19

Fuzzy Conceptual Data Analysis Applied to Knowledge Management

Carmen De Maio, Giuseppe Fenza, Vincenzo Loia, and Saverio Salerno

Abstract. Conceptual data analysis has been extensively exploited to support Ontology Learning, Information Retrieval, and so on. This work emphasizes the relevant role of uncertainty in the conceptual data analysis. Specifically, Fuzzy Conceptual Data Analysis has been exploited to address two Enterprise Knowledge Management methodologies: domain ontology learning and ontology merging.

19.1 Introduction and Related Works

Nowadays Semantic Web and Web 2.0 play a crucial role in the area of Knowledge Management. Last trend is definition of ontology-based Knowledge Management platforms [3], [4]. Ontology-based Knowledge Management platform raises new requirements in terms of preparation and maintenance of domain ontologies. Specifically, domain ontologies could provide a common access point to the linked data repository. So, there is a need to define methodologies able to support life cycle of large and heterogeneous knowledge bases. Nevertheless, Semantic Web tools are still under development and ontologies maintenance (e.g., preparation, update) requires a considerable effort.

This paper addresses these challenges defining methodologies for: domain ontology learning, taking into account User Generated Content (e.g., blog, wiki, etc.), and ontology merging, to update previously extracted domain ontologies. These methodologies leverage on Fuzzy extention of Conceptual Data Analysis. Conceptual Data Analysis mainly attains with Formal Concept Analysis (FCA) [8] algorithm. In particular, this paper argues that Fuzzy Theory [7] enable us to generalise Conceptual Data Analysis introducing uncertainty management with applied to FCA (FFCA) and data analysis techniques providing support to knowledge extraction and structuring. Conceptual Data Analysis has been extensively applied to Information Retrieval because browsing lattice according to the user's query enables query augmentation and refinement [13], [14]. Specifically, (Fuzzy) Conceptual Data Analysis has been applied also for Ontology Learning (analysing text and extracting lattice with FCA) and Ontology Merging (mashing different lattices to infer new concept hierarchies).

In literature, there are many works for Ontology Learning that analises domain data by using text-mining and machine learning techniques, some of these approaches exploit FCA. Specifically, [10] introduces the L-fuzzy context, as an endeavour to combine fuzzy logic with FCA but it seems to be not practicable for

R. Seising et al. (Eds.): *On Fuzziness: Volume 1*, STUDFUZZ 298, pp. 121–127.
DOI: 10.1007/978-3-642-35641-4_19

dealing with large data sets because a human support is required to define the fuzzy linguistic variables. In [9] the Fuzzy extension of FCA theory is exploited to build hierarchical classification of the collected resources.

As for Ontology Merging, the idea of using FCA was first proposed in [5], where the FCAMerge algorithm is described. FCAMerge is a bottom-up approach to ontology merging guided by application-specific instances of the given source ontologies. Specifically, formal context has obtained analysing documents representing the two input ontologies. Instead, FCAOntMerge [11] following approach defined in [6] translates each input ontology into attributes (i.e., columns) and objects (i.e., rows) of a formal context.

On the light of described scenario, this work defines: methodology for Ontology Learning orchestrating Fuzzy C-Means (FCM) and FCA; and methodology for semi-automatic Ontology Merging extending with fuzziness FCAOntMerge approach.

The paper is organised as follows: Section 19.2 introduces theoretical background of Fuzzy Conceptual Data Analysis; Section 19.3 describe workflows of defined methodologies for Ontology Learning and Merging, then, Section 19.4 gives some experimental results of the defined methodologies. Finally conclusion close the paper.

19.2 Fuzzy Conceptual Data Analysis

This section provides most relevant notions of FFCA and FCM that are exploited to perform Fuzzy Conceptual Data Analysis. These algorithms will be orchestrated in Sections 19.3.1 and 19.3.2 to support Ontology Learning and Ontology Merging, respectively.

19.2.1 Fuzzy C-Means – FCM

FCM [1] clustering is an unsupervised process, based on c-partition [2]. It takes as input a data matrix and it tries to get an "optimal" partitioning of the feature space (composed by the data matrix). FCM aims at maximizing the homogeneity, grouping into the same cluster the patterns which are closer. Each pattern is a row of matrix. FCM recognizes spherical "clouds of points" (clusters of data) in a multi dimensional data space (i.e. data matrix) and each cluster is represented by its center point (prototype or centroid). The function minimizes the weighted sum of the distances between data points x and the centroid v, according to this formula:

$$V(U) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{i,j}^{m} \left\| \underline{x}_j - \underline{v}_i \right\|^2 \tag{19.1}$$

where $c \geq 2$ is the number of clusters, $u_{i,j} \in [0,1]$ is the membership degree of x_i in the i-th cluster and $m > 1$ controls the quantity of fuzziness in the classification process.

After the FCM execution, data partitions are returned, in a prior fixed number c of clusters.

19.2.2 Fuzzy Formal Concept Analysis – FFCA

FCA is a technique of data analysis, which exploits the ordered lattice theory. Recently, FCA and fuzzy techniques are integrated in order to deal with uncertain and vague information. In particular, this approach exploits a fuzzy extension of FCA. Fuzzy FCA (FFCA) [9] combines fuzzy logic into FCA representing the uncertainty through membership values in the range [0, 1]. Through formal contexts, FFCA enables the representation of the relationships between objects and attributes in a given domain.

Some definitions about main concepts of Formal Concept Analysis extracted from [9] and its fuzzy extension are given.

Definition 1: *A* **Fuzzy Formal Context** *is a triple $K = (G, M, I = \varphi(G \times M))$, where G is a set of objects, M is a set of attributes, and I is a fuzzy set on domain $G \times M$. Each relation $(g, m) \in I$ has a membership value $\mu(g, m)$ in [0, 1].*

Definition 2: **Fuzzy Formal Concept**. *Given a fuzzy formal context $K=(G, M, I)$ and a confidence threshold T, we define $A^{*}= \{m \in M \mid \forall\ g \in A\colon \mu(g, m) \geq T\}$ for $A \subseteq G$ and $B^{*}=\{g \in G \mid \forall\ m \in B\colon \mu(g, m) \geq T\}$ for $B \subseteq M$. A fuzzy formal concept (or fuzzy concept) of a fuzzy formal context K with a confidence threshold T is a pair $(A_f = \varphi(A), B)$, where $A \subseteq G$, $B \subseteq M$, $A^{*}=B$ and $B^{*}=A$. Each object $g \in \varphi(A)$ has a membership μ_g defined as*

$$\mu_g = min_{m \in B}\ \mu(g,m)$$

where $\mu(g,m)$ is the membership value between object g and attribute m, which is defined in I. Note that if $B=\{\ \}$ then $\mu_g = 1$ for every g. A and B are the extent and intent of the formal concept $(\varphi(A),B)$ respectively.

The Fuzzy FCA takes into account the fuzzy formal context and performs a hierarchical arrangement of fuzzy formal concepts, so obtains fuzzy concept lattice.

Definition 3: ***A Fuzzy Concept Lattice*** *of a fuzzy formal context K with a confidence threshold T is a set F(K) of all fuzzy concepts of K with the partial order $\leq$ with the confidence threshold T .*

The fuzzy lattice evidences the membership associated to the objects and the class-subclass relationship [9]. Thanks to FCA theory, the concepts are arranged in a hierarchy, emphasizing semantic relationships like subsumption (i.e., ”is-a”).

19.3 Methodologies for Enterprise Knowledge Management

Following sections describe application of Fuzzy Conceptual Data Analysis algorihms to support Ontology Learning and Ontology Merging.

19.3.1 Case Study: Ontology Learning

This methodology analyses structured and unstructured (e.g., bolg, wiki, etc.) resources which are daily produced by employees in the organisation in order to extract unsupervised hiearchical conceptualisation (e.g., topics, and so on). The workflow and mapping on the exploited technological solutions is shown in Fig.19.1. It involves following phases:

- *Natural Language Processing*, that relies on several activities, such as: language detection (i.e., Apache TIKA), multiformat analysis, stopword removal, stemming and lemmatisation (i.e., Snowball), PoS tagging (i.e. Language Tool) and terms disambiguation (i.e., Wikipedia Miner), and so on. Specifically, this step exploits Wikipedia as external knowledge resource in order to enrich keywords extraction results with wikification of the main portion of text;

- *Vectorization*, that carries out feature set and term weighting of input text by mainly applying well known technique of TF-IDF;

- *Fuzzy Conceptual Data Analysis*, applies FCM in order to prune incoming data, then resulting partition is given as input to FFCA algorithm. At the end of this phase concept hierarchies have been extracted;

- *Semantic Technology Mapping*, that represents the extracted unsupervised conceptualisation (and their relationships) in a schema compliant with SemanticWeb technologies (i.e., SKOS and RDFS).

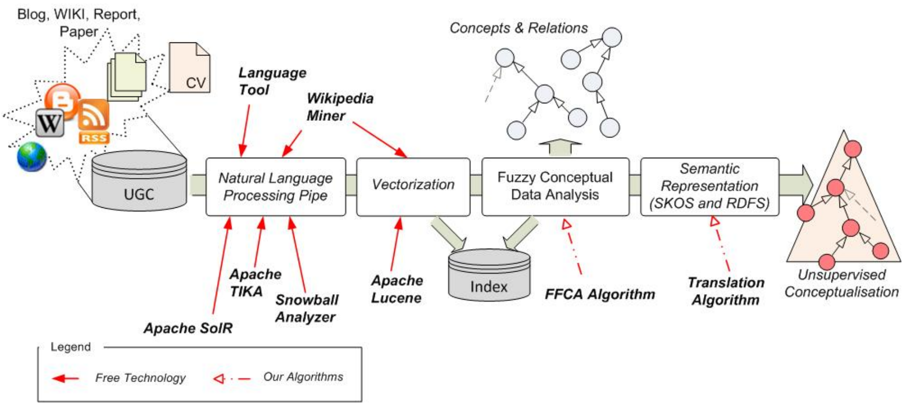

Fig. 19.1. Workflow of the Methdology for Ontology Learning

19.3.2 Case Study: Ontology Merging

This methodology is inspired to FCAOntMerge proposed in [11]. Specifically, this methodology is aimed to merge Unsupervised Conceptualisation and existing Domain Ontologies. The workflow and mapping on the exploited technological solutions is shown in Fig.19.2. It is composed of the following phases:

1. *Formal Context Creation*, input ontologies are transformed into two Formal Contexts. Each cell of the Formal Context represents instances-of relation (i.e., binary relation) between concept and individuals of input ontologies;

2. *Formal Contexts Merging*, this step exploit ontology matching results. Taking into account matching degree between different ontologies formal contexts will be merged in the union of them.

3. *Fuzzy Conceptual Data Analysis*, this step perform algorithm of FFCA on merged context in order to carry out a knowledge structure that integrates concepts of Unsupervised Conceptualisation and Domain Ontologies.

4. *Assessment and Semantic Technology Mapping*, interacting with expert user this step perform assessment in a semi-automatic manner of new knoweldge structure. After, the system translates knowledge structure into a SKOS and RDFS.

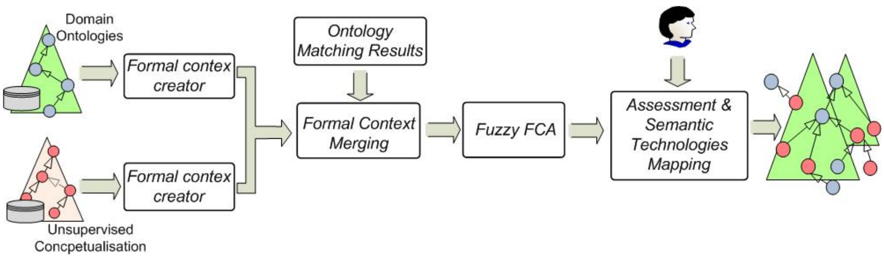

Fig. 19.2. Workflow of the Methdology for Ontology Merging

19.4 Experimental Results

In order to evaluate, the workflow of defined methodologies has been applied on subset of human classified repository of *Open Directory Project (ODP)*. Specifically, about 700 items of ODP have been analysed and classes of ODP have been used as *Domain Ontologies*(i.e., gold taxonomies). Let us note that only a brief text description of items has been exploited in the analysis process. The performances have been measured in terms of the micro-averaging of recall and precision [12]. The experimental results are shown in Fig.19.3.

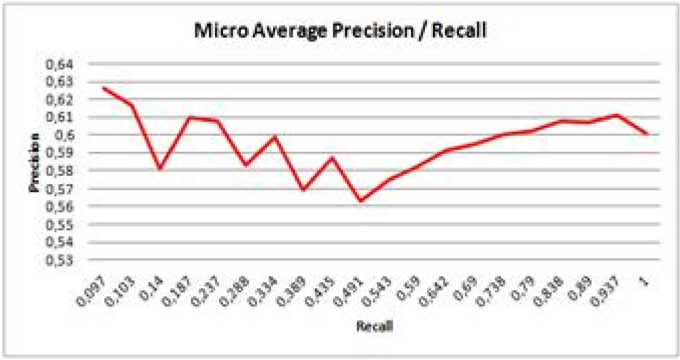

Fig. 19.3. Micro-averaging precision / recall results

From technological point of view, defined platform has been implemented exploiting existing software solutions, such as: Apache Solr, Apache Lucene, Wikipedia Miner, Sesame and OWLIM, and so on. Furthermore, services provided by the platform have been used in Ms Share Point 2010. Specifically, an existing connector framework (i.e., Manifold CF) has been instantiated in order to transparently acquire and up to date content indexes with data daily generated by the workers in Ms Share Point.

19.5 Conclusion

This contribution is aimed to describe the role of Fuzzy Conceptual Data Analysis in Knowledge Management. In particular, fuzziness enable us to reduce loosing of weak relations in the extracted unsupervised conceptualisation. Future works are aimed to decrease FFCA complexity for large data management exploiting incremental algorithms and emerging technologies (e.g., NoSQL DB).

Acknowledgement. This research is partially supported by the European Commission under the Collaborative Project ARISTOTELE "Personalised Learning & Collaborative Working Environments Fostering Social Creativity and Innovations Inside the Organisations", the Seventh Framework Programme (FP7), Theme ICT-2009.4.2 (Technology-Enhanced Learning), Grant Agreement n. 257886.

References

1. Bezdek, J.C.: Pattern Recognition and Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
2. Ruspini, E.: A new approach to clustering. Information and Control 15, 22–32 (1969)
3. Wang, Y., Guo, J., Hu, T., Wang, J.: An Ontology-Based Framework for Building Adaptable Knowledge Management Systems. In: Zhang, Z., Siekmann, J.H. (eds.) KSEM 2007. LNCS (LNAI), vol. 4798, pp. 655–660. Springer, Heidelberg (2007)

4. Buranarach, M., Supnithi, T., Chalortham, N., Khunthong, V., Varasai, P., Kawtrakul, A.: A Semantic Web Framework to Support Knowledge Management in Chronic Disease Healthcare. In: Sartori, F., Sicilia, M.Á., Manouselis, N. (eds.) MTSR 2009. CCIS, vol. 46, pp. 164–170. Springer, Heidelberg (2009)
5. Gerd, S., Maedche, A.: FCA-Merge: Bottom-Up Merging of Ontologies. In: IJCAI 2001 - Proceedings of the 17th International Joint Conference on Artificial Intelligent, Seattle, USA (2001)
6. Tao, G.: Using Formal Concept Analysis for Ontology Structuring and Building. ICIS, Nanyang Technological University (2003)
7. Zadeh, L.A.: Fuzzy Sets. Information and Control 8, 338–353 (1965)
8. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Heidelberg (1999)
9. De Maio, C., Fenza, G., Loia, V., Senatore, S.: Hierarchical Web Resources Retrieval by Exploiting Fuzzy Formal Concept Analysis. Information Processing & Management 48(3), 399–418 (2012)
10. Pollandt, S.: Fuzzy-Begriffe: Formale Begriffsanalyze unscharfer Daten. Springer, Heidelberg (1996)
11. Li, G.-Y., Liu, S.-P., Yan, Z.: Formal Concept Analysis Based Ontology Merging Method. In: 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), July 9-11, vol. 8, pp. 279–282 (2010)
12. Van Rijsbergen, C.J.: "Keith" Information Retrieval, Butterworth, London (1979)
13. Godin, R., Saunders, E., Jecsei, J.: Lattice Model of Browsable Data Spaces. Journal of Information Sciences 40, 89–116 (1986)
14. Carpineto, C., Romano, G.: Concept Data Analysis: Theory and Applications. John Wiley & Sons, Ltd., Atrium Southern Gate (2004)