that these factors are very domain-specific. In the restaurant example above, obviously the occasion of the meal is an important contextual variable, but also the time of day, the day of the week, and the user's location are all plausible candidates, and if the dataset included all of this information, it would be tempting to build all available variables into a contextual model. However, it is unlikely that all of these

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 88–99, 2012.

contextual factors will repeat exactly in peer users' profiles, so choosing contextual variables becomes crucial. [16]

Context-aware algorithms typically require that contextual features are selected in advance and applied throughout the recommendation computation. By contrast, we propose a novel way to apply contexts in recommendation. First, we decompose a recommendation algorithm into different functional components. Second, we create a formalism for contextual constraints, so that contextual features may be fully or partially matched in computing recommendations. Last, we identify appropriate relaxations of the contextual constraints for each algorithm component. Applying context to recommendation therefore becomes a task not only of finding the best subset of contextual variables to select, but of discovering the best way to use context in each part of the algorithm. We take the traditional user-based collaborative recommendation as an example, decompose it into three components, apply our hybrid contextual approach and demonstrate the effectiveness of contextual relaxation for each component by comparing against our other baseline algorithms.

2 Related Work

Context-aware recommendation has been the subject of intensive research over the past several years, with a number of workshops and challenges on different aspects of the topic [CARS, RecSys; CAMRa, RecSys; CaRR, IUI; CoMoRea, PerCom; ICAS, CISIS; IIiX; etc].

Surprisingly, the field has yet even to agree on the definition of a context. The most commonly used definition is the one given in 1999 by Abowd *et al*. "*Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.*" [1] This definition hardly limits the set of variables that can be considered when performing recommendation.

In [2], G. Adomavicius, et al. introduce a two-part classification of contextual information. Their taxonomy is based on two considerations: what a recommender system knows about a contextual factor and how the contextual factor changes over time. System knowledge can be subdivided into *fully observable*, *partially observable* and *unobservable*. The temporal aspect of a factor can be categorized as *static* or *dynamic*. This analysis yields six possible classes for contextual factors. In this paper, we are concerned with *static* and *fully observable* factors – the question is which such factors to incorporate into a recommendation model and which to ignore.

Typically, researchers working in a particular domain will identify contextual variables that they believe are important, and that are available in their data. The question then becomes one of discovering the importance of each of these variables. Baltrunas *et al* [6] conducted a survey asking participants to evaluate if a particular contextual factor influenced their ratings or not, in order to acquire contextual relevance from subjective judgements and further build predictive models for mobile recommender systems. Another attempt to explore contextual relevance is proposed by Huang *et al* [12], where they combine attributes of contexts with items directly and extract a set of significant contextual attributes to index a particular item based on rough set theory reduction. The process is similar to contextual feature selection [16] or attribute reduction, and it requires large datasets with contextual ratings under multi-dimensional contexts for training purposes.

However, few available datasets have rich contextual ratings – cases where users have rated the same item multiple times in different contexts – or explicit meta-information from users about the importance of context. This limits the applicability of the methods discussed above where the context is treated as a strict constraint. In this paper, we assume that relaxations of the contexts can also play influential roles in recommendation, especially when the contextual information is sparse in the dataset. For instance, if a user is going to take a business trip to Chicago, he may take advice from his friend who once went to Chicago although the two may travel for different purposes.

2.1 Travel Recommendation

For our experiments, we make use of a dataset crawled from Tripadvisor.com, which is one of the world's largest travel sites. The site supports travel planning with its extensive database of reviews of travel destinations and accommodations. Tripadvisor.com allows users to post ratings and reviews for hotels where they have lodged, and it provides an option for the users to explicitly indicate the type of trip using one of five categories: family, friend, couples, solo and business trip. Previous research into context-aware recommendation on Tripadvisor.com has used this trip type information as the key contextual variable [11,14].

Research on tourism behavior [9,13] indicates that geographical locations including both the origins and destinations of travel are considered as influential factors beyond user profiles, hotel amenities and prices. Klenosky and Gitelson [13] also pointed out that trip type appeared to have a primary influence and trip origin a secondary influence on travel recommendations. We decided to follow this line of inquiry and incorporate location of origin and destination as contextual variables in addition to trip type.

We quickly discovered that exact matches in origin city and destination city were fairly rare. So, we generalized the notion of location, creating a geographical hierachy for locations in the United States with three levels: city, state, and time zone.

3 Methodology

As described in [3], there are three basic approaches to context-aware recommendation: pre-filtering, post-filtering, and contextual modeling. The filtering approaches apply a context-dependent criterion to list of items, selecting only those appropriate to a given context. For example, in a pre-filtering system using location as a context, all items that are not nearby the user would be filtered out before recommendations are processed. A post-filtering approach applies the filter after recommendations have been computed. These approaches work best if the contextual data is relatively dense – most objects can be evaluated relative to most contexts. Contextual modeling, by contrast, builds contextual considerations into the recommendation process itself. The application of this technique therefore is intimately tied to the particular recommendation algorithm being employed. In this section, we show how a user-based collaborative algorithm can be adapted to form a contextual hybrid.

3.1 Algorithm Decomposition

User-based collaborative recommendation using Resnick's popular algorithm is treated as a prediction problem where the task is to predict the expected rating $P_{a,i}$ that a user a would give to an item i. The first step in this process is to identify a set of neighbors N who are similar to user a (often using Pearson correlation as a metric). Once the neighbors are identified, the prediction is given by

$$P_{a,i} = \bar{r}_a + \frac{\sum\limits_{u \in N} (r_{u,i} - \bar{r}_u) \times sim(a,u)}{\sum\limits_{u \in N} sim(a,u)} \tag{1}$$

where a is the target user, i is an item, N is a neighborhood of users u similar to a, and sim is a similarity measure between users. Correspondingly, $r_{u,i}$ is neighbor u's rating on item i, $\bar{r}_a$ is user a's average rating over all items, and $\bar{r}_u$ is user u's average rating.

We can think of this algorithm as weighting the variance of neighbors' ratings by their similarity to the user and then applying this calculated variance to the user's own average rating. The assumption is that every user has an average "baseline" rating that may differ from user to user and what is significant about a particular rating is how it deviates from the user's personal baseline.

We turn this algorithm into a context-sensitive one by incorporating context into the computation of each prediction. Assume that there exists some contextual constraint C that applies to the current situation and we are attempting to recommend items that would be appropriate to a in this context. There are three places in this algorithm where context can come into play, and three components that could be used to form a contextual recommender:

1. Instead of considering all users who have rated item i, the neighborhood calculation can choose only users who have rated item i in a context that matches C.
2. The neighbor's baseline that is used to calculate the variance for each neighbor $\bar{r}_u$ can be limited to those ratings associated with contexts matching C, rather than an overall average for that user. For example, a user might be less stringent in his rating of hotels he stays in for business because he generally has little choice, but rate more critically on family vacations where he is paying himself.
3. The user's baseline $\bar{r}_a$ can be a function of C for the same reason.

3.2 Context Relaxation

For our purposes, we will assume that each single context c consists of a set of contextual features and that there is a binary matching process between the contextual constraint C and any individual context c – either a context matches the constraint or it does not.[1] For this research, we are considering three possible types of constraints for

[1] We plan to consider the partial matching of constraints in future research.

each feature: `any`, which means no constraint for a given feature, `exact`, meaning that the feature must match exactly, and `contained` which applies to location information where one geographical designation (like city) might be contained in another (such as time zone).

For example, we may have the following contextual features: { `trip_type`, `trip_duration`, `origin_city`, `destination_city`, `month` }. If we know that Adam, a resident of Los Angeles, stayed at Club Quarters during his stay in Chicago on business for three days last summer and gave it a rating of 4 stars, the overall context of this rating would be C_{adam} = { "business trip", "3", "Los Angeles", "Chicago", "July" }. If Betsy is seeking a prediction for Club Quarters, but she is coming from Seattle on business and staying for a week in January, Adam's rating would not contribute to her prediction if the algorithm requires a strict match between contexts. This highlights a typical problem in applying context strictly with multiple context dimensions – the more features are used, the greater the sparsity of the recommendation data and the more often it will fail to find any neighbors at all. For this reason, researchers typically concentrate on only one or two contextual features, such as only the trip type.

Suppose we relax the context constraint somewhat, for example: C = { `(exact trip_type), (any duration), (contained time_zone origin_city), (exact destination_city), (any month)` }. We can see that this constraint relaxes the original context by ignoring `month` and `duration` and allowing any `origin_city` in a matching time zone. Now there is a match between the two contexts, and we can use Adam as a neighbor when predicting whether or not Betsy will want to stay at Club Quarters during her trip. The most extreme relaxation of the contextual constraint is to ignore context altogether, resulting in the original context-insensitive prediction algorithm given above.

In this research, we ask the question if we can get better predictions by using different relaxations of the context for different purposes in the same algorithm. In other words, instead of narrowing C to just a few variables, which is what researchers typically do to avoid excessive sparsity, we define relaxations of C applicable to different parts of the prediction function. For example, we might use `trip_type` to select neighbors, but `origin_city` to establish a user baseline. Above, we identified three functions of context in Resnick's algorithm: neighbor selection, neighbor baseline, and user baseline. We will notate the relaxed versions of C applied to each of these areas as C_1, C_2, and C_3, respectively.

As a result, the task of identifying influential contexts becomes the task of discovering the best choices as the context relaxation (C_1, C_2 and C_3) for the three components. Insufficient relaxation will not solve the problem of sparsity and may impair predictive performance. Excessive relaxation will fail to capture the contextual effects that make one circumstance different from another.

3.3 Hybrid Contextual Modeling

Based on the ideas above, we introduce the context relaxation to the three algorithm components mentioned previously and come up a hybrid contextual modeling approach.

We start by defining a new prediction algorithm in which these constraints play a role. We need three new terms:

– N_{C_1} is the set of neighbors of user a, filtered such that those neighbors have rated item i in a context that satisfies contraint C_1.
– $\bar{r}_{u,C_2}$ is the baseline for neighbor u, taking into account only those ratings given in contexts that satisfy C_2.
– $\bar{r}_{a,C_3}$ is the baseline for user a, the target user, using only those ratings given in contexts satifying C_3.

With these modifications in place, we can now state a context-sensitive version of the prediction formula as below, where C_1, C_2 and C_3 are relaxed versions of the original constraint C. Note that $P_{a,i,C}$ turns out to be the predicted rating for user a on item i under contexts C, where r_{u,i,C_2} is selected neighbor u's rating on item i under the relaxed contexts C_2.

$$P_{a,i,C} = \bar{r}_{a,C_3} + \frac{\sum\limits_{u \in N_{C_1}} (r_{u,i,C_2} - \bar{r}_{u,C_2}) \times sim(a,u)}{\sum\limits_{u \in N_{C_1}} sim(a,u)} \qquad (2)$$

Algorithm 1

Require: Context constraints C_1, C_2, C_3
Require: Ratings database R, training set S and testing set T
Ensure: Coverage score
Ensure: RMSE

$E \Leftarrow \emptyset$ (vector of errors)
$k \Leftarrow 0$
for each $t \in T$ **do**
  $a \Leftarrow t.user$
  $i \Leftarrow t.item$
  $c \Leftarrow t.context$
  $r \Leftarrow t.rating$
  $c_1 \Leftarrow$ Apply(c, C_1)
  $c_2 \Leftarrow$ Apply(c, C_2)
  $c_3 \Leftarrow$ Apply(c, C_3)
  $N \Leftarrow$ Neighbors(a, c_1, S)

  if $|N| > 0$ **then**
    $k \Leftarrow k + 1$
  end if
  $p \Leftarrow \bar{r}_{a,c_3} + \frac{\sum_{u \in N} (r_{u,i,c_2} - \bar{r}_{u,c_2}) \times sim(a,u)}{\sum_{u \in N} sim(a,u)}$
  $E = E.\text{append} < r, p >$
end for
$Coverage = k/|T|$
Calculate RMSE from the vector of errors E

Algorithm 1 shows how the experimental methodology is implemented. It assumes that we have a database of ratings R, where a rating consists of a 4-tuple: user, item, rating and contextual features associated with that rating, and that database is split into a training set S and a testing set T. The algorithm takes the three contextual constraints C_1, C_2, and C_3, and applies them to the specific context of the test rating, yielding an instantiation of each constraint. We assume that our neighborhood selection and baseline averaging functions make use of these constraints appropriately. We iterate through all items in the test set, evaluating the context-sensitive predictions with the three constraints each playing their individual role.

We consider this algorithm as a contextual hybrid. A hybrid recommendation algorithm [8] combines two or more recommendation techniques to provide recommendations. In our case, we decompose the user-based CF into three components where the 1^{st} component filters out neighbors who never rate item in matched contexts. This is a form of contextual pre-filtering. However, we also introduce contexts to the other two components in the model, where the selections for C_2 and C_3 are allowed to be different than C_1, thus it is a contextual modeling approach in view that contexts are introduced into the process of modeling. The combination of these three applications of context results in hybrid of contextual pre-filtering and contextual modeling.

3.4 Experimental Setup and Design

We selected hotels in the 120 largest cities in USA from Tripadvisor.com, and crawled ratings as well as geographical location information from user and hotel profiles. We assume that the origin location is the same as the user's geographical information as defined in his or her profile. As mentioned above, many profiles are incomplete, either without home city information or without state information. We opted only to include state and time zone as features for origin location. Users from outside of the USA were removed. The final dataset includes user id, user's state of residence with time zone information, hotel id, hotel city as well as the associated state and time zone information, trip type and user's rating. In order to conduct 3-fold cross-validations, we only include users who have at least 3 ratings in the dataset and then split the dataset into 3 folds using 1/3 of each user's profile. In the process of selecting neighbors, we selected neighbors from the users who have rated the same hotel in the training set, and user similarity is measured by Pearson correlation coefficient.

A closer look at the data revealed that users were selecting multiple contexts for a single review.[2] For example, a business traveler might annotate one stay at a hotel with both "business" and "solo". In these cases, we really have only one rating for the hotel, not multiple ones in different contexts. Therefore, we filtered the context information in this and similar cases. If a rating listed both "business" and "solo" as trip types, we kept only "business" under the assumption that "business" is a stronger descriptor of the traveling situation in those cases. We handled the pairs "business" / "couples" and "business" / "family" by selecting "couples" and "family" respectively, based on

[2] The current version of the Tripadvisor.com website only allows a single trip type to be chosen for each rating/review, but there is still legacy data in the system from when multiple trip type entries were permitted.

the understanding that any time other parties are along for the trip, their needs will influence the acceptability and hence the rating for a hotel. This filtering step ensures that we have only one trip type for each rating and no double counting.

After the pre-processing above, the full dataset before the split contains 2,562 users, 1,455 hotels and 9,251 ratings. 45% of users have multiple ratings in at least one city, and 13% of users had multiple reviews of the same hotel under different trip purposes.

In our experiments, we consider three different contextual variables for hotel recommendation. Trip type is the reason that the user gives for his or her trip. Destination city is the location of the hotel being rated, and origin city is the location from which the user's travel begins. We examine two different types of `contained` constraints for the geographic features (state and time zone) in addition to the `exact` and `any` possibilities that we consider for type type. So, the space of possible relaxations theoretically has 2 possibilities for trip type, 4 for destination, and 4 for origin, giving 32 possibilities for each of C_1, C_2, and C_3, a total space of $32^3 = 32k$ possible constraints to search for our optimal set of relaxations.

The characteristics of our data and the algorithm allow us to reduce the combinatorial complexity considerably. Recall that we select neighbors from users who have rated the same item, thus the item features in the contextual vector C_1 is unnecessary. Similarly, C_2 and C_3 are applied to ratings coming from one user, so user demographic information is redundant as a constraint. With these considerations, the search space becomes greatly reduced in size – $6 \times 8 \times 8 = 384$ possibilities – quite tractable for complete search.

We run an exhaustive search to fully explore the performance of different context relaxations in terms of RMSE and coverage, where coverage indicates the percentage of instances in which neighbors can be found for collaborative prediction.

In our experiments, we compare our algorithm to two classes of competitors. The first type is user-based collaborative recommendation without incorporating context. The second is a typical contextual recommendation approach using contextual pre-filtering. In our formalism, contextual pre-filtering is achieved by applying contextual constraints to C_1 only.

4 Experimental Results

In order to compare how context relaxation performs, we introduce three context strategies in the model: 1) use trip type only (as previous authors have done with TripAdvisor data); 2) use strict contexts without relaxation; or, 3) use relaxed contexts for each component – so that we can compare the performance of trip type, strict contexts and relaxed contexts. Keep in mind that the contextual pre-filtering in our experiments is the one which we only introduce contexts to the 1^{st} component for neighbor filtering. The results of the experiment are shown by Figure 1 and Table 1.

In terms of RMSE, experimental results in Figure 1 show that introducing contexts always outperform the standard CF without contexts. The differences in RMSE between all conditions are significant at the 10^{-3} level using paired Student's t-test, and a p-value of less than 0.01 if using the Wilcoxon signed-rank test. Applying context relaxation outperforms situations where we use trip type only or use strict contexts. Our

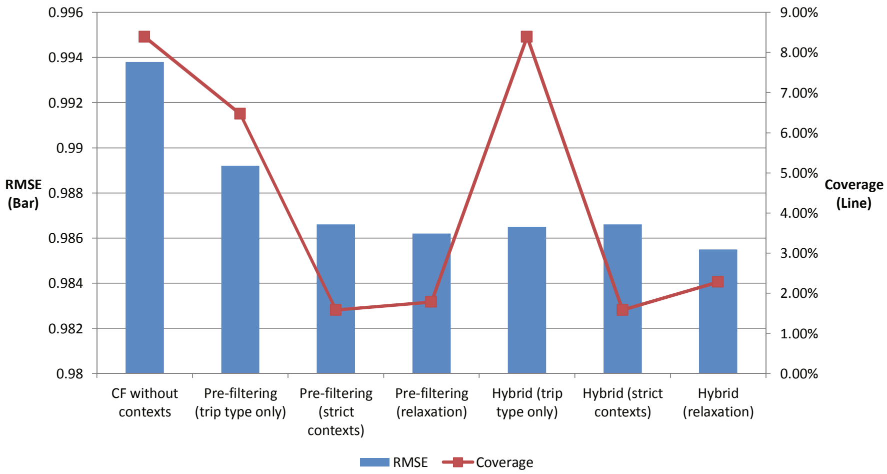


Fig. 1. Comparison of Algorithms in terms of RMSE

hybrid contextual model actually performs better than contextual pre-filtering when using the same context strategy. The best performing one is our hybrid contextual modeling approach using context relaxation, where the RMSE is improved from 0.9938 (the baseline without contexts) to 0.9855.

The coverage in all cases is low (see Figure 1 and Table 1), which means that the default rating is assigned in the vast majority of cases. Recall that C_3 is the constraint that controls which items are chosen to be averaged to create this default. This part of the calculation is the same for the baseline and the optimum algorithms, which means that improved accuracy on a relatively small number of possible predictions accounts for all of the difference in accuracy, which although small are significant.

Recall that when we cannot find a neighbor in our algorithm, we use default predicted ratings where no collaborative predictions and no contextual effects (because the best selection for C_3 is empty without contexts, see Table 1. Thus it is necessary to evaluate the effects contributed by the valid collaborative efforts only – the situation that we can find neighbors and use our hybrid contextual model to make predictions. If we compare the algorithms just on their collaborative predictions – rule out the default predictions and calculate the RMSE on the remaining predictions only, we see a much larger difference between the baseline (RMSE 1.1453) and the optimal relaxation (RMSE 1.0703), an improvement of 6.6%. This suggests that in a dataset with greater density would see an even greater benefit from this technique.

From the Table 1, we can see the optimum set of relaxed constraints can be summarized as follows.

– C_1 = Neighbors filtered based on the state of origin.
– C_2 = Neighbors' baselines computed based on trip type.
– C_3 = Users' baseline computed over all prior ratings.

Trip type and point of origin turn out to be useful aspects of the context in view of the optimal context relaxation for both contextual pre-filtering and our hybrid contextual

Table 1. Comparison of Algorithms

Algorithm	Optimal Contextual Constraints	RMSE	Coverage
CF Without contexts	{}	0.9938	8.39%
Pre-filtering (trip type only)	C_1= {exact trip_type}	0.9892	6.47%
Pre-filtering (strict contexts)	C_1= all contexts	0.9866	1.58%
Pre-filtering (relaxation)	C_1= {(exact trip_type), (contained origin_city state)}	0.9862	1.78%
Hybrid (trip type only)	C_1= {}, C_2={exact trip_type}, C_3={}	0.9865	8.39%
Hybrid (strict contexts)	C_1= all contexts, C_2={}, C_3={}	0.9866	1.58%
Hybrid (relaxation)	C_1= {contained origin_city state} C_2= {exact trip_type}, C_3= {}	0.9855	2.28%

modeling approach – the only difference is they are applied into different components. Our hybrid approach reveals that placing the state of origin in the 1^{st} component and trip type in the 2^{nd} component can help achieve the best RMSE. Destination information seems not to be important here, perhaps because the choice of hotel already subsumes this information. We suspect that if we were considering hotel chains (Hilton, Mariott, etc.) with multiple locations, destination would be much more useful. We also find that "state" is the most effective level of abstraction at which to consider origin information. This is the most specific level at which we have origin data.

To more fully understand these results and the dependencies between each aspect of the algorithm, we performed a sensitivity analysis of this optimum condition, looking at the results in which two of the constraints are held fixed and the other is varied. Holding C_2 and C_3 fixed at their optimal values allows us to see the shape of the optimization space as C_1 changes. See Table 2.

The RMSE difference is not large between these relaxation options. Because C_1 controls which neighbors are allowed to contribute to the recommendation, the coverage goes down significantly when the constraint is strict and increases at the more relaxed settings. In all, we see that there is benefit to using origin information, even though some coverage is sacrificed in doing so.

Table 2. Top 3 Relaxations for C_1 in Hybrid Approach

Optimal Relaxation for C_1	Contextual Constraints	RMSE	Coverage
Top-1	C_1 ={ (contained origin_city state)}	0.9855	2.28%
Top-2	C_1 ={ (exact trip_type), (contained origin_city state)}	0.9857	1.78%
Top-3	C_1 ={ (contained origin_city timezone)}	0.9859	5.48%

We performed a similar analysis for C_2 and C_3. The results are shown in Table 3. Coverage is not impacted by these constraints, so it is omitted from the tables. As suggested in prior research, matching trip type is important for establishing a useful neighbor baseline via C_2. The destination city has only a weak impact on accuracy.

For C_3, we find that there is a large difference between using all of the user's data (empty constraint) and more focused calculation of the user baseline. This constraint shows the sharpest RMSE "valley" in the constraint space.

Table 3. Top 3 Relaxations for C_2 and C_3 in Hybrid Approach

Optimal Relaxation for C_2	Contextual Constraints	RMSE
Top-1	C_2 ={ (exact trip_type)}	0.9855
Top-2	C_2 ={ (exact trip_type), (contained dest_city timezone)}	0.9857
Top-3	C_2 ={ (exact trip_type), (contained dest_city state)}	0.9858
Optimal Relaxation for C_3	**Contextual Constraints**	**RMSE**
Top-1	C_3 ={ }	0.9855
Top-2	C_3 ={ (exact trip_type), (exact dest_city)}	1.0172
Top-3	C_3 ={ (exact dest_city)}	1.0220

Therefore the remaining contexts in our optimal model are trip type and the state of origin, which is consistent with Klenosky and Gitelson [13]'s findings in their research, where they point out that origin is functionally related to the distance and the time users needed for travel. In this case, origin usually plays an important role for personal travel – the non-business trip, which occupies the largest percentage in our dataset.

The tradeoff between coverage and context-sensitivity appears in this study as it has in many others. Our baseline algorithm already has very poor coverage (below 10%) and our optimal variant has only about 1/4 of that. However, the fact that accuracy can be improved at that level of coverage (and without altering the user baseline C_3) means that we are greatly improving recommendations for the small number of cases in which neighbors matching the origin state can be found. We believe this is a very good sign because it suggests that even small increases in density (more neighbors) may yield substantial accuracy improvements for our contextual recommendation approach.

5 Conclusion

Researchers in context-aware recommendation have long known that the danger of sparsity arises if contextual information is applied too strictly. The usual response is to eliminate most contextual variables from consideration, and to focus on the one or two most salient.

In this work, we introduce the ideas of algorithm decomposition and context relaxation, and apply these ideas to a user-based collaborative recommendation algorithm. We break the algorithm into three components and show how contextual constraints can be applied independently in each component. Through exhaustive search, we locate optimal relaxations for our travel data set, and show that error can be significantly reduced. This benefit comes at the cost of reduced coverage, but coverage is not as low as it is when the context is applied strictly. Our approach can be easily applied to other algorithms and datasets, and we plan to do so in future research. We expect that, in other datasets, we will not be able to use exhaustive search over all possible combinations of constraints, and in our future work, we will need more efficient techniques to search the constraint space.

References

1. Abowd, G.D., Dey, A.K., Brown, P.J., Davies, N., Smith, M., Steggles, P.: Towards a Better Understanding of Context and Context-Awareness. In: Gellersen, H.-W. (ed.) HUC 1999. LNCS, vol. 1707, pp. 304–307. Springer, Heidelberg (1999)
2. Adomavicius, G., Mobasher, B., Ricci, F., Tuzhilin, A.: Context-aware recommender systems. AI Magazine 32(3), 67–80 (2011)
3. Adomavicius, G., Tuzhilin, A.: Context-aware recommender systems. In: Recommender Systems Handbook, pp. 217–253 (2011)
4. Anand, S.S., Mobasher, B.: Contextual Recommendation. In: Berendt, B., Hotho, A., Mladenic, D., Semeraro, G. (eds.) WebMine 2006. LNCS (LNAI), vol. 4737, pp. 142–160. Springer, Heidelberg (2007)
5. Baltrunas, L., Amatriain, X.: Towards time-dependant recommendation based on implicit feedback. In: ACM RecSys 2009, the 1st Workshop on Context-Aware Recommender Systems, CARS 2009 (2009)
6. Baltrunas, L., Ludwig, B., Peer, S., Ricci, F.: Context relevance assessment and exploitation in mobile recommender systems. Personal and Ubiquitous Computing, 1–20 (2011)
7. Böhmer, M., Bauer, G., Krüger, A.: Exploring the design space of context-aware recommender systems that suggest mobile applications. In: ACM RecSys 2010, the 2nd Workshop on Context-Aware Recommender Systems, CARS-2010 (2010)
8. Burke, R.: Hybrid recommender systems: Survey and experiments. User modeling and User-adapted Interaction 12(4), 331–370 (2002)
9. Chan, E., Wong, S.: Hotel selection: when price is not the issue. Journal of Vacation Marketing 12(2), 142–159 (2006)
10. Domingues, M., Jorge, A., Soares, C.: Using contextual information as virtual items on top-n recommender systems. In: ACM RecSys 2009, the 1st Workshop on Context-Aware Recommender Systems, CARS 2009 (2009)
11. Hariri, N., Mobasher, B., Burke, R., Zheng, Y.: Context-aware recommendation based on review mining. In: Proceedings of the 9th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems (ITWP 2011), p. 30 (2011)
12. Huang, Z., Lu, X., Duan, H.: Context-aware recommendation using rough set model and collaborative filtering. Artificial Intelligence Review, 1–15 (2011)
13. Klenosky, D., Gitelson, R.: Travel agents destination recommendations. Annals of Tourism Research 25(3), 661–674 (1998)
14. Liu, L., Lecue, F., Mehandjiev, N., Xu, L.: Using context similarity for service recommendation. In: 2010 IEEE Fourth International Conference on Semantic Computing (ICSC), pp. 277–284. IEEE (2010)
15. Lombardi, S., Anand, S.S, Gorgoglione, M.: Context and customer behavior in recommendation. In: ACM RecSys 2009, the 1st Workshop on Context-Aware Recommender Systems, CARS 2009 (2009)
16. Vargas-Govea, B., González-Serna, G., Ponce-Medellín, R.: Effects of relevant contextual features in the performance of a restaurant recommender system. In: ACM RecSys 2011, the 3rd Workshop on Context-Aware Recommender Systems, CARS 2011. ACM (2011)

Multi-criteria Ratings for Recommender Systems: An Empirical Analysis in the Tourism Domain

Matthias Fuchs[1] and Markus Zanker[2]

[1] Mid Sweden University,
83125 Östersund, Sweden
matthias.fuchs@miun.se
[2] Alpen-Adria-Universität Klagenfurt,
9020 Klagenfurt, Austria
markus.zanker@aau.at

Abstract. Most recommendation systems require some form of user feedback such as ratings in order to make personalized propositions of items. Typically ratings are unidimensional in the sense of consisting of a scalar value that represents the user's appreciation for the rated item. Multi-criteria ratings allow users to express more differentiated opinions by allowing separate ratings for different aspects or dimensions of an item. Recent approaches of multi-criteria recommender systems are able to exploit this multifaceted user feedback and make personalized propositions that are more accurate than recommendations based on unidimensional rating data. However, most proposed multi-criteria recommendation algorithms simply exploit the fact that a richer feature space allows building more accurate predictive models without considering the semantics and available domain expertise. This paper contributes on the latter aspects by analyzing multi-criteria ratings from the major etourism platform, TripAdvisor, and structuring raters' overall satisfaction with the help of a Penalty-Reward Contrast analysis. We identify that several a-priori user segments significantly differ in the way overall satisfaction can be explained by multi-criteria rating dimensions. This finding has implications for practical algorithm development that needs to consider different user segments.

1 Introduction

Recommender systems (RS) are tools for consumer decision support that help to overcome information overload in online environments. Their purpose is to point users to items that best match their presumed preferences and needs. Different basic paradigms of recommender systems exist: *collaborative filtering* builds on the assumption that peers with similar ratings and behavior in the past will also have comparable preferences in the future; *content-based filtering* assumes that users' tastes can be semantically described and therefore proposes items whose content descriptions are similar to what is already known that the user likes; *knowledge-based recommendation* systems try to mimic sales agents that exploit

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 100–111, 2012.

domain expertise in order to best possible match elicited customer needs to items. In the RS literature different interpretations of *knowledge* are common [1], for instance, explicitly encoded knowledge bases that contain sets of logical sentences [2] and case-based recommendation approaches that require domain knowledge for defining similarity functions [3]. The first two families of algorithms tackle recommendation as a Machine Learning task, where from known training data models are learned and subsequently employed to predict unseen or withheld data in a second step. In contrast knowledge-based recommendation in the sense of [2] does not follow an inductive learning approach but requires explicitly coded expertise, for instance, in the form of business rules and constraints that should constitute plausible heuristics and procedures in the eyes of experienced sales personnel. For more details on recommender systems and the genesis of the research field see [1,4].

The present work now tries to build bridges between recommendation systems that are mainly driven by codified knowledge such as [5,2] or [6] and majorly learning and data-driven approaches. Algorithms for multi-criteria recommender systems typically learn predictive models by exploiting the different rating dimensions without exploiting any domain specific knowledge [7,8,9]. However, we propose to explore models that are in accordance with theoretical findings in the respective application field. In our case the application domain is tourism where empirical findings about the structure of service quality judgments need to be considered. Therefore, we propose to consider user segmentation and to incorporate models from consumer (i.e. tourist) satisfaction research such as the Kano model [10]. Results from analyzing ratings on the major tourism platform, TripAdvisor.com, indicate users belonging to different market (i.e. travel) segments. We apply Brandt's [11] Penalty-Reward-Contrast analysis in order to explore if and how different rating criteria constitute dissatisfiers (i.e. hygiene factors) or excitement factors in different travel segments.

According to Compete, Inc. (2007) in 2006, 52% of U.S. online shoppers visited at least one community website before having bought their travel and tourism services. Moreover, 26% of U.S. tourists deliver feedback on a community website in connection with their trip. The TripAdvisor portal[1] is considered as the biggest and most famous tourism-related social network site worldwide. In numbers, the strongly interlinked portal counts more than 7,000 URLs and shows above 30 million unique users [12]. Visitors around the globe constantly use this web 2.0 portal to write and read assessments about service quality experiences concerning specific hotels. At TripAdvisor user evaluations are recorded both in the form of standardized items and free-texts. Although the online platform is specialized in hotel and accommodation products, a series of chat-room services related to travel services, restaurant services, as well as trip ideas are additionally provided [13].

Next, related work on multi-criteria recommendation is discussed. Section 3 describes the applied methodology, details empirical results and discusses implications for algorithm development. Finally, perspectives on future work are discussed and a conclusion section is provided.

[1] www.TripAdvisor.com

2 Related Work

Several lines of work have successfully exploited multi-criteria ratings to improve the accuracy of recommendations. An early and very encompassing article that proposes a *contextualized* view on ratings is Adomavicius et al. [7]. Although their users still provide unidimensional ratings, the situational context of users adds additional dimensionality to the ratings. This means that ratings are labeled with contextual parameters such as user's age, sex or weekday and that recommendations for a male user in the twenties on a Monday are presumably more accurate if the system preferably only exploits ratings that have been added by users in their twenties, who are males and who have rated the item on a Monday. Adomavicius & Kwon [8] proposed also a recommendation approach that can exploit real multi-criteria ratings, i.e. a user provides rating values that appreciate different aspects of the same item. In order to determine the overall rating of a specific user for an item their algorithm performs three steps: first, like in traditional recommender systems working with unidimensional ratings for each criterion a rating value is computed; second an aggregation function is estimated that allows computing an overall rating from the multiple criteria ratings; finally, third the overall rating value is computed and recommendable items are ranked according to the estimated overall rating value.

Jannach et al. [9] further developed the ideas of Adomavicius & Kwon. They also employed accommodation ratings from a major tourism platform in their evaluation scenario. They compare regression models that constitute specific aggregation functions for each user and each item. Their results are, for instance, that regression models learned with a classifier using a support vector machine perform better than linear least squares regression models and that a weighted combination of user and item specific regression models perform best in their case.

TripAdvisor data is particulary of interest in tourism research, for instance doing research on complaint management or aggregating review data to the destination level. However, the work of Graebner et al. [14] is also focusing on predicting users' rating values. In contrast to our work, Graebner et al. exploit the users' textual reviews in order to predict the overall rating value. Using both textual review and the multi-criteria ratings in order to predict the user's overall assessment value could be a future extension for this work.

3 Empirical Analysis

Today, TripAdvisor represents the world largest and most successful social networking and community site in tourism comprising over 25 Mio unique users [12]. The platform facilitates the reviewing of hotels around the world and brings together individuals in discussion forums and provides users with independent travel reviews and comments. Figure 1 depicts TripAdvisor's view on the rating feedback for an accommodation of an arbitrary user. Users can rate a hotel according to 7 different dimensions: value for money, quality of rooms, location of

Fig. 1. Detailed view on user rating

the hotel, cleanliness of the hotel, quality of check-in, overall quality of services and particular business services. In addition, users provide unidimensional overall ratings on hotels (not depicted in Figure 1). These standardized evaluation-items are consistently measured on the base of a 5-point scale (i.e. from excellent to terrible). Furthermore, users are explicitly asked if they would recommend the hotel to a friend. By contrast, the recommendation to visit a hotel is measured by two separate binary ratings (i.e. recommend: yes/1; no/0). Standardized assessments of 62,290 unique users concerning hotels from 14 touristic cities (e.g. Vienna, Munich, London, New York, Singapore, San Francisco, Hong Kong, Sydney, Orlando, etc.) were collected during January 2010 by processing web crawling data. Controlled by the user profile (e.g. age, travel motive, trip type, etc.) weighting schemes related to (e.g. hotel) service quality domains are considered as ideal input to parameterize electronic recommender systems ([6]). The goal of the subsequent data analysis is to identify empirical relationships between the users' willingness to recommend a hotel, their partial assessments of the different rating dimensions as well as user profile and context information about their stay.

As a first methodological step the strength of the empirical dependency between users' willingness to recommend the hotel to a friend and their overall rating has been quantified by using a logistic regression. As to be expected, the empirical results clearly show that the overall quality assessment related to a hotel is an excellent determinant for both positive and negative hotel recommendations. From Table 1 emerges that Nagelkerke's R^2's rank well above the threshold value of 0.2 and about 90% of cases are correctly classified by the estimated logistic function ([15]). The difference between the sum of YES and NO values for the dependent variable willingness to recommend and the overall N results from missing values, i.e. users can provide an overall rating and a textual review, but they are not forced to provide detailed ratings for the 7 dimensions or answer if they would recommend the hotel to a friend. As the strong dependency of willingness to recommend on the overall rating value has been confirmed, we will analyze next if the overall rating value strongly depends on the 7 rating dimensions.

Table 1. Results from Logistic Regression - Hotel recommendation and Overall Assessment

Variable in Equation	Regression Coefficient	Wald Statistic	Sig.	Exp(B)
Overall assessment	1.948	13,160.401	.000	7.016
Constant	-5.665	8,852.921	.000	.003
Nagelkerke's R^2	.574			
Log-Likelihood	34,198.27			
% correctly classified:	89.4%			
N	62,290			
Dependent Variable: Recommendation = YES (49.872)/1				
Overall assessment	-2.591	10,619.689	.000	.075
Constant	7.171	7,802.032	.000	1,301.034
Nagelkerke's R^2	.696			
Log-Likelihood	22,161.47			
% correctly classified:	92.4%			
N	62,290			
Dependent Variable: Recommendation = No (9.933)/0				

3.1 Segmentation of User Base

The above empirical confirmation of a strong and significant relationship between hotel recommendation and overall quality assessment leads to the next question about the role of the various antecedents (i.e. hotel quality domains) in affecting overall assessments. More precisely, in arriving at a holistic assessment concerning a particular (e.g. hotel) service experience, consumers typically 'weight' their overall assessment according to the relative importance of particular quality dimensions [16]. Applied to our TripAdvisor data, the relative level of determinance of the above mentioned seven hotel quality domains on overall assessment is identified by using multiple regressions [17,18]. Moreover, in order to show the adequateness of the proposed approach to generate useful input data for electronic recommender systems, regressions were run with respect to four tourist segments previously defined by TripAdvisor data: segment 1: senior couples (i.e. age above 50, leisure trip, staying with spouse in 4/5 star hotel), segment 2: business tourist solo (i.e. age between 35 and 50, business trip, staying alone in 4/5 star hotel), segment 3: budget family tourist (i.e. age between 35 and 50, leisure trip, staying with partner & children in 0-3 star hotel), and, finally, segment 4: youth tourists (i.e. age below 25, leisure trip, staying with friends in 0-3 star hotel). The emerging weighting schemes related to the quality domains in determining the overall assessment are shown in Table 2.

To start with, all models show a strong explanation power (Adj. R^2), are statistically significant (F-Value), and are free of auto-correlated residuals (Durbin Watson) or multi-correlated variables (Variance Inflation Factor). Thus, the quality of TripAdvisor data looks satisfactory for being used to identify how

Table 2. Multiple Regression Results - Determinance of Overall Assessment by Partial Quality Domains*

A-priori Segments	Segment 1: Senior Tourist Couples			Segment 2: Business Tourist Solo		
	Adj. R^2 = .787 F = 114.21 DW = 1.91 N = 1,284 Share = 8.8%			Adj. R^2 = .807 F = 260.24 DW = 1.85 N = 1,366 Share = 9.3%		
Quality domains	Beta	T-Value	VIF	Beta	T-Value	VIF
Value	0.384	8.274	2.169	0.332	9.211	2.927
Rooms	0.247	4.861	2.610	0.271	7.386	3.031
Locations	0.039	1.157	1.175	0.091	3.788	1.273
Cleanliness	0.128	2.692	2.276	0.122	3.264	3.138
Checkin	0.081	1.755	2.089	0.048	1.475	2.386
Service	0.096	1.742	3.057	0.161	4.242	3.221
Business	0.178	4.427	1.625	0.252	2.885	1.718
A-priori Segments	Segment 3: Budget Family Tourist			Segment 4: Youth tourist & friends		
	Adj. R^2 = .769 F = 183.36 DW = 2.23 N = 2,302 Share = 15.7%			Adj. R^2 = .698 F = 45.99 DW = 2.19 N = 875 Share = 6%		
Quality domains	Beta	T-Value	VIF	Beta	T-Value	VIF
Value	0.444	11.026	2.687	0.266	3.442	2.692
Rooms	0.203	4.684	3.103	0.459	5.411	3.245
Locations	0.055	2.010	1.236	0.185	3.443	1.308
Cleanliness	0.179	4.342	2.820	0.081	0.947	3.272
Checkin	0.107	1.991	1.990	0.021	0.287	2.458
Service	0.044	1.077	2.760	0.131	1.662	2.801
Business	0.058	1.952	1.455	0.093	1.514	1.684

* The (adj.) Coefficient of Determination R^2 is the proportion of variability in data accounted for by the statistical model;
Beta is a measure of how strongly a predictor influences the dependent variable;
An F- or a T-test are statistical tests in which the test statistic has an F or a T-distribution under the null hypothesis.
The null hypothesis is rejected if the F or T-value calculated from the data is greater than the critical value of the F- or T-distribution for some desired false-rejection probability (e.g. 0.05).
The Durbin Watson (DW) Test detects autocorrelation (i.e. residuals from a multiple regression model are independent).
Variance Inflation Factor (VIF) quantifies the degree of multicollinearity (i.e. correlated predictor variables) in regression analyses.

various (i.e. hotel) quality domains determine the overall assessment among different customer segments ([18]). For instance, the results show that the relatively strongest and most general determinance stems from both, the perceived 'value for money' and the 'room quality' (i.e. Beta, T-Value > 2; Tab. 2). However, for youth tourists 'room quality' becomes most important, while for the budget family segment 'value-for money' is the most critical unique quality domain.

Interestingly enough, the remaining quality domains are playing completely different roles in determining overall quality assessments among the hotel customer segments. For instance, not only for business tourists, but also for senior tourist couples a convenient 'business environment' seems to be a crucial quality domain, whereas the 'location' factor only plays an insignificant role. Moreover, for business tourists, in contrast to all other segments, the general 'service quality' of the hotel is the third most important quality domain. Furthermore, budget family tourists put the third biggest emphasis on the 'cleanliness of the hotel', while for the youth tourist segment the 'location factor' becomes a relatively important determinant of their overall quality judgement (Table 2).

3.2 Penalty-Reward-Model

Next to the purely quantitative role of quality domains in determining overall assessments, literature also discusses their relevance from a qualitative point of view [19]. Already since the nineties researchers have begun to tackle empirical problems of service quality perception with a multi-factor-structure of customer satisfaction [20]. This model has been adopted and empirically validated both, in a service marketing and tourism context, respectively [21,22,23,24,25]. The three-factor structure of customer satisfaction was first defined by Kano [10]. Based on his model, quality attributes may be grouped into three categories, each of which exerts a different impact on customer satisfaction:

- *Basic factors* are minimum requirements that cause dissatisfaction if not fulfilled but do not lead to customer satisfaction if fulfilled or exceeded; negative performances with these quality domains has a greater impact on overall satisfaction than a positive one. Hence, basic factors are expected by the customer (i.e. regarded as prerequisites).
- *Excitement factors* are factors that increase customer satisfaction if delivered but do not cause dissatisfaction if they are not delivered; positive performance on these quality dimensions has a greater impact on overall satisfaction than a negative one.
- *Performance factors* lead to satisfaction if performance is high and lead to dissatisfaction if performance is low. In this case, the attribute performance-overall satisfaction relationship is linear and symmetric [24].

Based on Brandt's [11] Penalty-Reward-Contrast analysis a method to empirically decipher the factor-structure of customer satisfaction is presented next. The method employs a dichotomised regression analysis using dummy variables ([18]). More precisely, one set of dummy variables exemplifies in quantitative form excitement factors, while a second set expresses basic factors. In order to carry out the analysis using our TripAdvisor data the 5 point scales of the independent variables (i.e. from 5= excellent-1= terrible) were recoded in a way that scores of 5 were used to form a first dummy variable (i.e. representing the quantification of the excitement factor with a value of 1). Due to the empirical distribution of the (i.e. independent) variables scores of 3, 2 and 1 (i.e.

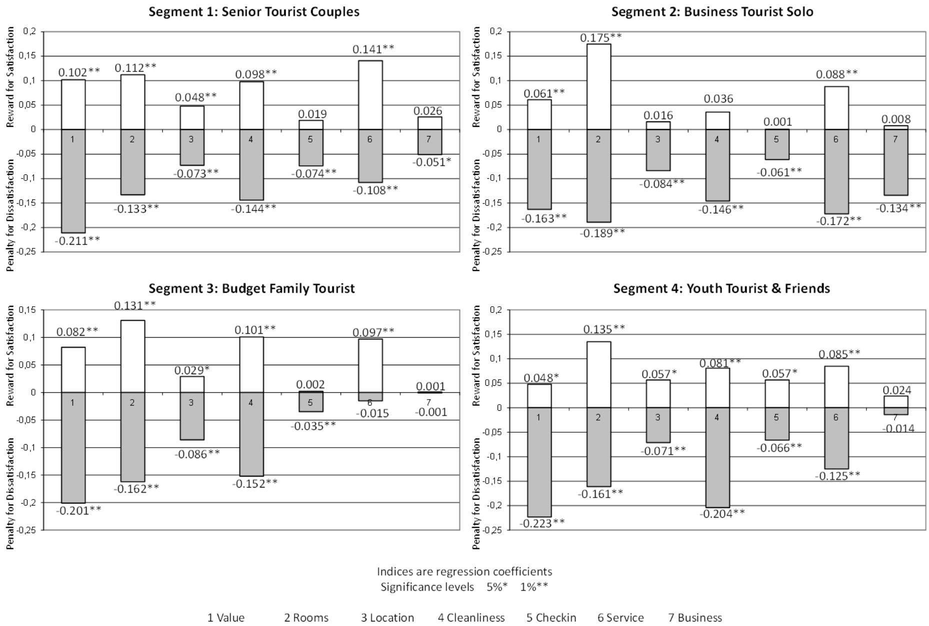


Fig. 2. Penalty Reward Contrast Analysis

relatively low satisfaction) were chosen to create a second dummy representing the quantification of the basic factor with a value of 1. Although this approach shows some degree of arbitrariness, the empirical distribution of the raw data is taken into consideration and is, thus, recommended by [19,22,23,24]. Finally, empty cells of both dummies were recoded with a value of zero. With the help of this recoding multiple regression analyses were carried out to quantify basic requirements and excitement factors using the *overall rating* assessment as the dependent variable and the two dummy variables for each of the seven quality domains as independent variables. 'Penalties' can now be expressed as the incremental decline associated with low levels of satisfaction, while 'rewards' become expressed as the incremental increase associated with high satisfaction to be observed within a certain hotel quality domain. Thus, if penalty levels surpass reward levels the respective quality domain is a basic factor. Otherwise, if the reward index surpasses the penalty value the quality dimension should be interpreted as an excitement factor. Finally, if reward and penalty values are rather similar, the quality domain will contribute to tourist satisfaction only when its level of performance is high. It will lead, at the same time to dissatisfaction when the performance is low (i.e. performance factor). Using the described Penalty-Reward approach and applying it to the seven hotel quality domains gathered from TripAdvisor data the following results emerged (Fig. 2).

As with the previous tests, again all regressions show a strong explanation power (Adj. R^2 between 0.681 and 0.723), are statistically significant (F-Value between 74.28 and 429.24), and are free of auto-correlated residuals (Durbin Watson between 1.87 and 2.02) and multi-correlated variables (Variance Inflation Factor between 1.224 and 2.087). Thus, the adequateness of TripAdvisor data for indentifying the factor-structure of customer satisfaction for (i.e. hotel) quality domains among various segments could be well confirmed ([18]).

Interestingly enough, for all customer segments results revealed a lack of pure delighting factors (i.e. the positive reward index surpassing the negative penalty value; Figure 2). Obviously, the (i.e. hotel) quality dimensions measured by TripAdvisor are generally perceived as performance or basic factors, respectively. More precisely, while for all customer segments the 'room quality' shows a relatively strong potential to increase overall satisfaction (i.e. if its performance level is high), for senior tourist couples also the 'general service quality' shows potentials to 'delight'. However, at the same time the quality domains 'value for money' 'room quality' and 'cleanliness' show large penalty potentials (i.e. to decrease overall assessment if performance is low). A completely different picture emerged for the business tourist segment, since both 'business convenience' and 'general service quality' emerged as quality dimensions with relatively large penalty potentials (Figure 2). Moreover, for the budget family tourists segment, fully consistent with previous results, next to 'room quality' and 'cleanliness' also the 'location factor' emerged as a quality dimension with strong penalty potentials. Finally, results for the youth tourist segment may be interpreted in analogous fashion.

To summarize, the proposed approach revealed significantly differing determinance profiles (Table 2) and penalty-reward profiles (Figure 2) between the examined customer segments. These insights are particularly valuable, since weighting schemes attached to (e.g. hotel) quality domains build the basis to form overall quality assessments [16,21]. Thus, segment-specific recommendation strategies might simultaneously consider determinance and penalty-reward profiles, consequently recommending those products (e.g. hotels) that show highest performance values in those quality domains that emerged as significant in respective weighting schemes.

3.3 Discussion and Implications

Obviously, the willingness to recommend a hotel to a friend strongly depends on the overall rating value as has been shown in Table 1. The 7 specific rating dimensions from "value for money" to "business services" capture most of the signal to determine the overall rating value (Adj. R^2 clearly above 0.7 for most regression models in Table 2). However, this relationship between multi-criteria ratings and overall rating is clearly moderated by the tourist segment (user profile data and travel context), i.e. the relative influence of the specific rating dimensions changes for the different segments. Furthermore, a penalty-reward-contrast analysis unveils a qualitative interpretation of multi-criteria ratings as basic and excitement factors and indicates differences for the different segments. To summarize, the proposed methods revealed both, plausible and statistically significant results that have implications for the development of future recommendation algorithms:

- Conversational and knowledge-based recommendation systems can explicitly consider these differing weights users attribute to different criteria when appreciating items and can elicit users' travel segment affiliation in order to adjust utility weights when ranking search results.
- Collaborative multi-criteria recommendation approaches could apply segment specific weights when determining overall ratings and recommended items. Furthermore, the existence of basic and excitement factors indicate non-linear relationships between factor ratings and overall ratings that could be better modeled when ratings are re-coded according to the Kano model.
- Knowing about the qualitative differences in the appreciation of different criteria can be used to generate segment-specific item descriptions and explanations [26,27] in order to not only more accurately predict items of interest but also to create more persuasive [28] interaction experiences.

Consequently, future research is needed to investigate whether recommendation strategies using segment-specific weighting schemes and a re-coding of ratings according to the Kano model that differentiates between basic and excitement factors outperforms traditional regression models. The approach of [9] actually learns user-specific and item-specific regression models, which means that weights are not adapted on the segment level but instead on the more fine-granular user and item level. However, the approach of [9] cannot be applied to personalize offerings to novel or *cold-start* users and ignores that the same user identity might travel in different contexts, e.g. business trip vs. family.

As a sidenote authors also would like to mention that TripAdvisor has modified its rating criteria since data extraction for this paper took place. The two dimensions *Checkin* and *Business Services* have been deleted and the dimension *Sleep Quality* has been introduced. This partly corresponds to our findings (Figure 2) as Business Services only insignificantly influences overall rating assessments for Segments 3 and 4. However, Checkin has been shown to be a basic factor for all four segments.

4 Conclusions

The paper presented an empirical analysis of multi-criteria ratings harvested from the TripAdvisor portal. The results indicate a significant and strong moderating effect of travel segment in multiple regression models where dimensional ratings predict the users overall rating. Furthermore, a Kano model that qualitatively differentiates rating dimensions into basic and excitement factors indicates non-linear relationships between multi-criteria ratings and overall rating. The paper outlines how the theoretically plausible and statistically significant findings can serve as a basis for further refinements of future recommendation algorithms.

Acknowledgment. Authors acknowledge the financial support from the European Union (EU), the European Regional Development Fund (ERDF), the Austrian Federal Government and the State of Carinthia in the Interreg IV Italien-Österreich programme (project acronym O-STAR). Furthermore, the authors would like to thank Srinivas Donepudi for performing major parts of the data extraction task.

References

1. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: Recommender Systems An Introduction. Cambridge University Press (2010)
2. Zanker, M., Jessenitschnig, M., Schmid, W.: Preference reasoning with soft constraints in constraint-based recommender systems. Constraints 15(4), 574–595 (2010)
3. Ricci, F., Venturini, A., Cavada, D., Mirzadeh, N., Blaas, D., Nones, M.: Product Recommendation with Interactive Query Management and Twofold Similarity. In: Ashley, K.D., Bridge, D.G. (eds.) ICCBR 2003. LNCS, vol. 2689, pp. 479–493. Springer, Heidelberg (2003)
4. Ricci, F., Rokach, L., Shapira, B., Kantor, P. (eds.): Handbook on Recommender Systems. Springer (2011)
5. Zanker, M., Bricman, M., Gordea, S., Jannach, D., Jessenitschnig, M.: Persuasive Online-Selling in Quality and Taste Domains. In: Bauknecht, K., Pröll, B., Werthner, H. (eds.) EC-Web 2006. LNCS, vol. 4082, pp. 51–60. Springer, Heidelberg (2006)
6. Jannach, D., Zanker, M., Fuchs, M.: Constraint-based recommendation in tourism: A multi-perspective case study. Information Technology & Tourism 11(2), 139–155 (2009)
7. Adomavicius, G., Sankaranarayanan, R., Sen, S., Tuzhilin, A.: Incorporating contextual information in recommender systems using a multidimensional approach. ACM Transactions on Information Systems 23(1), 103–145 (2005)
8. Adomavicius, G., Kwon, Y.: New recommendation techniques for multicriteria rating systems. IEEE Intelligent Systems 22, 48–55 (2007)
9. Jannach, D., Gedikli, F., Karakaya, Z., Juwig, O.: Recommending hotels based on multi-dimensional customer ratings. In: International Conference on Information and Communication Technologies in Tourism (ENTER), pp. 320–331. Springer (2012)
10. Kano, N.: Attractive Quality and Must-be Quality. Hinshitsu: The Journal of the Japanese Society for Quality Control 14(2), 39–48 (1984)
11. Brandt, R.D.: How service marketers can identify value enhancing service elements. Journal of Services Marketing 2(3), 35–41 (1988)
12. O'Connor, P.: User-generated content and travel - a case study on tripadvisor.com. In: International Conference on Information and Communication Technologies in Tourism (ENTER), pp. 47–58. Springer (2008)
13. Dippelreiter, B., Gruen, C., Poettler, M., Seidel, I., Berger, H., Dittenbach, M., Pesenhofer, A.: Online tourism communities on the path to web 2.0 - an evaluation of virtual communities in travel and tourism. Information Technology & Tourism 10(4), 329–353 (2007)
14. Graebner, D., Zanker, M., Fliedl, G., Fuchs, M.: Classification of customer reviews based on sentiment analysis. In: 19th Conference on Information and Communication Technologies in Tourism (ENTER), pp. 460–470. Springer (2012)

15. Hosmer, D., Lemeshow, S.: Applied Logistic Regression, 2nd edn. Wiley, New York (2000)
16. Klaus, P.: Quality Epiphenomenon: The Conceptual Understanding of Quality in Face-to-Face Service Encounters. In: The Service Encounter: Managing Employee Customer Interaction in Service Business, Lexington, pp. 17–33 (1985)
17. Weiermair, K., Fuchs, M.: Measuring tourist judgments on service quality. Annals of Tourism Research 26(4), 1004–1021 (1999)
18. Hair, J.E., Anderson, R.E., Bubin, B.J., Tatham, R.L., Black, W.C.: Multivariate Data Analysis, 6th edn. Prentice-Hall, New York (2006)
19. Busacca, B., Padula, G.: Understanding the relationship between attribute performance and overall satisfaction: Theory, measurement and implications. Marketing Intelligence & Planning 23(6), 543–561 (2005)
20. Johnston, R.: The determinants of service quality: Satisfiers and dis-satisfiers. International Journal of Service Industry Management 6(1), 53–71 (1995)
21. Fuchs, M., Weiermair, K.: New perspectives on satisfaction research in tourism destinations. Tourism Review 58(3), 6–14 (2003)
22. Matzler, K., Sauerwein, E.: The factor structure of customer satisfaction: An empirical test of the importance grid and the penalty-reward-contrast analysis. International Journal of Service Industry Management 13(4), 314–332 (2002)
23. Matzler, K., Bailom, F., Hinterhuber, H., Renzl, B., Pichler, J.: The asymmetric relationship between attribute-level performance and overall customer satisfaction: A reconsideration of the importance-performance analysis. Industrial Marketing Management 33, 271–277 (2004)
24. Fuchs, M., Weiermair, K.: Destination benchmarking: An indicator-system's potential for exploring guest satisfaction. Journal of Travel Research 42, 212–225 (2004)
25. Mikulic, J., Prebežac, D.: Prioritizing improvement of service attributes using impact range-performance analysis and impact-asymmetry analysis. Managing Service Quality 18(6), 559–576 (2008)
26. Zanker, M., Ninaus, D.: Knowledgable explanations for recommender systems. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI/IAT), pp. 657–660. IEEE (2010)
27. Friedrich, G., Zanker, M.: A taxonomy for generating explanations in recommender systems. AI Magazine 32(3), 90–98 (2011)
28. Yoo, K.H., Gretzel, U., Zanker, M.: Persuasive Recommender Systems - Conceptual Background and Implications. Springer (2013)

Leveraging Social Media Sources to Generate Personalized Music Playlists

Cataldo Musto, Giovanni Semeraro, Pasquale Lops,
Marco de Gemmis, and Fedelucio Narducci

Department of Computer Science,
University of Bari Aldo Moro, Italy
{cataldomusto,semeraro,lops,degemmis,narducci}@di.uniba.it

Abstract. This paper presents MyMusic, a system that exploits social media sources for generating personalized music playlists. This work is based on the idea that information extracted from social networks, such as Facebook and Last.fm, might be effectively exploited for personalization tasks. Indeed, information related to music preferences of users can be easily gathered from social platforms and used to define a model of user interests. The use of social media is a very cheap and effective way to overcome the classical *cold start* problem of recommender systems. In this work we enriched social media-based playlists with new artists related to those the user already likes. Specifically, we compare two different enrichment techniques: the first leverages the knowledge stored on DBpedia, the structured version of Wikipedia, while the second is based on the content-based similarity between descriptions of artists. The final playlist is ranked and finally presented to the user that can listen to the songs and express her feedbacks. A prototype version of MyMusic was made available online in order to carry out a preliminary user study to evaluate the best enrichment strategy. The preliminary results encouraged keeping on this research.

Keywords: Music Recommendation, Social Media, Personalization, DBPedia.

1 Introduction

The concept of Information Overload describes a state where the efficiency is jeopardized by the amount of available information [7]. This definition perfectly fits with the Web navigation scenario, where users are overwhelmed by the continuous flow of data. Hence, Information Filtering (IF) tools, such as Recommender Systems (RS) are more and more needed, since their main goal is to optimize the access to data sources and provide users just with the relevant information according to their preferences. It is common to refer to these systems as *personalization systems* [17].

Even if the recent evolution of the Web is universally considered as one of the causes of information overload [10], the Web 2.0 phenomenon further worsened

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 112–123, 2012.

the problem and changed the rules for personalization. Indeed, the recent spread of social networks and collaborative platforms makes physiologically impossible to follow the information flow in real-time. Recent studies showed that more than 20 hours of video are uploaded every minute to YouTube[1]. So, despite users spent 22% of their web navigation time on these platforms[2], the current scenario made the problem of information overload today felt also for images, video and audio contents. Consequently, it is necessary to adapt IF tools and techniques in order to make them able to properly handle multimedia contents as well.

However, it is not correct to consider Web 2.0 solely as a source of problems. The other side of the coin is that social networks can be a rich source of information useful to automatically infer user preferences, and this can be helpful to mitigate the *cold start* problem of RS.

The main contribution of this work is MyMusic, a system that leverages different social media sources for generating personalized music playlists. The filtering model behind MyMusic is based on the assumption that information about music preferences can be easily gathered from Facebook profiles. Next, playlists built using explicit Facebook preferences may be enriched with new artists somehow related to those the user already likes. We propose and compare two different enrichment techniques: the first leverages the knowledge stored on DBpedia, the structured version of Wikipedia, while the second is based on the content-based similarity between descriptions of artists. The final playlist is then ranked and finally presented to the user that can express her feedback. A prototype version of MyMusic was made available online and a preliminary user study to detect the best enrichment technique was performed.

The paper is organized as follows. Section 2 gives a general overview of the most relevant related work in the area of music recommendation. The architecture of the systems is sketched in Section 3, together with the approaches for building personalized music playlists. Finally, Section 4 presents the results of the preliminary experimental evaluation, while Section 5 contains the conclusions and the future directions of this research.

2 Related Work

The topic of music recommendation has been widely covered in literature, and a good overview of the state of the art is given in [3] by Oscar Celma.

In the music domain, the commonly used technique for providing recommendations is collaborative filtering, implemented in very well known services, such as MyStrands[3], Last.fm[4] or iTunes Genius. MyStrands is a great service for music discovery and recommendation, based on songs/artists uploaded either from

[1] http://www.youtube.com/t/press_timeline
[2] http://blog.nielsen.com/nielsenwire/social/
[3] http://www.mystrands.com
[4] http://www.last.fm

iTunes playlists or added as favorites on the site. It allows the recommendations of similar songs, albums, and artists. Last.fm is another interesting and widely adopted music recommendation service, whose features are improved by the Last.fm Scrobbler, focusing on the music already played to help users to discover more music. Finally, iTunes Genius is the recommender system implemented within iTunes. It sifts the iTunes library to find songs that might be combined together, in order to obtain compelling compilations, such as personal channels that broadcast music just for that specific user.

An early attempt to recommend music using collaborative filtering was done by Shardanand [22]. However, collaborative filtering methods are not generally able to tackle the problem of finding items within the *long tail* of music production, those for which the amount of taste data is limited. This limits the recommendation of novel and serendipitous items, that is a fundamental feature of music recommender systems.

The recent trend is thus to use content-based recommendation strategies, which analyze diverse sets of low-level features (e.g. harmony, rhythm, melody) [15], or high-level features (metadata or content-based data available in social media) [9] to provide recommendations. Unlike collaborative filtering, content-based models can effectively tackle the problem of *new or unpopular items*, typical of real use cases, where the information flow is continuous and uncontrolled, with new artists and songs continuously published. The most noticeable system using (manual) content-based descriptions to recommend music is Pandora[5], with 100 million registered users, more than 900 thousands songs and 90 thousands artists in catalog. The main problem of the system is the scalability, because the music annotation process is entirely done manually.

Conversely, *FOAFing the music* [5,6] is able to recommend, discover and explore music content, based on user profiling via *Friend of a Friend* (FOAF)[6] descriptions, context-based information extracted from music related RSS feeds, and content-based descriptions automatically extracted from the audio itself. The use of FOAF and in general of Linked Data is investigated in [18], that reports a wide range of music-related data sources that have been interlinked within the Linking Open Data initiative. For example, the DBTune project[7] exports several datasets encompassing detailed editorial information, geolocalization of artists, social networking information amongst artists and listeners, listening habits, and content-based data in RDF format. Such interlinking provides an open *social graph* that can be queried and processed in an uniform way. Inspired by this work, we decided to use DBPedia as a Linked Data source for enriching initial music preferences of users with related artists and songs.

Recently, Bu et al. [2] followed the recent trend of harvesting information coming from social media for personalization tasks and proposed its application for music recommendation. In [12], Lamere investigated the use of textual contents, such as tags, as source for music recommendation as well, and Wang et al. [23]

[5] `http://www.pandora.com`
[6] `http://www.foaf-project.org`
[7] `http://dbtune.org`

showed the usefulness of tags with respect to other content-based sources. Finally, a recent approach which exploits tags to obtain a semantic representation is proposed by Levy [14]. Thus, we decided to evaluate an alternative way that exploits tags for enriching initial music preferences of users with related artists and songs.

Another important feature of music recommender systems is represented by the context, in order to adapt recommendations to different situations. This aspect is out of the scope of the present work, but there is a body of work related to the problem of context-aware music recommendation, taking into account different contextual variables, such as time, weather data, traffic condition or driver's mood for users when in a car [13,19,1].

Finally, a complete picture of recent trends in the music recommendation area can be found in [4].

3 MyMusic: Personalized Playlists Generator

The general architecture of MyMusic is depicted in Figure 1. The playlist generation is performed in different steps, each handled by a specific component. The process is directly triggered by the user, who invokes the PLAYLIST GENERATOR module. The set of her favourite artists is built by mapping her preferences gathered from the Facebook profile with a set of artists extracted from Last.fm. Given this preliminary set, the PLAYLIST ENRICHER adds new artists using different enrichment strategies. Finally, for each artist in that set, the most popular tracks are extracted and ranked. The final playlist is then built and shown to the target user who can express her feedback on the proposed tracks.

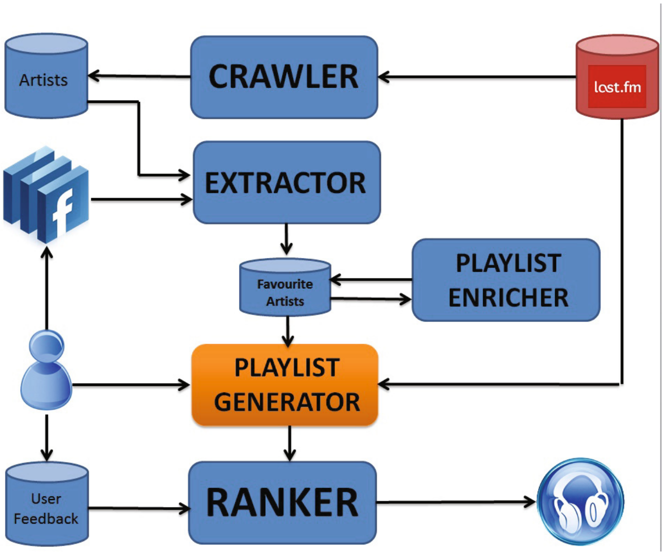


Fig. 1. MyMusic architecture

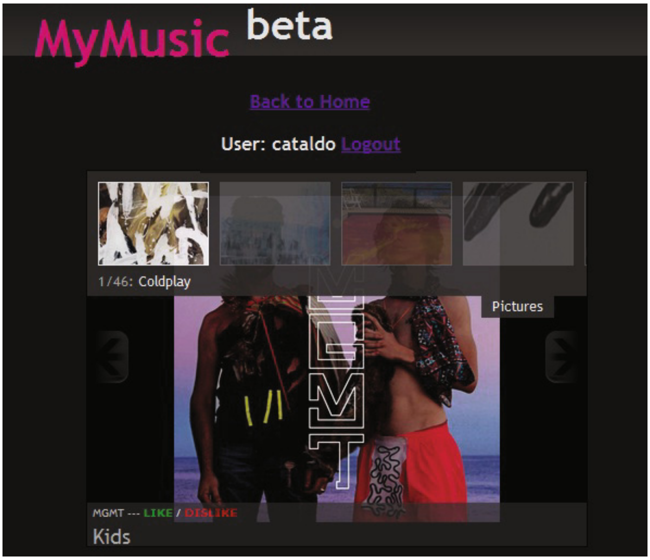


Fig. 2. MyMusic screenshot

A working implementation of MyMusic is available online[8], while a screenshot of the user interface is provided in Figure 2. There is a slideshow which allows the user to move among the tracks in the playlist and start listening some of them with a simple click. The lower part of the interface reports the title of the song, the name of the artist and the like/dislike button to allow users express a binary feedback.

More details about the components and the whole generation process follow.

Crawling Data from Last.fm. MyMusic needs a set of artists to feed the playlist generation process. The CRAWLER module queries Last.fm through its public APIs in order to build a corpus of artists. This process is performed in a batch way, and it is scheduled in order to have an up to date set of data. For each artist in Last.fm, the name, a picture, the title of the most popular tracks, their playcount and a set of tags that describe that artist (see for example Figure 5) are crawled.

Extracting Preferences from Facebook. A common weakness of personalization systems is the need of explicit preferences to learn profile of user interests in order to provide recommendations. Thanks to Web 2.0 applications, and specificically with the advent of social networks, gathering explicit preferences is becoming increasingly simple. Facebook profiles, for example, contain explicit information about the artists preferred by a user (Figure 3). The EXTRACTOR

[8] http://193.204.187.223:8080/sssc

module of MyMusic connects to Facebook profiles, extracts user preferences related to music, and maps them with the data gathered from Last.fm in order to build a preliminary set of artists preferred by that user.

Fig. 3. User Preferences from a Facebook profile

Even if Facebook profiles also contain a lot of implicit data that might be analyzed to infer music preferences, such as links, attended events, content of posts, . . . , the current version of MyMusic only relies on explicit data.

Playlist Enrichment. The main goal of the PLAYLIST ENRICHER module is to add new artists to the set of those preferred by the user and explicitly expressed in her Facebook profile. This should also help to discover new artists.

Formally, given a user u, we define the set of favourite artists extracted from the Facebook profile as $F_u = \{f_1, f_2, ..., f_n\}$. For each f_i, the PLAYLIST ENRICHER builds a set $E_{f_i} = \{e_1, e_2, ..., e_m\}$ that contains the m most similar artists among those stored in the corpus built using Last.fm. Finally, the final set of artists for user u is defined as:

$$E_u = \bigcup_{i=1}^{n} E_{f_i} \tag{1}$$

In this work we have evaluated two different techniques for enriching the playlists with new artists. The first technique is based on the exploitation of DBPedia, while the second one uses the content-based similarity between artist representations.

Enrichment Based on DBpedia. The first approach for enriching the playlist with new artists relies on the exploitation of DBpedia[9], a project whose goal is to represent information stored in Wikipedia in a structured form by means of RDF triples [11].

Nowadays DBpedia represents the nucleus of the so-called Web of Data, since the information is made freely available online and it is possible to submit queries to get complex information by exploiting the structured nature of this data and the relationships among them. In DBpedia, each Wikipedia entry is mapped in a DBpedia concept, and each concept is assigned with a unique Uniform Resource

[9] http://dbpedia.org

Identifier (URI), e.g. HTTP://DBPEDIA.ORG/RESOURCE/COLDPLAY, and each concept can be imagined as a node in a graph. Node pairs are connected by means of relations, called *properties*. As shown in Figure 4, thanks to DBpedia we can represent complex information in a structured form. For example, the following *triple* encodes the fact that Coldplay is a music band whose genre is Alternative Rock.

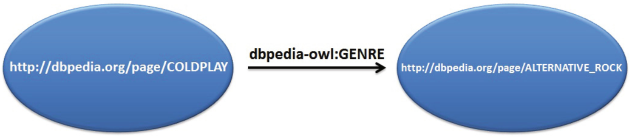


Fig. 4. An example of property in DBPedia

Our approach is based on the assumption that each artist can be mapped to a DBpedia node. The inceptive idea is that the similarity between two artists can be computed according to the number of properties they share (e.g. two Italian bands playing rock music are probably more similar than an Italian and an English band that play different genres). Consequently, we looked for properties that could be useful for computing similarity. We decided to use DBPEDIA-OWL:GENRE (describing the genre of the artist) and DCTERMS:SUBJECT that provides information about the musical category of the artist.

Operationally, we used a powerful query language, namely SPARQL – SPARQL Protocol and RDF Query Language – an RDF query language, similar to a query language for databases, able to retrieve and manipulate data stored in Resource Description Framework format. By simply querying a SPARQL endpoint containing information stored in DBpedia we can extract the set of artists related to those already liked by a user. For example, we extracted the set of artists related to Coldplay by using the following query:

```
select distinct ? artist where {
?artist dbpedia-owl:genre dbpedia:Alternative_Rock.
?artist dcterms:subject category:English_singer-songwriters.
};
```

Since DBPedia returns an unordered set of results, returned artists are ranked according to their playcount in Last.fm. The first m artists returned by the SPARQL endpoint are considered as relevant (and related) and added to the set of the favourite artists.

Enrichment Based on a Content-Based Model. As stated in Section 3, each artist in Last.fm is described through a set of tags as in classical collaborative tagging systems [8]. Each tag provides information about the genre played by the artist, such as *rock*, or about typical features of her songs, such as *melanchonic* (Figure 5). By following the classical Vector Space Model (VSM) [20], each artist can be represented as a point in a *n-dimensional* vector space, whose dimensions are all the n different tags used to describe all the artists (vocabulary).

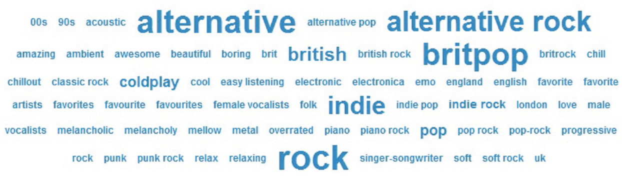


Fig. 5. Coldplay tag cloud from Last.fm

The rationale behind this enrichment strategy is that it might be easy to understand the similarity/relatedness of two artists by comparing tags used to describe them. Besides, we can suppose that artists most related to those the user already likes can be considered as relevant for her, as well. Thus, given a VSM-based representation, we can use *cosine similarity* to compute how much similar or related two artists are. More formally, given two artists a and b, their correlation is computed as follows:

$$cosSim(a,b) = \frac{a \cdot b}{\|a\| \cdot \|b\|} = \frac{\sum_{i=1}^{n} a_i \times b_i}{\sqrt{\sum_{i=1}^{n}(a_i)^2} \times \sqrt{\sum_{i=1}^{n}(b_i)^2}} \tag{2}$$

In this scenario, for each artist in F_u (i.e. extracted from the Facebook profile) we compute the cosine similarity between the artist and all the others extracted from Last.fm. The m artists with the highest cosine similarity are added by the PLAYLIST ENRICHER module to the list of favourite ones.

3.1 Playlist Generation and Ranking

The PLAYLIST GENERATOR module builds the set of candidate tracks by merging the songs played by the artists extracted from the Facebook profile with those played by artists added through the enrichment process. Finally, since the set of candidate tracks need to be ranked, the RANKER exploits a scoring function to sort the tracks in a descending relevance order. For each track in the candidate track list, the score is calculated as follows:

$$score(s_i) = play(s_i) * source(s_i) \tag{3}$$

where $play(s_i)$ is the normalized playcount of the song s_i in the Last.fm community, while $source(s_i)$ is a weight assigned to the source of the song. $source(s_i)$ is set to α for tracks played by an artist in the Facebook profile, and $1 - \alpha$ for tracks played by an artist produced by the enrichment process, in order to weigh differently explicit and implicit preferences.

4 Experimental Evaluation

The goal of the experimental evaluation is to identify the most effective enrichment technique able to generate the most relevant playlists for final users. We carried out the experiment by involving 30 users under 30, heterogeneously distributed by sex and education, according to the availability sampling strategy. None of them performed studies in the music field. The final crawl of Last.fm was performed at the end of November, 2011, and data about 228,878 artists were extracted. For each artist we got also the top-5 tracks, so the final MyMusic dataset contained information about more than 1 million tracks. Each user explicitly granted the access to her Facebook profile to extract data about favourite artists. At the end of the extraction step 325 artists were extracted, so each user had about 10 favourite artists on average. Users left on the platform 462 feedbacks, almost 15 per user on average. For the enrichment technique based on DBpedia the SPARQL endpoint located at `http://dbpedia.org` was queried. It contained 3.64 million nodes and more than 100,000 are related to musical artists or albums. For the cosine similarity-based technique we exploited Last.fm APIs to extract the most popular tags associated to each artist. The less significant and meaningful tags (such as *seenlive*, *cool*, and so on) were considered as noisy and filtered out. In order to identify the best enrichment technique, we asked users to use the application for three weeks. In the first two weeks the system was set with a different enrichment technique, while in the last one a simple baseline based on the most popular artists on Last.fm was used, thus playlists were enriched with tracks played by the most popular artists regardless user preferences.

For each round of the experimental evaluation the users granted the access to their Facebook preferences. Given the playlist generated by the system, we asked them to explicity express their feedback on the tracks played by the artists generated by the enrichment techniques. For each track it was possible to express a binary feedback (whether they liked the song or not) or to simply ignore the suggestion (See Figure 2). Users were not aware of the enrichment technique adopted in that round. Regarding the parameters, the value of α was set to 0.5, so to favourite artists and to those added by the enrichment process was given an equal weight. This value was set through a rough heuristic. The parameter m indicates the number of similar artists extracted for each favourite one, and we compared user behavior in three different configurations, with m set to 1,2 and 3. This means that we enriched each artist extracted from Facebook with respectively 1,2 or 3 new artists. The maximum size of the playlist was set to 50. To get the final results, we computed the *precision* of the system as the ratio between the number of positive feedbacks and the total number of suggested tracks. The overall results are reported in the Table 1.

It is worth to notice that the two enrichment strategies are able to outperform the baseline. This means that data extracted from social networks actually reflect user preferences and the intuition of modeling users according to the information gathered from social sources is valid. The enrichment technique that gained the best performance is that based on the *tag-based similarity*.

Table 1. Results of Experiments

	Artists		
Enrichment strategy	m=1	m=2	m=3
DBpedia	65.9%	64.6%	63.2%
Tag-based similarity	**76.3%**	**75.2%**	**69.7%**
Popularity	58%		

It overcomes the DBpedia-based strategy of roughly 10 points with *m=1* and *m=2* and 6% with *m=3*, and the baseline between 11 and 18 points %. Even though this technique gained the best results, a deeper analysis can provide different outcomes. Indeed, with *m=3* the gap between DBpedia and tag-based similarity decreases of almost 5 points. This means that with higher values of m a pure content-based representation introduces much more noise than DBpedia, whose effectiveness remains constant (only 2 points % of difference among the different configurations). In other terms, with low similarity values, it is likely that the enrichment technique based on tag-based similarity suggests artists not actually interesting for the target user. The introduction of some thresholding strategy might be useful to avoid this problem. The good results obtained by the baseline based on popularity can be justified by the low diversity of the users involved in the evaluation. Since most of them like very common artists, such as U2 and Coldplay, a simple popularity-based approach is very accurate. It is likely that, by involving users with a more specific musical knowledge and uncommon preferences, results obtained by the baseline may get worse.

Despite its results, the DBpedia-based enrichment technique may represent a valuable alternative to avoid the typical drawbacks of pure content-based representations. Indeed, it might be helpful for providing *explanations* about the produced recommendations by analyzing the relations among the favourite and suggested artists. This is a fundamental feature of a good music recommender system, that should be *transparent*, i.e. it should be able to provide a convincing explanation of recommendations. Finally, DBpedia might also help to face the other typical problem of content-based filtering approaches, such as the *overspecialization* [16]. Indeed, suggested artists might be relevant but too similar to those the user likes. This means that suggestions might be accurate but obvious, thus not useful. Hence, we decided to further investigate on the use of DBPedia and Linked Data in order to exploit the wealth of data and relations to obtain more serendipitous (unexpected) results. This allows to focus on some other important aspects besides accuracy for obtaining valuable music recommendations, as devised in a very recent work [21].

5 Conclusions and Future Work

In this work we presented a preliminary version of MyMusic, a system for building personalized music playlists based on social media sources. This work is based on the idea that information extracted from social networks such as Facebook and Last.fm is very meaningful and can be exploited for personalization tasks. We developed a prototype application able to model user preferences in music and to generate personalized playlists. Moreover, we compared two different techniques for enriching the playlist with new artists related to those the user already likes, the first based on DBpedia and the second based on similarity calculations in vector spaces. In the experimental session we asked users to use the system in order to understand which enrichment technique gained the best results: VSM-based approach was the preferred one. Even if this version of the system is only a simple prototype, results emerged from this preliminary evaluation encouraged keeping on this research. In general, there is still space for future work: first, a deeper evaluation with more users and different values of the parameter α might be useful to evaluate the effectiveness of the systems in building playlists totally based on new singers. Next, the enrichment approaches might be extended by analyzing, respectively, different DBpedia properties or different processing techniques for tags in order to avoid the typical problems of synctatical-based representation. Finally, regarding the extraction step, it might be interesting to introduce the analysis of implicit data extracted from Facebook such as attended events, links, groups and so on.

References

1. Baltrunas, L., Kaminskas, M., Ludwig, B., Moling, O., Ricci, F., Aydin, A., Lüke, K.-H., Schwaiger, R.: InCarMusic: Context-Aware Music Recommendations in a Car. In: Huemer, C., Setzer, T. (eds.) EC-Web 2011. LNBIP, vol. 85, pp. 89–100. Springer, Heidelberg (2011)
2. Bu, J., Tan, S., Chen, C., Wang, C., Wu, H., Zhang, L., He, X.: Music recommendation by unified hypergraph: combining social media information and music content. In: Proceedings of the International Conference on Multimedia, MM 2010, pp. 391–400. ACM, New York (2010)
3. Celma, O.: Music Recommendation and Discovery - The Long Tail, Long Fail, and Long Play in the Digital Music Space, pp. 1–194. Springer (2010)
4. Celma, O., Lamere, P.: Music recommendation and discovery revisited. In: Mobasher, B., Burke, R.D., Jannach, D., Adomavicius, G. (eds.) Proceedings of the 5th ACM Conference on Recommender Systems, RecSys 2011, pp. 7–8. ACM (2011)
5. Celma, O., Ramírez, M., Herrera, P.: Foafing the Music: A Music Recommendation System based on RSS Feeds and User Preferences. In: 6th International Conference on Music Information Retrieval (ISMIR), London, UK, pp. 464–467 (2005)
6. Celma, O., Serra, X.: FOAFing the Music: Bridging the Semantic Gap in Music Recommendation. Journal of Web Semantics 6(4), 250–256 (2008)
7. Eppler, M.J., Mengis, J.: The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. Information Society 20(5), 325–344 (2004)
8. Golder, S., Huberman, B.A.: The Structure of Collaborative Tagging Systems. Journal of Information Science 32(2), 198–208 (2006)
9. Hahn, C., Turlier, S., Liebig, T., Gebhardt, S., Roelle, C.: Metadata Aggregation for Personalized Music Playlists. HCI in Work and Learning, Life and Leisure, 427–442 (2010)

10. Ho, J., Tang, R.: Towards an optimal resolution to information overload: an infomediary approach. In: Proceedings of the, International ACM SIGGROUP Conference on Supporting Group Work, GROUP 2001, pp. 91–96. ACM, New York (2001)
11. Kobilarov, G., Bizer, C., Auer, S., Lehmann, J.: DBpedia - a linked data hub and data source for web applications and enterprises. In: Proceedings of Developers Track of 18th International World Wide Web Conference (WWW 2009), Madrid, Spain, April 20-24 (2009)
12. Lamere, P.: Social Tagging and Music Information Retrieval. Journal of New Music Research 37(2), 101–114 (2008)
13. Lee, J.S., Lee, J.C.: Context Awareness by Case-Based Reasoning in a Music Recommendation System. In: Ichikawa, H., Cho, W.-D., Satoh, I., Youn, H.Y. (eds.) UCS 2007. LNCS, vol. 4836, pp. 45–58. Springer, Heidelberg (2007)
14. Levy, M., Sandler, M.: Learning latent semantic models for music from social tags. Journal of New Music Research 37(2), 137–150 (2008)
15. Li, Q., Kim, B.M., Guan, D.H., Whan Oh, D.: A music recommender based on audio features. In: Sanderson, M., Jrvelin, K., Allan, J., Bruza, P. (eds.) SIGIR, pp. 532–533. ACM (2004)
16. Lops, P., de Gemmis, M., Semeraro, G.: Content-based recommender systems: State of the art and trends. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 73–105. Springer (2011)
17. Mulvenna, M.D., Anand, S.S., Büchner, A.G.: Personalization on the net using web mining: introduction. Communications of the ACM 43(8), 122–125 (2000)
18. Passant, A., Raimond, Y.: Combining Social Music and Semantic Web for Music-Related Recommender Systems. In: Social Data on the Web, Workshop of the 7th International Semantic Web Conference, Karlsruhe, Deutschland (October 2008)
19. Reddy, S., Mascia, J.: Lifetrak: music in tune with your life. In: Proceedings of the 1st ACM International Workshop on Human-Centered Multimedia, pp. 25–34 (2006)
20. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Commun. ACM 18(11), 613–620 (1975)
21. Schedl, M., Hauger, D., Schnitzer, D.: A Model for Serendipitous Music Retrieval. In: Proceedings of the 2nd Workshop on Context-Awareness in Retrieval and Recoomendation (2012)
22. Shardanand, U.: Social information filtering for music recommendation. Master's thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science (1994)
23. Wang, D., Li, T., Ogihara, M.: Are tags better than audio? the effect of joint use of tags and audio content features for artistic style clustering. In: ISMIR, pp. 57–62 (2010)

Computational Commerce: A Vision for the Future

Alegria Baquero and Richard N. Taylor

Institute for Software Research,
University of California, Irvine
{abaquero,taylor}@ics.uci.edu
http://www.isr.uci.edu

Abstract. The proliferation of goods and services offered online and the growing number of e-consumers are catalysts for the ongoing burgeoning of e-commerce. Many industries have adopted e-commerce technologies to optimize and automate business processes. Despite the co-dependent relation between fundamental e-commerce components—negotiation, contracts, and business workflow—research and development is greatly done in isolation, conveying divergent and disconnected technologies. We provide a vision of the future of e-commerce along with a model of decentralized computation exchange—grounded on the COAST architectural style—which consolidates negotiation, contracts, and business workflow.

Keywords: Software architecture, E-commerce, Contracts, Workflow.

1 Introduction

E-commerce has come a long way since its beginnings in the mid-1990's, now being a leading activity in the Internet. Purchasing an array of goods and services—such as books, clothes, and airline tickets—is today a matter of a few clicks. Consumers need not to concern about the machinery of inter-organization processes set in motion. Behind user interfaces, market analysis, negotiation, procurement, production, alliances, and other interactions among industries occur.

Business-to-business (B2B) e-commerce technology attempts to optimize intra- and inter-organizational activities, and to provide security mechanisms, and in so doing reduce administrative and operational costs, and increase trust.

However, integration within current e-commerce applications is a challenging problem that comes at a high price, not only among organizations, but is a well-known in-house problem as well. *Despite the existence of successful e-commerce applications, full exploitation of co-dependent business components—negotiation, contracts, and workflow—is elusive. Their smooth integration, automation, and dynamic co-adaptation is a goal, not a current reality.*

In our vision, e-commerce between business and consumer and between businesses will be driven by a secure, dynamic, and decentralized exchange of computations between actors [1]—peers deliberately enabled to execute computations—where a computation is a live, stateful process that is collaboratively evolved.

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 124–136, 2012.

We explore the possibility of achieving the fluid architectural and technical integration of negotiation mechanisms, contract formalisms, and business workflow by providing a model where decentralized actors—on behalf of organizations—communicate and cooperate by exchanging secure mobile computations to accomplish mutual business goals.

COmputAtional State Transfer (COAST) [2][1] is an architectural style which subsumes content under computational exchange among autonomous and distributed peers. COAST provides the design principles for building decentralized, secure, and highly adaptive systems. By describing a novel model for building e-commerce systems our intention is coming closer to this vision of the future.

Following, we present a motivating scenario (Sect. 2). We describe the existing gap between negotiation, contracts, and workflow (Sect. 3), and provide a novel computation-based perspective (Sect. 4). We present COAST's foundations (Sect. 5), and propose a model for computation-based e-commerce (Sect. 6). We describe related work (Sect. 7), and provide conclusion remarks (Sect. 8).

2 A Motivating Scenario

In a B2B e-commerce scenario, a pet store sends computations to a group of providers to inquire the price and delivery time of pet food. The result of the computations executions are the corresponding suppliers' quotes. Following evaluation, a digital contract is established with the selected supplier. The agreement shapes parties' workflows, where proprietary and external actors cooperate through a dynamic exchange of computations to comply with the contract terms.

The pet store's procurement actor sends a computation to the bank requesting a funds transfer to the provider's account for 50% of the purchase price. A credit is observed by a supplier-controlled remote actor—which monitors the account activity—and notifies accounting and logistics actors. Logistics coordinates the order fulfillment with production and delivery actors.

A long-running computation controls an RFID reader which regularly scans tags on the ready-to-ship pallets. Based on this data, the delivery actor coordinates with a shipping service the pet food dispatch. At the pet store's warehouse, an RFID reader notifies procurement about incoming supplies so that the inventory is updated and the outstanding balance is paid according to the contract.

In a more complex scenario, the inventory computes the required food supplies by periodically querying the kennel. Identifying a food shortage, a procurement actor is notified so that suppliers are invited to bid for the next pet food batch. Suppliers' bidding actors—executing at the pet store—adjust their negotiation strategy based on operational cost and capability data gathered on-the-spot. At the biding deadline, an e-contract is established with a supplier.

When multiple parties cooperate, unexpected situations arise such as a delay in the attainment of pet food packing sacks. In the effort to comply with the delivery date, procurement uses the Better Business Bureau's remote services to

[1] Formerly known as Computational Representation State Transfer (CREST).

find alternative accredited, highly rated suppliers. After a particular company claims product availability, negotiations on price and delivery time begin.

The pet food store's negotiation strategy adapts according to the urgency of product need, reputation and previous relations with the supplier, and the penalties for delayed delivery to its customer. The packaging company's leads a yielding and compromising negotiation given the customer's high market value—retrieved through a remote service—anticipating consideration on future business. Negotiation persists until demand and supply equilibrium is reached.

The new commercial partnership requires the workflow's dynamic adaption to cope with the new just-in-time-delivery of supplies. The logistics actor evaluates the new workflow—involving the automated interaction of autonomous parties—and predicts no further delays on the delivery of pet food to the customer.

Later on, a broken part causes the malfunction of the pet feed bin. Multiple computations are sent to part suppliers for quotes. Despite the most expedited delivery offering, it is concluded that delivering the pet food within the agreed time frame will not be possible. Alternatives to deal with contract breach are evaluated. The first alternative is to pay the penalties stipulated in the contract. The final dues—including penalties—are calculated by an accounting actor once the supplies are delivered. A second alternative is to go back to negotiation where, for example, the supplier offers to double the amount of delivered product in compensation. If a new agreement between consumer and supplier is reached, both contract and cross-organizational workflow are dynamically adapted.

Negotiation is not limited to two parties. The pet store might commit to purchase a certain amount from both a cat food and a dog food suppliers if they agree to deliver a combined shipment. Further negotiation takes place among suppliers to evaluate if the proposal is feasible and profitable. If an agreement is reached, their workflows are dynamically linked on the delivery phase.

The described flexibility can be extended to business to consumer markets to achieve a higher degree of personalization. A customer can negotiate a discount on dog toys based on the purchase quantity, or the pet store might refund shipping expenses if a customer's product recommendation leads to further sales.

Security is imperative in a model based on computation exchange. At checkout, the customer sends a computation to the pet store containing credit card data. Consumer and provider exchange self-certified and cryptographically unforgeable messages, allowing authentication and message integrity verification. This way, credit card fraud and compromised customers' data are deterred. Furthermore, computations execute within environments with restricted access to the host's functionality, thus impeding malicious computations to damage access the host's file system. This mechanism can be also leveraged to provide customers' differential access to a service. For example, a service may offer augmented functionality—access to pet training courses—to premium customers.

The secure integration and dynamic co-adaptation of negotiation, contracts, and business workflow that we describe is difficult to achieve with current technologies, which tend to treat these e-commerce phases individually, without considering their co-dependance and integration requirements.

3 Negotiation, Contracts, and Workflow: A Pragmatic Gap

3.1 Contracts

A business contract is a legally-binding agreement among parties engaging in economic exchanges and transactions. Commercial contracts involve a sequence of statements describing permissions and obligations held by each party.

With the increasing number of online transactions, the need to formalize and execute business contracts has become apparent. Formalizing complex contracts in expressive machine languages is still a challenging problem subject to many research questions [3], for example regarding: a) the essential information–and detail level—to be captured in a contract abounding in legal jargon; b) existence of important clauses not prone to formal modeling; c) providing flexibility to handle contingencies unforeseen at contracting time.

The expected benefits of e-contracts are eliminating natural language ambiguity, enabling automation of contract execution, identifying conflicting or contradictory clauses, inter-agency accountability and awareness, hypothetical reasoning, and guidance to workflow, among others.

Research on contracts for e-commerce have focused on the explicit formalization of contracts, exception handling, contract monitoring, enforcement, and execution, as well as development of research prototypes.

3.2 Negotiation

Negotiation is the ongoing conversation between self-governing parties to reach a mutually beneficial agreement. Most often, the main criteria to negotiate in a commercial transaction are the nature of the exchanged goods, quantity or use period, price, and time for delivery.

Negotiation is an important component of a new generation e-commerce systems, mostly for B2B markets where there are complex interactions among supply chain partners. Although, B2B negotiation scenarios are common, auctions are a popular B2C and C2C negotiation mechanisms, where a product is sold to the highest offer in a defined time period.

A goal long desired, and focus of research work in the last 15 years, has been the automation of negotiations. Digital agents negotiate on behalf of humans and organizations by collecting data, evaluating offerings, searching for deals, and making utility-maximizing decisions. However, in today's e-commerce, humans are still responsible of evaluating offers and making decisions. Negotiation is a complex problem—both for humans and systems—where social, economic, legal, and management factors interplay.

Research in computer-based negotiations involve negotiation mechanisms—protocols and strategies for auctions, and bilateral and multilateral negotiations—communication languages, agent-based negotiation, and auctioning and negotiation prototypes [4]. Agents are considered appropriate for negotiation as they capture the autonomous, self-interested nature of the represented individuals and organizations.

3.3 Workflow

To fulfill business, commercial, and organizational goals, agencies share information, carry out co-dependent activities, and aim for coactive business units. Workflow technology attempts capturing cooperation of humans and software to efficiently support business activities, increase activity awareness, and reduce operational costs. Workflow broadly refers to either business processes, process specifications, process automation, or coordination/collaboration systems [5].

Some of the field challenges are: a) mapping high-level business concepts to tasks performed by a network of systems and people; b) the smooth integration of heterogeneous systems and data models; c) the automation of intra- and inter-organizational business processes; d) a decentralized and collaborative fine-grained workflow specification; e) dynamic adaptation to changing requirements.

Research in this domain involves workflow specification, reengineering, and automation technologies [5], as well as business process management systems and e-commerce frameworks—successors of workflow management systems (WfMS)—which focus on high-level business concerns as opposed to operational processes.

3.4 The Pragmatic Disconnect

A commercial activity relies upon the synergy between negotiation, contracts and inter-organizational workflow. Research and development in both academia and industry strive for interoperability and automation within these domains. However, research in these areas is largely done in isolation—few studies address all three concerns—when in practice they are tightly related and co-dependent.

The consequence of this fragmented body of literature is an assortment of technologies specific to a particular business phase, with insufficient or no effort to provide a holistic architecture which integrates negotiation, specification, and execution of commercial contracts. Research on these fields is carried out mostly in different academic and industrial settings. Communities such as the S-Cube network share our concerns towards research fragmentation and integration issues, and call for more cohesive research agendas.

Leveraging these contributions suggests nontrivial effort to semantically and syntactically bridge these technologies, which in most cases are specific for either negotiation, contract formalisms, or workflow. Space constraints impede an extended discussion on the limitations of B2B technologies, but these are mainly the difficulty to compare, manage, and make sense of this technology, and the lack of dynamicity, flexibility, and widely-adopted standards for business interoperation [6]. Also, dynamically including new participants is difficult, market hubs are centralized, lack critical mass, do not support differential offerings, or interact with other markets resulting in fragmented markets [7]. Lastly, web services are not readily customizable to user needs and offer restrictive data exchange.

Undoubtedly, these are complex domains with specific concerns which require a specialized and in-depth approach to research. However, our intention is to provide the architectural foundations for building e-commerce systems—involving negotiation, contracts, and workflow phases—in practice.

4 Negotiate, Contract, and Work: A Computations World

Negotiation, contracting, and business processing are axiomatic commerce phases. In both human and digital realms, commerce is carried out through sequential and concurrent activities or computations which result in some business state.

When parties negotiate, an offer is accepted, rejected or a counteroffer is made. The analogy between this process and a computation's input/output is evident. This conceptual similarity suggests that negotiations can be carried out through the dynamic exchange of computations among agencies. The state of the computation embodies the state of the negotiation, and parties collaboratively advance the negotiation by executing and sending computations. Agencies leverage other computations' result to adapt their strategy. For example, a company acquiring chocolate products uses commodities market's data on cocoa beans.

The result of negotiation is a contract which specifies permissions, obligations, and parties' roles, and how business processes—computations themselves—should be carried out. For example, a contract may read "the pet store pays 50% of the price to the provider, after which the provider delivers the pet food to the customer. Upon receipt, the pet store pays the provider the remanding balance."

Business processes can be enacted by the exchange of computations between business units. For example, procurement sends raw materials to production, and the product is then forwarded to packaging and delivery. Each actor executes a computation whose result is dispatched to the next unit in the business chain.

Since the underlying commonality of these propositions are computations, the integration of negotiation, contracts, and workflow is instinctively reduced to a uniform computation execution and exchange between domain specific peers.

There are several benefits to the aforementioned integration. First, transitioning from negotiation to contract formation is smooth—as well as going back to negotiation contingent upon contract nonconformity—when there are shared underlying principles. Second, workflows can be derived from contracts to ensure that inter-agency agreements are followed. Third, a contract is a tool for monitoring and—if convenient—enforcing compliance with collective business goals with the flexibility to fulfill tasks in the most convenient way. Lastly, adopting a unifying architectural style affords a higher degree of co-dependent adaptation. For example, contract renegotiation propagates changes to business processes, and a workflow activity that hinders contract compliance reactivate negotiations.

Interoperation involves communicating, accessing remote data, and using the obtained information successfully with low overhead. The advantage of leveraging computations over current integration efforts is subsuming and generalizing these communications—independently of their nature—through the exchange of computations. The distinctive characteristic that sets apart this model from common request/response protocols are the power and expressiveness granted by function composition which allows tailoring required information to individual needs, as well as innately supporting scalability and decentralization.

We sustain that *the key principle for a smooth integration of negotiation, contracts, and workflow technologies is the bidirectional flow of computations among self-interested, decentralized, and autonomous actors.*

5 COAST Architectural Foundations

COAST is an architectural style for the design and development of distributed, decentralized, adaptive, secure, collaborative, and customizable applications. The key principle is peer interaction driven by the bilateral exchange of computations. We provide a concise description—due to space constraints—of COAST's principled approach towards application concerns. For a more in-depth technical discussion about the style and its supporting infrastructure, please refer to [2].

Topology. Departing from client-server architectures, COAST builds on hierarchical arrangements of *actors* grouped by *clans* within *islands*—unique IP address/port pairs. Actors—computations themselves—map incoming messages to sending messages to other actors, executing some behavior, or to creating new actors [1]. A chieftain—a distinguished actor—creates, terminates, and grants capabilities to actors in the clan, and controls the clan's resource consumption.

Communication. Actors—within the same or different clan—communicate uniquely through asynchronous messages. Mutable, shared memory is prohibited. Messages may contain primitive types, data structures, and mobile computations—closures, continuations, and binding environments. Content is simply a side effect of computation exchange. The execution of a computation might result in the creation or *spawning* of a new actor.

Addressability. *Capability URLs (CURLs)* name computations—*binding environment* and *execution engine* tuples—with authority-to-execute semantics. This capability model grants differential access privileges to actors' environments and islands' fungible resources—processor cycles, bandwidth, and memory.

Application state is driven by computation exchange, where control flows among actors. Continuations can be shipped and later resumed at another actor's execution environment, thus state progresses in a decentralized fashion.

Application adaptation can be achieved through on-the-fly actor spawning and *environment sculpting*—namely, deriving new binding environments from existing ones. Actors can dynamically gain or modify their capabilities by transitively obtaining new binding environments in a message.

System scalability is promoted by decentralized actors, message passing, actor spawning, and environment sculpting. Moreover, sandboxes—a security model which limits execution privileges—regulate actors' resource consumption and chieftains can terminate actors in violation.

Security. CURLs authorization and capability-based security provides controlled access to execution environments. Capabilities conferred to an actor can not be greater than those of its parent actor. Furthermore, the interpretation of computations sent in messages are CURL-specific, thus no computation is allowed outside the scope of the authorized binding environment. CURLs are unforgeable, self-certifying, encrypted, and signed. Moreover, clan- and actor-specific sandboxes provide an extra layer of security.

Customization is fostered by CURLs, which endorse custom services through the deliberate access or restriction to execution environments' capabilities.

6 A Model for Computation-Based E-Commerce

By capitalizing on a model of mobile code, self-certified CURLs, a hierarchy of processing actors, and URL-specific capabilities, COAST enables building secure, decentralized, and proactively evolving e-commerce applications.

COAST supports multi-organization cooperation through the fluid migration of computations. The hierarchical topology of self-interested processing peers spontaneously maps to the dynamics of commercial associations and activities.

Actor decentralization, distribution, and loose coupling allows scalability of e-commerce systems, where parties join and leave business liaisons when convenient at different points in time. Furthermore, COAST's provisions for dynamic adaptation are optimal for modification, diversification, and augmentation of business services. Additionally, authority-based CURLs allow tailoring services for different kind of customers, achieving fine-grained personalization.

COAST enables the design of secure e-commerce systems. Unforgeable, self-certified connections and encrypted messages are sufficient to create secure connections capable of handling sensitive data, such as financial information.

E-commerce components—negotiation, contracts, and business workflow—can be modeled as an assembly of specialized, interacting, and collaborating actors. The underlying architectural structure—islands, clans, actors, and binding environments—and communication mechanisms—asynchronous messages—are the same whether actors perform negotiation, accounting, or production roles.

6.1 Negotiation as Exchanged Computations

Computations that negotiating actors exchange might contain—at the simplest—a product description, price, and quantity. For example, actor A may send actor B a computation `(evaluate-contract pet-food $400 50 initial-proposal)` to be executed in B's binding environment. B interprets `evaluate-contract` in the context of its own environment, for A and B might evaluate contracts in a different way. B might test the offer against the organization's operational capacity and adaptation costs, while A independently tests the offer against B's reputation, retrieving this information by sending a computation to a third party. B's response to A may be a computation `(evaluate-contract pet-food $450 50 counter-offer)`, hence evolving the negotiation until an equilibrium is reached or a party withdraws from the negotiation. Multiple issues can be concurrently negotiated and negotiation strategies can adapt dynamically according to newly gathered information. For example, B might change its negotiation strategy after being notified about a more significant incoming product order.

Continuations—namely snapshots of computation states (including stack values)—can be leveraged to implement negotiation mechanisms. After an offer is evaluated—resulting in a rejection, approval, or counteroffer—the state of the computation is captured, sent, and resumed at the negotiating counterpart. A counteroffer may be the modification or augmentation of the offer computation.

6.2 Contracts as Computations

A contract can be described by a data structure which holds parties' roles, product descriptions, prices, payment and delivery methods, deadlines, and other clauses. Obligations and permissions in this data structure are described as computation descriptions, hence as processes. For instance, the statement "the pet store shall pay the food provider $450 on December 31st" can be translated to `(pay pet-store food-provider $450 "12/31/2012")`. The contract can be delivered as an argument to `evaluate-contract`, and thus exchanged between the negotiating parties.

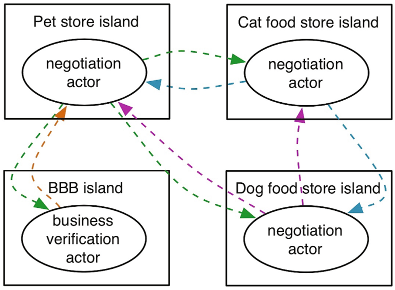


Fig. 1. Multi-party contract negotiation

6.3 Computation-Driven Workflow

At the partaking agreement each party evaluates its obligations, and derives a set of alternative workflows to achieve the required goals. For example, several process alternatives can fulfill the clause *"deliver 300 pounds of pet food to customer"*. A specialized actor evaluates the alternatives based on cost estimations.

Alike negotiation, cross-organizational workflow can be driven by business processing actors that collaborate to complete tasks. For instance, an actor in control of production can complete some task, capture the state of the overall process and send it to the packaging actor, which continues with its part of the process before passing this live computation—the business process—to a shipping actor. An auditor actor constantly monitors the process schedule and checks it against the contract terms to report potential delays.

Computation exchange is not only unilateral—from negotiation to workflow—but bilateral, and encouraging information retrofit and permanent integration. For example, workflow management actors inform negotiation actors of production and shipping costs in order to lead profitable negotiations. Also, if contract non-compliance is predicted, computations can flow back to a negotiation stage.

Dynamic adaptation to new requirements is a big aspiration in workflow research. Environment sculpting, actor spawning, and extending actors' capabilities—by receiving a binding environment or an actor's CURL in a message—are all supported forms of adaptation in COAST. With these mechanisms, workflows can scale, be modified, migrate completely or partially, and be dynamically linked among them according to changes in operations or business goals.

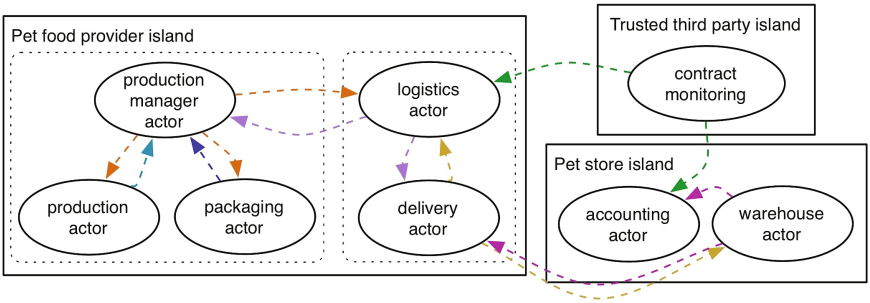


Fig. 2. Cross-organizational workflow

6.4 Supporting Infrastructure

The first generation of infrastructure for building COAST applications has been developed and described in [2]. *Motile*—a domain-specific, functional mobile code language—enables computation exchange through network serialization and recompilation. When a computation is sent to another peer, free variables within the lexical scope are included on the serialized closure, while the global binding environment is left behind. Free global variables are rebound to the receiving actor's binding environment. Deserialization, recompilation, and execution take place in a clan-specific sandbox and an actor-specific binding environment. Additional security mechanisms such as connection self-certification and message encryption, as well as actor's Capability URLs are provided at the infrastructure level. This means that applications built with this infrastructure need not to worry about security mechanisms, but security comes for free.

This infrastructure provides a testbed for experimentation on building COAST-based e-commerce applications where negotiation, contracting, and workflow co-evolve and are instinctively integrated through computation-exchanging actors.

7 Related Work

Few studies are concerned with the synergic integration of negotiation, contracting, and workflow processes. Milosevic et. al [8] provide a framework of contract templates, contract negotiation, validation, monitoring, and enforcement.

Judge et.al [9] present a distributed process management architecture, where agents negotiate task assignment contracts and coordinate workflow adaptation.

COSMOS [10] supports contract execution by deriving Petri Nets from contracts, and provides adaptors for existing WfMS. Limitedly supports negotiation.

In MEMO [11], a marketplace mediates negotiation and contracting. A contract repository informs a workflow manager for linking distributed workflows.

The ER-EC framework [3] derives executable workflows from XML-based contracts, while E-ADOME [12] models, monitors, and enforces contracts based on workflow views which balance inter-agency trust and security, and allow holding contracts with many partners. Both address dynamic workflow adaptation.

In Cheung et. al [13] a negotiation plan is derived from negotiable issues in a contract template. The plan specifies the order and co-dependency of negotiables.

CrossFlow [14] supports cross-agency workflow linkage through dynamic service outsourcing. Services are specified in contract templates and the execution infrastructure is dynamically set up according to a contract specification.

These studies have limitations with respect to the symbiotic relation of negotiation, contracts, and workflow, where—in some cases [3,12,13,14]—at least one of them is not supported. Automation is not fully supported [8,9,10,13], as well as exception handling and run-time adaptation of negotiation, contracts, and/or workflow [8,3,11,14]. Lastly, multi-party negotiation scenarios are mostly ignored and proposed architectures require substantial effort to integrate systems addressing different e-commerce stages.

An orthogonal but related body of work is that of mobile-agents, leveraged for example in advertising applications such as in Mahmoud et. al [15].

The distinction between agents—popular in negotiation—and COAST's actors is that agents exist at a higher, application level, and preserve their identity as they migrate between hosts. In contrast, actors do not migrate but enclose expressions within messages, and spawned actors have their own identity. Moreover, agents allow shared memory, while actors only communicate through messages.

Service Oriented Architectures have been advocated as an integration alternative by exposing business processes as services. The fundamental difference is that while in SOA the service provider has full control over fixed services which only return content, COAST empowers users to tailor services to their needs through environments—actors—where custom function compositions execute. The benefits are desired, fine grained services, with substantially reduced bandwidth use (only useful data is returned), and significantly increased security.

8 Conclusions and Future Work

E-commerce is—as of now—tightly woven into the fibers of society and increasingly making a more significant impact on the world economy. Needless to say, research on improving e-commerce processes and security is essential.

Negotiation, contract, and business workflow research and development have been carried out mostly in isolation, resulting in a myriad of fractured technologies that pose difficulties for mutual integration. We propose a pragmatic approach for building e-commerce systems that integrate negotiation, contracts, and inter-organizational workflow under a model of computational exchange.

COAST's architectural principles are appropriate for the design of a new generation of e-commerce systems where decentralization, security, dynamic adaptation, and fluid interoperation of business components pertaining to different facets of e-commerce are essential to conduct efficient and trusted commerce. COAST spontaneously captures the complex interactions that involve, not only bi-party cooperation, but multiple parties participating in business endeavors.

This preliminary work presents a conceptual model for the aforementioned integration within e-commerce systems. Future work includes experimentation with the COAST infrastructure in developing e-commerce prototypes. A prototype application has been developed and presented as part of [2], where an encoder, for example, migrates between camera-provided islands capturing video streams. Subscribed decoders can migrate or be copied to other video-displaying islands. A variety of other dynamic adaptations have been demonstrated.

Many questions remain to be explored regarding the representation of contracts as computations, and how transactions—crucial in e-commerce—can be achieved with computation exchange. Also, challenging scenarios need to be devised for evaluating COAST's performance against other architectural solutions.

Acknowledgment. Thanks to Kyle Strasser for his insights. This material is based upon work supported by the National Science Foundation, Grant Nos. 0820222 and 0917129. Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Agha, G.: Actors: a model of concurrent computation in distributed systems (1985)
2. Gorlick, M., Strasser, K., Baquero, A., Taylor, R.N.: CREST: principled foundations for decentralized systems. In: SPLASH Proceedings, pp. 193–194 (2011)
3. Krishna, P., Karlapalem, K., Dani, A.: From contracts to e-contracts: Modeling and enactment. Information Technology and Management 6(4), 363–387 (2005)
4. Guttman, R., Moukas, A., Maes, P.: Agent-mediated electronic commerce: a survey. The Knowledge Engineering Review 13(02), 147–159 (1998)
5. Georgakopoulos, D., Hornick, M., Sheth, A.: An overview of workflow management: From process modeling to workflow automation infrastructure. Distributed and Parallel Databases 3(2), 119–153 (1995)
6. Bussler, C.: B2B integration: Concepts and architecture. Springer (2003)
7. Albrecht, C., Dean, D., Hansen, J.: Marketplace and technology standards for B2B e-commerce: progress, challenges, and the state of the art. Information & Management 42(6), 865–875 (2005)
8. Milosevic, Z., Bond, A.: Electronic commerce on the internet: what is still missing? In: Proc. of the 5th Annual Conference of the Internet Society, Hawaii, USA (1995)
9. Judge, D., Odgers, B., Shepherdson, J., Cui, Z.: Agent-enhanced workflow. BT Technology Journal 16(3), 79–85 (1998)
10. Griffel, F., Boger, M., Weinreich, H., Lamersdorf, W., Merz, M.: Electronic contracting with COSMOS-how to establish, negotiate and execute electronic contracts on the internet. In: EDOC 1998 Proceedings, pp. 46–55 (1998)
11. Papazoglou, M.P., Jeusfeld, M.A., Weigand, H., Jarke, M.: Distributed, Interoperable Workflow Support for Electronic Commerce. In: Lamersdorf, W., Merz, M. (eds.) TREC 1998. LNCS, vol. 1402, pp. 192–204. Springer, Heidelberg (1998)
12. Chiu, D., Karlapalem, K., Li, Q., Kafeza, E.: Workflow view based e-contracts in a cross-organizational e-services environment. Distributed and Parallel Databases 12(2), 193–216 (2002)

13. Cheung, S.C., Hung, P.C.K., Chiu, D.K.W.: A Meta-model for e-Contract Template Variable Dependencies Facilitating e-Negotiation. In: Spaccapietra, S., March, S.T., Kambayashi, Y. (eds.) ER 2002. LNCS, vol. 2503, pp. 50–64. Springer, Heidelberg (2002)
14. Grefen, P., Aberer, K., Hoffner, Y., Ludwig, H.: CrossFlow: cross-organizational workflow management in dynamic virtual enterprises (2000)
15. Mahmoud, Q., Yu, L.: Mobile agents for Location-Aware advertising in mobile wireless environments (2004)

Usage Analysis of a Mobile Bargain Finder Application

Stephan Karpischek, Darshan Santani, and Florian Michahelles

Chair of Information Management, ETH Zürich, Switzerland

Abstract. Mobile shopping applications for smartphones are popular among consumers. While mobile commerce research has focused on experimental prototypes and evaluation in small groups, only little is known about the real-world usage of these applications. Established tools and methods for analysis are missing. In this paper, we present the usage analysis of a mobile bargain finder application based on server logs of more than 98,000 users over a period of 6 months. We show that plots of the cumulative distribution function (CDF) are well suited to analyze the distribution of relevant parameters and present simple heuristics to identify active users. We can show that Pareto's law applies to the distribution of user requests. We also propose metrics to measure usage focus and find that active users tend to become more focused with increased usage. Finally, we combine the results from the log analysis with demographics from an online user survey.

Keywords: Mobile, Shopping, Consumer, Application, Log, Usage, Analysis.

1 Introduction

Mobile shopping assistants help consumers to make better buying decisions and have been a topic of interest in mobile commerce research for many years. Today shopping applications for mobile phones are used on a daily basis by millions of consumers worldwide. While research so far has mostly focused on experimental prototypes and evaluation in small scale user studies or laboratory environments, only little is known about the actual usage of these applications. Application developers and other stakeholders in the area of mobile commerce are interested in finding out how mobile applications are used in real-world scenarios. The goal of this paper is to better understand the real-world usage of a popular mobile commerce application for finding bargains in nearby supermarkets and to present methods and tools which are well suited for analyzing mobile commerce applications.

This paper follows a previous analysis of an earlier version of the mobile bargain finder application [8]. In this paper we analyze an updated version of the application and combine the log analysis with demographics from an online survey conducted among the application's users. We extend our analysis methodology and show that cumulative distribution function (CDF) plots are well suited to analyze the distribution of relevant parameters such as user requests or sessions. We also describe simple heuristics to identify active users. For the bargain finder application we can show that the distribution of user requests follows the Pareto principle. Then we propose metrics to measure a user's focus when using the application and show that the focus of active users increases with usage duration.

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 137–148, 2012.

2 Related Work

Mobile applications which help consumers to make better buying decisions have been a topic of interest in research for many years: Early prototypes were customized hardware devices [1], later software prototypes were implemented on Personal Digital Assistants (PDAs) [5,9,13,10], and beginning in 2003, software prototypes were implemented on mobile phones [15].

In 2009, Deng and Cox presented LiveCompare, a prototype using mobile camera phones for grocery bargain hunting through participatory sensing [4]. The system focused on crowdsourcing price information for grocery products and discussed the problem of data scarcity and data integrity. Obtaining data from retailers directly would help overcome these limitations. Our work studies the usage of a mobile application for finding bargains which gets its information on bargains directly from retailers.

Research on mobile shopping assistants so far has focused on prototypes which have not been widely deployed and evaluated on a large scale. While other mobile applications have been researched in the large [11,12,2], the evaluation in the reviewed work on mobile commerce applications used relatively small user groups and took place in controlled lab environments. Findings about the real-world usage of mobile applications by consumers are relevant for mobile commerce research and practitioners like retail companies and application developers but are missing so far.

Today the distribution channels for mobile applications to consumers on smartphones offer an interesting opportunity to deploy mobile shopping applications to large user groups and also to analyze real-world usage over a longer period of time. Our work focuses on this in-the-wild approach which has not yet been applied to mobile shopping applications for consumers. The contributions of this paper are a set of methods to analyze mobile commerce applications and findings about the real-world usage of a mobile bargain finder application.

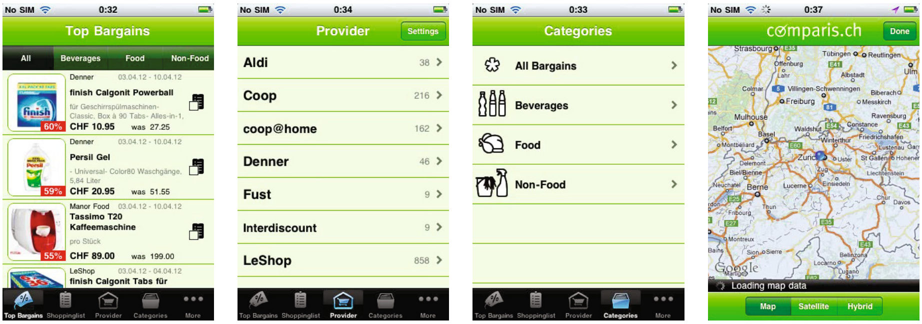
3 Background and Data

3.1 Mobile Application Overview

Comparis Shopper is a mobile commerce application for the iPhone which was first released in the iTunes App Store under the name Bargain Finder in 2009. Users can inform themselves about bargains and special offers in supermarkets nearby. They can choose to display bargains from specific product categories or retailers. Figure 1 shows some screenshots of the iPhone application.

3.2 Log Data

In this section we give an overview of the anonymized query logs comprising of over 5 million requests which were collected from over 98,000 users over 191 days, from February 2011 until August 2011. As a user interacts with the application, hypertext transfer protocol (HTTP) requests are sent by the iPhone to the backend server application fetching relevant information for the user. Since there is no local caching on the iPhone, the server logs provide a detailed and precise representation of the spatio-temporal usage patterns of individual users.



(a) (b) (c) (d)

Fig. 1. Comparis Shopper Screenshots (a) Top Bargains (b) Retailer (c) Product categories (d) Shop Finder

Each request in the query logs contains a unique user identifier, originating IP address, the URI containing the application's request for a specific function call, session identifier and the respective timestamp when the request was generated. With every fresh install of the iPhone application, the backend server generates a unique random number as user identifier which is stored locally on the device and appended every time the device sends a request. Thus every installed instance of the application can be identified in the query logs, and requests coming from the same instance and user can be grouped together. Similarly a session ID is generated to differentiate user sessions, where a session is defined as a series of requests from the same device within a certain period of time, commonly known as session delta [16]. If the device does not send a request within session delta of the last request, the session expires and next subsequent request is assigned a new session ID.

Request Types. One of the interesting data embedded in the query logs is the URI string, which represents the mobile application's request for a specific function of the server backend, and in most cases a user's request for a specific information. In total, the query logs data contain 17 different request types, each representing a particular functionality. We can differentiate two basic groups of requests:

- **System Requests**: System requests are operational in nature and requested implicitly by the application in the background without any explicit action of the user, e.g., on startup the application checks if a newer version exists.
- **User Requests**: User requests are explicitly triggered by the user to satisfy a specific information need, e.g., the search for bargains from a specific retailer or for a specific product category.

Over 41% of the total requests are system requests, while the rest of 59% are user requests.

3.3 Survey Data

In addition to analyzing the query logs an online survey was conducted among the users of the application in July and August 2011. When starting the application during this time, users were presented with an invitation to fill out the survey online using the web browser of their iPhone. When accepting the invitation the iPhone's web browser launched the online survey with the unique user identifier from the application as parameter. Survey participants remained anonymous over the whole time. The only personal information collected was the email adress which participants could enter after completing the survey if they chose to participate in a raffle. Email adresses were only used for the raffle and not in the analysis to preserve the participants privacy. In this paper we report results from combining the log analysis with demographics from 1,009 completed online surveys. More details about the survey and more results can be found in [6].

4 Analysis and Results

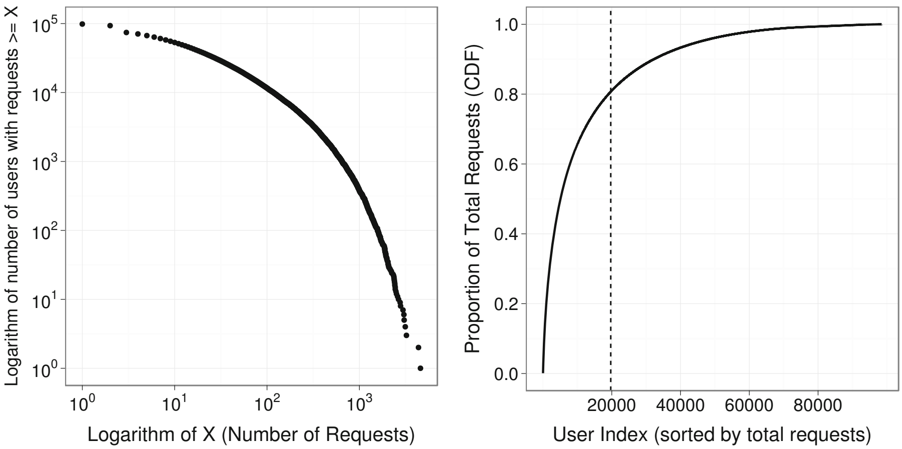
4.1 Preliminary

In this section we specify definitions related to users request types which will be used throughout the remainder of the paper. For a user i we denote the distribution of its requests across M request types $(q_1, q_2, \ldots, q_M)$ as a vector $(r_{i1}, r_{i2}, \ldots, r_{iM})$. And we define the total queries requested by user i across all the request types during the entire period of analysis T as $R_i = \sum_{k=1}^{M} r_{ik}$.

With the given premise we now examine the fundamental properties and characteristics of usage behavior in the subsequent sections. The main goals of our analysis are (a) to understand how the application is used, (b) to explore whether application usage changes as users spend more time using the application, and (c) to develop a set of methods and tools which are well-suited to analyze mobile commerce applications.

4.2 Requests

First we analyze user activity by examining the total number of requests across all request types. Intuitively one would expect that different users have different access patterns – some are active users who frequently use the application on a regular basis, while others have sporadic application usage over their lifetime. The given dataset matches our intuition and we observe a similar behavior. Figure 2a shows a Complementary Cumulative Distribution Function (CCDF) of total requests R_i by users. The plot shows a heavy tail distribution where a handful of extremely active users send the most number of requests. In fact, 50% of all the users send less than 10 requests in total. Note that in the given figure (and subsequent analysis) we chose to use cumulative distribution function (CDF) or complementary cumulative distribution function (CCDF) as they show the relevant statistics and the underlying (probability) distribution more clearly and succinctly, in addition to providing a better visual aid to compare multiple distributions in the same plot.



(a) CCDF of requests per user (b) CDF of requests per user

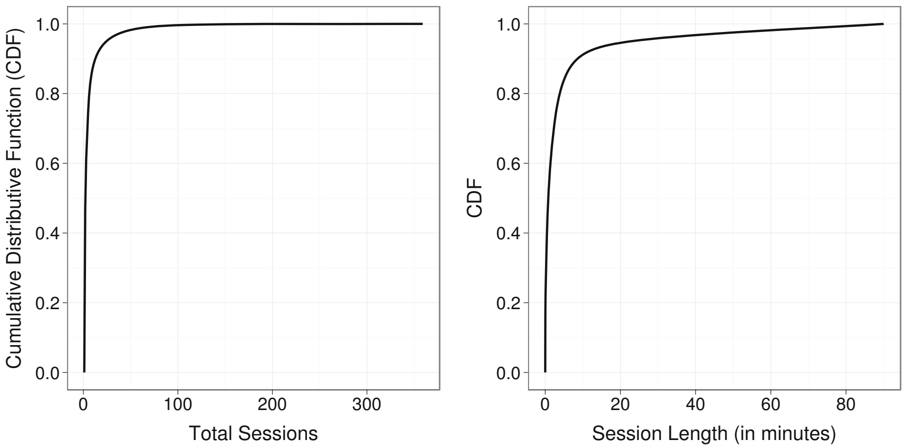
Fig. 2. Requests

To further investigate the skewness in the distribution of queries, we examine the applicability of the Pareto Principle, commonly known as the 80-20 rule. We analyzed the total queries for all the users, as shown in the cumulative distribution function (CDF) plot in Figure 2b, with the user rank on the horizontal axis and fraction of total requests from the corresponding set of users on the vertical. Users are ordered by the number of total requests. The distribution closely follows the Pareto Law, i.e., more than 75% of all requests are generated by top 20% of users.

4.3 Sessions

Now we turn our attention to investigating the user sessions, in particularly focusing on the number of sessions performed by each user and its respective session length. As explained in Section 3, a session is a period of constant activity where the user sends a series of queries within a certain time interval, usually termed as session delta [16]. If the user does not send a query within session delta of the last query, the current session expires and with the next subsequent request a new session is started. For our analysis, we have set the session delta to be 90 minutes.

In our dataset, we have observed a total of 561,707 sessions for all users. Similar to the heavy tailed distribution of request queries, total sessions per user also exhibit a heavy tail distribution, as shown in Figure 3a. Over 75% of users have less than 5 sessions in total, with the most active user having 359 sessions. It indicates that majority of the users are dormant users who just install the application for curiosity and exploration and then never use it again or even uninstall it. It is also interesting to note that in addition to total sessions, individual session lengths follow a similar distribution as shown in Figure 3b, where we have plotted individual session lengths on the horizontal axis and their corresponding CDF on the vertical axis. Over 75% of total sessions



(a) CDF of sessions per user (b) CDF of session lengths

Fig. 3. Sessions

lasted for less than a minute, clearly suggesting low application usage and interactivity for majority of users.

4.4 User Classification

In addition to looking into individual session lengths, we also analyze the total session duration, i.e. the sum of the lengths of all sessions for one user. We observe that over 72% of users have used the application for less than 10 minutes in total. These cumulative patterns reveal significant variability in application usage and suggest a possible categorization of users as per their activity into "Inactive" and "Active" users. Inactive users are further classified as either "Dormant" or "One-Timers" users with heuristics shown in Table 1. Now we describe these user segments in more detail:

- **Dormant Users:** Dormant users typically install the application, but never interacts with it beyond opening the application's landing page once in a while (which explains them having multiple sessions but with all sessions having length of 0)
- **One-Time Users:** One-time users are the ones who communicates with the application once in their lifetime but resulting in prolonged interactivity. In our dataset, one-timers constitute of over 31% of total users, who sends an average of 6 requests over their only session.
- **Active Users:** The rest of the user base i.e., which are not inactive, are classified as active users. Active users constitute of over half of the user population.

Table 2 shows basic statistics for the three user segments and in addition for the survey participants. Moreover, we perform a median split of the active user segment and list the relevant summary statistics for these two categories in Table 3.

Table 1. User Segments

User segment	Number of sessions	Usage duration
Dormant	≥1	0
One Timers	1	>0
Active	>1	>0

Table 2. User Segments and Survey Participants Statistics

User segment	% of Total users	% of Total requests	Average requests per user	Average sessions per user
Dormant	16.15%	0.59%	1.87	1.08
One-Timers	31.35%	3.93%	6.42	1.0
Active	52.50%	95.48%	93.14	9.94
Survey	1.02%	6.65%	332.34	31.75

Table 3. Active Users Median Split

User segment	Requests per user			Sessions per user		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
Active Top 50%	164.2	86.0	228.91	16.47	9.00	22.34
Active Bottom 50%	22.13	18.00	16.97	3.40	3.00	2.35

4.5 Usage Metrics

Our goals are to better understand how the application is used, and if application usage changes with increasing time of usage. In this section we propose metrics to measure the focus of users and the variability of user requests. We begin by formally defining two key measures for each user i as: (a) **Request Type Focus** F_i (b) **Request Type Entropy** E_i as

$$F_i = \frac{1}{R_i} \max\ r_{ik} \tag{1}$$

$$E_i = -\sum_{k=1}^{M} \frac{r_{ik}}{R_i}\ \log \frac{r_{ik}}{R_i} \tag{2}$$

The idea of defining these metrics is mainly inspired by research work in the domain of information retrieval [7], large scale analysis of a popular online video service [3] and empirical analysis to understand content access behavior based on video-on-demand service [17]. It is important to note that while defining the focus and entropy measures for users, we take into account only the user specific requests and ignore the system requests. Now we describe these usage metrics in more detail.

Request Type Focus. Given the distribution of a user's requests $(r_{i1}, r_{i2}, \ldots, r_{iM})$ across all user specific request types, focus is the highest fraction of queries a user has sent for a single request showing a specific information need. In other words, focus metric measures how focused users are for their varied information needs, whether a user tends to access one request type category repeatedly than others or his access patterns are more spread across different request types.

Intuitively the focus value always lies between 0 and 1 as per definition, with higher values indicating low variability and more predictability towards user behavior, while low values describing the opposite. An F_i value of 1 shows that the user is interested in only one request type indicating a targeted and narrow application usage, while a value of 0 indicates that users requests are spread uniformly across all request types $(1/k, 1/k, \ldots, 1/k)$, hence highlighting high variability in application usage and providing little information about specific information needs. Figure 4a shows the CDF of focus values with 50% of total users having focus values less than 0.5.

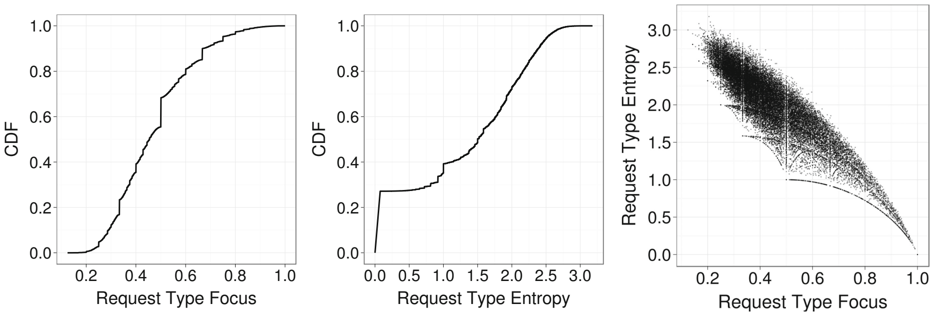
Request Type Entropy. The idea of request type entropy is primarily inspired by Shannon's work in information entropy [14]. Higher values of entropy indicate a usage pattern which is spread uniformly, while users with focused (and hence more predictable) usage are characterized by lower entropy values. Figure 4b shows the CDF of entropy values.

To give an example for how request type focus and request type entropy represent a user's behavior consider two users: user A has 3139 total requests (with 1334 being user specific requests while the rest are system specific) in 267 sessions consisting of 4 different user request types: 1281 shopping list requests, 45 shopping list alert requests, 5 top bargains requests, and 3 retailer requests. This highly skewed distribution of request types results in a focus value of 0.96 and an entropy value of 0.27, representing a highly focused user. In contrast, user B has 3231 total requests (with 1725 being user specific ones) in 176 sessions. Her sessions consist of 8 different request types with a more even distribution (532 category bargains requests, 514 shopping list requests, 258 retailer requests, 223 shopping list alerts, 124 bargains given product identifier requests, 59 top bargains, 9 shop locator and 6 subcategory requests), resulting in a focus value of 0.39 and an entropy value of 2.16, representing a relatively less focused user.

As per definition focus and entropy values are inversely correlated. Higher values of focus relate to lower values of entropy and vice versa; the inverse correlation is clearly visible in Figure 4c, which has focus values on the horizontal axis and entropy on the vertical axis across all user segments. For the subsequent analysis and discussion, we choose to use request type focus as our usage metric because of the more intuitive range between 0 and 1.

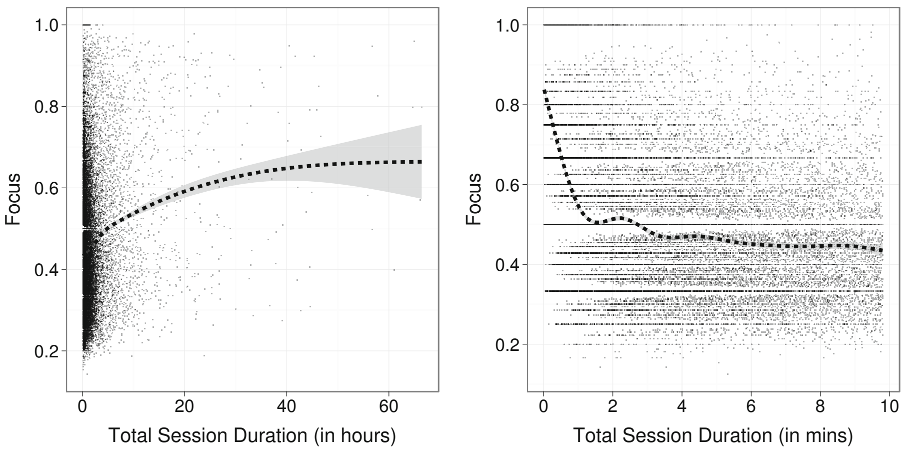
4.6 Request Type Focus and Session Duration

In order to understand whether the focus of users changes with increasing application usage, we examine the relationship between the request type focus values and total session duration for the active user segment. Figure 5a shows a scatter plot of total session duration, i.e. the total time of application usage in hours, for a given user on the horizontal axis and corresponding focus values on the vertical axis for the active top 50% users. Figure 5b shows the same plot with the usage time in minutes for the bottom 50% of active users.



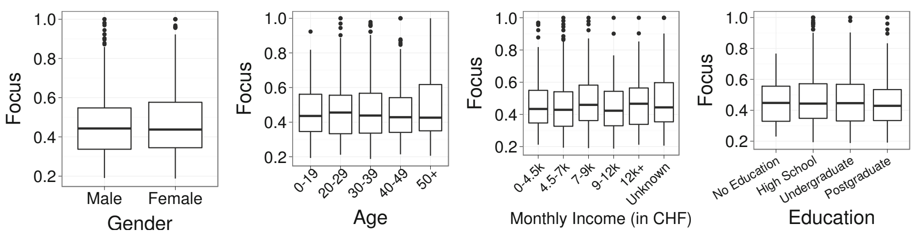
(a) CDF of focus (b) CDF of entropy (c) Correlation of Focus and Entropy

Fig. 4. Usage metrics



(a) (b)

Fig. 5. Total session duration and corresponding focus values for active users (a) Top 50% (b) Bottom 50%. Note the different time units.



(a) (b) (c) (d)

Fig. 6. Distribution of focus values for different user demographics (a) Gender (b) Age (c) Income (d) Education

4.7 User Demographics

From an online survey conducted amongst the users we collected demographics data including gender, age, education status and income levels. Since survey users are amongst the existing user base and the unique user identifier from the query logs is one parameter in the survey, we can link the survey with the log analysis in order to investigate the application usage of survey users. In this section, we analyze the request type focus across different demographic groups.

Of all 1,009 survey users, 253 of them are first time users, while the rest 756 users have used the application before. Even though we have a significant fraction of survey users (75%) who are familiar with the application, we don't know to which user segment the survey group belongs to. To investigate it, we compare the probability distribution of focus values for survey users with different user segment and as it turns out, survey users primarily consist of top 50% of active user base (plot omitted due to space constraints). Close to 90% of survey users are part of the active top 50% user group.

Now to examine whether the focus values vary across different demographics, we analyze the distribution of focus values for different gender, age, income levels and education status. The box plots in Figure 6 show that the focus values do not vary significantly across these demographic groups.

5 Discussion

The total number of users is likely misleading when talking about real-world application usage as only a smaller fraction of users actually use the application. The CDF of user requests shows this very clearly. In our case we also find that the distribution of user requests follows the Pareto principle. It would be interesting to investigate if this is also the case with other mobile applications.

We consider the identification of active users an important part of usage analysis. We classified users into segments based on simple heuristics using the number of sessions and the total session duration. In the following we concentrated the analysis on the segment of active users, i.e. users with more than one session and a session duration greater than zero. The top 50% of active users is the user segment where most of the activity happens, and also the usage patterns in this group differ from the rest of users. As a result, we aim to understand their application usage in more detail.

We proposed two metrics to measure a user's focus when using the application. For us the request type focus seems to be more intuitive due to its range from 0 to 1. Figure 5a shows the trend that the top 50% of active users become more focused with increasing usage of the application. As active users interact more with the application, they seem to become more aware of which application functionalities best suit their respective needs and indulge in using only a few functions.

The same analysis for the bottom 50% of active users indicates a different trend – the focus values don't vary significantly with increased application usage. One of the reasons to attribute this behavior to is that the bottom 50% active users spend a considerable less amount of time with the application compared to the active top 50% users: While the top 50% active users have an average of 92 minutes of total session duration, the bottom 50% users spend an average of 3.5 minutes with the application.

Combining quantitative log analysis with a qualitative user survey offers interesting opportunities for research, especially when survey results of individuals can be combined with corresponding usage logs and given that the privacy of users is preserved at all times. However, our analysis did not result in relevant findings as the user focus does not vary significantly over different user demographics.

6 Conclusions

In this paper we have analyzed the usage of a mobile bargain finder application for the iPhone using query logs from a period of 6 months. Using CDF plots we could show that the distribution of user requests follows the Pareto principle. We proposed simple heuristics to identify active users on which we concentrated our analysis. We also proposed metrics for measuring a user's focus when using the app and showed that active users tend to become more focused with increasing usage. We also combined the analysis of user focus with demographic data from an online survey and found no significant differences in user focus across demographic user segments.

In future work we want to apply the same user segmentation and metrics to other mobile commerce applications and compare measurements and CDF plots for different applications. In the long term mobile commerce research could benefit from well established ways to measure active usage and user focus when analyzing mobile applications.

Acknowledgements. We are grateful to Comparis for providing access to their dataset. We also want to thank Gilad Geron for conducting the online survey.

References

1. Asthana, A., Crauatts, M., Krzyzanowski, P.: An indoor wireless system for personalized shopping assistance. In: IEEE Workshop on Mobile Computing Systems and Applications, pp. 69–74 (1994)
2. Böhmer, M., Hecht, B., Schöning, J., Krüger, A., Bauer, G.: Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage. In: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI 2011, pp. 47–56. ACM, New York (2011)
3. Brodersen, A., Scellato, S., Wattenhofer, M.: Youtube around the world: geographic popularity of videos (2012)
4. Deng, L., Cox, L.P.: Livecompare: grocery bargain hunting through participatory sensing. In: Proceedings of the 10th Workshop on Mobile Computing Systems and Applications, HotMobile 2009, pp. 4:1–4:6. ACM, New York (2009)
5. Fano, A.E.: Shopper's eye: using location-based filtering for a shopping agent in the physical world. In: Proceedings of the Second International Conference on Autonomous Agents, AGENTS 1998, pp. 416–421. ACM, New York (1998)
6. Geron, G.: Mobile shopping assistant usage analysis - analysis of comparis smartshopper iPhone application. Master Thesis, ETH Zürich (2011)
7. Kamvar, M., Kellar, M., Patel, R., Xu, Y.: Computers and iphones and mobile phones, oh my!: a logs-based comparison of search users on different devices. In: Proceedings of the 18th International Conference on World Wide Web, WWW 2009, pp. 801–810. ACM, New York (2009)

8. Karpischek, S., Geron, G., Michahelles, F.: Towards a better understanding of mobile shopping assistants - a large scale usage analysis of a mobile bargain finder application. In: Second Workshop on Research in the Large at UbiComp 2011, Beijing, China, September 18 (2011)
9. Kotlyar, V., Viveros, M., Duri, S.S., Lawrence, R.D., Almási, G.S.: A Case Study in Information Delivery to Mass Retail Markets. In: Bench-Capon, T., Soda, G., Tjoa, A.M. (eds.) DEXA 1999. LNCS, vol. 1677, pp. 842–851. Springer, Heidelberg (1999)
10. Kourouthanassis, P., Roussos, G.: Developing consumer-friendly pervasive retail systems. IEEE Pervasive Computing 2(2), 32–39 (2003)
11. McMillan, D., Morrison, A., Brown, O., Hall, M., Chalmers, M.: Further into the Wild: Running Worldwide Trials of Mobile Systems. In: Floréen, P., Krüger, A., Spasojevic, M. (eds.) Pervasive 2010. LNCS, vol. 6030, pp. 210–227. Springer, Heidelberg (2010)
12. Miluzzo, E., Lane, N.D., Lu, H., Campbell, A.T.: Research in the app store era: Experiences from the cenceme app deployment on the iphone. In: First Workshop on Research in the Large at UbiComp 2010, Copenhagen, Denmark, September 26 (2010)
13. Roussos, G., Kourouthanasis, P., Moussouri, T.: Designing appliances for mobile commerce and retailtainment. Personal Ubiquitous Comput. 7, 203–209 (2003)
14. Shannon, C.E.: A mathematical theory of communication. Bell Systems Technical Journal 27, 623–656 (1948)
15. Shekar, S., Nair, P., Helal, A.S.: iGrocer: a ubiquitous and pervasive smart grocery shopping system. In: Proceedings of the, ACM Symposium on Applied Computing, SAC 2003, pp. 645–652. ACM, New York (2003)
16. Silverstein, C., Marais, H., Henzinger, M., Moricz, M.: Analysis of a very large web search engine query log. SIGIR Forum 33, 6–12 (1999)
17. Yu, H., Zheng, D., Zhao, B.Y., Zheng, W.: Understanding user behavior in large-scale video-on-demand systems (2006)

Argumentation–Based Negotiation? Negotiation–Based Argumentation!

Jürgen Landes and Ricardo Buettner*

FOM Hochschule für Oekonomie & Management, University of Applied Sciences,
Chair of Information Systems, Organizational Behavior and Human Resource Management,
Arnulfstrasse 30, 80335 Munich, Germany
juergen_landes@yahoo.de, ricardo.buettner@fom.de

Abstract. We design a protocol for two autonomous negotiating agents to incorporate Dung-style argumentation into an ongoing bargaining dialogue. Previous approaches considered bargaining and Dung-Style Argumentation as separated components, we show that intertwining these approaches increases the agents scope of action. In our framework the acceptance of an argument or attack uttered by self-interested agents is conditional on the acceptance by the negotiating partner. Our protocol thus enables autonomous agents to engage in a variety of human negotiation behaviours and thereby increase the agents capabilities to come to mutually satisfactory agreements.

Keywords: Dung-Style Argumentation, Negotiation, Multi-Agent System, Autonomous Agents, Protocol.

1 Introduction

In any society consisting of self-motivated inhabitants successful cooperation is by no means a certainty. Negotiation [31,37,46,68,79] is one common mean to solve complex coordination problems involving multiple entities represented by agents with private agendas [75]. Humans have been negotiating for a long time and have developed a multitude of negotiating tactics and strategies. The development of negotiation/bargaining protocols [41,61,71] has made it possible for computer agents to engage in negotiations.

However not every negotiation is successful, in particular in case of an empty zone of possible agreement. Failure to come to agreements causes at least one agent to end up in a non-optimal state and leads in general globally to less than ideal allocations of resources. Termination of negotiation without an agreement can be due to a number of reasons, one of the most important reason is that agents conceive the world in different ways.

Arguments may be introduced into a negotiation dialogue to put the negotiation on a more rational footing and to narrow or even bridge the gap between the agents perspectives. The seminal work of Dung [32,33] puts argumentation as a formal sub-field of AI on solid foundations. Dung-Style Argumentation (DSA) can thus be used in negotiations between electronic agents to get private valuations closer together thereby

* The authors gratefully acknowledge support from grant 17103X10 from the German Federal Ministry of Education and Research.

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 149–162, 2012.

creating a possibly larger (or not empty) zone of possible agreement. Much work has been put into studying Dung's framework; see [13, 66] for recent overviews; popular lines of investigations have been argumentation semantics [9,21,30,33,34], extensions of the original framework [1,10,35] and more recently dynamic argumentation [17,47] and judgment aggregation [23,29].

However almost all these approaches have in common that the introduction of an argument is taken at face value and is not questioned. In the absence of an omniscient oracle that could be queried to decide the status of an argument negotiating agents have to *agree* on whether to collectively accept an argument or not. Thus negotiating agents require a coordination mechanism to successfully incorporate DSA into an ongoing negotiation.

In this paper we design a negotiation protocol enabling agents to engage in multi-agent argumentation. We aim at designing a manageable framework that treats argumentation and bargaining as strongly connected components which extends the agents room to maneuver. This allows sophisticated agents to display a large spectrum of human negotiation behaviors. We thus support agents autonomously solving complex coordination problems by expanding their scope of action.

The rest of the paper is organized as follows. Next we discuss related work, then we develop our protocol and give an example. Afterwards we touch on strategic behavior and on real world considerations before we conclude.

2 Related Work

2.1 Negotiation

Negotiations have played a major part in human societies dating back to ancient times [15]. The quest to (semi-) formalize human dialogues [77] has come a long way leading to today's negotiations between computer agents [5, 48, 50, 80]. A recurring theme in (electronic) negotiations is the notion of negotiation as a dialogue [42, 52, 53, 82].

As we here aim at designing a protocol we will not focus on the agents using our protocol. We postpone the development of strategies and tactics to further work and content ourselves here with stating such agents may be modeled formally, for instance via a BDI [60,69], BOID [20]. See [73,74] for models of BDI agents able to engage in ABN and refer to [72] for a model of single(!) agent argumentation. See [45] for BDIG agents able to engage in negotiation dialogues with (not Dung-style) arguments.

Argumentation Based Negotiation. Influenced by studies of human negotiations the following definition of an argument may be found in [65, p 347]: "an argument as a piece of information that may allow an agent to: (a) justify its negotiation stance; or (b) influence another agent's negotiation stance." This definition proved to be very influential. Indeed negotiations among autonomous agents that allow the exchange of arguments in the above sense are known as Argumentation Based Negotiations (ABN) [40,65,78]. See [67] for comparisons of ABN and bargaining. A format for the exchange of arguments between different agents has been introduced in [26].

Engaging in a dialogue incurs a cost in time and effort on the agents part, thus generating transaction costs [27, 28, 56, 58]. Result presented in [43] suggest that ABN is an efficient mean of resolving conflicts in Multi Agent Systems (MAS) in case of rare resources, however if resources are abundant re-planning and evasion maybe more efficient means of resolving conflicting interests.

2.2 Dung-Style Argumentation

Argumentation was introduced by Dung [32, 33] as a proper subfield of AI by means of an ordered pair $\langle \mathcal{A}, \mathcal{R} \rangle$. $\mathcal{A}$ is taken to be a finite set of *arguments* and $\mathcal{R}$ is a binary *attack* relation on the arguments. The ordered pair $\langle \mathcal{A}, \mathcal{R} \rangle$ is called an argumentation framework. Determining the justified (or justifiable) arguments in a given argumentation framework is a key problem in argumentation which is normally addressed by using an argumentation semantics $\mathcal{S}$. Extension-based semantics pick out sets of extensions $\mathcal{E}_{\mathcal{S}}(AF) \subseteq 2^{\mathcal{A}}$, such that $\mathcal{E}_{\mathcal{S}}(AF)$ contains all and only all the sets of arguments, which satisfy certain properties. Such an extension semantics is called single-status if and only if $\mathcal{E}_{\mathcal{S}}(AF) \subseteq \mathcal{A}$ holds for all argumentation frameworks AF.

Numerous extension-based semantics have been proposed in the literature [21, 30, 32–34] with no clear front-runner emerging. The choice of a semantics has become dependent on personal preferences and circumstances. Among the crucial principles a semantics may satisfy or fail to satisfy are the following principles: Free of conflict P, admissibility P, reinstatement P, language independence P, I-maximality P and the directionality P [7, 8]. Further such principles have been discussed [22, 38].

Judgment Aggregation in Argumentation. Multi-agent DSA has been considered in [6, 23, 29, 81]. There the authors solve the problem of merging different argumentation systems from different agents by considering voting procedures. However a simple voting mechanism does not allow the agents to engage in meaningful dialogues. Furthermore, votes with only two eligible voters with equal weights end in stalemates, if votes are cast for different alternatives.

Dynamic Argumentation. Building on the work of [4] and [25] Liao et al. [47] studied dynamic change in argumentation frameworks and their extensions. [4, 25] and [14] only considered altering the argumentation framework by one instance. [17] investigated if extensions of argumentation frameworks remain the same in case arguments and/or attacks are removed in the argumentation framework.

Further Recent Developments in Argumentation Research. Going beyond mere attacks to also include support of arguments was studied in [1, 10], while in [10] also *strength* of arguments was considered. Further advances in argumentation introducing degrees of strength of arguments were reported in [11, 35].Higher order issues of argumentation have recently been scrutinized, among them meta-argumentation [16], equivalence [55] and mechanism design [63]."Probabilistic Argumentation" as been put forward by Haenni et al., we exemplary mention [39].

Argumentation Dialogues. Formal dialogue models where participants aim to persuade or to resolve a difference of opinion have received considerable interest in the literature [62]. In [3] a dialogue protocol for two agent argumentation is developed.

In this approach no connections between negotiation and argumentation are possible, arguments from the negotiation partner can only be rejected; if the agent it has an acceptable argument for their negation. Furthermore attacks follow immediately from the syntactical structure of the arguments and an agent can only make assertions for which it has support.

A recent approach to incorporate DSA into the fold of negotiation in AI may be found in [2]. In [57] argumentation dialogues in uncertain domains are studied where agents are self-interested but do not lie. While in [19] a mediator is brought in argumentation dialogues to integrate information from different agents.

In general, it has to be noted that these dialogue approaches fail to fulfill their potential since arguments and attacks brought forward by a negotiation partner are always accepted without questioning.[1] Further restricting assumptions are often made, for instance in [2] claimed arguments cannot be retracted and the set of all possible arguments is a priori known. Furthermore, agents are able to a priori specify a preference relation on all arguments which remains fixed throughout. Generally, the narrow focus on DSA, that is not considering the DSA embedded into an ongoing negotiation, makes it impossible to intertwine negotiation and argumentation.

Different appreciations of arguments by agents with different internal states were considered in [11, 12], where the authors introduced the notion of an audience. An audience amounts to a partial preference order on arguments. The authors study a two-player dialogue game on arguments. Such a game is won by the proponent player w.r.t to an argument x and a VAF at move m, if the opponent cannot find an audience and/or a new argument according to intricate rules such that for this audience and the updated VAF x is in the preferred extension.

3 Negotiation-Based Argumentation

3.1 Agents as Negotiation Partners

For the remainder we consider two agents $agent_1, agent_2$ which have already entered a negotiation dialogue. We furthermore assume that these agents have agreed to use DSA to obtain a more coherent view of the world, thus enhancing the chances of a successful conclusion of the negotiation.

We make the convention that the internal state of $agent_i$ also contains a representation of what $agent_i$ thinks the internal state of the other agent is, which in turn contains a representation of the perceived internal state of $agent_i$; and so on. Furthermore, we assume that the agents come equipped with a sufficiently powerful reasoning module enabling them to produce a dialogue according to the rules we here establish.

Realistic models of agents will not fully describe all for the negotiation possibly relevant aspects. Rather the agents are equipped with a mechanism that facilitates the incorporation of new information (possibly obtained from the outside world) and they then perform an update of their internal states. This mechanism enables agents to deal with unforeseen (or even unforeseeable) moves by the other agent.

[1] The notable exception is [18] where also illegal and meta-level illocutions are considered.

3.2 Pre-argumentation Arrangements

The agents consent to, that once the argumentation stage has been concluded, accepting the outcome of the argumentation as factual and to base further negotiation upon this established common ground. In order to determine the outcome of the argumentation procedure, the agents have agreed to use an argumentation semantics $\mathcal{AS}$ (complete, grounded, stable, preferred, stage, semi-stable, ideal, CF2 or a prudent semantics[2]). The outcome of the argumentation is formally given by the extension $\mathcal{E}_{\mathcal{AS}}(AF_{END})$ where AF_{END} denotes the final stage of the argumentation framework.

Since there is no omniscient oracle available, these agents have to determine the validity of arguments themselves. During the conversation the agents might agree to collectively discard a previously accepted argument or to discard an attack. Thus the validity of an argument is in general a dynamic property. However there may be certain parts of an argumentation framework which are by both agents perceived as: "valid beyond the shadow of a doubt" (possibly because there exists unassailable evidence supporting it).[3] Let AF_0 denote this argumentation framework of jointly perceived genuine truth; note that AF_0 may be empty.

Having agreed on $AF_0 = \langle \mathcal{A}_0, \mathcal{R}_0 \rangle$ the agents engage in a back-and-forth dialogue exchanging proposal and counter-proposal. Whenever the agents agree to introduce or delete a set of arguments and attacks the argumentation framework is updated yielding $AF_{N+1} = \langle \mathcal{A}_{N+1}, \mathcal{R}_{N+1} \rangle$ where $\mathcal{R}_i \subset \mathcal{A}_i \times \mathcal{A}_i$. Note that only pathological agents could possibly violate $\mathcal{A}_0 \subseteq \mathcal{A}_N$ and $\mathcal{R}_0 \subseteq \mathcal{R}_N$. Contrary, note that even for a single-status extension semantics $\mathcal{E}_{\mathcal{AS}}(AF_{END}) \subseteq \mathcal{E}_{\mathcal{AS}}(AF_0)$ and $\mathcal{E}_{\mathcal{AS}}(AF_0) \subseteq \mathcal{E}_{\mathcal{AS}}(AF_{END})$ both fail to hold in general. In the terminology of [51] $\mathcal{E}_{\mathcal{AS}}(AF_{END})$ can be understood as a joint commitment store. In case the set of all possible arguments that can be made is known, the semantics for partial argumentation frameworks developed in [24] can be used to model AF_0. If this set is not assumed to be known a priori a modest generalization of said framework is needed.

3.3 The Protocol

The argumentation protocol proceeds by taking turns, w.l.o.g. assume that $agent_1$ goes first. The game starts with turn $t = 1$. Consider turn a t and the current state of the argumentation framework AF_N, now $agent_{1+[(t+1) mod\, 2])}$ has one of the following argumentation illocutions or simply moves M_t available

- P ENDAR: propose to end argumentation,
- ACC PEA: accept proposal to end argumentation, if M_{t-1} =P ENDAR,
- END ARG: unilaterally end argumentation,
- END NEG: unilaterally end the whole negotiation,
- P UPDAT$(\cdot)$: propose update of AF_N

[2] This list of possible semantics is not intended to be exhaustive. The agents might even to agree to employ some abstract labeling semantics, however we will assume that an extension based semantics has been selected.

[3] A rationale for an agent to accept an unfavorable instance in AF_0 may be that, if the matter was settled in court, then every judge or juror would hold this instance to be true.

- Acc Off(O): accept updating offer O, if M_{t-1} =P Updat & $O \in M_{t-1}$,
- I Updat($\cdot$): suggest incomplete update of AF_N,
- Req Inf($\cdot,\cdot$): request information,
- Pro Inf($\cdot,\cdot$): provide information,
- Skip Tu: skip turn.

Due to our assumption that $agent_1$ and $agent_2$ are already engaged in a negotiation an agent may also play a (legal) move from the underlying negotiation protocol instead of an argumentation move.

At any point an agent might see no further advantage from continuing to argue. To end the argumentation phase and (re-)enter the bargaining-only phase the agent can either force the end of the argumentation protocol by playing the End Arg move or propose to end the argumentation protocol by choosing P Endar. This proposal may (by playing Acc Pea) or may not be accepted by the negotiation partner. Should the argumentation end via End Arg or Acc Pea, then $AF_{END} := AF_N$.

If it however becomes evident to an agent that there is no possible deal to be hammered out, this agent may simply decide to quit and play move End Neg. The moves End Arg and End Neg enables an agent to avoid to get stuck in (lengthy) loops.

To propose an update of the argumentation framework $(\mathcal{A}_N, \mathcal{R}_N) \rightsquigarrow (\mathcal{A}_{N+1}, \mathcal{R}_{N+1})$ $agent_{1+[(t+1) mod\, 2]}$ may play the P Updat move. An update of the argumentation framework amounts to the introduction and/or erasion of arguments and/or attack relations. With move $t+1$ $agent_{1+[t\, mod\, 2]}$ has the option to accept the suggested update by playing Acc Off. An offered update, which is not immediately accepted, is considered as rejected and thus retracted.

In complex (multi-issue) negotiations finding a comprise is a complex undertaking. Instead of trying to find a compromise directly, it has proven helpful to address the problem step by step [70]. We thus introduce a non-binding I Updat move, which puts agents into a position to gradually build a mutually acceptable update. Such an incomplete update consists of eight pairwise disjoint components $Del_{arg}^{fix} \subset \mathcal{A}_N, Del_{arg}^{mod} \subset \mathcal{A}_N, Int_{arg}^{fix} \cap \mathcal{A}_N = \emptyset, Int_{arg}^{mod} \cap \mathcal{A}_N = \emptyset, Del_{att}^{fix} \subset \mathcal{R}_N \times \mathcal{R}_N, Del_{att}^{mod} \subset \mathcal{R}_N \times \mathcal{R}_N, Int_{att}^{fix} \cap \mathcal{R}_N \times \mathcal{R}_N = \emptyset, Int_{att}^{mod} \cap \mathcal{R}_N \times \mathcal{R}_N = \emptyset$.

The intended interpretation is that the negotiation partner comes up with a counter-proposal to update AF_N which deletes the arguments in Del_{arg}^{fix} from $\mathcal{A}_N$ and introduces the arguments in Int_{arg}^{fix} into $\mathcal{A}_N$. The arguments in $Del_{arg}^{mod}/Int_{arg}^{mod}$ to be deleted from/introduced to $\mathcal{A}_N$ may be modified (as little as possible) in a counter-proposal; and similarly for the attacks in $Del_{att}^{fix}, Del_{att}^{mod}, Int_{att}^{fix}, Int_{att}^{mod}$.

An I Updat move can be understood as a formalization of a Request For Proposal (RFP): "Which update of AF_N can you agree to that contains all the changes in $Int_{arg}^{fix}, Del_{arg}^{fix}, Int_{att}^{fix}, Del_{att}^{fix}$?" Note that if the sets $Del_{arg}^{mod}, Int_{arg}^{mod}, Del_{att}^{mod}, Int_{att}^{mod}$ are all empty, then an I Updat move constitutes a non-binding P Updat move.

The move Req Inf enables an agent to formally pose the following question "What do I have to provide arguments for/against, if I want you to accept my previous update offer in P Update? The Pro Inf move provides a mean for an agent to communicate "If you want me to accept your update offer in P Update then you should provide arguments for/against this/these argument/s." That is updates of AF_N respectively updates of $\mathcal{E}_{\mathcal{AS}}(AF_N)$ are discussed. Observe that we do not require agents to answer requests nor that a PRO INF move is preceded by a request move.

The move SKIP TU is self-explanatory.

Proposition 1. Every consequence c of the jointly perceived truth AF_0 is also a consequence after the argumentation protocol terminates, unless both agents have mutually agreed to change the status of c.

Proof. Consider a proposition c which follows in previously agreed manner from $\mathcal{E}_{\mathcal{AS}}(AF_0)$. Since every update is computed on the bases of mutual agreement, any change in the status of c has been endorsed by both agents.

Choosing a single-status argumentation semantics $\mathcal{AS}$ allows a straight forward definition of "result of the argumentation phase", i.e. for all decided propositions p either $p \in \mathcal{E}_{\mathcal{AS}}(AF_{END})^{\models}$ or $\neg p \in \mathcal{E}_{\mathcal{AS}}(AF_{END})^{\models}$ and for undecided propositions u it holds that $\{u, \neg u\} \cap \mathcal{E}_{\mathcal{AS}}(AF_{END})^{\models} = \emptyset$. However such a semantics does not leave room for any ambiguity, every agent will draw the same conclusions from the argumentation framework, cf. [17, p. 100].

3.4 Intertwining Negotiation and Argumentation

So far, the negotiation (bargaining) and the argumentation protocol had little concrete interactions. We now want to create synergies by intertwining these protocols.

The move REQ INO$(\cdot,\cdot)$ enables an agent to formally pose the following question "What do I have to successfully argue for/against, if I want you to accept my previous offer O in the bargaining protocol? The PRO INO$(\cdot,\cdot)$ move provides a mean for an agent to communicate "If you want me to accept an offer O, then you should provide arguments for/against this/these argument/s."[4]

The move MOD OFF$(\cdot,\cdot)$ allows an agent to ask the following question: "Which offer O is acceptable to you, if I successfully argue for/against this/these argument/s." An answer move ANS MOF$(\cdot,\cdot)$ enables agent to provide an answer to an MOD OFF query.

Now suppose there is a notion of a distance on the space of possible offers, as for instance in many one-issue distributive negotiations. We then design the moves REQ INO', PRO INO', MOD OFF', ANS MOF' which are obtained from the above by replacing "offer O" by "offer O' close to O" respectively by replacing "offer O" by "neighborhood of offers".

Let $f : \{1, 2, 3, \ldots\} \rightarrow \{0, 1, 2, \ldots\}$ be the function that assigns every time t the number $f(t)$ such that $AF_{f(t)}$ is the at time t agreed upon argumentation framework. Introducing the move OFF ARG$(\cdot,\cdot,\cdot,\cdot,\cdot)$ into the bargaining protocol enables an agent to express "I now offer O_t because since my offer O_τ $(\tau < t)$ we have agreed to the following change(s) in $AF_{f(\tau)}$ and/or $\mathcal{E}_{\mathcal{AS}}(AF_{f(\tau)})$ and I proposed the following updates since $t = 1$. If all but the first argument of a OFF ARG move are empty, then there are no explanations given, and the OFF ARG move is in fact a simple offer in the bargaining

[4] In complex (e.g. multi-issue) negotiations an offer may only be partial by fixing only certain aspects (issues) of the negotiation item while keeping a more open mind about other aspects. To keep the notation tractable, we use the term offer to refer to a complete or a partial offer.

protocol. The introduction of OFF ARG into the bargaining protocol facilitates the explanation of offers. Analogously one may define a move ROF ARG asking for an explanation for a received offer.

The above protocol does not always terminate, which is entirely intentional. If both agents consider it beneficiary to communicate until the end of time, see for exemplary [36] for an analysis of time constraints on negotiations, then why should the protocol force them to end their conversation?[5] By the same token the protocol does not rule out loops. Typical negotiation models in the literature such [46,76] distinguish information phase(s) and negotiation phase(s).

4 An Illustrative Example

We now give our version of a well known example of negotiating agents [3,59,60,64]. Calvin, represented by $agent_{son}$, wants to nail a few twigs to his tree house. His father, represented by $agent_{dad}$, has the hammer locked away and must be persuaded by Calvin to let him carry out his home improvement. Using our framework these two agents might have the following conversation:

$agent_{son}$: "Please give me the hammer." - MAKE OFFER
$agent_{dad}$: "No." - REJECT OFFER
$agent_{son}$: "If you give me the hammer, I will do the dishes tonight." - MAKE OFFER
$agent_{dad}$: "No." - REJECT OFFER
$agent_{son}$: "What do I have to argue for, that you'll give me the hammer?" - REQ INO
$agent_{dad}$: "Prove that you will use the hammer safely." - PRO INO
$agent_{son}$: "I will have a woodwork class in school next week, this proves that children my age can safely use a hammer." - P UPDAT
$agent_{dad}$: "You are not safe in your tree house, because at school you are under supervision of an adult." - P UPDAT
$agent_{son}$: "If you help me with my home improvement, then I will be safe." - P UPDAT
$agent_{dad}$: "Yes." - ACC UPD
$agent_{son}$: "Daddy, pleaaaase come help me with my home improvement." - MAKE OFFER
$agent_{dad}$: "I help you, if you do the dishes tonight." - MAKE OFFER
$agent_{son}$: "Deal. Let's GO!!" - ACCEPT OFFER

The moves MAKE OFFER, REJECT OFFER and ACCEPT OFFER are basic moves in the underlying bargaining protocol. After the initial bargaining failed, $agent_{son}$ inquired how $agent_{dad}$ might be persuaded. Given that information the agent set out to reach AF_{END}, which implies that Calvin will be safe. To persuade the negotiating partner $agent_{son}$ has to make on last concession of doing the dishes.

This example shows why arguments may be rejected by a negotiation partner and that getting an argument accepted may require some bargaining. Furthermore, we see how an agent may obtain information to formulate (more) persuasive arguments. The example also illustrates how new information may modify the perceived set of possible

[5] We are designers of a protocol and not the (thought) police. Recall that both agents are at all times free to unilaterally end all interaction.

reasonable actions. In general, our protocol thus enables agents to (more) efficiently explore the space of possible arguments as well as the space of possible actions. Creative explorations of such spaces are key traits of sophisticated integrative human negotiation behaviors.

5 Strategic Behavior Considerations

At the outset we indicated that participation is beneficial to both agents and they have thus incentives to engage in negotiation. In reality there is no constraint that forces the agents to negotiate nor are they forced to complete the negotiation process. They are thus free to abandon the negotiation at any time. Leaving the negotiation table is a rational choice for an agent, if the Best Alternative To a Negotiated Agreement (BATNA) appears more preferable than a continued negotiation, see [43] for a discussion of this point. Note however that the perceived BATNA and the most preferred alternative in the perceived zone of agreement and the appreciation of such alternatives are subject to dynamic change as new information comes to light during the negotiation process and from communications with the outside world.

Recall that we set out to study autonomous agents engaging in a negotiation. So although an agent may agree to accept certain argumentation frameworks as a bases for further negotiations there is no outside party that forces the agent to actually do so. A genuine interest in solving the coordination problem and/or the possibility of looming legal action (time consuming and costly) do provide incentives to play by the rules. Without any impetus to solve the underlying coordination problem an agent might simply play random moves without ever entering a binding agreement for pure amusement.

Selecting a successful (or even optimal) negotiation strategy is in general a complex problem [49]. An agent might use the following heuristics whether to accept an update: i) accept a proposed update if and only if the agent does not have evidence to the contrary, ii) accept a proposed update if and only if the agent believes that after the update the resulting argumentation framework is closer to the truth/the full truth/more preferable. Finding an optimal strategy runs into the well-known and in general unsolvable problem of how to assess the private internal state of another agent. To evaluate argumentation strategies a testing platform has been developed which is described in [44].

The effects of an update of the argumentation framework on $\mathcal{E}_{\mathcal{AS}}$ can be challenging to compute for large argumentation frameworks. However recent advances in Dynamic Argumentation show that these effects may in some cases effectively be computed.

6 Conclusions

As advertised we have introduced a protocol that supports negotiating agents by enabling them to engage in (integrative) multi-agent argumentation. One key ingredient in our approach is to consider the ongoing negotiation and the argumentation dialogues as integral conjoint parts which may interact in numerous ways.

One major problem in evaluating designed applications is to find an appropriate comparison benchmark. An evaluation with respect to the desiderata for agent argumentation protocols proposed in [54] is given in Table 1. Summarizing, besides the desiderata we deliberately violated our protocol only violates the third and the last desiderata. An extension to more than two agents appears to be straight forward. Any protocol aiming to model a considerable part of human-to-human negotiation behaviors puts invariably high demands on the computing power of the participants.

Table 1. Protocol Desiderata

Desiderata	Our framework	Satisfied?
Stated dialogue purpose	Solve coordination problem	✓
Diversity of individual purposes	Agents inherently self-interested	✓
Inclusiveness	At the moment only 2 agents	No
Transparency	Agents even make rules themselves	✓
Fairness	Agents are treated as equals	✓
Clarity of Argumentation Theory	Dung-style, semantics mutually agreed upon	✓
Separation of Syntax and Semantics	Syntax and Semantics separated	✓
Rule-Consistency	Dialogue may contain loops, but there is always a legal move	Partially
Encouragement of Resolution	No incentives mechanism	No
Discouragement of Disruption	No incentives mechanism	No
Enablement of Self-Transformation	Dynamic internal states	✓
System simplicity	Locutions serve defined purposes	✓
Computational Simplicity	Move selection highly complex	No

Limitations of the presented work are inherited from the limitations of the frameworks used. Any imperfections/limitations of the negotiation model/protocol and of the model of the internal states of agents and their reasoning capabilities is also present in our approach. However such limitations are not system immanent but rather call for further developments addressing these limitations.

Limitations on the argumentation side are the (somewhat) arbitrary choice of an argumentation semantics and the problem of determining the conclusions to be drawn from the outcome of the argumentation in case of non single-status semantics. Furthermore the standard DSA framework is binary in the sense that either there is an attack between arguments or there is not. There is no way of expressing strength of arguments nor can we model uncertainty of an attack or represent support for an argument. These last limitations follow from our goal to keep the whole framework manageable and our subsequent choice to use basic DSA here. Finally, we want to reiterate that computing strategically sensible behavior is a nightmare in terms of computational complexity. This issue cannot be helped as we aimed at developing a framework enabling agents to adopt a wide range of human negotiation behaviors.

Further research avenues are the design of new (more expressive or powerful) moves. Prime candidates are moves that query which offers [set of offers] are not (never) acceptable after certain (even after all possible) updates of the argumentation framework, moves that allow an advanced formal incorporation of explanations ["I make this move, because I want to achieve goal *X*] and moves that incorporate threats, rewards and appeals. And finally an extension of the protocol to enable more than two agents to interact could be developed.

We also want to mention that soft concepts such as affect and emotions play a prominent role in human negotiations. A representation of these concepts would make it possible to design soft moves such as "How would your internal soft state change, if I successfully argue for/against X?"

Furthermore, strategic behavior and the computational complexity of such behavior are promising topics for further research. Agents need a formal logical system to reason about negotiation tactics and strategies, thus the development of such logics seems like a logical next step for further work. Clearly, designing sophisticated agents able to maximize the potential in our protocol is a promising research path, which we are currently following. Designing such agents will enable us to evaluate the protocol and such agent's strategies with computer simulations.

References

1. Amgoud, L., Cayrol, C., Lagasquie-Schiex, M.: On the bipolarity in argumentation frameworks. In: Proceedings of NMR-UF, pp. 1–9 (2004)
2. Amgoud, L., Dimopoulos, Y., Moraitis, P.: A unified and general framework for argumentation-based negotiation. In: Proceedings of AAMAS, pp. 158:1–158:8. ACM (2007)
3. Amgoud, L., Parsons, S., Maudet, N.: Arguments, Dialogue, and Negotiation. In: Horn, W. (ed.) Proceedings of ECAI, pp. 338–342. IOS Press (2000)
4. Amgoud, L., Vesic, S.: On Revising Argumentation-Based Decision Systems. In: Sossai, C., Chemello, G. (eds.) ECSQARU 2009. LNCS, vol. 5590, pp. 71–82. Springer, Heidelberg (2009)
5. Anandalingam, G., Day, R.W., Raghavan, S.: The Landscape of Electronic Market Design. Manage Sci. 51(3), 316–327 (2005)
6. Atkinson, K.: Value-Based Argumentation for Democratic Decision Support. In: Proceedings of COMMA, pp. 47–58. IOS Press (2006)
7. Baroni, P., Giacomin, M.: On principle-based evaluation of extension-based argumentation semantics. Artif. Intell. 171(10-15), 675–700 (2007)
8. Baroni, P., Giacomin, M.: Semantics of Abstract Argument Systems. In: Simari, G., Rahwan, I. (eds.) Argumentation in Artificial Intelligence, pp. 25–44. Springer (2009)
9. Baroni, P., Giacomin, M., Guida, G.: SCC-recursiveness: a general schema for argumentation semantics. Artif. Intell. 168(1-2), 162–210 (2005)
10. Barringer, H., Gabbay, D., Woods, J.: Temporal Dynamics of Support and Attack Networks: From Argumentation to Zoology. In: Hutter, D., Stephan, W. (eds.) Mechanizing Mathematical Reasoning. LNCS (LNAI), vol. 2605, pp. 59–98. Springer, Heidelberg (2005)
11. Bench-Capon, T.: Agreeing to Differ: Modelling Persuasive Dialogue Between Parties With Different Values. Informal Logic 22(3), 231–245 (2002)
12. Bench-Capon, T.J., Doutre, S., Dunne, P.E.: Audiences in argumentation frameworks. Artif. Intell. 171(1), 42–71 (2007)
13. Bench-Capon, T.J., Dunne, P.E.: Argumentation in artificial intelligence. Artif. Intell. 171, 619–641 (2007)
14. Bisquert, P., Cayrol, C., de Saint-Cyr, F., Lagasquie-Schiex, M.C.: Change in Argumentation Systems: Exploring the Interest of Removing an Argument. In: Benferhat, S., Grant, J. (eds.) SUM 2011. LNCS, vol. 6929, pp. 275–288. Springer, Heidelberg (2011)
15. Blum-Kulka, S., Blondheim, M., Hacohen, G.: Traditions of dispute: from negotiations of talmudic texts to the arena of political discourse in the media. J. Pragmatics 34(10-11), 1569–1594 (2002)

16. Boella, G., Gabbay, D., van der Torre, L., Villata, S.: Meta-Argumentation Modelling I: Methodology and Techniques. Studia Logica 93, 297–355 (2009)
17. Boella, G., Kaci, S., van der Torre, L.: Dynamics in Argumentation with Single Extensions: Attack Refinement and the Grounded Extension (Extended Version). In: McBurney, P., Rahwan, I., Parsons, S., Maudet, N. (eds.) ArgMAS 2009. LNCS (LNAI), vol. 6057, pp. 150–159. Springer, Heidelberg (2010)
18. Brewka, G.: Dynamic Argument Systems: A Formal Model of Argumentation Processes Based on Situation Calculus. J. Logic Comput. 11(2), 257–282 (2001)
19. Brewka, G., Eiter, T.: Argumentation Context Systems: A Framework for Abstract Group Argumentation. In: Erdem, E., Lin, F., Schaub, T. (eds.) LPNMR 2009. LNCS, vol. 5753, pp. 44–57. Springer, Heidelberg (2009)
20. Broersen, J., Dastani, M., Hulstijn, J., Huang, Z., van der Torre, L.: The BOID architecture: Conflicts between beliefs, obligations, intentions, and desires. In: Proceedings of Autonomous Agents, pp. 9–16. ACM (2001)
21. Caminada, M.: Semi-Stable Semantics. In: Dunne, P.E., Bench-Capon, T.J. (eds.) Proceedings of COMMA, vol. 144, pp. 121–130. IOS Press (2006)
22. Caminada, M., Amgoud, L.: On the evaluation of argumentation formalisms. Artif. Intell. 171(5-6), 286–310 (2007)
23. Caminada, M., Pigozzi, G.: On judgment aggregation in abstract argumentation. Auton Agent Multi Agent Syst. 22, 64–102 (2011)
24. Cayrol, C., Devred, C., Lagasquie-Schiex, M.: Handling Ignorance in Argumentation: Semantics of Partial Argumentation Frameworks. In: Mellouli, K. (ed.) ECSQARU 2007. LNCS (LNAI), vol. 4724, pp. 259–270. Springer, Heidelberg (2007)
25. Cayrol, C., de Saint-Cyr, F., Lagasquie-Schiex, M.: Revision of an Argumentation System. In: Proceedings of KR, pp. 124–134 (2008)
26. Chesñevar, C.I., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G.R., South, M., Vreeswijk, G., Willmott, S.: Towards an Argument Interchange Format. Knowl. Eng. Rev. 21(4), 293–316 (2006)
27. Coase, R.H.: The Nature of the Firm. Economica 4(16), 386–405 (1937)
28. Coase, R.H.: The Problem of Social Cost. J. Law Econ. 3(1), 1–44 (1960)
29. Coste-Marquis, S., Devred, C., Konieczny, S., Lagasquie-Schiex, M.C., Marquis, P.: On the merging of Dung's argumentation systems. Artif. Intell. 171(10-15), 730–753 (2007)
30. Coste-Marquis, S., Devred, C., Marquis, P.: Prudent Semantics for Argumentation Frameworks. In: Proceedings of ICTAI, pp. 568–572 (2005)
31. Davis, R., Smith, R.G.: Negotiation as a metaphor for distributed problem solving. Artif. Intell. 20(1), 63–109 (1983)
32. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning and logic programming. In: Proceedings of IJCAI, pp. 852–857. Morgan Kaufmann (1993)
33. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. 77(2), 321–357 (1995)
34. Dung, P., Mancarella, P., Toni, F.: A dialectic procedure for sceptical, assumption-based argumentation. In: Proceedings of COMMA, pp. 145–156. IOS Press (2006)
35. Dunne, P.E., Hunter, A., McBurney, P., Parsons, S., Wooldridge, M.: Weighted argument systems: Basic definitions, algorithms, and complexity results. Artif. Intell. 175(2), 457–486 (2011)
36. Fatima, S.S., Wooldridge, M., Jennings, N.R.: An agenda-based framework for multi-issue negotiation. Artif. Intell. 152(1), 1–45 (2004)
37. Fisher, R., Ury, W.: Getting to Yes: Negotiating agreement without giving in. Houghton Mifflin (1981)

38. Gorogiannis, N., Hunter, A.: Instantiating abstract argumentation with classical logic arguments: Postulates and properties. Artif. Intell. 175(9-10), 1479–1497 (2011)
39. Haenni, R., Romeijn, J.W., Wheeler, G., Williamson, J.: Probabilistic Argumentation. In: Probabilistic Logics and Probabilistic Networks, Synthese Library, vol. 350. Springer (2011)
40. Jennings, N.R., Parsons, S., Noriega, P., Sierra, C.: On Argumentation-Based Negotiation. In: Proceedings of IWMAS, pp. 1–7 (1998)
41. Jennings, N., Faratin, P., Lomuscio, A., Parsons, S., Wooldridge, M., Sierra, C.: Automated Negotiation: Prospects, Methods and Challenges. Group Decis. Negot. 10, 199–215 (2001)
42. Johnson, M.W., McBurney, P., Parsons, S.: A Mathematical Model of Dialog. Electronic Notes in Theoretical Computer Science 141(5), 33–48 (2005)
43. Karunatillake, N.C., Jennings, N.R.: Is It Worth Arguing? In: Rahwan, I., Moraïtis, P., Reed, C. (eds.) ArgMAS 2004. LNCS (LNAI), vol. 3366, pp. 234–250. Springer, Heidelberg (2005)
44. Kośmicki, P.: A Platform for the Evaluation of Automated Argumentation Strategies. In: Szczuka, M., Kryszkiewicz, M., Ramanna, S., Jensen, R., Hu, Q. (eds.) RSCTC 2010. LNCS, vol. 6086, pp. 494–503. Springer, Heidelberg (2010)
45. Kraus, S., Sycara, K., Evenchik, A.: Reaching agreements through argumentation: a logical model and implementation. Artif. Intell. 104(1-2), 1–69 (1998)
46. Lewicki, R., Saunders, D., Minton, J., Roy, J., Lewicki, N.: Negotiation, 6th edn. McGraw-Hill (2010)
47. Liao, B., Jin, L., Koons, R.C.: Dynamics of argumentation systems: A division-based method. Artif. Intell. 175(11), 1790–1814 (2011)
48. Lomuscio, A.R., Wooldridge, M., Jennings, N.R.: A Classification Scheme for Negotiation in Electronic Commerce. Group Decis. Negot. 12, 31–56 (2003)
49. Lopes, F., Coelho, H.: Strategic and Tactical Behaviour in Automated Negotiation. Int. J. Artif. Intel.l 4(S 10), 35–63 (2010)
50. Lopes, F., Wooldridge, M., Novais, A.: Negotiation among autonomous computational agents: principles, analysis and challenges. Artif. Intell. Rev. 29, 1–44 (2008)
51. Mackenzie, J.D.: Question-Begging in Non-Cumulative Systems. J. Philos Logic 8(1), 117–133 (1979)
52. McBurney, P., Hitchcock, D., Parsons, S.: The Eightfold Way of Deliberation Dialogue. Int. J. Intell. Syst. 22(1), 95–132 (2007)
53. McBurney, P., Parsons, S.: Games That Agents Play: A Formal Framework for Dialogues between Autonomous Agents. J. Logic Lang. Inf. 11, 315–334 (2002)
54. McBurney, P., Parsons, S., Wooldridge, M.: Desiderata for Agent Argumentation Protocols. In: Proceedings of AAMAS, pp. 402–409 (2002)
55. Oikarinen, E., Woltran, S.: Characterizing strong equivalence for argumentation frameworks. Artif. Intell. 175(14-15), 1985–2009 (2011)
56. Oren, N., Norman, T., Preece, A.: Information Based Argumentation Heuristics. In: Maudet, N., Parsons, S., Rahwan, I. (eds.) ArgMAS 2006. LNCS (LNAI), vol. 4766, pp. 161–174. Springer, Heidelberg (2007)
57. Oren, N., Norman, T.J., Precce, A.: Subjective logic and arguing with evidence. Artif. Intell. 171(10-15), 838–854 (2007)
58. Paglieri, F., Castelfranchi, C.: Why argue? Towards a cost-benefit analysis of argumentation. Argument & Computation 1(1), 71–91 (2010)
59. Parsons, S., Jennings, N.: Negotiation through argumentation - a preliminary report. In: Proceedings of ICMAS, pp. 267–274 (1996)
60. Parsons, S., Sierra, C., Jennings, N.R.: Agents That Reason and Negotiate by Arguing. J. Logic Comput. 8(3), 261–292 (1998)
61. Prakken, H.: Relating Protocols For Dynamic Dispute With Logics For Defeasible Argumentation. Synthese 127, 187–219 (2001)

62. Prakken, H.: Formal systems for persuasion dialogue. Knowl. Eng. Rev. 21(2), 163–188 (2006)
63. Rahwan, I., Larson, K.: Mechanism design for abstract argumentation. In: Proceedings of AAMAS, pp. 1031–1038 (2008)
64. Rahwan, I., Pasquier, P., Sonenberg, L., Dignum, F.: A formal analysis of interest-based negotiation. Ann. Math. Artif. Intell. 55, 253–276 (2009)
65. Rahwan, I., Ramchurn, S., Jennings, N.R., Mcburney, P., Parsons, S., Sonenberg, L.: Argumentation-based negotiation. Knowl. Eng. Rev. 18(04), 343–375 (2003)
66. Rahwan, I., Simari, G.R.: Argumentation in Artificial Intelligence. Springer (2009)
67. Rahwan, I., Sonenberg, L., McBurney, P.: Bargaining and Argument-Based Negotiation: *Some Preliminary Comparisons*. In: Rahwan, I., Moraïtis, P., Reed, C. (eds.) ArgMAS 2004. LNCS (LNAI), vol. 3366, pp. 176–191. Springer, Heidelberg (2005)
68. Raiffa, H., Richardson, J., Metcalfe, D.: Negotiation analysis: the science and art of collaborative decision making. Belknap Press (2002)
69. Rao, A., Georgeff, M., et al.: BDI agents: From theory to practice. In: Proceedings ICMAS, pp. 312–319 (1995)
70. Reyes-Moro, A., Rodríguez-Aguilar, J.A., López-Sánchez, M., Cerquides, J., Gutierrez-Magallanes, D.: Embedding Decision Support in E-Sourcing Tools: Quotes, A Case Study. Group Decis. Negot. 12(4), 347–355 (2003)
71. Rosenschein, J.S., Zlotkin, G.: Rules of Encounter: Designing Conventions for Automated Negotiation among Computers. MIT Press (1994)
72. Rotstein, N.D., García, A.J., Simari, G.R.: Defeasible Argumentation Support for an Extended BDI Architecture. In: Rahwan, I., Parsons, S., Reed, C. (eds.) Argumentation in Multi-Agent Systems. LNCS (LNAI), vol. 4946, pp. 145–163. Springer, Heidelberg (2008)
73. Rueda, S.V., García, A.J., Simari, G.R.: Argument-based Negotiation among BDI agents. J. Comp. Sci. & Technol. 2(7), 1–8 (2002)
74. Rueda, S.V., Martinez, M.V.: A framework for deliberation and negotiation among BDI agents. J. Comp. Sci. & Technol. 5(4), 334–341 (2005)
75. Sandholm, T.W.: Distributed Rational Decision Making. In: Weiss, G. (ed.) Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence, ch.5, pp. 201–258. MIT Press (1999)
76. Schmid, B., Lindemann, M.A.: Elements of a Reference Model for Electronic Markets. In: Proceedings of HICSS, pp. 193–201 (1998)
77. Searle, J.R.: Speech acts: An essay in the philosophy of language. Cambridge Univerity Press (1969)
78. Sierra, C., Jennings, N., Noriega, P., Parsons, S.: A Framework for Argumentation-Based Negotiation. In: Rao, A., Singh, M.P., Wooldridge, M.J. (eds.) ATAL 1997. LNCS, vol. 1365, pp. 177–192. Springer, Heidelberg (1998)
79. Smith, R.: The Contract Net Protocol: High-Level Communication and Control in a Distributed Problem Solver. IEEE_J_C 100(12), 1104–1113 (1980)
80. Ströbel, M., Weinhardt, C.: The Montreal Taxonomy for Electronic Negotiations. Group Decis. Negot. 12, 143–164 (2003)
81. Tohmé, F., Bodanza, G., Simari, G.: Aggregation of Attack Relations: A Social-Choice Theoretical Analysis of Defeasibility Criteria. In: Hartmann, S., Kern-Isberner, G. (eds.) FoIKS 2008. LNCS, vol. 4932, pp. 8–23. Springer, Heidelberg (2008)
82. Walton, D.N., Krabbe, E.C.: Commitment in dialogue: basic concepts of interpersonal reasoning. State University of New York (1995)

A Systematic Success Factor Analysis in the Context of Enterprise 2.0: Results of an Exploratory Analysis Comprising Digital Immigrants and Digital Natives

Dietmar Nedbal[1], Andreas Auinger[1], Alexander Hochmeier[1], and Andreas Holzinger[2]

[1] Research Unit Digital Economy, Department for Marketing and Electronic Business, University of Applied Sciences Upper Austria, Wehrgrabengasse 1-3, 4400 Steyr/Austria
{dietmar.nedbal,andreas.auinger,alexander.hochmeier}@fh-steyr.at
[2] Research Unit Human–Computer Interaction, Institute for Medical Informatics, Statistics and Documentation, Medical University Graz, Auenbruggerplatz 2/V, 8036 Graz/Austria
andreas.holzinger@medunigraz.at

Abstract. Organizations are increasingly investing in social collaboration and communication platforms for integrated exchange of information within and between enterprises. These Enterprise 2.0 projects always have a deep impact on organizational and cultural changes and need a critical mass of user involvement across all different groups. Users that grew up in the digital age and use new forms of collaborative platforms within their daily activities are often more technologically adept and more willing to share information. This leads to a digital divide between Digital Natives and Digital Immigrants, which needs to be addressed within such projects. The main objective of this paper is to investigate the perceived differences in success factors for Enterprise 2.0 seen by Digital Natives and Digital Immigrants and its implications on the implementation of a process oriented methodology for Enterprise 2.0 projects.

Keywords: Success Factor Analysis, Process orientation, Innovation Management, Enterprise 2.0, Technology Experience, Digital Immigrants, Digital Natives.

1 Introduction

Effective collaboration of organizations cooperating in a flexible business network is one of the competitive advantages on the global market [12] and is especially important in today's challenging economic situation [11]. Globalization has caused a significant shift in business processes, from static solutions to flexible processes that can address rapidly changing business needs, also considering virtual supply chains, where business partners change seamlessly as new business opportunities arise [1]. The term "Enterprise 2.0" is defined in this context as the use of interactive and collaborative Web 2.0 concepts and technologies within and between enterprises [20]. The focus of this research lies on the shift of user paradigms of Web 2.0 concepts and technologies like blogs, wikis, tagging, rating, social networking, etc. which provide the foundation for user-generated content [25]. This offers great opportunities for

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 163–175, 2012.

more flexible ways of communication, loosely-coupled process integration, ad-hoc information exchange and improved possibilities in idea generation, when embedded and correctly executed within business process to guarantee the success of such a platform [28]. Especially social networks (i.e. one Enterprise 2.0 concept) focusing on research and development are able to drive an enterprise's success and innovation, as Web 2.0 based solutions offer better ways to make tacit knowledge transparent than traditional, standardized IT solutions, because they enable new means of communication and collaboration [24]. But the adoption of Enterprise 2.0 differs from common IT projects by their nature for the following reasons [4, 7, 17]: They always have a (i) *deep impact on organizational and cultural changes* by enabling employees to pro-actively enlarge their own role, (ii) mandatorily *need a critical mass of user involvement*, (iii) *have to face the fact of missing best practices and reputation*, (iv) are *not yet an established part of a company's state-of-the-art IT portfolio*, and (v) *confront the users with unused ways of working with IT systems* (e.g. the use of tagging, the syntax of enterprise wikis,...). This implies that Enterprise 2.0 needs to be user centered and needs to target at both technical (e.g. usability of the system) and organizational (e.g. support business processes) success factors.

As Enterprise 2.0 introduces new IT systems within organizations, technology acceptance is a crucial aspect to be considered. A previous study showed that there is a strong correlation with a factor called Previous Exposure to Technology (PET) [14]. From the users' perspective, it can be stated that users that grew up in the digital age and use new forms of collaborative IT systems within their daily work activities and private (internet) life are often more technologically adept (see the above mentioned study on PET, which provides a proof). Subsequently, these people have a wealth of virtual experience from visiting, using and interacting with the Web (2.0) [19]. This leads to a new form of the digital divide, which is also referred to as "social divide" or "Digital Divide 2.0" [30]. Applied to the organizational context this implies that (i) *younger employees have more affinity to such tools* whereas (ii) *older employees on the other hand tend to have more knowledge*. This results in a gap in knowledge sharing when those who are willing to share do not have the knowledge ("Digital Natives") and those who are experts are not willing to share this knowledge ("Digital Immigrants") in a (semi-)public forum or environment [22] such as an enterprise. Additionally, Hoberg and Gohlke identified the challenge for enterprises, that older employees tend to be less open to use new technologies (e.g. Why should I use these new Enterprise 2.0 tools?) and willing to learn how to use new technologies (e.g. How to use new Enterprise 2.0 tools?)[13]. To investigate into this willingness and ability to use Enterprise 2.0 from the users' perspective is therefore an important issue that is still in initial stages [9]. This is also underlined by Renken et al. by stating that further research regarding the influence of socio-demographic aspects like the users' age on acceptance of Enterprise 2.0 is needed [24].

The *objective* of this paper is to shade some light onto success factors when implementing Enterprise 2.0 projects and explore their perception from the viewpoint of Digital Natives and Digital Immigrants. The main research question addressed accordingly was: *What is the perceived difference in success factors for Enterprise 2.0 seen by Digital Natives and Digital Immigrants and what are the success factors to be addressed in conjunction with Enterprise 2.0 projects to match the different* *requirements of Digital Natives and Digital Immigrants?* To answer this question this paper first briefly introduces a typical methodology for process oriented implementation of Enterprise 2.0 in companies in section 2. Section 3 shows a success factor analysis method for identifying critical success factors. Via literature research, critical success factors for the context of this paper are identified. An exploratory online survey was undertaken to answer the main research question. With it, insight into the findings that need to be addressed from the viewpoint of the two groups of Digital Natives and Digital Immigrants are given.

2 Process Oriented Methodology for Enterprise 2.0 Projects

Within a three years research project a participative, evolutionary approach for implementing Enterprise 2.0 platforms was created [2], which is a necessity for the success of such projects [17]. The methodology was practically evaluated it in pilot projects carried out with three Austrian mid-sized companies within their organizations and selected supply chain partners. The overall methodology in the Enterprise 2.0 projects included five phases, that are both common and well-established within IT projects: Assessment ("Whether to start the Enterprise 2.0 project"), Analysis ("What are the requirements?"), Design ("How can the requirements be realized?"), Realization ("Do the implementation and roll it out"), and Operation ("Support and evaluate the productive information system"). The activities within these phases are especially tailored for Enterprise 2.0 projects, as specific methods were used to address the success factors for Enterprise 2.0 and change projects [3]. This includes evaluating a company's organizational structure, its business processes and recent pains and needs, as well as its organizational experience (e.g. projects that failed in the past) [10, 21]. Having analyzed the key factors, an adequate approach addressing them from a process oriented view was developed. Within the overall approach the following specific methods are used in this context:

1. *Standardized questionnaires* were used to identify basic needs of the users regarding the current situation in communication, documentation, project and innovation management and collaboration with the supply chain partners;
2. A *stakeholder analysis* was carried out to find out the attitude of the involved employees towards the project;
3. *Workshops with semi-structured interviews* were undertaken to identify and document relevant information systems and involved business processes that could be supported by Enterprise 2.0;
4. An additional *success factor analysis* was carried out to identify issues of high priority that are supported insufficiently, which is on the main focus in the following. On this basis, concepts for Enterprise 2.0 tools addressing the mentioned issues were developed;
5. The Enterprise 2.0 platform was implemented using the concepts of *perpetual beta*;
6. Besides *training* of the end users at an early stage the IT department and admin users were trained to enable them to *maintain and further develop* the platform themselves;

7. *Usability evaluation* of the beta releases was conducted by *eye tracking* methodology and *heuristic evaluation*, and
8. Continuous *feedback* was collected using a project blog. The feedback, the usability and heuristic evaluation results were important inputs for the *continuous improvement* of the platform.

In the course of steps (1), (2), (3), (4), (6) and (7) the skills and demands of Digital Natives and Digital Immigrants may differ in highly challenging dimensions. Therefore, it is important to investigate their specific requirements and address the underlying success factors.

3 Success Factor Analysis in the Context of Enterprise 2.0

For the purpose of this paper, special attention is now drawn to the success factor analysis within the Analysis phase (cf. section 2), because within this phase the critical success factors regarding Digital Natives and Digital Immigrants shall be made transparent. The analysis' results are crucial for a tailored design and realization phase, as addressing the identified success factors for both user groups should ease the adoption of the solution.

3.1 Related Work and Methodology

Related work for this research can be found in studies for technology adoption as well as on the digital divide. Most related studies about the digital divide address the criterion of *access* to new information technologies, particularly as embodied in the Internet [26]. Robinson et al. show that those who have made it online are also unequal with respect to the ways they use the medium, especially the content they access from the Internet. Factors like education, income, age and marital status are also associated with more long-term technology use [26]. There is also already research in success factors for the adoption of Enterprise 2.0 [21], intra-organizational IT projects [29], change management [16], and knowledge management in organizations [18]. Because it proofed as suitable framework to assess the success of knowledge management and knowledge transfer in organizations analysis, KnowMetrix [18] was used as the prime basis for this research. Table 1 summarizes the success factors from this related work.

Franken et al. point out that organizations "have limited time and resources that they can devote to executing strategic change; hence, it is critical that change programs are prioritized. This requires an effective aligning and filtering process, as the number of suggested change programs is typically too great for an organization to pursue" [10]. This requirement especially can be met by a success factor analysis, as it allows arranging important factors according to their perceived importance. Beyond that, the authors intended to raise awareness and participation for the Enterprise 2.0 project among all users and to point out possible differences between the group of Digital Natives and Digital Immigrants. To distinguish the two groups we follow Palfrey and Gasser [22]: They characterize Digital Natives as born after

1980, when social digital technologies came online. They all have access to and the skills to use those technologies. And they are much more willing to share information over the Internet than Digital Immigrants. Digital Immigrants on the other side were born before 1980 and therefore have learned how to use email and other social technologies later in life.

Table 1. Success factors for project success and adoption

Success Factor	Source
Need for change and feasibility analysis of the new system	Sutanto et al. [29]
Top management support	Sutanto et al. [29]
Shared vision for system-related change	Sutanto et al. [29]
Systematic plan for project and change management	Sutanto et al. [29]
Institutionalization of system-related change	Sutanto et al. [29]
Energy for system-related change	Sutanto et al. [29]
Promote a balanced change culture	Ibbs et al. [16]
Recognize change	Ibbs et al. [16]
Evaluate change	Ibbs et al. [16]
Implement change	Ibbs et al. [16]
Continuously Improve from Lessons Learned	Ibbs et al. [16]
Determine desired results, then deploy appropriate emergent social software platforms	McAfee [21]
Prepare for the long haul	McAfee [21]
Communicate, educate, and evangelize	McAfee [21]
Move emergent social software platforms into the Flow	McAfee [21]
Measure progress, not ROI	McAfee [21]
Show that Enterprise 2.0 is valued	McAfee [21]
Knowledge management as service and cross-divisional function within organization (containing 12 specific items)	Lehner et al. [18]
Knowledge transfer (containing 13 specific items)	Lehner et al. [18]

Table 2. User anxiety and resulting metrics [15]

Question	Resulting Metric
Can I trust it?	How can passive technology be made more trustworthy?
Can I switch it off/on?	How can we make it more controllable?
Can I understand it?	How can we improve understanding of the principles and functionality, without too many confusing details?
Will it obey me?	Can we remove the Frankenstein element; turn it from “magic” to machine, thereby inspiring confidence?
Who can see me?	Can we counteract the Big Brother element; Replace the fear of being controlled with a feeling of being in control?
Do I really need this?	Explanation of benefits and purposes, appropriateness of the measures taken.

The question of why a certain technology is adopted by individuals often is based on the Technology Acceptance Model (TAM) by Davis [8]. TAM posits that “perceived usefulness” and “perceived ease of use” are the fundamental determinants of an individual's intention to use a system. Also in this context the usability metrics found by Holzinger et al. [15] are here very important: They show that metrics for the evaluation of trustworthiness and acceptance of passive technology for the elderly must be approached from the viewpoint of the elderly. There is a strong analogy between user anxiety and metrics (cf. Table 2). To meet these demands questions addressing these aspects were integrated into the questionnaire.

3.2 Setting, Subjects and Instrument

To involve all necessary stakeholders in the analysis (cf. step (iv) mentioned in section 2), a questionnaire focusing on the priority of relevant processes and the recent satisfaction with its efficiency was issued. The authors conducted the success factor analysis within an Austrian enterprise in the energy sector. The company has grown since its foundation in the 1940s to a multinational organization with over 3200 employees situated across five production sites located in Austria, Czech Republic, and Ukraine. The R&D and related departments are mainly scattered around the production sites, working together on their three strategic business areas battery charging, welding, and solar. To emphasize on their strategy to foster employees' enthusiasm for customer-oriented activities and innovations, the top-management set up a project in the Enterprise 2.0 field.

To assure that all important factors were included in the questionnaire, the analysis was initially based on KnowMetrix, which contains success factors for knowledge management consistently identified in literature [18]. As indicated by KnowMetrix, the success factors were adjusted to the needs of the organization via workshops. The workshops especially helped to identify specific information systems and involved business processes (items in block 1 "software support") and to recognize the importance of Enterprise 2.0 for innovation management activities. The questionnaire was finally extended with factors from relevant literature and addressing user anxiety and the viewpoint of the elderly (cf. Table 1) in order to achieve perceived usefulness and ease of use in the following. The factors were clustered into the following five blocks:

1. *Software support and overall usability*: How would you rate the general software support of the following processes in your company?
   (b) Knowledge documentation
   (c) Social networking (within thematic networks)
   (d) Search and find knowledge carriers ("Who knows what…")
   (e) Idea management: generation, discussion, evaluation, and selection of creative new ideas (innovations)
   (f) Communication support across departments, divisions, teams and projects
   (g) Alternative communication channels: pull instead of push
   (h) FAQs: Documentation of frequently asked questions and solutions
   (i) Rapid decision-making support for a certain topic
   (j) Document management
2. *Organizational culture*: How would you rate the following cultural aspects in your company?

(a) Incorporation of all employees – regardless of hierarchy or function – in the innovation process
(b) Expression of feedback and criticism across hierarchies
(c) Quick and unfiltered information sharing across hierarchy levels
(d) Joint solutions for complex tasks, challenges and problems through direct communication, knowledge sharing and mutual support (no egoisms)
(e) Access to data across departments or projects (permissions)
(f) Open innovation culture in the sense of a clear commitment to new ideas
(g) Willingness to share and trust the knowledge shared ("Culture of trust")
(h) Possibility to make mistakes and learn from them ("Fault tolerance")
(i) Colleagues actively seek to solve your problem if asked for help
(j) Feel free to ask questions to colleagues

3. *General requirements for innovation*: How would you rate the following general conditions for innovation in your company?
(a) Shared corporate vision, shared goals and values within the organization
(b) Top management support of innovation projects
(c) Availability of sufficient resources (time, money, personnel, tools) for innovation projects
(d) Capability of the organization to realize own innovative ideas independently
(e) Identify and evaluate trends and megatrends as the basis for competitiveness and strategic alignment
(f) Speed and quality of decision-making within the organization

4. *Management of knowledge and innovation*: How would you rate the following aspects concerning the use of knowledge and innovation in your company?
(a) Existence of awareness, motivation for knowledge sharing
(b) Possibility to access new and exchange existing knowledge
(c) Overview of knowledge and skills within the organization
(d) Integration of the environment (eg. customers, suppliers, benchmarking, alternative industries, other sources such as literature, events and platforms, evaluation and feedback) into innovation projects
(e) Cooperation with external partners in innovation process (especially in terms of establishing long-term, trust-based partnerships)
(f) Use of all possibilities for identification of customer needs (eg. market research, customer surveys, direct involvement of customers in the development)
(g) Existing possibilities for idea management to systematically collect, categorize, evaluate and select ideas of all employees
(h) Spawning of radical innovations (in addition to the improvement or adaptation of existing products, processes, or services)

5. *Personal situation*: How would you rate the following aspects concerning the personal situation of employees in your company?
(a) Availability of incentives for innovation (monetary / non-monetary)
(b) Sufficient time to develop and carry out new ideas
(c) Availability of sufficient know-how support from other departments for new ideas

(d) Employees have enough decision-making competencies (i.e. not hindered by rules, excessive control, or lack of trust)
(e) Existence of a trusting relationship between employees and supervisors

The factors were queried (in German language) in a standardized questionnaire according to their performance and priority in Austria's school grading system from 1 (best) to 5 (worst): The respondents had to rate the current performance of the factor in question and should provide a priority for that factor. As a result, a factor rated with a low performance is only a problem when the priority for this factor is high, and vice versa. Factors with high priority and low satisfaction are to be targeted first.

The questionnaire was undertaken online via Qualtrics (www.qualtrics.com). The link to the online survey was issued to a small group of beta-users (38 employees) via email. The probands who took part were from the R&D, IT and human resources departments. 33 people (23 "Digital Immigrants" and 10 "Digital Natives") responded to the survey representing a response rate of 87%. Besides demographic data (position within company, department, period of employment, usage of Web 2.0, Digital Immigrant or Digital Native) the identified relevant success factors were queried.

3.3 Survey Results and Discussion

As already mentioned, the success factors should include user anxiety and the viewpoint of the elderly (cf. Table 1). Addressing the perceived usefulness and ease of use right from the beginning of the project by involving the users should also help to increase the actual usage of the system according to TAM. Therefore each of the blocks considers different parts of them. For example block 2 (organizational culture) and block 5 (personal situation) contain questions addressing the mentioned user anxiety aspects "Can I trust it?" and "Who can see me?" by questions concerning the "Culture of Trust (2g)", "Feel free to ask questions (2j)", and "Trusting relationship between employees and supervisors (5e)" (cf. Figure 1). These questions target on the differences regarding Digital Immigrants and Digital Natives and their perceived level of trust and feeling of security. Block 1 (software support and overall usability) and block 4 (management of knowledge and innovation) include questions considering the anxiety aspect "Do I really need this?" as, amongst other things, they help to make gaps between existing software support and the users' demand transparent, which can be met by Enterprise 2.0 tools: e.g. "Knowledge documentation (1a)", and "Overview of knowledge and skills (4c)".

Figure 1 shows the evaluation of the relevant success factors both for Digital Immigrants and Digital Natives. In general, factors having a high priority need to be examined in relation to the measures and whose performance has to be improved. The main focus of the discussion is now on the similarities and differences in the rating of the factors between the two groups of Digital Immigrants and Digital Natives:

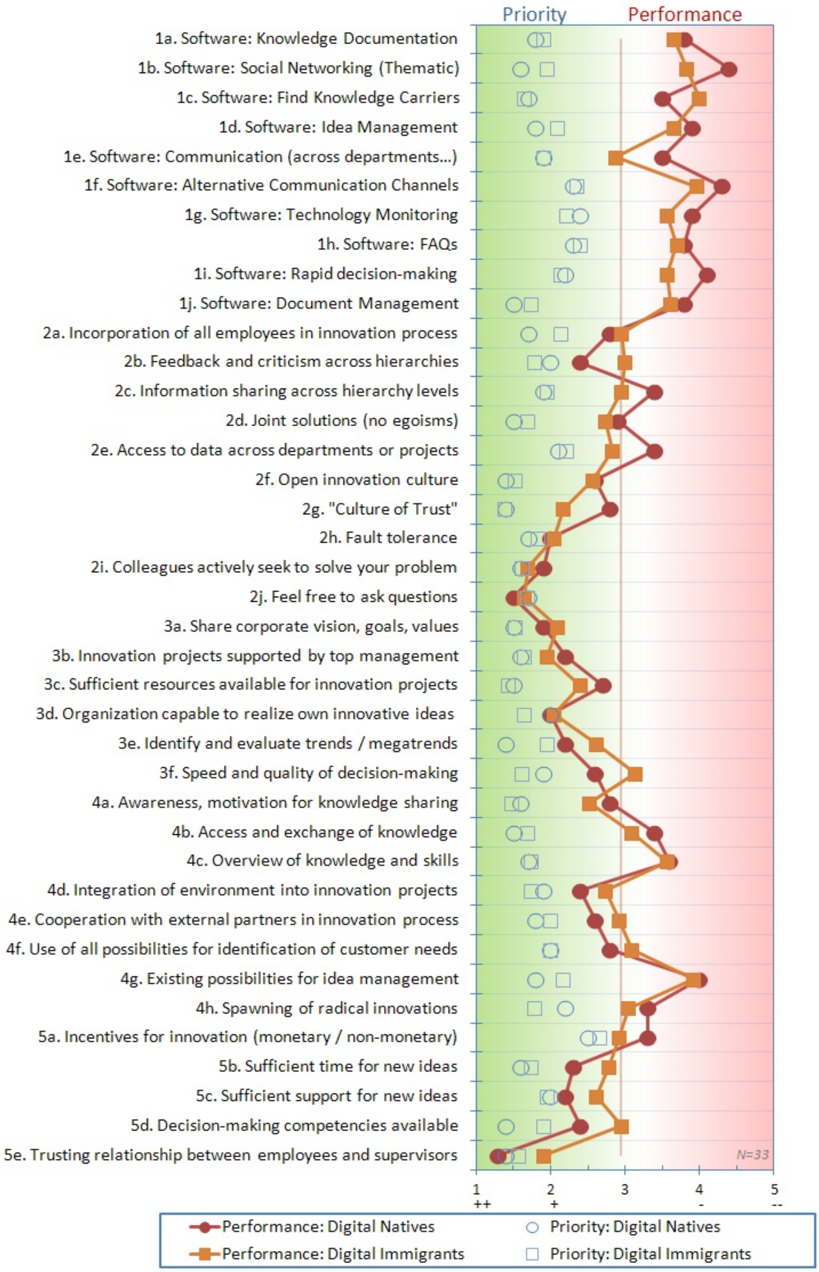


Fig. 1. Success factor analysis: Digital immigrants (blank squares = priority, filled squares = performance) and Digital Natives (blank circles = priority, filled circles = performance)

(i) Quite a lot of factors showed a consensus in both their performance and priority. This could be observed both for factors with rather good performance (e.g. 2h, 2i, 2j) and lower performance (e.g. 4g). Especially a lot of organizational or “soft” factors (block 2 to 5), that are vital for an Enterprise 2.0 project proved as important, regardless of the group. Most factors also had broad agreement in their priority within the two groups. But, ignoring these factors will lead to the Digital Divide 2.0, or “social divide”, which is about the ability and the willingness to get involved in Enterprise 2.0.

(ii) Digital Natives tend to rate the software and usability support (block 1) lower than the Digital Immigrants: Although Web 2.0 tools are more and more accessible and easy to use this may be caused by a better knowledge of tools available by the Digital Natives. E.g. Digital Natives rated the software support for communication across departments (1e) lower than Digital Immigrants. One explanation could be that Digital Immigrants are not aware of other means for communication than email, telephone, or face-to-face conversation. These differences related to technology shows that especially usability issues need to be considered for both groups. This is in line to previous mentioned research on usability. Especially for Enterprise 2.0 this is very important. Prensky pointed out that “Digital Natives are used to receiving information really fast. They like to parallel process and multi-task. They prefer their graphics before their text rather than the opposite. They prefer random access (like hypertext). They function best when networked.” [23]

(iii) Digital Natives tend to rate their personal situation (block 5) better than the other group. These factors may be more influenced by other types of attributes like their seniority level (e.g. 2b), gender, physical location in the foundation, or personal interests [9] . More investigation into this topic is needed, e.g. theories in IS research based on the adoption of technology on the organizational level like the Diffusion of Innovation theory [27].

4 Conclusion

The paper presents the success factor analysis as a suitable method in the context of Enterprise 2.0 to identify and prioritize needs for action in a transparent and participative way. It shades light on critical success factors from the users’ and organizations’ perspective and combines it with usability aspects for two groups of employees: the Digital Natives, and the Digital Immigrants.

The results show that Enterprise 2.0 projects are multi-faceted: Besides technological aspects, social factors of the different stakeholders need to be considered right from the beginning. This combination of the organizations’ business-oriented view and usability aspects are seen as important factors for the success of such projects. Future research also needs to include factors like the gender, seniority level, physical location in the foundation, and personal interests. E.g. maybe the position within the company or amount of years in the company may weight more than the age or the gender of the employee.

Although the high response rate shows a general interest for this topic in the organization where it was carried out, the very low sample rate and amount of data could only serve as an exploratory analysis in this context. Some other factors were also queried for this research, but the low number of probands did not allow the authors to undertake additional statistical tests, like correlation analysis, or t-test, etc. Nevertheless, the paper should raise awareness for this topic and with the methodology shown and the success factors it may serve for future research in this field.

The implementation of Web 2.0 concepts and technologies in enterprises and supply chains opens the mind of both groups, Digital Natives and Digital Immigrants for new technologies. Matchmaking and intelligent reasoning [5] opens the opportunity to exploit Web 2.0 / Enterprise 2.0 collective knowledge (together with the individual knowledge of the employees) in order to achieve the vision of Web 3.0. Web 3.0, also called the Intelligent Web, refers to the provision of a more productive, personalized and intuitive environment through the integration of Semantic Web and in general Artificial Intelligence technologies emphasizing the information understanding. Semantics is a necessary part of the next generation of the Web [6] but seems also to be a precondition for handling the huge amounts of unstructured knowledge within and among enterprises for the future.

References

1. Ardagna, D., Baresi, L., Comai, S., Comuzzi, M.: A Service-Based Framework for Flexible Business Processes. IEEE Software (2011), doi:10.1109/MS.2011.28
2. Auinger, A., Nedbal, D.: Effective Supply Chain Information Management in Supply Networks using Enterprise 2.0. In: Proceedings of the 2009 International Conference on e-Learning, e-Business Enterprise Information Systems, and e-Government, EEE 2009 (2009)
3. Auinger, A., Hochmeier, A., Nedbal, D.: An Approach For Implementing Enterprise 2.0 Platforms: Methodology And Selected Results Of Pilot Projects. In: Proceedings of the IADIS International Conference on Information Systems 2011 (IS 2011), Avila, Spain, pp. 179–185 (2011)
4. Back, A., Gronau, N., Tochtermann, K.: Web 2.0 in der Unternehmenspraxis: Grundlagen, Fallstudien und Trends zum Einsatz von Social Software, 2nd edn. Oldenbourg, München (2009)
5. Benassi, M., Di Noia, T., Marino, A.: A Matchmaking Architecture to Support Innovation by Fostering Supply and Demand of Venture Capital. In: Psaila, G., Wagner, R. (eds.) EC-Web 2008. LNCS, vol. 5183, pp. 61–70. Springer, Heidelberg (2008)
6. Cena, F., Farzan, R.: & Lops, P. Web 3.0: merging semantic web with social web. In: Proceedings of the 20th ACM Conference on Hypertext and Hypermedia, HT 2009, pp. 385–386. ACM, New York (2009), doi:10.1145/1557914.1558002
7. Chui, M., Miller, A., Roberts, R.P.: Six ways to make Web 2.0 work. McKinsey Quarterly (2009)
8. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly (1989)

9. Ferron, M., Massa, P., Odella, F.: Analyzing collaborative networks emerging in Enterprise 2.0: the Taolin Platform. In: Procedia - Social and Behavioral Sciences (2011), doi:10.1016/j.sbspro.2011.01.010.
10. Franken, A., Edwards, C., Lambert, R.: Executing Strategic Change: Understanding the critical management elements that lead to success. California Management Review (2009)
11. Furdík, K., Mach, M., Sabol, T.: Towards Semantic Modelling of Business Processes for Networked Enterprises. In: Di Noia, T., Buccafurri, F. (eds.) EC-Web 2009. LNCS, vol. 5692, pp. 96–107. Springer, Heidelberg (2009)
12. Gong, R., Li, Q., Ning, K., Chen, Y., O'Sullivan, D.: Business Process Collaboration Using Semantic Interoperability: Review and Framework. In: Mizoguchi, R., Shi, Z., Giunchiglia, F. (eds.) ASWC 2006. LNCS, vol. 4185, pp. 191–204. Springer, Heidelberg (2006)
13. Hoberg, A., Gohlke, P.: Selbstorganisiertes Lernen 2.0: Ein neues Lernkonzept für die berufliche Weiterbildung. HMD - Praxis der Wirtschaftsinformatik (2011)
14. Holzinger, A., Searle, G., Wernbacher, M.: The effect of previous exposure to technology on acceptance and its importance in usability and accessibility engineering. Universal Access in the Information Society (2011)
15. Holzinger, A., Searle, G., Kleinberger, T., Seffah, A., Javahery, H.: Investigating Usability Metrics for the Design and Development of Applications for the Elderly. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 98–105. Springer, Heidelberg (2008)
16. Ibbs, C.W., Wong, C.K., Kwak, Y.H.: Project Change Management System. Journal of Management in Engineering (2001)
17. Koch, M., Richter, A.: Enterprise 2.0: Planung, Einführung und erfolgreicher Einsatz von Social Software in Unternehmen, 2nd edn. Oldenbourg, München (2009)
18. Lehner, F., Amende, N., Wildner, S., Haas, N.: KnowMetrix - Erfahrungen mit der Erfolgsbewertung im Wissensmanagement in einem mittelständischen Unternehmen. In: Hinkelmann, K., Wache, H. (eds.) WM 2009: 5th Conference on Professional Knowledge Management, Solothurn, Switzerland, March 25–27, pp. 470–479 (2009)
19. Lorenzo, C., Constantinides, E., Geurts, P., Gómez, M.A.: Impact of Web Experience on e-Consumer Responses. In: Psaila, G., Wagner, R. (eds.) EC-Web 2007. LNCS, vol. 4655, pp. 191–200. Springer, Heidelberg (2007)
20. McAfee, A.P.: Enterprise 2.0: The Dawn Of Emergent Collaboration. MIT Sloan Management Review (2006)
21. McAfee, A.P.: Enterprise 2.0: New collaborative tools for your organization's toughest challenges. Harvard Business Press, Boston (2009)
22. Palfrey, J.G., Gasser, U.: Born digital: Understanding the first generation of digital natives. Basic Books, New York (2008)
23. Prensky, M.: Digital Natives, Digital Immigrants Part 1. On the Horizon (2001)
24. Renken, U., Bullinger, A.C., Möslein, K.M.: Webbasierte Werkzeuge für Wissensarbeiter. HMD - Praxis der Wirtschaftsinformatik (2011)
25. Rinner, C., Keßler, C., Andrulis, S.: The use of Web 2.0 concepts to support deliberation in spatial decision-making. In: Computers, Environment and Urban Systems (2008), doi:10.1016/j.compenvurbsys.2008.08.004.
26. Robinson, J.P., DiMaggio, P., Hargittai, E.: New social survey perspectives on the digital divide. It & Society (2003)
27. Rogers, E.M.: Diffusion of innovations, 5th edn. Free Press, New York (2003)

28. Rossi, G., Schmid, H., Lyardet, F.: Customizing Business Processes in Web Applications. In: Bauknecht, K., Min Tjoa, A., Quirchmayr, G. (eds.) EC-Web 2003. LNCS, vol. 2738, pp. 359–368. Springer, Heidelberg (2003)
29. Sutanto, J., Kankanhalli, A., Tay, J., Raman, K.S., Tan, B.C.Y.: Change Management in Interorganizational Systems for the Public. Journal of Management Information Systems (2008)
30. Vie, S.: Digital Divide 2.0: "Generation M" and Online Social Networking Sites in the Composition Classroom. Computers and Composition (2008), doi: 10.1016/j.compcom, 09.004

An Experimental Analysis of Online Unidirectional Conversion Problem

Iftikhar Ahmad[1,⋆] and Günter Schmidt[1,2]

[1] Chair of Information and Technology Management, University of Saarland, Saarbrücken, Germany
[2] Department of Statistical Sciences,University of Cape Town, South Africa
{ia,gs}@itm.uni-sb.de

Abstract. Financial markets are highly volatile and decision making in these markets is highly risky. With the introduction of automated trading, a number of techniques are developed to facilitate the automation of financial markets. We consider a set of preemptive as well as non-preemptive online algorithms and evaluate them on real world as well as synthetically produced data. We present extensive computational results based on the observed performance of algorithms in terms of experimentally achieved competitive ratio, number of transactions performed and consistency of the results. We also investigate the gap between the worst case competitive ratio and experimentally achieved competitive ratio and conclude that algorithms perform better than their performance guarantee suggest. We conclude by highlighting a number of open questions.

Keywords: Online algorithms, Experimental evaluation, Gap between theory and practice.

1 Introduction

With the rapid development of e-commerce technologies, a number of techniques are developed to facilitate automated trading in financial markets. We consider the methods proposed in theoretical computer science and evaluate their applicability in financial markets. These strategies are called online algorithms for conversion problem. Unlike other approaches (such as Artificial Neural Networks), online algorithms do not rely on past data. Thus the performance of online algorithms are not effected by the choice of parameters such as past data or forecasts.

In an online unidirectional conversion problem, the aim is to convert an asset D into another asset Y with the objective to maximize the amount of Y after time T. On each day t, the player is offered a price q_t; the player either accepts the offered price and converts whole/portion of her remaining wealth at offered price or alternatively rejects the offered price and waits for a better price. The game ends when the player converts her whole wealth D into Y.

⋆ Corresponding author.

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 176–187, 2012.

A number of online algorithms are proposed to address the unidirectional conversion problem [4,5,6,7,9]. Online algorithms are evaluated using competitive ratio. Competitive ratio measures the performance of an online algorithm against optimum offline algorithm. Let ON be an online algorithm for some maximization problem P and $\mathcal{I}$ be the set of problem instances. Let $ON(I)$ be the performance of ON on input sequence I and $OPT(I)$ be the performance of optimum offline algorithm. The algorithm ON is *c-competitive* if $\forall\ I \in \mathcal{I}$

$$ON(I) \geq \frac{1}{c} \cdot OPT(I). \tag{1}$$

1.1 Motivation

Although a number of solutions are proposed to solve *unidirectional conversion* problems [4,5,6,7,9], there are very few studies to investigate the applicability of these solutions to real world problems - for instance trading in financial markets. Similarly, the variety of solutions proposed are all based on different assumptions such as a priori knowledge about the upper bound of offered prices or fluctuation ratio etc. Chen et al. [4] and Hu et al. [7] compared their proposed solutions to classical buy and hold and dollar average strategy. Mohr and Schmidt [10] compared only a single online algorithm to classical techniques like moving average and buy and hold. To our knowledge, there is no comprehensive study in literature that investigates the applicability of online algorithms to real world problems.

Our aim is to conduct an extensive experimental study to evaluate the performance of online algorithms for *unidirectional conversion* problem and report the findings based on the competitive ratio. Our focus is to find out, how these algorithms fit in the real world scenario. We will identify a set of algorithms that performs better than others and will reason about their performance edge. Moreover we identify a set of problems that needs to be addressed in order to improve the applicability of *online conversion* algorithms in real world applications.

2 Related Work

Experimental analysis of online conversion algorithms has not received much attention, so far. Mohr and Schmidt [10] investigated the empirical and worst case performance of reservation price policy [6] and compared it with buy and hold. Chen et al. [4] and Hu et al.[7] compared their proposed solutions to buy and hold and dollar average strategy. Schmidt, Mohr and Kersch [12] compared threat based algorithm of El-Yaniv et al. [6] to reservation price algorithm, average price algorithm and buy and hold.

In contrast to experimental study of online algorithms, there is a significant amount of experimental studies on heuristic trading algorithms like Moving Average Crossover and Trading Range Breakout (TRB). Buy and hold (BH) is used as benchmark in these studies. Brock et al. [2] conducted an extensive experimental study of Dow Jones Industrial Index (DJIA) from 1897 to 1986. They introduced Moving Average Cross over and Trading Range Breakout (TRB), which are of great interest in literature. They compared the returns of buy (sell) signal on *DJIA* to that generated by autoregressive (AR), generalized autoregressive conditional heteroskedasticity in mean (GARCH-M) and an exponential GARCH. The results found that technical trading rules are superior to BH, AR(1), generalized autoregressive conditional heteroskedasticity in mean (GARCH-M) and an exponential GARCH. Kwon and Kish [8] extended Brock et al. [2] work by studying the predictive ability of Variable Moving Average (VMA), Fixed Moving Average (FMA) and TRB on New York Stock Exchange as well as NASDAQ indices. Other related works include [3,11,13,14].

3 Basic Definitions

We define a set of standard definitions which are used in the remaining of this paper:

i. *Duration (T)*: The length of the investment horizon in which all transactions must be carried out.
ii. *Upper Bound (M)* : The upper bound of prices in the investment horizon.
iii. *Lower Bound (m)* : The lower bound of prices in the investment horizon.
iv. *Fluctuation Ratio (ϕ)*: The predicted maximum fluctuation of prices that can possibly be observed during the time interval, calculated by M/m.
v. *Threat Duration (k)* : Number of days after which the adversary may drop the offered price to some minimum level m and will keep the offered price at minimum level for the rest of the investment horizon, $k \leq T$.
vi. *Amount converted (s_t)* : Specifies which fraction of the amount available is converted at price q_t on day t, $0 \leq s_t \leq 1$.
vii. *Price Function ($g(q_t)$)* : Models price q_t based on some predefined function, e.g., the current price q_t is a function of previous price, i.e; $q_t = g(q_{t-1})$.

4 The Implemented Algorithms

In the following, we provide an overview of the algorithms selected for our experimental study. We briefly describe the algorithms and their competitive ratios. For proof of the competitive ratio the reader is referred to the respective paper.

4.1 Unidirectional Non-preemptive

In *unidirectional non-preemptive* solution (also called as *Reservation Price* algorithms), the player converts only once in an investment horizon. The player computes a reservation price q^* and compares each offered price q_t with q^*. The player accepts the first offered price q_t which is at least q^* and converts whole of D into Y in one transaction. We consider two such algorithms for our experimental study namely *RPMm* [6] and *RPMT* [5].

Algorithm 1. *(RPMm)*
Accept the first price greater than or equal to $q^* = \sqrt{M \cdot m}$.

Theorem 1. *Algorithm 1 is $\sqrt{M/m}$ competitive.*

Algorithm 2. *(RPMT)*
Accept the first price greater than or equal to $q^* = M/\sqrt{T}$.

Theorem 2. *Algorithm 2 is $\sqrt{T}$ competitive.*

4.2 Unidirectional Preemptive

In *unidirectional preemptive* solution, the player does not convert only once in the investment horizon but depending on the priced offered q_t (or time t) converts a portion s_t of D into Y. For our experiments, we consider the algorithms proposed by El-Yaniv et al. [6], Hu et al. [7], Chen et al. [4] and Lorenz et al. [9]. We briefly describe each algorithm and the competitive ratio as follows;
El-Yaniv et al [6] Threat based Algorithm: El-Yaniv et al. [6] proposed a threat based algorithm based on the assumption that there exists a threat that on day $k \leq T$, the adversary may drop the offered price to minimum level m and keep it there for the remaining period of the investment horizon.

Algorithm 3. *The basic rules of the threat-based algorithm are:*

1. *Consider a conversion from asset D to asset Y only if the price offered is the highest seen so far.*
2. *Whenever convert asset D to asset Y, convert just enough D to ensure that a competitive ratio c would be obtained if an adversary drops the price to the minimum possible price m, and keeps it there afterwards.*
3. *On the last trading day T, all remaining D must be converted to Y, possibly at price m.*

El-Yaniv et al [6] presented four variants of Algorithm 3, each assuming different a priori knowledge. We restrict our study to two variants of Algorithm 3.

i. Variant 1 (YFKTMm): With known M and m

Theorem 3. *Variant 1 of Algorithm 3 has a competitive ratio c as:*

$$c = \ln\left(\frac{\frac{M}{m} - 1}{c - 1}\right). \tag{2}$$

ii. Variant 2 (YFKTMmk) : With known M , m and k

Theorem 4. *Variant 2 of Algorithm 3 has a competitive ratio c of:*

$$c = k\left(1 - \left(\frac{m(c-1)}{M-m}\right)^{1/k}\right). \tag{3}$$

Hu et al [7] with known $g(q_t)$ and T: Hu et al. [7] presented two algorithms to achieve optimal competitive ratio under worst case assumptions, namely *Static Mixed Strategy* and *Dynamic Mixed Strategy*, where the player has the knowledge of length of investment horizon T and price function $g(q_t)$. Hu et al. [7] assumed that the current day price q_t satisfies $(1-\gamma)q_{t-1} \leq q_t \leq (1+\gamma)q_{t-1}$, where $\gamma \leq 1$

Static Mixed Strategy: The static mixed strategy allocates the amount to be converted based on the worst-case input sequence of prices.

Algorithm 4. *(HGLSMS): Amount converted on day t is determined by the following rules;*

$$s_t = \begin{cases} \left(\frac{1+\gamma}{(T-1)\gamma+2}\right) & t = 1, \\ \left(\frac{\gamma}{(T-1)\gamma+2}\right) & t \in [2, T-1], \\ \left(\frac{1}{(T-1)\gamma+2}\right) & t = T. \end{cases} \tag{4}$$

Theorem 5. *The competitive ratio c achieved by Algorithm 4 is*

$$c = 1 + \frac{\gamma}{2}(T-1). \tag{5}$$

Dynamic Mixed Strategy: The worst-case scenario does not occur that frequently as assumed by the static mixed strategy. The dynamic mixed strategy allocates s_t based on the remaining number of days T' in the time interval.

Algorithm 5. *(HGLDMS): Amount converted on day t is determined by the following rules;*

$$s_t = \begin{cases} \left(\frac{1+\gamma}{(T'-1)\gamma+2}\right) W'_t & t = 1, \\ \left(\frac{\gamma}{(T'-1)\gamma+2}\right) W'_t & t \in [2, T-1], \\ \left(\frac{1}{(T'-1)\gamma+2}\right) W'_t & t = T. \end{cases} \tag{6}$$

where W'_t denotes the remaining amount of wealth at day t and $T' = T - t + 1$.

Theorem 6. *The competitive ratio c achieved by Algorithm 5 based on the remaining number of days T' is*

$$c = 1 + \frac{(T'-1)\gamma}{2}. \tag{7}$$

Chen et al [4] with known $g(q_t)$ and T: Chen et al. [4] assume prior knowledge of the duration T, and the price function $g(q_t)$. The constants α and β $(\alpha, \beta \geq 1)$ determine the prices offered on a day t, and q_t satisfies $q_{t-1}/\beta \leq q_t \leq \alpha \cdot q_{t-1}$. The algorithm and the amount invested s_t on day t is described as follows:

Algorithm 6. *(CKLW): Determine the amount to be converted at time t by the following rules*

$$s_t = \begin{cases} \frac{\alpha(\beta-1)}{T\alpha\beta-(T-1)(\alpha+\beta)+(T-2)} & t = 1, \\ \frac{(\alpha-1)(\beta-1)}{T\alpha\beta-(T-1)(\alpha+\beta)+(T-2)} & t \in [2, T-1], \\ \frac{(\alpha-1)\beta}{T\alpha\beta-(T-1)(\alpha+\beta)+(T-2)} & t = T. \end{cases} \tag{8}$$

Theorem 7. *The competitive ratio c achieved by Algorithm 6 is*

$$c = \frac{T\alpha\beta - (T-1)(\alpha+\beta) + (T-2)}{\alpha\beta - 1}. \quad (9)$$

Lorenz et al [9] with known m and ϕ: Lorenz et al [9] proposed an algorithm with known m and ϕ. We discuss the strategy for max search (selling).

Algorithm 7. *(LPS):* ***Max-search Problem:*** *At the start of the game compute reservation prices $q_i^* = (q_1^*, q_2^*, ...q_u^*)$, where $i = 1, .., u$. As the adversary unfolds the prices, the algorithm accepts the first price which is at least q_1^*. The player then waits for the next price which is at least q_2^*, and so on. If there are still some units of asset left on day T, then all remaining units must be sold at the last offered price, which may be at the lowest price m.*

$$q_i^* = m\left[1 + (c-1)\left(1 + \frac{c}{u}\right)^{i-1}\right]. \quad (10)$$

Where c is the competitive ratio for the max-search problem.

Theorem 8. *Let $u \in N$, $\phi > 1$, there exists a c-competitive deterministic algorithm for u max-search problem where $c = c(u, \phi)$ is the unique solution of*

$$\frac{(\phi - 1)}{(c-1)} = \left(1 + \frac{c}{u}\right)^u. \quad (11)$$

5 Experiments

We consider the set of algorithms as described in Section 4 and execute them on two different types of dataset, real world data and synthetic data (bootstrap data). We evaluate performance and the consistency of performance. The performance of algorithms is measured in terms of competitive ratio and variance of competitive ratio is used as consistency measure. We also record the number of transactions performed by each algorithm. In the following we describe the dataset, experimental settings and results.

5.1 Dataset

We consider the following two types of datasets for our experiments.

Real World Data: Two datasets *DAX30* (1.1.2001 to 31.12.2010) and *S&P500* (1.1.2001 to 31.12.2010) are considered.

Synthetic Data: We employed bootstrap method to generate additional datasets. Bootstrap is useful technique to produce additional data where original data sample size is small [14]. Using the moving block bootstrap, we generated 15 additional samples for each year for DAX30 and S&P500 (2001-2010) datasets. So, for each dataset, we generate 150 synthetic time series.

5.2 Experimental Settings

Each algorithm is executed on yearly data of *DAX30* and *S&P500*. Competitive ratio for all algorithms on yearly data is calculated. Variance is used as consistency measure. Further, we recorded the number of transactions by each algorithm, as in real world each transaction has an associated cost, it will be helpful to identify if an algorithm performing well on the basis of competitive ratio also has fewer number of transaction or vice versa.

Assumptions. For the sake of simplicity, we make the following assumptions.

i. Each transaction has associated cost of 0.025% of volume transacted.
ii. The yearly interest rate is zero.
iii. The prices considered are all closing day prices.
iv. Any amount of wealth left on last trading day is converted at the last day offered price. This is inline with the rules of threat based algorithms [6] and reservation price algorithms [5,6].
v. For algorithm $YFKTMmk$, we consider $k = T$, similarly for LPS, we assume $u = T$.
vi. For all algorithms, required a priori parameters such as m, M and/or ϕ etc are derived from the time series before execution of algorithm begins. This is inline with the working of algorithms as every algorithm assumes the exact a priori information about the future.

5.3 Results

Real World Data
DAX30 (2001-2010)
Table 1 summarizes the results for the *DAX30* and *S&P500* datasets for the years 2001 to 2010. The column *"Ave CR"* represents the average of the competitive ratio calculated over the yearly data for *DAX30* (2001-2010) and the column *"Var"* shows the variance of the competitive ratio. An average competitive ratio closer to 1 reflects the better performance of algorithms while a low variance shows the consistency of the algorithm.

In our experiments, we observed that unidirectional preemptive algorithm $YFKTMm$ suggested by El-Yaniv et al [6] performs the best among all set of algorithms considered with an average competitive ratio of 1.0873. The worst performance is observed for $HGLSMS$ [7] with an average competitive ratio of 1.1771. The most consistent algorithm is $YFKTMm$ which has a variance of $2.91 * 10^{-3}$, whereas the most inconsistent performance behavior is observed for LPS which has a variance of $12.79 * 10^{-3}$. The next closest (worst) algorithm in terms of consistency is reservation price algorithm $RPMT$ by Damaschke et al [5] with variance of $10.13 * 10^{-3}$.

S&P500 (2001-2010)
The unidirectional preemptive algorithm $YFKTMm$ of El-Yaniv et al. [6] with known M and m performance is found the best with an average competitive ratio of 1.0606 and is the most consistently performing algorithm as well with

Table 1. Avg CR ,Variance and number of transactions on real world data

Algorithm	*DAX30*			*S&P500*		
	AveCR	$Var(10^{-3})$	*#Tx*	*AveCR*	$Var(10^{-3})$	*#Tx*
RPMm	1.1178	6.32	1	1.0844	4.28	1
RPMT	1.1848	10.13	1	1.1297	5.59	1
YFKT Mm	1.0873	2.91	24.2	1.0606	2.14	23
YFKT Mmk	1.1651	3.46	24.6	1.1256	2.14	23.6
HGL SMS	1.1771	6.34	254.2	1.1192	2.86	251.5
HGL DMS	1.1729	7.53	254.2	1.1162	3.23	251.5
CKLW	1.1766	6.47	254.2	1.1192	2.87	251.5
LPS	1.1532	12.79	158	1.099	4.18	150.5

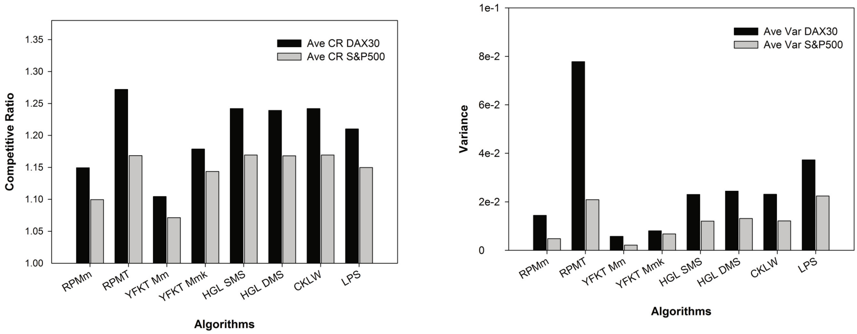
variance of $2.14 * 10^{-3}$. The worst competitive ratio is observed for $CKLW$ [4] and $HGLSMS$ [7] with an average competitive ratio of 1.1192. The most inconsistent algorithm is reservation price policy $RPMT$ of Damaschke et al. [5] with variance of $5.59 * 10^{-3}$. Table 1 summarizes the results.

Synthetic Datasets: Table 2 summarizes the results on bootstrap data. Although the individual performance on algorithm varies on different datasets, (for instance, on bootstrap DAX30 dataset, the average competitive ratio of $RPMm$ is 1.14 whereas on bootstrap S&P500 the average competitive ratio is 1.09) there is little change in overall performance order. For example, $YFKTMm$ is the best performing algorithm on both datasets. Similarly, the algorithms' behavior remains the same in terms of performance consistency as well, as depicted by Fig. 1(b). The performance consistency of $YFKTMm$ is found the best among all algorithms on both synthetic datasets, whereas the performance of $RPMT$ is found the most inconsistent. Fig. 1 depicts the performance and consistency pattern of algorithms on bootstrap data.

Number of Transactions: As each transaction has an associated cost, thus an algorithm with large number of transactions may not be a viable option. We discuss the number of transactions for each algorithm on both DAX30 and

Table 2. Avg CR, Variance and number of transactions on bootstrap data

Algorithm	*DAX30*			*S&P500*		
	AveCR	$Var(10^{-3})$	*#Tx*	*AveCR*	$Var(10^{-3})$	*#Tx*
RPMm	1.1495	14.4	1	1.099	4.8	1
RPMT	1.3045	81.8	1	1.1982	22.1	1
YFKT Mm	1.1044	5.7	21.25	1.071	2.0	18.48
YFKT Mmk	1.1784	8.05	21.6	1.1436	6.77	18.94
HGL SMS	1.2420	23.0	254.1	1.169	12	251.5
HGL DMS	1.2391	24.4	254.1	1.167	13.1	251.5
CKLW	1.2420	23.1	254.1	1.17	12.1	251.5
LPS	1.2099	37.3	135.96	1.149	22.4	130.67



(a) Performance - bootstrap data (b) Consistency - bootstrap data

Fig. 1. Performance and consistency on bootstrap data

S&P500. Table 1 indicates that the non-preemptive algorithms carry only a single transaction in each investment horizon. This holds true for all unidirectional non-preemptive strategies as they convert at a single point of time in the investment horizon, but not for unidirectional preemptive solutions, where the conversion amount is calculated based on the price offered and time in the investment horizon. Table 1 reflects that for DAX30, the algorithm $YFKTMm$ by El-Yaniv et al. [6] has the least number of transactions in all unidirectional preemptive solutions, as the algorithm only invests when the price offered is the highest seen so far, thus the algorithm does not convert at all offered prices but does so on local maxima. HGL [7] and CKLW [4] have the highest number of transactions, which is the same as the number of days in the investment horizon, as they invest on each day of the investment horizon. The same pattern is found when transactions on bootstrap data are considered. Table 2 also summarizes the resultant number of transactions on bootstrap data.

6 Discussion

From the outcome of the experiments and based on the criterion of competitive ratio, we observe that unidirectional preemptive algorithm $YFKTMm$ [6] performs better than other algorithms. On DAX30 and S&P500 datasets, $YFKTMm$ performs 6% and 4% better than the average performance of the remaining algorithms, whereas on bootstrap data the corresponding numbers are 9% and 7% respectively. Similarly, the $YFKTMm$ also proves to be more consistent in terms of variance in the competitive ratio. On DAX30 and S&P500 datasets, the variance of $YFKTMm$ is on average 62% and 67% less than the average variance of all other considered algorithms. $YFKTMm$ remains the most consistent algorithm on bootstrap data as well. There is no clear worst performing algorithm as a number of algorithms perform poorly on different datasets. For instance on DAX30 dataset, $HGLSMS$, $HGLDMS$ and $CKLW$ performance are among the worst three, they have an average competitive ratio of approximately 1.17, the same holds for S&P500. On DAX30 and S&P500 bootstrap data, the worst performing algorithm is $RPMT$ with an average competitive ratio of 1.3 and 1.19 respectively.

The reason for the better performance of $YFKTMm$ [6] is that the algorithm converts only when it finds a new maximum, this not only results in better performance but also reduces the number of transactions. Although LPS [9] also converts only when it encounters a new maximum, it is not as competitive as $YFKTMm$, this can be attributed to the amount of wealth converted. $YFKTMm$ considers the offered price q_t when calculating s_t but LPS does not consider the offered price. Another significant result, we observed, is the performance of non-preemptive algorithm of El-Yaniv et al. [6]. On dataset S&P500, the average competitive ratio of non-preemptive algorithm of El-Yaniv et al [6] is 1.0844 which is second only to $YFKTMm$, the same results holds for DAX30 dataset and for bootstrap data (DAX30, S&P500). Another aspect of the study is that intuitively, the more information available to (use by) an algorithm, the better it must perform, but this however may not happen. On both datasets DAX30 and S&P500, the preemptive algorithm $YFKTMm$ performs better than $YFKTMmk$, this can be attributed to the 'luckily behaving data' which results in better performance of $YFKTMm$.

An important consideration of any experimental study is to observe the gap between theory and practice. For all algorithms and for each yearly dataset, we calculate the worst case competitive ratio (c_{wc}) that an algorithm can achieve with the given setting and after the algorithm is executed on yearly data, we record the experimental competitive ratio (c_{ec}). For instance, consider yearly data of DAX30 for 2001 and algorithm $YFKTMm$, before the algorithm begins execution, we calculate the c_{wc} using Theorem 3, similarly when the algorithm is executed on data, we record the c_{ec} achieved by algorithm. The process is repeated for all algorithms and for all real world data of DAX30 and S&P500 datasets. We limit it only to real world data and do not include the bootstrap data as we are investigating the gap between theory and practice, hence synthetic (bootstrap) data is not considered. It can be seen from Table 3 that the algorithms suggested by El-Yaniv et al. [6] ($RPMm, YFKTMm, YFKTMmk$) have the least gap between c_{wc} and c_{ec} whereas other algorithms have considerable gap between c_{wc} and c_{ec}. For instance, $RPMT$ [5] on DAX30 dataset, has (average) c_{wc} of 387.19 whereas the c_{ec} is 1.155, this is because of the reservation price calculation of $RPMT$ (Theorem 2) which only considers M and T and not the relative fluctuation in the prices. For HGL, $CKLW$ and LPS, the gap is not as wide as of $RPMT$ but is considerably more than that of El-Yaniv et al. [6]. The gap between c_{wc} and c_{ec} of $HGLSMS$, $HGLDMS$ and $CKLW$ is based on the fact that length of investment horizon T has significant contribution in determining the worst case competitive ratio (see Theorem 5 and 7). For LPS the gap is the result of our choice of parameter u, as we consider $u = T$, thus it results in higher c_{wc}. It is interesting to see that the performance of $HGLSMS$, $HGLDMS$ and $CKLW$ does not differ a great deal, it is because of the fact

Table 3. Gap between theory and practice

Algorithm	$DAX30$		$S\&P500$	
	c_{wc}	c_{ec}	c_{wc}	c_{ec}
RPMm	1.2403	1.1178	1.1660	1.0844
RPMT	387.1989	1.1848	81.8012	1.1297
YFKTMm	1.188	1.0873	1.1408	1.0606
YFKTMmk	1.1812	1.1651	1.1382	1.1256
HGLSMS	8.3821	1.1771	6.8383	1.1192
HGLDMS	8.3821	1.1729	6.8383	1.1162
CKLW	8.3821	1.1766	6.8383	1.1192
LPS	7.6397	1.1532	6.5921	1.099

that these algorithms considers only price function ($g(q_t)$) and the length of investment horizon T. In addition, the price function considered by Hu et al. [7] is identical to that of Chen et al. [4] (only the mathematical formulation differs), thus resulting in similar performance behavior. Considering the performance of LPS, it is important to mention that for $u = 1$, the algorithm is similar to unidirectional preemptive algorithm $RPMm$ but as we consider $u = T$ thus the performance varies.

A major drawback in preemptive algorithms is the large number of transactions. On yearly real world data, with approximately 250 trading days, the least number of transactions performed by pre-emptive algorithms is 24 by $YFKTMm$. Although the number of transaction of $YFKTMm$ are less than other pre-emptive algorithms like $HGLSMS$, $HGLDMS$ and $CKLW$, which trades on every day, it still is significantly higher number when the impact of transaction cost on performance is considered. However, the ideal number of transactions per year is hard to envisage and depends on the amount of wealth available to the player.

7 Future Work and Conclusion

We presented an extensive experimental study to evaluate the applicability of online conversion algorithms in real world scenario such as trading in financial markets. We observed that although, a good number of algorithms are proposed to deal with unidirectional conversion problem, there are still a considerable number of open questions. An important factor for the designing new conversion algorithms must be to reduce the number of transactions, as in real world each transaction has an associated cost, thus reducing the number of transactions can be useful. However, the optimum number of transactions cannot explicitly be defined. Another open question will be to develop algorithms that provide risk management for the investors, as in real world the investors want to manage risk but online algorithms are designed based on risk mitigation paradigm. Albinali [1] proposed a risk-reward framework, which can be used to incorporate risk management in online conversion problems.

References

1. Al-Binali, S.: A risk-reward framework for the competitive analysis of financial games. Algorithmica 25, 99–115 (1999)
2. Brock, W.A., Lakonishok, J., LeBaron, B.: Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance 47, 1731–1764 (1992)
3. Chang, E.J., Lima, E.J.A., Tabak, B.M.: Testing for predictability in emerging equity markets. Emerging Markets Review 5(3), 295–316 (2004)
4. Chen, G.-H., Kao, M.-Y., Lyuu, Y.-D., Wong, H.-K.: Optimal buy-and-hold strategies for financial markets with bounded daily returns. SIAM Journal on Computing 31(2), 447–459 (2001)
5. Damaschke, P., Ha, P.H., Tsigas, P.: Online search with time-varying price bounds. Algorithmica 55(4), 619–642 (2009)
6. El-Yaniv, R., Fiat, A., Karp, R.M., Turpin, G.: Optimal search and one-way trading algorithm. Algorithmica 30, 101–139 (2001)
7. Hu, S., Guo, Q., Li, H.: Competitive Analysis of On-line Securities Investment. In: Megiddo, N., Xu, Y., Zhu, B. (eds.) AAIM 2005. LNCS, vol. 3521, pp. 224–232. Springer, Heidelberg (2005)
8. Kwon, K.-Y., Kish, R.J.: A comparative study of technical trading strategies and return predictability: An extension of brock, lakonishok, and lebaron (1992) using NYSE and NASDAQ indices. Quarterly Review of Economics and Finance 42(3), 613–633 (2002)
9. Lorenz, J., Panagiotou, K., Steger, A.: Optimal algorithms for k-search with application in option pricing. Algorithmica 55(2), 311–328 (2009)
10. Mohr, E., Schmidt, G.: Empirical Analysis of an Online Algorithm for Multiple Trading Problems. In: Le Thi, H.A., Bouvry, P., Pham Dinh, T. (eds.) MCO 2008. CCIS, vol. 14, pp. 299–308. Springer, Heidelberg (2008)
11. Ratner, M., Leal, R.P.C.: Tests of technical trading strategies in the emerging equity markets of latin america and asia. Journal of Banking and Finance 23, 1887–1905 (1999)
12. Schmidt, G., Mohr, E., Kersch, M.: Experimental analysis of an online trading algorithm. Electronic Notes in Discrete Mathematics 36, 519–526 (2010)
13. Shen, P.: Market-timing strategies that worked. Jounal of Portfolio Management 29(2), 57–68 (2003)
14. Tabak, B.M., Lima, E.J.A.: Market efficiency of brazilian exchange rate: Evidence from variance ratio statistics and technical trading rules. European Journal of Operational Research 194, 814–820 (2009)

Multichannel Sales Services for Enterprise Cloud Vendors

Jih-Shyr Yih

IBM T. J. Watson Research Center
Hawthorne, New York 10532, USA
jyih@us.ibm.com

Abstract. For fast delivery and pay-as-you-go cost, corporations have begun to run application development and deployment projects on servers from cloud vendors. Value-add resellers (VARs) have even built dynamic, higher-level computing services using public cloud offerings as infrastructure. Amazon, as a cloud offerings pioneer deeply rooted in consumer web commerce, has led public cloud vendors in serving IT needs in a B2C direct fashion. This paper describes the overlooked importance of understanding enterprise buying practices and multichannel enterprise sales mechanisms. It is shown how existing sales services can be reused and modified to sell cloud computing resources in B2B direct and indirect ways. More important, the role of dealers/distributors, who manage their own customer relationships, is emphasized. To accommodate the various procurement systems being used by the different distributors, B2B gateway architecture for standard protocols and transactions is illustrated. This paper also suggests business models on how distributors can collaborate with a vendor in cloud resource planning by applying optimized inventory control.

Keywords: Multichannel Commerce, Sales Services, Cloud Computing, B2B Gateway, Buy Analysis, Inventory Control, Cloud Capacity Planning.

1 Introduction

Electronic commerce has recently seen a new height in research, development, mergers, and acquisitions as commerce 1) adds social media and mobile apps into multimodal interactions and 2) integrates customer access across multiple channels. Many leading commerce specialists have been bought by major software companies, e.g., IBM acquired Sterling Commerce in May '10, Coremetrics in Jun. '10, and Unica in Aug. '10, and DemandTec in Dec. '11, and Oracle acquired Art Technology Group in Nov. '10 and Endeca in Oct. '11, and many more by Google.

However, a survey of public infrastructure cloud vendors shows that the commerce consumption aspect is often overlooked in comparison to the IT delivery aspect such as server provisioning and backend system management automation. Cloud computing customers are often shown a catalog as a flat list of offerings, each in cryptic one-line description. Customers are to buy one image on a virtual machine at a time, and must follow through the fulfillment process to complete the buying. These

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 188–194, 2012.

deficiencies may be fixable with the use of a better commerce engine. But, the B2C-oriented mindset of major vendors is ignoring the key characteristics and requirements of selling information technology to enterprise customers. As a result, major penetration into enterprise computing environment has yet to be seen by any cloud vendor. Cloud vendors need to learn from the established enterprise buying practices and make necessary changes in commerce support for a better business model.

This paper describes an enterprise sales services transformation, from selling traditional hardware and software offerings to Infrastructure as a Service (IaaS) cloud computing. Section 2 uses industry sales figures as references, to illustrate the direct and indirect sales potentials. Specific examples of established sales channels of how enterprise customers buy IT products and services are also given. In Section 3, challenges in selling infrastructure cloud offerings are discussed, as vendors are often tempted to build another one-of-a-kind, isolated customer portal with commerce capability tailor-made for the cloud. We will show how existing commerce systems and service desks can be integrated to support cloud sales, so that customers can shop for traditional discrete technology pieces side by side with the integrated cloud offerings. We will also discuss B2B gateway architecture for accommodating various dealer systems. In Section 4, the cloud capacity reservation as a supply chain inventory problem is described. A cloud vendor and its distributors can collaborate to reduce cost for cloud end users. A summary with future directions is given in Section 5.

2 Enterprise Buying: Relationship, Process and Tools

To promote the use of shared cloud infrastructure, let's examine how customers are reached by the traditional server hardware vendors. The IBM x86 server hardware is used here as a reference.

2.1 Types of Relationship Buying

IBM has built 31 commerce enabled country portals (www.ibm.com) around the globe for consumers. Separately, there are over 650 private enterprise relationship portals with commerce capability (e-Sites) dedicated to IBM's largest enterprise clients [1, 2]. There are contracts to offer entitled offerings with entitled prices for the employees in such a large enterprise company. An e-Site private store is also integrated with the client environment, so that an employee can use corporate accounting and payment methods. Purchase orders at an e-Site can be approved by the line managers and the finance department. These customers basically know what they want to buy and appreciate the value for the money. The corresponding buying pattern is represented by the lower left section in Fig. 1.

The public and private web stores also allow customers to interact with telesales representatives (reps) over different communication media. Some of these reps are for coverage purposes, as they respond to in-coming requests. Contrarily, and much more important to IBM are the *inside sales* representatives, who work from within IBM and reach out to corporate buyers to identify/validate opportunities, and serve as the single point of contact to coordinate all the technical and financial supporters from the sell side for the customer needs [3]. The inside sales target customer buying patterns depicted in the middle section in Fig. 1. Besides to be knowledgeable in products and corporate buyers, sales rep typically masters several dozens of tools and systems for the opportunity-to-order process, e.g., to learn client background including past deals, locate the proper expertise, compose proposals, and place order entries.

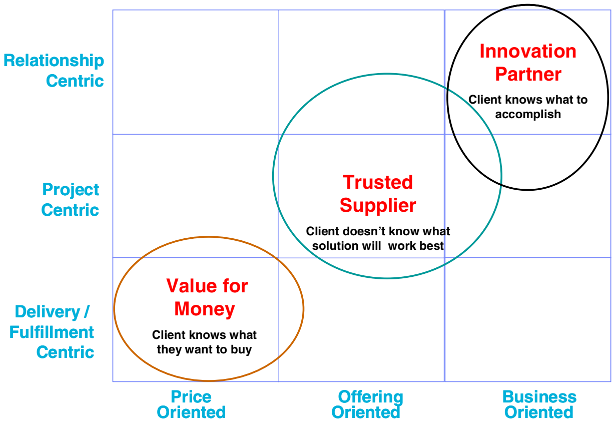


Fig. 1. Customer buying patterns

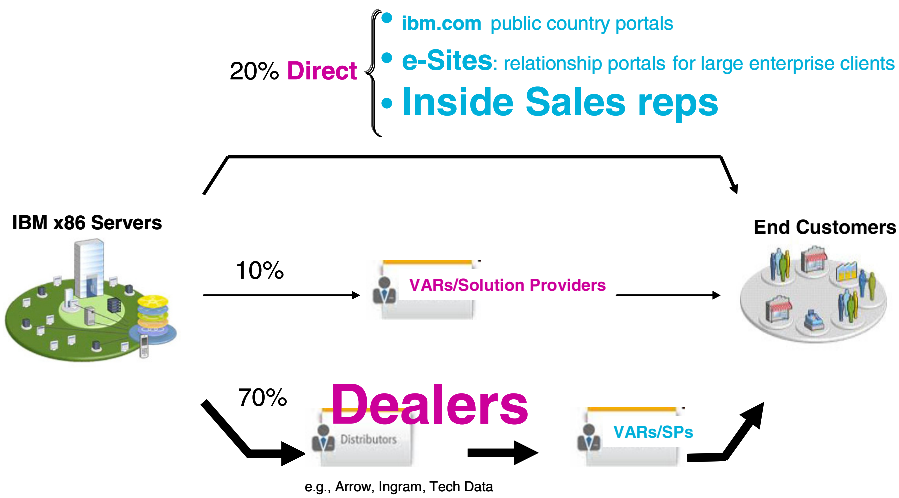


Fig. 2. IBM x86 server sales volume distribution

Surprisingly, with all the IBM direct sales forces, systems, and tools mentioned above, they only capture about 20% of the x86 servers sales volume. Fig. 2 shows a breakdown on the total sales, where 80% of the selling are done indirectly by going through value-add resellers (VARs) and solution providers, who have working relationship with the end customers. More interestingly, most (in 7 to 1 ratio) VARs and solution providers order x86 servers through dealers (distributors), who hold inventory for fast delivery and provide value-add services. The distributors often have their own web stores, procurement methods, sales support systems, and sales representatives. The distributors are considered efficient business partners as they are relieved from the cost of product research, development, marketing, and so on.

3 Multichannel Sales Services for Cloud Vendors

3.1 B2C and B2B Sales Channel Synergy

Fig. 3 shows a possible multichannel implementation for the IBM SmartCloud Enterprise [4] sales services. Inside the SmartCloud Enterprise, a B2C customer portal is implemented, where a customer can perform personal account management, catalog shopping, and all the special cloud asset management functions. We will describe how enterprise customers can collaborate with IBM across the different channels for the access to SmartCloud Enterprise in the following paragraphs.

For the e-Site or B2B direct customers, the laborious steps of first-time account set up and credit check can be waived, by propagating existing customer credentials and preferences from e-Sites. Customers can shop for cloud offerings on e-Sites and, equally important, get management approvals and payment arrangements. After placing orders, customers can be informed by email about the fulfillment status and the SmartCloud Enterprise customer portal access link, through which they could participate in the cloud asset management if needed.

For the B2B indirect customers, who work with a VAR or a solution provider and buy resource via a distributor, the purchase orders can be done using a procurement system available at the dealership, such as Ariba Solutions for Buyers, Oracle Buyer, or any proprietary system. These dealer procurement systems can be preloaded with offering catalogs or use standardized punch-out protocols to compose shop carts. The rest of the collaboration between an end customer and the vendor is similar to the description given for the e-Site customers.

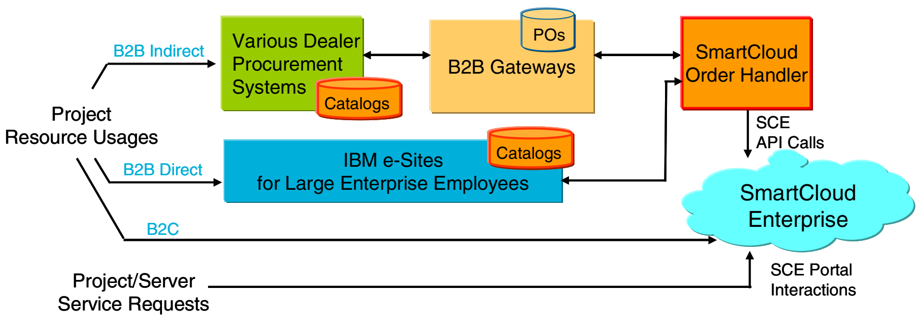


Fig. 3. Multichannel commerce for a cloud vendor

3.2 B2B Gateways and Protocols

These various dealer procurement systems are usually required to support standard protocols and transactions. Fig. 4 shows a reference implementation of a B2B gateway which supports a list of transmission protocols, such as EDI, RosettaNet, and the web. In addition to the interactions with trading partners, a B2B gateway needs to provide various support functions, such as trading partner registry, content/message repository, virus scan, and other security functions. A B2B gateway not only provides non-repudiation between the trading partners, and can actively filtering the transaction content for policy-based management enforcement.

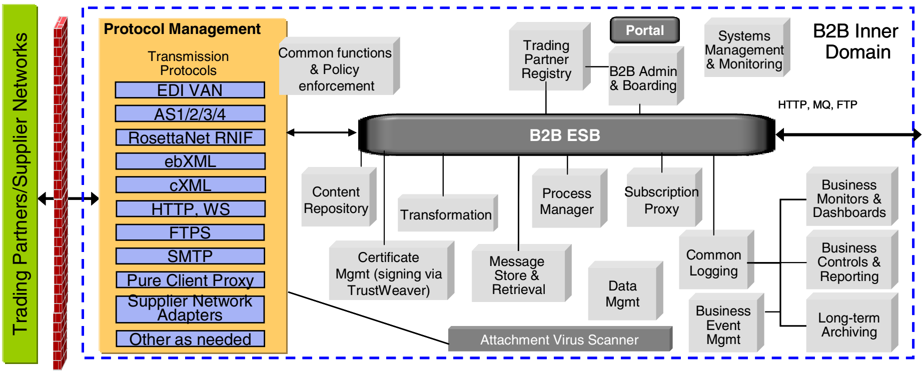


Fig. 4. B2B gateway and infrastructure architecture

4 Cloud Capacity Reservation Schemes

A traditional enterprise hosting provider often will not begin the specific capital request for servers before a firm work contract is secured. The follow up steps in capital approvals, hardware purchase, machine assembly, and connecting to the networks could easily drag on into months [5, 6]. For cloud computing infrastructure, advanced capacity planning ahead of actual usages is simply a must to provide on-demand services.

There can be many different models for capacity planning. For example, the Vendor Managed Inventory (VMI) has been well practiced for the traditional consumer goods and can be referenced by the cloud computing needs [7]. A cloud vendor could take advantage of the demand visibility (the customer project needs), along with financial constraints in the contracts and, perform optimization analysis in adjusting resource pool sizes. The benefits of VMI are well documented, and the case of Wal-Mart and its suppliers is a much publicized story. However, because of the inherent undertaking of the financial risk by owning the inventory, vendors tend to under estimate the inventory needs, compared with Distributor Managed Inventory approach.

On the other hand, the Amazon cloud is known for reusing hardware retired from amazon.com's large commerce operations to avoid advanced new capital commitment for the public cloud business. However, enterprise customers with business applications to run often work with physical deployment architects to determine exact computing requirements down to machine types. For enterprise use, dated or unknown hardware for computing power may be simply not acceptable.

One of the main advantages of selling enterprise cloud resources through distributors is the possibility of sharing the capital risk between the vendor and the distributors. For cloud resource B2B indirect sales, IBM has capacity contracts with its distributors to help determine the size of a resource pool. Under contract, there can be two levels of buying from a distributor to IBM, as depicted in Fig. 5. Distributors can place Level 1 block orders from time to time, and end customers acquire the cloud resources with Level 2 orders for project level consumption. There are incentives for a distributor to avoid end customers placing orders beyond capacity guaranteed by the advanced block orders. To help distributors determine suitable block orders, there are the Buy Analysis Tools, shown in Fig. 5, for this purpose [8, 9]. The IBM Buy Analysis Tools monitor the Level 2 orders in quantity and timing as well as the remaining resource pool capacity, and make recommendations to distributors. Notice that the Level 1 block orders can be made at the SmartCloud Enterprise portal or through the dealer procurement systems.

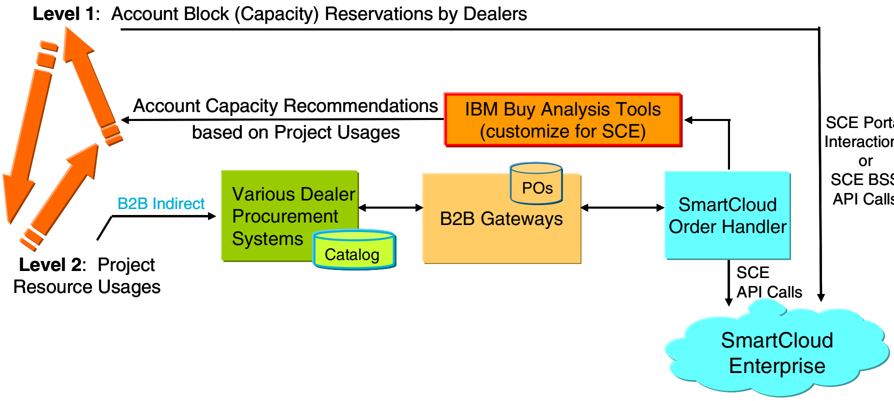


Fig. 5. Multichannel commerce with advanced capacity reservations

5 Concluding Remarks

Cloud computing is a new way of consuming and delivering IT. This paper describes the role of multichannel sales services in enabling the enterprise users to the consumption of cloud computing, using IBM SmartCloud Enterprise sales services as an example. The established enterprise procurement processes, systems, and collaborations with distributors still dictate how corporate users buy IT, as corporate users seldom use credit cards directly at the cloud customer portal for consumption.

As cloud offerings become more standardized, the sales services can rely more on the distributors. We have identified the issues in adding enterprise cloud computing offering as a new brand to the channels and in extending the customer relationships to the new cloud customer portal.

Cloud computing has shifted the burden of capital equipment allocations to the on-demand vendors. The cloud capacity planning can be assisted by advanced reservation block orders from channels. There are pros and cons in Distributor Managed Inventory and Vendor Managed Inventory models. We have adopted a hybrid model of asking 1) distributors to take risk in capital investment and 2) the vendor to provide incentives for advanced capacity block orders. More incentive models are currently in plan for future works.

References

1. Chieu, T., Pinel, F., Yih, J.: Unified Commerce Server Architecture for Large Number of Enterprise Stores. In: IEEE Conference on E-Commerce Technology, San Diego, CA (2004)
2. Chieu, T., Fu, S., Pinel, F., Yih, J.: Unified Solution for Procurement Integration and B2B Stores. In: Fifth International Conference on Electronic Commerce, Pittsburgh, PA (2003)
3. Chieu, T., Chowdhary, F.S., Pinel, F., Yih, J.: Enterprise Telesales Opportunity Pipelines Performance Management. In: IEEE International Conference on e-Business Engineering, Shanghai, China (2006)
4. Michael, H., Kochut, A., Schulz, C., Shaikh, H.: Virtual Hypervisor: Enabling Fair and Economical Resource Partitioning in Cloud Environments. In: IEEE/IFIP Network Operations and Management Symposium (2010)
5. Yih, J.-S., Liu, Y.-H.: Data Center Hosting Services Governance Portal and Google Map-Based Collaborations. In: Vossen, G., Long, D.D.E., Yu, J.X. (eds.) WISE 2009. LNCS, vol. 5802, pp. 455–462. Springer, Heidelberg (2009)
6. Liu, Y., Wu, F., Yih, J.: Enterprise Data Center Governance Using Web 2.0 Portal and Services Integration. In: IEEE Service-Oriented Computing and Applications, Taipei, Taiwan (2009)
7. Zipkin, P.: Foundations of Inventory Management. McGraw-Hill (2000)
8. Ettl, M., Kapuscinski, R.: Modeling Price Protection Contracts in IBM's Extended Server Supply Chain. In: M&SOM SCM SIG Conference, Boston, MA (2009)
9. Ervolina, Ettl, M., Lee, Peters: Managing Product Availability in an Assemble-To-Order Supply Chain with Multiple Customer Segments. OR Spectrum (2008)
10. Fu, S., Chen, S., Yih, J., Pinel, F., Chieu, T.: Web-Based Sell-Side Commerce Aggregation. In: Third International Conference on Internet Computing, Las Vegas, NV (2002)

Applying Contextual Advertising to MultiModal Information Content

Giuliano Armano[1], Alessandro Giuliani[1], Alberto Messina[2], Maurizio Montagnuolo[2], and Eloisa Vargiu[1,3]

[1] University of Cagliari, Dept. of Electrical and Electronic Engineering, Piazza d'Armi, 09123 Cagliari, Italy
{armano,alessandro.giuliani}@diee.unica.it
[2] RAI Centre for Research and Technological Innovation, C.so Giambone, 68, 10135 Torino, Italy
{a.messina,maurizio.montagnuolo}@rai.it
[3] Barcelona Digital, C. Roc Baronat, 117, 08018 Barcelona, Spain
evargiu@bdigital.org

Abstract. Contextual Advertising, a major sources of income for a large number of websites, is aimed at suggesting products and services to the ever growing population of Internet users. In this paper, we focus on the problem of suggesting suitable advertisements to news aggregation from television and from the Internet. To our best knowledge, this is the first attempt to perform this task in the field of multimodal aggregation. The proposed system suggests from 1 to 5 advertisements related to the main topic of aggregated news items. 15 users were asked to evaluate the relevance of the suggested advertisements. Preliminary results are encouraging for further development and application of contextual advertising in the field of multimodal aggregation.

1 Introduction

Modern technology for multimedia content production allows decisive strategic developments in media, information and communication services, with the capability of representing information into various forms and different devices. As a consequence, common ad serving technologies treat TV watchers and Internet surfers as passive recipients of ads. Moreover, these messages are often uncorrelated to the consumed content, thus making ad serving extremely inaccurate and unprofitable. A new challenge for media industry consists of using the intrinsic semantics of the produced contents to plan computer-assisted advertising strategies. This opens the way to new business opportunities, such as which automatic, content-based Web and TV advertising.

Web advertising is a major sources of income for a large number of websites. Its main goal is to suggest products and services to the ever growing population of Internet users. There are two primary channels for distributing ads: Sponsored Search (or Paid Search Advertising) and Contextual Advertising (or Content Match). Sponsored Search displays ads on the page returned from a

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 195–202, 2012.

Web search engine following a query [1], whereas Contextual Advertising (CA) displays ads within the content of a generic, third party, webpage. The state-of-the-art approaches on CA adopt syntactic and semantic techniques [2,3]. Syntax is usually exploited by adopting a suitable text summarization technique aimed at reducing the size of data while preserving the original meaning. Semantics is exploited by using a centroid-based classifier devoted to capture the main topics concerning a given news aggregation.

This paper presents a preliminary study for contextually associating advertisements (ads) with MultiModal Aggregations (MMAs) of news items. Multimodal Aggregation is the process of merging content from different data sources (e.g., Web portals, IPTV, etc.) to produce new hybrid data that was not originally provided. The main contribution of this work is the development of a suitable CA system for MMAs. To our best knowledge, this is the first proposal aimed at automatically suggesting ads in a multi-modal mass-production setting. News come from two inputs: (i) automatically extracted, chaptered, and transcribed TV news stories, and (ii) RSS feed items from online newspapers and press agencies.

The rest of the paper is organized as follows: Section 2 illustrates the implemented system; Section 3 presents experiments and results; and Section 4 concludes the paper with a sketch of future work.

2 The Proposed System

As we are interested in suggesting ads that match with the content of a given news aggregation, we adopted a solution compliant with state-of-the-art CA approaches, such as those proposed in [2,3,4].

2.1 Scenario

Let us consider a family composed by Bob, Alice, and their son Edinson. Bob is an investment broker. Due to his job, he is interested in both economy/finance news and transportation services. Alice is a housewife who cares with the health of her child. Edinson is a sport lover and his favorite sport is football. Each family member is encountering severe problems for fulfilling personal information needs. From the Internet point of view, the availability on a daily basis of a wide variety of information sources generates a disproportionately high amount of content (e.g., newspaper articles and news agency releases) that makes it impossible for each of them to read everything that is produced. Furthermore, it is obvious that the Internet is not (yet) the only source of news information, with traditional media based on television channels still far from being left out in the near future. Due to the heterogeneity of individual interests, classical newscast programmes and TV advertising can be extremely inefficient, making viewers annoyed and upset. To fullfill the users' needs, a wish of a media industry would be to have an application able to aggregate data produced from different sources, to give a short description of them in form of keywords, and to associate them with advertising messages according to the content of the generated aggregations. In this way, the map is complete and each user can be informed about completeness and efficacy. For example, Bob might stay tuned on last stock market news through his tablet while being advised on journeys and transports. Alice might watch healthcare news stories on her television while being recommended on the healing properties of herbs and natural remedies. Finally, Edinson might browse his favorite team articles while getting hints on new sport furniture.

2.2 The Approach

News aggregations are automatically generated from online newspaper articles and TV newscasts by the RAI Hyper Media News aggregator (as detailed in [5]). Each news aggregation, also called *subject*, integrates items coming from both contributions. From the set of attributes that describes each subject we take into account: (i) *info*, the general information, title and description included; (ii) *categories*, the set of most relevant categories to which the news aggregation belong; and (iii) *items*, the set of Web articles that compose the aggregation. The categories are automatically assigned by AI:Categorizer[1], a system trained with radio programme transcriptions, according to a set of journalistic categories.

Figure 1 illustrates the main components of the proposed system:

- *News Extractor*. It is aimed at extracting all the news that compose it. In order to transform the news content into an easy-to-process document, any given news is also parsed to remove stop-words, tokenize it and stem each term;
- *Text Summarizer*. For each news, it calculates a vector representation of its summary as Bag of Words (BoW), each word being weighted by TF-IDF [6];
- *BoW Aggregator*. Takes as input the list of $\langle news, BoW \rangle$ pairs, and it is devoted to calculate the BoW of the whole news aggregation. The aggregated BoW vector is obtained by considering the occurrences in the whole set of news, weighted by TF-IDF. As CA systems work with a sole webpage this module is absent in classical CA systems. Its goal is to allow us to work with aggregated data.
- *Classifier*. To infer the topics of each news, it analyzes them according to a given set of classes based on a taxonomy of journalistic categories. First, each class is represented by its centroid, calculated starting from the training set. Then, each document is classified by the Rocchio classifier [7] with only positive examples and no relevance feedback. Each centroid component is defined as a sum of TF-IDF values of each term, normalized by the number of documents in the class. The classification is based on the cosine of the angle between the news and the centroid of each class, normalized by the news and class lengths to produce a comparable score. The output of this module is a list of $\langle news, CF \rangle$ pairs. In accordance with the work of Broder

[1] http://search.cpan.org/ kwilliams/AI-Categorizer-0.09/lib/AI/Categorizer.pm

et al. [2], CF are the Classification Features extracted by the classifier. The CF are weighted and represented as a vector.
- *CF Aggregator*. It is devoted to calculate the CF of the whole news aggregation. The aggregated CF is obtained considering the scores given by the classifier. It is worth noting that this module, absent in classical CA systems, allows us to work with aggregated data (similarly to the *BoW Aggregator*).
- *Matcher*. Each ad, which in our work is represented by the webpage of a product or service company, is processed in a similar way and it is represented by suitable BoW and CF. To choose the ads relevant to a news aggregation, the module assigns a score s to each ad according to its similarity with the given news aggregation:

$$s(n, a) = \alpha \cdot sim_{BoW}(n, a) + (1 - \alpha) \cdot sim_{CF}(n, a) \quad (1)$$

in which α is a global parameter that permits to control the impact of BoW with respect to CF, whereas $sim_{BoW}(n, a)$ and $sim_{CF}(n, a)$ are cosine similarity scores between the news (n) and the ad (a) using BoW and CF, respectively. For $\alpha = 0$ only semantic analysis is considered, whereas for $\alpha = 1$ only syntactic analysis is considered.

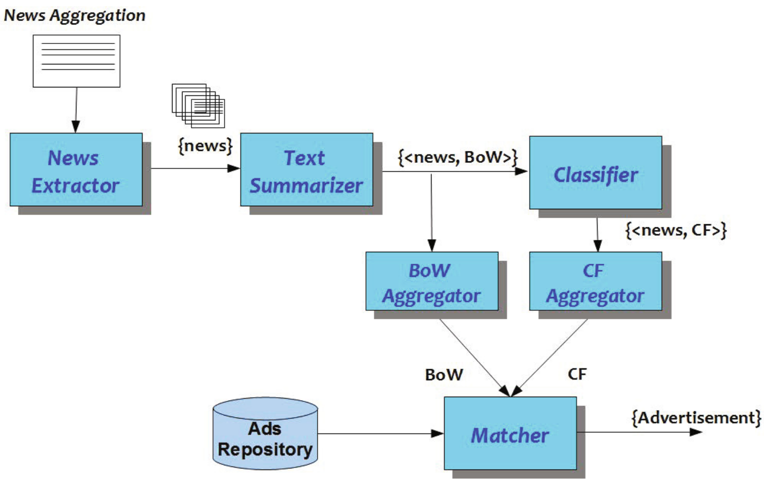


Fig. 1. The process of BoW and CF extraction

3 Experimental Results

Experiments were performed on a set of about 600 news aggregations classified according to 8 categories: *Economy and Finance*, *Environment Nature and Territory*, *Health and Health Services*, *Music and Shows*, *Publishing Printing and Mass Media*, *Religious Culture*, *Sports*, and *Transportation*. For each category, we selected 11 ads as webpages of product-service companies.

3.1 System Set Up

According to [8], we implemented the text summarizer by adopting an extraction-based technique, called TF2P, which takes into account information belonging to the Title and the First Two Paragraphs of the news. This choice is motivated by a previous work in which the adopted technique showed the best performances in the field of multimodal aggregation [9]. Each aggregation summary contains about 300 meaningful terms.

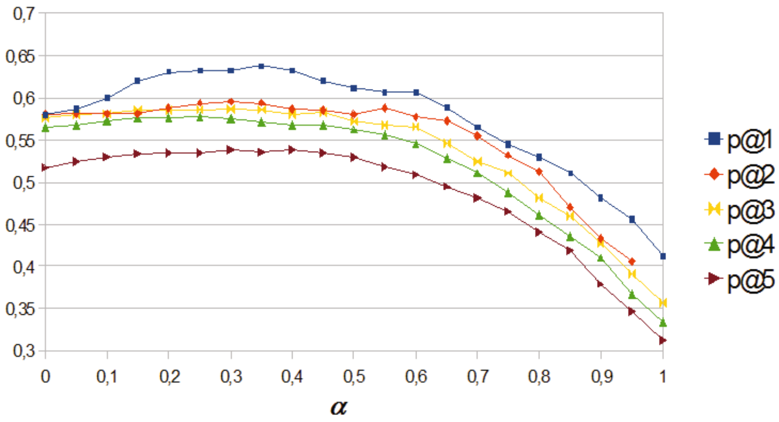


Fig. 2. *Precision at k*, varying α

Preliminary experiments have been performed to calculate the best value of α in Equation 1 to maximize the number of correct proposed ads. Figure 2 shows the results obtained comparing, for each suggested $\langle newsaggregation, ad\rangle$ pair, the category to which the news aggregation belongs with the one to which the ad belongs, varying α. The best results are obtained with a value of α in the range $0.25 - 0.40$, meaning that the impact given by the semantic contribution is greater.

3.2 System Evaluation

Being this the first proposal aimed at applying CA to multimodal real-world aggregated data, a comparison with existing baseline systems is not feasible. To assess its performances, we asked 15 assessors[2] to evaluate the relevance of the suggested ads with respect to a set of 10 randomly selected news aggregations for each category. Each assessor evaluated the relevance of 5 ads for each aggregation, for a total of 400 $\langle newsaggregation,\ ad\rangle$ pairs. The assessors have

[2] Assessors have been selected among students and young researchers of the Department of Electrical and Electronic Engineering, as well as among workers at RAI centre of Research and Technological Innovation.

no particular specialization related to the set of the selected categories. We set $\alpha = 0.35$ in Equation 1 and each ad was evaluated using a degree of relevance, i.e., relevant (1), somewhat relevant (2), or irrelevant (3). According to [10], the assessor scores were averaged to produce a composite score and converted in a binary score by assuming as irrelevant (i.e., false positives) $\langle newsaggregation, ad \rangle$ pairs with a composite score higher or equal to 2.4, as done in [2]. Performances were calculated in terms of precision (p) in suggesting k ads, with k varying from 1 to 5.

As we rely on a graded relevance scale of evaluation, to measure the effectiveness of the approach we adopt further evaluation metrics, i.e., the *Normalized Discounted Cumulative Gain (nDCG)*[11], the *Expected Reciprocal Rank (ERR)* [12], and the *Disagreement Variance* (Dis_k)[13]. We also calculated the average value of the assessors judgements, μ_k, where k is the number of ads suggested by the proposed CA system for each news aggregation.

Table 1. Overall results of the proposed CA system according to the assessor evaluation

k	**p@k**	**nDCG@k**	**ERR@k**	Dis_k	μ_k
1	0.632	0.504	0.304	0.390	2.136
2	0.639	0.548	0.385	0.411	2.123
3	0.594	0.593	0.401	0.397	2.154
4	0.573	0.650	0.407	0.383	2.179
5	0.542	0.705	0.409	0.375	2.224

Table 1 reports the results in terms of $p@k$, $nDCG@k$, $ERR@k$, Dis_k, and μ_k. Those results show that the $p@k$ is on average around 0.6. This depends on several issues. First, let us note that some noise might be introduced by the fully automatic process of aggregation building. Furthermore, news aggregation descriptions are often too short and not enough informative. The scarce information associated to each aggregation leads to a difficult automatic ad suggestion (but such issue could be true also for a human). Hence, the irrelevant score given by the assessors could be often due to the difficult identification of the right topic of the aggregation. Moreover, some categories are very generic (e.g., *Economy and Finance*), whereas others are very specific (e.g., *Religious Culture*). In particular, for generic categories, the test set might be too small to contain a significant vocabulary, whereas for categories with a specific vocabulary (e.g., *Religious Culture*) the matching could be easier. Moreover, if a category has a heterogeneous content, the choice of an appropriate ad is difficult not only for an automatic system, but also for human advertiser. It is plausible that the system provides worst performance for categories like *Economy and Finance*. A further issue is the limited number of assessors, due to the difficulty to experiment with a large customer base. However, we are currently planning to evaluate the system with a larger number of assessors.

4 Conclusions and Future Work

This paper presented a preliminary solution that applies classical contextual advertising solutions to suggest ads to multimodal aggregations of news stories from television and from the Internet. As for each aggregation the included news stories are fully cross-referenced, ads can be automatically placed in both the webpages linked by the RSS articles and the TV news stories associated to them.

The proposed solution, compliant with state-of-the-art approaches, has been experimented on a set of about 600 news aggregations. To calculate the overall performance, we made several experiments aimed at measuring the performance in suggesting k ads, with k from 1 to 5. Performance has been calculated for two different settings: (i) measuring precision according to previously classified news and ads and (ii) asking 15 users to give a degree of relevance. Results showed that the system reaches good performances in terms of precision, nDCG, Expected Reciprocal Rank, and Disagreement Variance. The system performs very well with some specific categories, whereas poor performances are obtained for more generic categories, mainly due to the heterogeneity of their contents and the difficulty to experiment with a large customer base.

To extensively validate the model system, we plan to set up new experiments with a richer ad dataset and with further users. We will also investigate the impact of a taxonomy of categories and the adoption of a hierarchical classifier on the performance of the system.

References

1. Feldman, J., Muthukrishnan, S.: Algorithmic methods for sponsored search advertising. CoRR abs/0805.1759 (2008)
2. Broder, A., Fontoura, M., Josifovski, V., Riedel, L.: A semantic approach to contextual advertising. In: SIGIR 2007: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 559–566. ACM, New York (2007)
3. Anagnostopoulos, A., Broder, A.Z., Gabrilovich, E., Josifovski, V., Riedel, L.: Just-in-time contextual advertising. In: CIKM 2007: Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, pp. 331–340. ACM, New York (2007)
4. Armano, G., Giuliani, A., Vargiu, E.: Studying the impact of text summarization on contextual advertising. In: 8th International Workshop on Text-based Information Retrieval (2011)
5. Messina, A., Montagnuolo, M.: Multimodal Aggregation and Recommendation Technologies Applied to Informative Content Distribution and Retrieval. In: Soro, A., Vargiu, E., Armano, G., Paddeu, G. (eds.) Information Retrieval and Mining in Distributed Environments. SCI, vol. 324, pp. 213–232. Springer, Heidelberg (2010)
6. Salton, G., McGill, M.: Introduction to Modern Information Retrieval. McGraw-Hill Book Company (1984)
7. Rocchio, J.: Relevance feedback in information retrieval. In: The SMART Retrieval System: Experiments in Automatic Document Processing, pp. 313–323. Prentice Hall (1971)

8. Armano, G., Giuliani, A., Vargiu, E.: Experimenting text summarization techniques for contextual advertising. In: IIR 2011: Proceedings of the 2nd Italian Information Retrieval (IIR) Workshop (2011)
9. Armano, G., Giuliani, A., Messina, A., Montagnuolo, M., Vargiu, E.: Experimenting text summarization on multimodal aggregation. In: Lai, C., Semeraro, G., Vargiu, E. (eds.) 5th International Workshop DART 2011, New Challenges on Information Retrieval and Filtering. CEUR Workshop Proceedings, vol. 771 (2011)
10. Ciaramita, M., Murdock, V., Plachouras, V.: Semantic associations for contextual advertising. Journal of Electronic Commerce Research 9(1), 1–15 (2008); Special Issue on Online Advertising and Sponsored Search
11. Järvelin, K., Kekäläinen, J.: Ir evaluation methods for retrieving highly relevant documents. In: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2000, pp. 41–48. ACM, New York (2000)
12. Chapelle, O., Metlzer, D., Zhang, Y., Grinspan, P.: Expected reciprocal rank for graded relevance. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009, pp. 621–630. ACM, New York (2009)
13. Amer-Yahia, S., Roy, S.B., Chawlat, A., Das, G., Yu, C.: Group recommendation: semantics and efficiency. Proc. VLDB Endow. 2(1), 754–765 (2009)

User Semantic Preferences for Collaborative Recommendations

Sonia Ben Ticha[1,2], Azim Roussanaly[2], Anne Boyer[2], and Khaled Bsaies[1]

[1] URPAH Teams Tunis, Tunisia
[2] KIWI Team LORIA laboratory Nancy, France

Abstract. Personalized recommender systems provide relevant items to users from huge catalogue. Collaborative filtering (CF) and content-based (CB) filtering are the most widely used techniques in personalized recommender systems. CF uses only the user-rating data to make predictions, while CB filtering relies on semantic information of items for recommendation. In this paper we present a new approach taking into account the semantic information of items in a CF system. Many works have addressed this problem by proposing hybrid solutions. In this paper, we present another hybridization technique that predicts users 'preferences for items based on their inferred preferences for semantic information. With this aim, we propose a new approach to build user semantic profile to model users' preferences for semantic information of items. Then, we use this model in a user-based CF algorithm to calculate the similarity between users. We apply our approach to real data, the MoviesLens dataset, and compare our results to standards user-based and item-based CF algorithms.

Keywords: Recommender systems, collaborative filtering, semantic information, user modeling.

1 Introduction

Collaborative filtering (CF) and content-based (CB) filtering are the most widely used techniques in Personalized Recommender Systems (RS). The fundamental assumption of CF is that if users X and Y rate n items similarly and hence will rate or act on other items similarly [7]. In CB, user will be recommended items similar to the ones he preferred in the past. However, CF and CB techniques must face many challenges [9] like the data sparsity problem, the scalability problem for big database with the increasing numbers of users and items and the cold start problem. To overcome the disadvantages of both techniques and benefit from their strengths, hybrid solutions have emerged. In this paper, we present a new approach taking into account the semantic information of items in a CF system. In our approach, we design a new hybridization technique, called User Semantic CF (USCF), which predicts user preferences for items based on their inferred preferences for semantic information.

Our contribution is summarized as follows: (i) we propose a new approach for building *user semantic model*, that inferred the user preferences for semantic information of items, (ii) we define a classification of attributes and propose a suited algorithm for each class, (iii) for each relevant attribute, we build the *user semantic*

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 203–211, 2012.

attribute model using the suited algorithm, (iv) we provide predictions and recommendations by using the user semantic model in a user-based CF algorithm [5], (iv) we perform several experiments with real data from the MoviesLens data sets which showed improvement in the quality of predictions compared to only usage CF.

2 Related Work

RS have become an independent research area in the middle 1990s. CF is the most widespread used technique in RS, it was the subject of several researches [5][6][**7**]. In CF user will be recommended items that people with similar tastes and preferences liked in the past. CB is another important technique; it uses techniques developed in information filtering research [15][16]. CB assumes that each user operates independently and recommends items similar to the ones he preferred in the past. The major difference between them is that CF only uses the user-item ratings data to make recommendations, while CB rely on the features of items for predictions.

To overcome the disadvantages of both techniques and benefit from their strengths several RS use a hybrid approach by combining CF and CB techniques. The Fab System [1] counts among the first hybrid RS. Many systems have been developed since [3][10]. In [2], authors integrate semantic similarities of items with item rating similarities and used it in item based CF algorithm to generate recommendations. Most of these hybrid systems ignore the dependency between users' ratings and items' features in their recommendation process; taking account of this link can improve the accuracy of recommendation. In [8], this dependency was computed by using TF/IDF measure to calculate the weight of item feature for each user. In [14], authors are inferring user preferences for item 'tags by using several measures. This work is suitable only for item 'tags and cannot be used for others kinds of attributes.

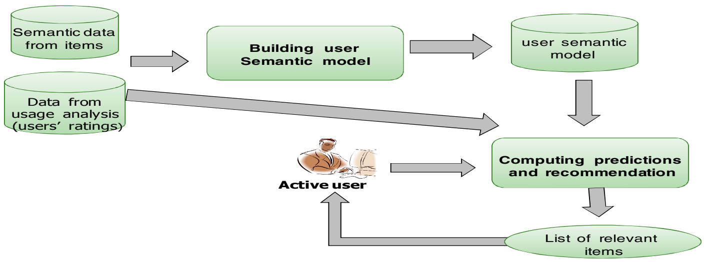


Fig. 1. Architecture of our system: USCF approach

3 User Semantic Collaborative Filtering (USCF) Approach

Our system consists of two components as shown in Fig. 1. The first builds the user semantic model by inferring user semantic preferences from user ratings and item features. The second predicts for each user a list of relevant items based on the user-based CF algorithm and using the user semantic model for computing similarities between users. USCF uses only data from usage analysis and semantic information of items. Table 1 describes all used symbols. So, we define:

- **Data from usage analysis**: the usage analysis profile of item i is given by the following ratings vector $I_i=(r_{1,i},r_{2,i},...r_{u,i},...,r_{N,i})$, $r_{u,i}$ is the rating of user u on item i; it can be either a missing value or a number on a specific scale if user u rated item i.
- **Semantic information from items**: we assume that item is represented by *structured data* [16] in which there is a small number of attributes, each item is described by the same set of attributes, and there is a known set of values that each attribute may have. In the following, we will use the terms *feature* to designate a value of an attribute. The semantic attribute based profile of item i on attribute A is given by the features vector: $F_{Ai}=(b_{Ai,1},..., b_{Ai,f}, ...,b_{Ai,LA})$, where:

$$b_{\mathrm{Ai},f} = \begin{cases} 0 \; if \; f \; is \; not \; a \; feature \; of \; item \; i \\ 1 \; if \; f \; is \; a \; feature \; of \; item \; i \end{cases} \tag{1}$$

Table 1. Description of the used symbols

Symbol	Meaning	Symbol	Meaning
N	number of users	I	item-user ratings matrix, $(r_{u,i})_{1..M,1..N}$
M	number of items	F_{Ai}	semantic attribute based profile of item i
S_I	set of items described by I_i	S_{FA}	set of items defined by F_{Ai}
I_i	usage analysis profile of item i	F_A	item semantic attribute matrix (M,L_A)
L_A	number of features of A	Q	user semantic model (matrix $(q_{u,k})_{1..N,1..L}$)
K_A	Number of clusters associated to A	U	User-item rating matrix (I transposed)
Q_A	user semantic attribute model (matrix $(q_{u,Ak})_{1..N,1..KA}$)	$q_{u,Ak}$	inferred preference of user u on feature(s) of A labeling the cluster k
A	Relevant attribute	$q_{u,k}$	inferred preference of u on feature(s) k

Otherwise, we must distinguish between two kinds of attributes: *multi-valued* and *mono-valued* attribute. For a same item, if an attribute can have many values, then it is a multi-valued attribute (a *movie* can have many *genres)*; while if it must have only one feature it is called mono-valued attribute (a *movie* has only one *director*).

Furthermore, all item attributes do not have the same degrees of importance to users. There are attributes more relevant than others. For instance, the *movie genre* can be more important, in the evaluation criteria of user, than the *release date*. Experiments that we have conducted (see section 4) confirmed this hypothesis. In this paper, we assume that relevant attributes will be provided by a human expert.

3.1 Building the Users Semantic Model

In our approach we have defined two classes of attributes: *dependent attribute* which having very variable number of features. This number is directly correlated to the number of items. Thus, when the number of items is increasing, the number of features is increasing also (*actors of movies; user tags*). *Non dependent attribute* which having a very few variable number of features that is no correlated to the number of items. Thus, the increasing number of items has no effect on the number of features (*movie genres*). For each class, we have defined a suited inferring user semantic preferences algorithm. For the dependent attribute, we propose techniques issues from information filtering research like TF/IDF. For non dependent attribute, we use machine learning algorithms. The aim of this paper is to present our solution for non dependent attributes, dependent attributes will be addressed in future works.

For each relevant attribute A, we have built the corresponding user semantic attribute model Q_A that provides the inferring user preferences for its features. The user semantic model Q is so the horizontal concatenation of all users semantic attributes models. For example, assume that we have a movies Data set with users ratings and we want to infer the preference $q_{u,action}$ of user u on the *action movies*. This means computing an aggregation overall ratings of user u on all action movies (eq. 2). The aggregation function can be a simple function like the average (AVG), or more complicated mathematical function like TF/IDF, or special user-defined function. For non dependent attribute, we choose to define a special user function, so we use a clustering algorithm to learn the user semantic attribute model.

$$q_{u,genre=action} = AGGR_{i.genre=action}\, r_{u,i} \tag{2}$$

3.2 User Semantic Attribute Model for Non Dependent Attribute

The idea is to partition S_I in K clusters; each cluster is labeled by a feature or a set of features of A ($K<=L_A$). Thus, the cluster center $C_{A,k}=(q_{1,Ak},\ldots, q_{u,Ak},\ldots,q_{N,Ak})$ modeled the inferred users preferences for the feature(s) associated to cluster k. For example, assume that we have a movies dataset and we want to infer users' preferences on *movie genre*. The attribute *genre* has L_{genre} features, if each cluster is labeled by a feature, then we will have L_{genre} clusters. Assume that the feature *action* is labeled the cluster l, then after running the clustering algorithm, the center of cluster l provides the *action*-users profile $C_{genre,1}=(q_{1,genre1},\ldots,q_{u,genre1},\ldots,q_{N,genre1})$ where $q_{u,\ genre1}$ provides the inferring preference of user u on *action movie*. Matrix Q_A is so obtained by calculating the transposed matrix of C_A. However, the question is what clustering algorithm to use? As we have already said, we have two kinds of attributes, multi-valued attribute and mono-valued attribute. For multi-valued attribute, a same item can belongs to many clusters, so the clustering algorithm must provide non disjointed clusters, while, for mono-valued attribute, an item must belong to only one cluster so the clustering must provide disjointed clusters. In previous work [11] we addressed the multi-valued attribute and we choose the Fuzzy C Mean as a fuzzy clustering algorithm. In this paper, we present our solution for mono-valued attribute.

After a study of several clustering algorithms, we have chosen the K-Mean clustering algorithm for its simplicity. The result of K-mean is depending on the number K of clusters, and the initial set of cluster centers. In this paper, we design an algorithm for the initialization step of the K-mean algorithm.

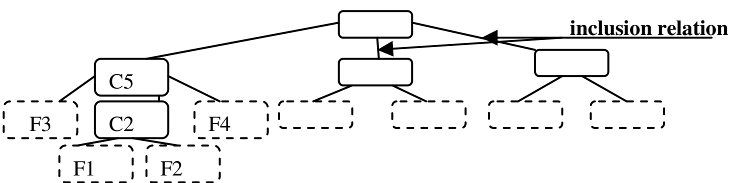


Fig. 2. Ontology of an attribute

Algorithm of the K-Mean initialization step

To determine the number of clusters and their respective initial centers, we have defined two thresholds: *MinNbRaIt* that defines the minimum number of item ratings and *MinNbItClust* which indicates the minimum number of items by cluster in the initialization step. Each cluster is labeled by a feature and is created according to formula 3, its initial center is the mean value of its items. Among all clusters created, only those checking the selection criteria described by formula 4 are preserved. Thus, user preferences cannot be inferred for features assigned to non selected clusters. Because of the data sparsity, the number of these features can be important. To solve this problem, we use an ontology describing the non dependent attribute, thus a cluster can be labeled by a single or several features.

$$C_k = \{item\ i\ \in\ S_I /\ ratings\_Number(\ i._{attribute=f_k}\ ) \geq \mathrm{MinNbRaIt}\} \tag{3}$$

$$|C_k| >= \mathrm{MinNbItClust} \tag{4}$$

We assume having an ontology describing the attribute. The concepts of the ontology (solid line in Fig. 2) are interconnected hierarchically, and the leaf nodes describe the features of the attribute (dashed in Fig. 2). For example, features F1 and F2 are included in the concept C2; features F3 and F4 and concept C2 are included in concept C5. Each feature does not check the selection criteria defined above, will be replaced by its closest ancestor meeting the criteria in the ontology. In the example described in Table 2, F1 and F3 satisfy the selection criterion, so a cluster will be assigned to each. However, as F2 does not satisfy the criteria, it will be replaced by its father C2; Similarly, C2 does not satisfy the criteria itself, it will be replaced by C5. In addition, F4 does not check the criteria; it will also be replaced by C5. The number of items assigned to the concept C5 is equal to 8 (5+3) and it's greater than MinNbItClust. As, C5 satisfies the criterion, a cluster will be associated to it. Using this initialization algorithm, we will be able to infer user preferences for the concept C5 which groups features F2 and F4.

Table 2. Example, *MinNbItClust = 6*

Feature	Nb items with *ratings_Number* >= *MinNbRaIt*
F1	10
F2	5
F3	12
F4	3

3.3 Computing Predictions and Recommendation

To compute predictions we use the user semantic model Q in a user-based CF algorithm [5] for computing similarities between users. User-based CF is based on the k-Nearest-Neighbors algorithm. Formula (5) computes similarities between two users with the Pearson correlation introduced by Resnick et al. [5]; $\bar{q}_v$ is the average of inferred preferences of user v on features. Then, the prediction of rating value of active user u_a on non rated item i was computed by formula 6; V denotes the set of the nearest neighbors that have rated item i.

$$sim(u_a, v) = \frac{\sum_k (q_{u_a,k} - \bar{q}_{u_a})(q_{v,k} - \bar{q}_v)}{\sqrt{\sum_k (q_{u_a,k} - \bar{q}_{u_a})^2}\sqrt{\sum_k (q_{v,k} - \bar{q}_v)^2}} \quad (5)$$

$$pr(u_a, i) = \bar{r}_{u_a} + \frac{1}{\sum_{v \in V} |sim(u_a, v)|} \sum_{v \in V} sim(u_a, v)(r_{v,i} - \bar{r}_v) \quad (6)$$

3.4 USCF Algorithm

Our approach resolves the scalability problem for several reasons. First, the building process of user-semantic model is fully parallelizable and can be done offline. Second, this model allows a dimension reduction since the number of columns in user semantic model Q is much lower than those of user rating matrix U. Third, the computing of similarities between users can be done offline. In addition, USCF alleviates the data sparsity problem by providing solution to the neighbor transitivity problem in which users with similar preferences may not be identifies as such if they haven't any items rated in common. Indeed, the number of missing values is much lower in Q than in U; thus, all similarities can be computed. This is not the case with U, because similarities between users who have no co rated items cannot be computed.

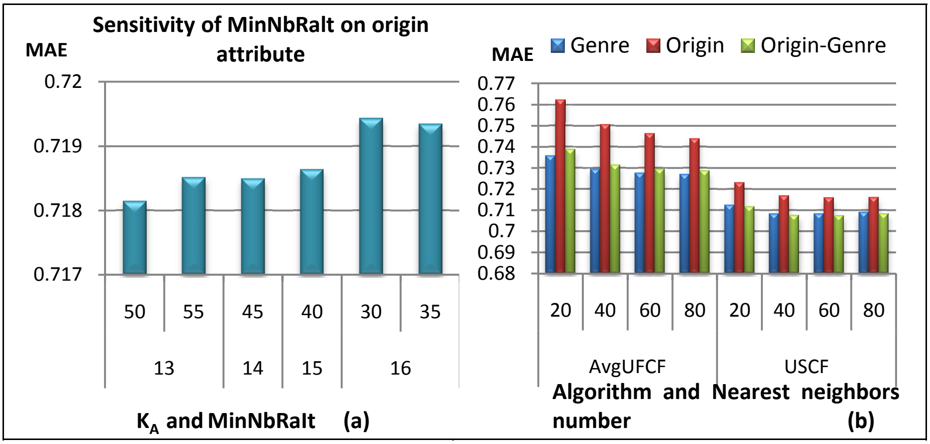


Fig. 3. Comparative results for USCF: (a) by varying the MinNbRaIt; (b) between USCF and AvgUFCF on all relevant attributes.

4 Performance Study

In this section, we study the performance of the USCF algorithm against the standard User-Based CF [5] (UBCF), the standard Item-Based CF (IBCF) [6] and Average User Feature CF algorithm (AvgUFCF). For IBCF, predictions have been computed using the Adjusted Cosine correlation measure which provides, according to [6], best prediction accuracy. In AvgUFCF user semantic model has been built by using the average (AVG) as an aggregation function (formula 2). We evaluate these algorithms in terms of predictions relevancy by using the Mean Absolute Error (MAE) (7).

$$MAE = \frac{\sum_{u,i} \left| pr_{u,i} - r_{u,i} \right|}{d} \quad (7)$$

d is the total number of ratings over all users, $pr_{u,i}$ is the predicted rating for user u on item i, and $r_{u,i}$ is the actual rating. Lower the MAE is, better is the prediction.

4.1 Experimental Datasets

We have experimented our approach on real data from the MovieLens1M dataset [4]. For the semantic information of items, we have used the HetRec 2011 dataset [12]. The *genre* and the *origin country* of movies have been used as non dependent attributes. Movie' genre is a multi-valued attribute whereas origin country is mono-valued. W3C movie ontology [13] has been used for describing the origin of movie.

We have filtered the data by maintaining only users with at least 20 ratings and the movies origins existing in the ontology. After the filtering process, we have obtained a data set with 6027 users, 3559 movies, 19 genres, 44 origins. The usage data set has been sorted by the timestamps, in ascending order, and has been divided into a training set (including the first 80% of all ratings) and a test set (the last 20%). We have tried several distance measures in the clustering algorithm; the cosines distance has provided the best result.

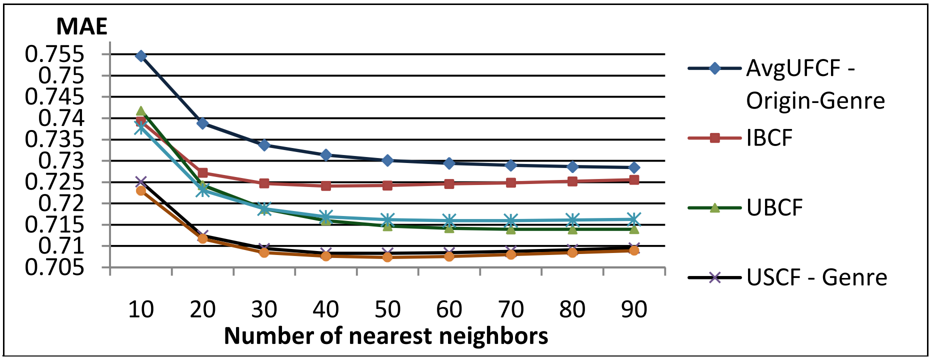


Fig. 4. Prediction accuracy for USCF v. IBCF, UBCF and AvgUFCF

4.2 Results

It should be noted that the inferring user preferences for the attribute *genre* have been addressed in a previous work [11]. Therefore, we will not detail the experiments conducted for this attribute in this paper. In Fig. 3 (a), the MAE has been plotted with respect to the MinNbRaIt parameter. It compares the K-Mean initialization algorithms on the attribute *origin* for MinNbItClust =9 and 60 neighbors. We note that the accuracy of recommendations improves with the decreasing number of clusters K_A. In addition, the MAE converges for 50 ratings; this shows the impact of MinNbRaIt on the accuracy of the recommendations. The plots in Fig. 3 (b) show that the *genre* provides better results than the attribute *origin*, for both algorithms USCF and AvgUFCF and regardless of the number of neighbors. Therefore, we can conclude that the *genre* is more relevant than the *origin.* Fig. 4 depicts the prediction accuracy of USCF, in contrast to those produced by IBCF, UBCF and AvgUFCF using the best parameters of each algorithm. In all cases, the MAE converges between 60 and 70 neighbors, however, our algorithm results in an overall improvement in accuracy. This improvement can be explained by many reasons. First, the use of semantic information of items in CF. Second, user semantic model is built according to a collaborative principle; ratings of all users are used to compute the semantic profile of each user. It is not the case of the AvgUFCF algorithm; this may explain its results despite taking into account the semantic aspect. Third, the choice of the attribute can have significant influence on improving the accuracy. Lastly, matrix Q has few missing values, so, it allows inferring similarity between all users.

5 Conclusion and Future Work

In this paper, we have designed a new hybridization technique, which predicts users' preferences for items based on their inferred preferences for semantic information. We have defined two classes of attributes, the *dependent attribute* and the *non dependent attribute* and we have proposed an approach for inferring user semantic preferences for each class. Our approach provides solutions to the scalability problem, and alleviates the data sparsity problem by reducing the dimensionality of data. The experimental results show that the USCF algorithm improves the prediction accuracy compared to usage only approach (UBCF and IBCF) and hybrid algorithm (AvgUFCF). Furthermore, we have experimentally shown that, all the attributes don't have the same importance to users.

An interesting area of future work is to use machine learning techniques to automatically determine the relevant attributes. We will also further study the extension of the user semantic model to the dependent attribute and non structured data; study the use of this model in case-based RS to solve the cold start problem; and lastly, study the impact of using others machine learning algorithms for building the user semantic attribute model for non dependent attribute and comparing their results.

References

1. Balabanovic, M., Shoham, Y.: Fab: content-based, collaborative recommendation. Comm. ACM 40(3), 66–72 (1997)
2. Mobasher, B., Jin, X., Zhou, Y.: Semantically Enhanced Collaborative Filtering on the Web. In: Berendt, B., Hotho, A., Mladenič, D., van Someren, M., Spiliopoulou, M., Stumme, G. (eds.) EWMF 2003. LNCS (LNAI), vol. 3209, pp. 57–76. Springer, Heidelberg (2004)
3. Burke, R.: Hybrid Web Recommender Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web. LNCS, vol. 4321, pp. 377–408. Springer, Heidelberg (2007)
4. MovieLens, `http://www.movielens.org/` (retrieved: January 2012)
5. Resnick, P., Iacovou, N., Suchak, M., Bergstorm, P., Riedl, J.: Grouplens: an open architecture for collaborative filtering of netnews. In: Proc. of Conf. on Computer Supported Cooperative Work, pp. 175–186. ACM, North Carolina (1994)

6. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based collaborative filtering recommendation algorithms. In: Proc. 10th Int'l WWW Conf. (2001)
7. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. In: Adv. in Artif. Intell. (January 2009)
8. Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: Feature-Weighted User Model for Recommender Systems. In: Conati, C., McCoy, K., Paliouras, G. (eds.) UM 2007. LNCS (LNAI), vol. 4511, pp. 97–106. Springer, Heidelberg (2007)
9. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. on Knowledge and Data Engineering 17(6), 734–749 (2005)
10. Belén Barragáns-Martínez, A., Costa-Montenegro, E., Burguillo, J.C., Rey-López, M., Mikic-Fonteand, F.A., Peleteiro, A.: A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition. Information Sciences 15, 4290–4311 (2010)
11. Ben Ticha, S., Roussanaly, A., Boyer, A.: User Semantic Model for Hybrid Recommender System. In: The First Int'l Conf. SOTICS, Barcelona, Spain (October 2011)
12. 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011) at the 5th ACM Conf. on Recommender Systems (RecSys 2011)
13. Bouza, A.: MO – the Movie Ontology (2010), http://movieontology.org
14. Sen, S., Vig., J., Riedl, J.: Tagommenders: connecting users to items through tags. In: Proc. of the 18th Int'l Conf. on WWW, Madrid, Spain (April 2009)
15. Lops, P., de Gemmis, M., Semeraro, G.: Content-based Recommender Systems: State of the Art and Trends. In: Recommender Systems Handbook, pp. 73–105 (2011)
16. Pazzani, M.J., Billsus, D.: Content-Based Recommendation Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007)

A Multidimensional Model of Trust in Recommender Systems

Martina Maida, Konradin Maier, Nikolaus Obwegeser, and Volker Stix

Vienna University of Economics and Business, Institute for Information Business, Augasse 2-6, 1090 Wien, Austria
{martina.maida,konradin.maier,nikolaus.obwegeser,volker.stix}@wu.ac.at

Abstract. As e-commerce providers increasingly utilize recommender systems (RS) to support their customers, trust is emerging as a key factor for the design of such technologies. Although a considerable number of researchers addressed the issue of trust towards RS, there is still no common understanding of how trust relates to the acceptance of RS and what factors influence the perception of trustworthiness in RS. After a discussion on the peculiarities of RS, we build on a literature review of trust in order to analytically distinguish basic concepts of trustworthiness. We propose an integrated model of trustworthiness that accounts for the multiple dimensions and perspectives on trustworthiness in RS. Additionally, we will point out several implications for practice and conclude with suggestions for further research in this area.

Keywords: Trust, trustworthiness, recommender systems, black box.

1 Introduction

The scientific body of knowledge on trust in RS has grown considerably over the last decade. Contributions include numerous empirical as well as theoretical/conceptional studies. From the huge amount of factors which contribute to the acceptance of decision support technologies (see [1]), trust is one of the most important drivers for the success of a RS in an e-commerce environment [2,3,4]. The concept of trust as we use it in our everyday life can be considered rather fuzzy and not properly defined [5]. Scientific research though, in its essential need for clarity and defined concepts, has tried to come up with definitions and attributions of trust on several occasions [6,7,8]. While all of these contributions represent valuable advancements of the field, a unified model of trust in RS remains missing. Thus, the main goal of this study is to consolidate the findings from a broad literature review to generate an integrated, multidimensional model of trust in RS. This model can then be used by both academics and practitioners to derive specific guidelines on how to challenge the problem of designing trustworthy RS for e-commerce applications.

In the context of e-commerce typical RS are web-based tools that provide certain operations for the user in order to support the decision making process in the online shopping process. Usually, RS use information about the customer

C. Huemer and P. Lops (Eds.): EC-Web 2012, LNBIP 123, pp. 212–219, 2012.

such as preferences, previous shopping habits or expectations to give recommendations of a certain kind [9]. They are especially helpful in terms of reducing the often enormous amount of products and choices that customers are facing in online shopping nowadays [10]. In general, RS are employing a more or less complex decision support technology, depending on the type and field of application. Thus, in many cases, the inner working principles of how RS process the given information in order to come up with recommendations are nontransparent to the customer [11]. Trust is therefore much more relevant and important for RS then for conventional information systems (IS).

According to most researchers, trustworthiness of a technology is composed of three trusting beliefs: competence, benevolence and integrity [12]. In the case of RS in e-commerce, a customer might ask himself the following questions: Does the technology act in my best interest? Is the RS truthful in the operations performed? Is the RS competent enough to help me in my decision making? Thus, in order to come up with valuable propositions on how to increase the perceived trustworthiness, we have to investigate on the actual composition of trust as well as on the sources of trust. In a recent study, Wang et. al. [13] presented a laboratory experiment which successfully demonstrated that trust is based on several reasons, which they derived beforehand from a literature review. Their research is based on what has been named the computers as social actors (CaSA) paradigm, which states that people developed the habit of treating computers as social actors [14]. This is particularly important for research on trust in RS, as trust is a concept originally found in sociology to describe interpersonal relationships.

In this study we will build on the reasons to trust to model trust towards RS supported by a graphical representation. While research on the initial formation of trust before usage or at the time of first use is clearly promising, this model will depict a perspective on the perception of trust in the context of RS that is independent of time or usage history but rather presents a generalized view. We will present and clearly define two different dimensions of trust and propose possibilities for designers to effectively implement trustworthy RS.

2 Trust in Recommender Systems

There has been much debate in the literature on what constitutes "trust". Though different research fields have developed different perspectives on trust, a common element underlying most of these viewpoints is the trustor's "willingness to be vulnerable" by relying on the actions of another agent (the trustee) [15]. To give an example, the most frequently cited definition of trust is the "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" [6]. As trust implies vulnerability and the risk of abuse [5,8], the question how humans decide whether or not to trust other agents is an important one. In their integrative model of trust, Mayer et al. [6] propose that trust is formed by two factors: perceived trustworthiness of the trustee and the trustor's propensity to trust. Pertaining to the former, trustors judge an agent's trustworthiness based on perceived ability, perceived benevolence and perceived integrity of the agent. Mayer et al. describe the ability belief as whether an agent is perceived as having the necessary skills, knowledge and cognitive resources to perform a certain behavior on behalf of the trustor; the benevolence belief as the perceived willingness of the trustee to act in accordance with the trustor's interests and integrity as whether the trustee is perceived as adhering to a set of principles which are approved by the trustor. Pertaining to the latter, the degree to which perceived trustworthiness leads to trust is moderated by the trustor's propensity to trust, which is best described as his general stance to trust others. Whether a trustor actually commits to a trust relationship further depends on the level of risk associated with the specific task the trustee should perform on behalf of the trustor.

While the perspective on trust as the willingness to be vulnerable has gained some popularity in sociology, the research in RS and in IS usually defines trust as "an individual's beliefs in an agents competence [ability], benevolence, and integrity" [4]. The reason for this discrepancy is that there is no agreement whether trust should be conceptualized as a belief, attitude, behavioral intention or behavior of an agent [16]. While sociology focuses on trust as a behavioral intention, IS and RS research mainly conceptualize trust as a belief [17]. Here, we follow the idea that trust manifests itself on all levels of the theory of planned behavior [18] and that trustworthiness and trust are conceptually distinct constructs [19]. In that sense, trustworthiness is the foundation for trust which is again the basis for trusting behavior. Consequently, in contrast to most IS and RS research, we refer to the triad of trusting beliefs - ability, benevolence and integrity - as *perceived trustworthiness*, which is consistent with the sociological perspective on trust.

Independent from the level of conceptualization, most research accounts on trust have in common that they interpret trust as a function which enables people to rely on agents about which they do not have all relevant knowledge or which are too complex to be fully understood by a bounded rational decision maker [15]. That is, in situations where rational considerations are not feasible trust acts a mechanism which reduces complexity and thereby allows us to act based on little information and on intuitive decisions [8]. As users of RS usually have neither all relevant information about an RS nor - given the unlikely case of complete information - the necessary resources and skills to fully understand an RS, trust seems to be a precondition for RS usage. Indeed, there is much empirical evidence that trust is of major influence to RS usage and thus a prerequisite to support human decision making by providing RS (e.g. [2,4]).

Given the above relationship of trust and usage, the question how people assess the trustworthiness of an agent is of much interest for RS research. Only if we understand the bases of trust, we are able to systematically develop RS which are trustworthy and therefore able to gain user acceptance. By reviewing the trust literature Wang and Benbasat [13] identified six reasons on which people build to assess the ability, benevolence and integrity of others: dispositional, institutional, heuristic, calculative, interactive and knowledge-based reasons. Their empirical results indicate that especially interactive and knowledge-based reasons and to a lesser extent calculative and dispositional reasons are relevant for assessing the trustworthiness of RS.

3 Development of a Multidimensional Model

In the following section we will consolidate our findings from both RS and trust literature to develop an integrated, multidimensional model of trust. Thus, after introducing the basics concepts in the next subsection, we will present the model in Sect. 3.2 and elaborate on practical implications in Sect. 4.

3.1 Dimensions of Trustworthiness

RS in e-commerce, especially when employing complex MCDM methods, are likely to be perceived as black boxes [20]. The term black box is used to express a system of any kind which inner working principles are not known or not of interest. The only known characteristics of a black box are defined through its inputs and outputs. The opposite of a black box is a white box, a system which is completely understood and transparent in its working process to the user.

Based on this findings, Sinha et.al. found that users generally prefer transparent over not transparent systems and feel more confident in recommendations made by transparent systems [11]. Thus, many researchers approached this issue by proposing RS with some kind of explanation facilities or justification support [16,20].

Under the assumption that the RS under consideration would influence the decision making process positively, which is a general precondition for our model, only two possible sources exist why a customer would perceive such a technology as trustworthy. Based on the six reasons to trust found in [13] it is possible to analytically distinguish these two sources of trustworthiness without losing the unity of trust as a social experience [5]. Thus, we can identify two basic forms of how trustworthiness an be influenced: directly and indirectly.

Knowledge-based trustworthiness is grounded on the trustor's knowledge, skills and cognitive resources which enable to directly evaluate the trustee's ability, benevolence and integrity. In other words, knowledge-based trustworthiness emerges from a trustor's perceived understanding of the trustee's inner working principles and thereby from a perceived predictability of the specific trustee. In practice, this way to build trust changes the customers perception of the RS as a black box towards a trustworthy white box, based on the perceived insight into the RS. The trusting reason which mainly makes up this from of trustworthiness is knowledge, but also interactive reasons may lead to knowledge-based trustworthiness if the past interactions are used by the trustor to model the internal working principles of the trustee.

Inferred trustworthiness is based on a number of informational cues wich allow to estimate the agent's trustworthiness in an indirect way. That is, the trustor builds on rather general and agent-unspecific information to gain an impression of an agent's trustworthiness. For example, trust is inferred from different sources, like group memberships, roles, stereotypes, certificates or opinions of others, to the specific trustee at hand. This form of trust-building changes the users perception of the RS towards a trustworthy black box, because it does not require the trustor to gain knowledge about the trustee's inner information processing. The trusting reasons assigned to this kind of trustworthiness are dispositional, institutional, heuristic and calculative reasons. Interactive reasons are also assigned to this cluster as far as past experiences with the trustee are used only to classify the trustee (e.g. into a performance level) and not to reveal his internal working principles.

3.2 The Simplex of Trustworthiness

The 2-simplex presented in Fig. 1 is a comprehensive depiction of the problem at hand. It is spanned by three vertices: black box (BB), trustworthy black box (TBB) and trustworthy white box (TWB). The customers perception of the trustworthiness of the RS itself is a convex combination of the three vertices that span the simplex. Thus, the centroid of the simplex can be interpreted as the point where the users perceives equal parts of the system as untrustworthy, inferred (indirectly) trustworthy and knowledge-based (directly) trustworthy.

Any point within the simplex is therefore an evaluation of trustworthiness of a RS that consists of three dimensions: inferred trustworthiness, knowledge-based trustworthiness and untrustworthiness. Moving a point in one of these directions towards one of the vertices means that the user's assessment of trustworthiness of the system has changed, which is the challenge for designers of RS in the

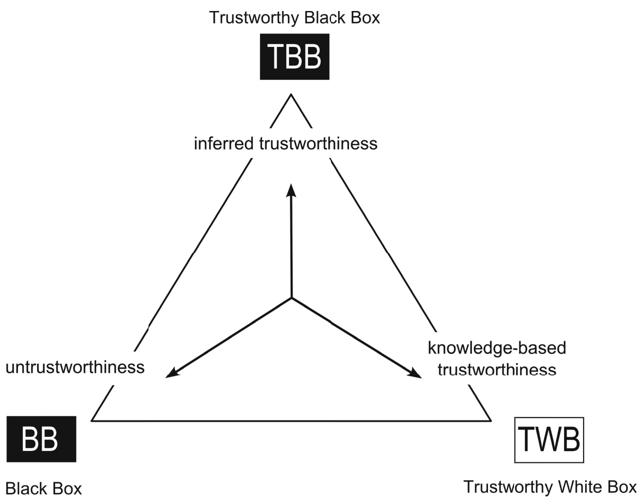


Fig. 1. Simplex model of trustworthiness in RS

practical field. The edge of the simplex connecting the TBB with the TWB depicts all points with full trustworthiness (in other words, the RS is perceived as completely trustworthy). If the RS is perceived on this edge, its location on the edge is an indication of the composition of trustworthiness.

4 Implications for RS in Practice

The conceptual model presented in the previous section enables not only the classification of measures to increase the use of RS but also the evaluation of these measures regarding their practicability. In essential, we can distinguish three different strategies to foster RS use: (1) increase users' knowledge regarding the RS (2) increase the inferred trustworthiness (3) increase the willingness to use RS by risk reduction. In the following we will provide descriptions of these strategies as well as discussions of each strategy in the light of bounded rationality.

4.1 Increase Knowledge-Based Trustworthiness

The first strategy - to increase the knowledge of users - is a common approach to foster the usage of RS. This includes not only measures which deliberate information about the RS like explanation facilities or tutorials, but also measures which help users to understand the RS in a more indirect fashion like reducing the complexity of the algorithm underlying the RS. The empirical results of Wang and Benbasat [13] show that the knowledge-strategy is indeed a powerful tool to increase positive and to reduce negative user experiences with RS.

Nevertheless, the strategy to increase the knowledge of users is limited in two ways: The amount of information which can be processed by cognitively limited users and by the willingness of the RS provider to disclose information. Pertaining to the former, the attention of a bounded rational decision maker is a scarce resource which he has to carefully allocate to different problems [21]. As RS are typically complex systems which are build on diverse research areas like mathematics, statistics, psychology, marketing or computer science [9,22,23] the amount of resources needed to fully understand how and why a RS works are usually prohibitive. Pertaining to the willingness of RS providers to disclose information, one has to keep in mind that a good RS represents a competitive advantage which the RS provider wants to protect [20]. As the deliberation of information about a RS potentially leads to an erosion of this competitive advantage, the knowledge-strategy of making RS trustworthy conflicts with the provider's aim of economic success.

4.2 Increase Inferred Trustworthiness

The second strategy is to increase the trustworthiness of the RS by providing cues which enable the decision maker to assess the trustworthiness without deliberating information directly related to the ability, benevolence and integrity of the RS. Cues on the trustworthiness of an agent include for instance reputation, certificates or general stance to trust others (cf. the reasons to trust assigned to inferred trustworthiness and [13]). From the perspective of RS design, only those cues are of interest which can be manipulated by the RS provider. For example, the ease of using a system is an important signal of trustworthiness [4]. Furthermore, the interface can be used to induce a feeling of sociability (social presence), which increases trust in RS [24].

The approach to increase inferred trustworthiness to gain user acceptance comes with the advantage that it does not require the user to perform a rational, knowledge-based evaluation of the RS. Thus, scarce resources of cognitively limited customers are freed up for more valuable purposes.

5 Conclusion

The underlying paper deals with the issue of trust in RS. Based on a broad literature review in social sciences, IS and especially RS, we tried to illustrate the forces that influence the perception of RS as trustworthy. Therefore we discussed different conceptualizations of trust and trustworthiness and defined a common understanding. We argued that RS in e-commerce are often perceived as black boxes by customers and that there are two basic forms of trustworthiness, knowledge-based trustworthiness and inferred trustworthiness, that can either transform the perception of the system towards a trustworthy white box or a trustworthy black box, respectively. Additionally, we proposed a graphical model to depict the perception of trustworthiness of RS and highlighted several practical implications. Designers in the practical field are recommended to build on our findings and implement RS according to the proposed strategies to convey a perception of trustworthiness to their customers.

Acknowledgments. This research has been funded by the Austrian Science Fund (FWF): project number TRP 111-G11.

References

1. Maida, M., Maier, K., Obwegeser, N., Stix, V.: Explaining mcdm acceptance: a conceptual model of influencing factors. In: 2011 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 297–303. IEEE (2011)
2. Komiak, S., Benbasat, I.: The effects of personalization and familiarity on trust and adoption of recommendation agents. Mis Quarterly, 941–960 (2006)
3. Gefen, D., Karahanna, E., Straub, D.: Trust and tam in online shopping: An integrated model. MIS Quarterly, 51–90 (2003)
4. Wang, W., Benbasat, I.: Trust in and adoption of online recommendation agents. Journal of the Association for Information Systems 6(3), 72–101 (2005)
5. Lewis, J., Weigert, A.: Trust as a social reality. Social Forces 63(4), 967–985 (1985)
6. Mayer, R., Davis, J., Schoorman, F.: An integrative model of organizational trust. Academy of Management Review, 709–734 (1995)

7. Kim, D., Ferrin, D., Rao, H.: A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. Decision Support Systems 44(2), 544–564 (2008)
8. Luhmann, N.: Trust and power. John Wiley and sons (1979)
9. Ansari, A., Essegaier, S., Kohli, R.: Internet recommendation systems. Journal of Marketing Research 37(3), 363–375 (2000)
10. Häubl, G., Trifts, V.: Consumer decision making in online shopping environments: The effects of interactive decision aids. Marketing Science, 4–21 (2000)
11. Sinha, R., Swearingen, K.: The role of transparency in recommender systems. In: CHI 2002 Extended Abstracts on Human Factors in Computing Systems, pp. 830–831. ACM (2002)
12. Harrison McKnight, D., Choudhury, V., Kacmar, C.: The impact of initial consumer trust on intentions to transact with a web site: a trust building model. The Journal of Strategic Information Systems 11(3-4), 297–323 (2002)
13. Wang, W., Benbasat, I.: Attributions of trust in decision support technologies: A study of recommendation agents for e-commerce. Journal of Management Information Systems 24(4), 249–273 (2008)
14. Nass, C., Steuer, J., Tauber, E.: Computers are social actors. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence, pp. 72–78. ACM (1994)
15. Rousseau, D., Sitkin, S., Burt, R., Camerer, C., et al.: Not so different after all: A cross-discipline view of trust. Academy of Management Review 23(3), 393–404 (1998)
16. Wang, W., Benbasat, I.: Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. Journal of Management Information Systems 23(4), 217–246 (2007)
17. Lee, J., See, K.: Trust in automation: Designing for appropriate reliance. Human Factors: The Journal of the Human Factors and Ergonomics Society 46(1), 50–80 (2004)
18. Ajzen, I.: The theory of planned behavior. Organizational Behavior and Human Decision Processes 50(2), 179–211 (1991)
19. McKnight, D., Cummings, L., Chervany, N.: Initial trust formation in new organizational relationships. Academy of Management Review, 473–490 (1998)
20. Herlocker, J., Konstan, J., Riedl, J.: Explaining collaborative filtering recommendations. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, pp. 241–250. ACM (2000)
21. Simon, H.: Designing organizations for an information-rich world. International Library of Critical Writings in Economics 70, 187–202 (1996)
22. Wei, K., Huang, J., Fu, S.: A survey of e-commerce recommender systems. In: International Conference on Service Systems and Service Management, pp. 1–5. IEEE (2007)
23. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17(6), 734–749 (2005)
24. Hess, T., Fuller, M., Campbell, D.: Designing interfaces with social presence: Using vividness and extraversion to create social recommendation agents. Journal of the Association for Information Systems 10(12), 889–919 (2009)

Author Index

Ahmad, Iftikhar 176
Armano, Giuliano 195
Auinger, Andreas 163

Baquero, Alegria 124
Ben Ticha, Sonia 203
Bhiri, Sami 50
Boyer, Anne 203
Bsaies, Khaled 203
Buettner, Ricardo 149
Burke, Robin 88

Cai, Shun 13
Cfarku, Denada 25
Chu, Kyounghee 13

Debenham, John 1
de Gemmis, Marco 112
Derguech, Wassim 50

Fuchs, Matthias 100

Gaaloul, Walid 63
Gao, Feng 50
Ge, Mouzhi 76
Giuliani, Alessandro 195
Gröning, Marian 76
Guerra, Francesco 38

Hanachi, Chihab 63
Haque, Rafiqul 25
Hochmeier, Alexander 163
Holzinger, Andreas 163

Jannach, Dietmar 76

Karpischek, Stephan 137

Landes, Jürgen 149
Lops, Pasquale 112

Maida, Martina 212
Maier, Konradin 212
Messina, Alberto 195
Michahelles, Florian 137
Mobasher, Bamshad 88
Mondi, Ravi 63
Montagnuolo, Maurizio 195
Musto, Cataldo 112

Narducci, Fedelucio 112
Nedbal, Dietmar 163

Obwegeser, Nikolaus 212
Olaru, Marius-Octavian 38

Papazoglou, Michael P. 25

Roussanaly, Azim 203

Santani, Darshan 137
Schmidt, Günter 176
Semeraro, Giovanni 112
Sierra, Carles 1
Stix, Volker 212

Taher, Yéhia 25
Taylor, Richard N. 124

van den Heuvel, Willem-Jan 25
Vargiu, Eloisa 195
Vincini, Maurizio 38
Vitvar, Tomas 50

Yih, Jih-Shyr 188
Yuan, Xina 13

Zanker, Markus 76, 100
Zaremba, Maciej 50
Zheng, Yong 88