Chapter 21
CV-Muzar - The Virtual Community Environment that Uses Multiagent Systems for Formation of Groups

Ana Carolina Bertoletti De Marchi and Márcia Cristina Moraes

Abstract The purpose of this chapter is to present two agents' societies responsible for group formation (sub-communities) in CV-Muzar (Augusto Ruschi Zoobotanical Museum Virtual Community of the University of Passo Fundo). These societies are integrated to execute a data mining classification process. The first society is a static society that intends preprocessing data, investigating the information about groups in the CV-Muzar. The second society is a dynamical society that will make a classification process by analyzing the existing groups and look for participants that have common subjects in order to constitute a sub-community. The formation of sub-communities is a new functionality within the CV-Muzar that intends to bring the participants together according to two scopes: interest similarity and knowledge complementarities.

21.1 Introduction

Over the last years, we were able to notice people's increasing interest in making use of the available resources on the Internet to improve their knowledge and interact with others. The virtual learning communities have proved to be suitable environments for this practice, because their participants are related to the construction of knowledge based on common goals. According to Pallof and Pratt [1] the virtual learning communities are dynamical components that emerge when a group of people shares certain practices, they are interdependent, make joint decisions, identify themselves with something larger than the total sum of their individual relationships and establish a long term commitment with the well being (theirs, others' and the

Prof. Dr. Ana Carolina Bertoletti De Marchi
Universidade de Passo Fundo, Faculty of Informatics, BR 285, Bairro So Jose, Zipcode: 99052-900, Passo Fundo, Brazil e-mail: carolina@upf.br

Prof. Dr. Marcia Cristina Moraes
Pontificia Universidade Catolica do Rio Grande do Sul, Faculty of Informatics, Av. Ipiranga 6681, Prdio 32, Zipcode: 90619-900, Porto Alegre, Brazil e-mail: marcia.moraes@pucrs.br

L. Cao (ed.), Data Mining and Multi-agent Integration, DOI: 10.1007/978-1-4419-0522-2_21,

group's in all their inter-relationships). The idea of groups formation inside a virtual learning community is interesting, because learning in groups aims to develop and to improve individual skills, to accept responsibilities for individual and group learning process, to develop abilities of reflecting about its own suppositions (expressing its ideas to the group) and to develop social and group abilities.

This chapter intends to present the groups formation within the CV-Muzar (Augusto Ruschi Zoobotanical Museum Virtual Community of the University of Passo Fundo). The groups are called sub-communities and are formed from two concepts: interest similarity and knowledge complementarities. In order to automate sub-communities construction we use predictive modeling, more specifically classification. Two agents' societies are used in this process. The first society is a static society that intends to investigate the information about groups in CV-Muzar. The second society is a dynamical society that will analyze the existing groups and look for participants that have common subjects in order to constitute a sub-community. So, the first society is responsible for preprocessing data using search algorithms to collect information for sub-communities establishment and the second society uses the preprocessing data to executes a classification technique that is based on the Dependence-Based Coalition Model, established on the Social Reasoning Mechanism and Contractual Network, based on Sichman's Economic Market Theory [2].

The chapter is organized in four sections. The second section presents the background related to museums and virtual learning communities and data mining as well as related works. The third section describes the group formation proposal for CV-Muzar. The fourth section presents some final considerations and future works.

21.2 Background and Related Work

21.2.1 Interactive Museums and Virtual Learning Communities

Among the proposals for the educational work in museums, Silva [3] suggests the use of new technologies in the study techniques of exhibitions, allowing new forms of interaction with experiments. The interactive museums are rich, attractive and motivating places that involve visitors in the process of scientific investigation. To improve this process, virtual communities are being used.

Rehingold [4] describes virtual communities as "social aggregations that emerge from the Net when enough people carry on public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace". Nowadays there are several kinds of virtual communities, such as: learning communities, practice communities and entertainment communities [5]. In this chapter we are going to work on virtual learning communities, that are communities which intends to promote an environment that favors knowledge construction, where the members of the community are related to common learning objectives.

Virtual learning communities can promote learning in two different ways: a formal one and an informal one. The formal way of learning considers that the community is going to be formed based on a real physical structure, where we have a teacher as a mediator of learners [1]. The informal way of learning considers that a community is going to be formed based on the personal interest of their members. These interests are going to define a net of self-organized relationships which have common goals that lead the members of a community into a continuous and permanent process of learning.

We approach in our work the informal way of learning in virtual communities. We share the idea of Souza [6] that learning occurs even apart of formal programs. One way to improve the informal learning process is to allow the members of a community to form groups or sub-communities. These groups will focus theirs discussions on specific subjects that are considered particularly interesting for their members. According to Vygotsky's [7] work, the use of groups is relevant because through discussions it is possible to have a knowledge consolidation and the findings of new solutions.

21.2.2 Data Mining

According to Han and Kamber [8], data mining refers to extracting or "mining" knowledge from large amounts of data. Data mining is an integral part of knowledge discovery in databases (KDD), which is the overall process of converting data into useful information. This process consists of a series of transformations steps, from data preprocessing to postprocessing of data mining results [9], as showed in Fig. 21.1. The input data can be stored in a variety of forms and can be located in a

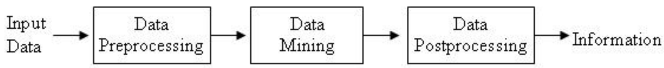

Fig. 21.1: The process of knowledge discovery in databases [9]

centralized repository or distributed in multiple sites. The objective of preprocessing is to transform the input data into an appropriate format to be used in data mining process. The postprocessing phase is responsible for integrating data mining results into a decision support system.

There are two kinds of data mining tasks: predictive and descriptive. The purpose of predictive task is to predict the value of a particular attribute based on the values of other attributes, one sample is predictive modeling. Predictive modeling intends to build a model for a target variable as a function of explanatory variables. There are two kinds of predictive modeling: classification (used for discrete target variables) and regression (used for continuous target variables). The purpose of the descriptive task is to derive patterns that summarize the underlying relationships in data [9], samples of this kind of task are: association analysis, cluster analysis and anomaly detection. Association analysis is used to discover patterns that describe associated features in the data. The discovered patterns are usually represented by implication rules. Cluster analysis aims to find groups that have closely related observations. Anomaly detection aims to identify observations whose characteristics are significantly different from the rest of the data.

In this work we use predictive modeling, more specifically classification. Classification is the task of learning a target function f that maps each attribute set x to one of the predefined class labels y. The target function is also known as a classification model. A classification technique (or classifier) is a systematic approach to building classification models from an input data set. Samples of classifiers include decision tree, rule-based classifiers and neural networks. Each technique employs a learning algorithm to identify a model that best fits the relationship between the attribute set and class label of the input data. Our classification technique is implemented through a multiagent system that uses an algorithm which implements Dependence-Based Coalition Model [2]. This algorithm is going to be presented in detail in section Sub-Communities Formations Assisted by a Multiagent System.

21.2.3 Related Works on Data Mining and Virtual Community

Silva et al [10] present distributed data minig algorithms focus on one class of distributed problem solving task executed by multiagent systems, analysis and modeling of distributed data. Gorodetskiy et al [11] describes an agent-mediated protocol based collaboration of distributed designers performing learning of agents of multiagent distributed classification system. Lina and Hsuehb [12] propose a knowledge map management system (based on information retrieval and data mining) to facilitate knowledge management in virtual communities of practice.

21.3 CV-Muzar (Augusto Ruschi Zoobotanical Museum Virtual Community of the University of Passo Fundo)

The CV-Muzar (http://inf.upf.br/comunidade) was developed with the main purpose of involving the museum visitors more and make them part of the experience, putting an end to the passive receiver of the expositive speech that was established unilaterally [13]. For the environment development we made use of the virtual communities' concepts to promote the exchanges among the visitors and the Learning Objects (LOs) that comprise materials developed for the experiments, materials kept in the Museum and users' productions. In order to promote the exchanges among visitors and LOs, CV-Muzar is organized into two main modules: the repository manager of teaching resources (that comprises all the LOs available) and the learning environment support (that is responsible for organizing the didactic activities). Inside the learning environment support is the sub-module of sub-communities formation.

21.3.1 Sub-Communities Formations Assisted by a Multiagent System

The sub-community term represents the formation of small groups within the CV-Muzar. In these groups will take place discussions about subjects of common interest among the participants. A sub-community can be created by any participant previously registered in the environment and its formation (constitution of its components) occurs in light of two needs: interest similarity (group formed by participants that have similar profiles) and knowledge complementarities (group formed by participants, that are gathered to accomplish complex tasks which require the composition of abilities for problems solving).

The sub-communities formation is undertaken through a multiagent system. A multiagent systems (MAS) is a society formed by agents that coexist in the same environment and interact in order to accomplish a common goal [14]. In this work the multiagent system is composed by two societies: static (aims to investigate the information about groups), and dynamical (aims to analyze the existing groups and try to look for participants that have similar subjects to participate). Both societies are used in a data mining classification method. The static society, named Investigating Society of Sub-Community (SIS-C) is responsible for doing the preprocessing phase, preparing data do be used during the data mining process. The dynamical society, named Investigating Society of Participants (SIP) is responsible for data mining, executing a classification method that implements an algorithm based on the Dependence-Based Coalition Model, found on the Social Reasoning Mechanism and Contractual Network, based on Sichman's Economic Market Theory [2].

The next sections are going to explain how Investigating Society of Sub-Community (SIS-C) and Investigating Society of Participants (SIP) works.

21.3.1.1 Investigating Society of Sub-Community (SIS-C)

The Investigating Society of Sub-Community (SIS-C) is characterized as a kind of static organization, because the roles that each agent will play within the society are already pre-defined. The roles that an agent can execute inside this society are a service provider agent and a leader agent. The service provider agent is the one responsible to provide service to others. The leader agent is responsible to find out which services agents can fulfill the necessary requirements to execute a task. One of the society agents is defined as the leader agent, regardless of his knowledge. The first task that will be accomplished by the leader agent in SIS-C is the search for the group profile. This search is composed by the following steps: to verify all the information provided by the group coordinator (this information includes objectives, keywords, area of interest and communication tools used by the group); to verify if the proposed profile is not similar to any other existing profile and to verify the group's central theme, if it is in accordance with the environment central idea. If all the listed requirements above are in accordance, the other activities are carried out; otherwise, the leader agent sends a message to the group coordinator advising him about the items that must be reviewed. The other activities are the search within the concentration area for the sub-communities and the search within the interest area for the possible participants. In order to do these tasks, the leader agent has to distribute tasks among service provider agents. He does that establishing a communication with a list of services provider agents that could have the capability to fulfill the requirements demanded by a specific task.

The option to use a leader agent enables the interoperability among the heterogeneous agents that are part of the society. After the communication cycle among the SIS-C society agents is over, the obtained information is stored in a database for a possible migration of some static society agents to the dynamical society. Thus, if it is necessary to migration, the coalition formation can occur for the search of participants that have profile similar to the group.

With the first part of information about the stored groups in a database, it is necessary to obtain information about the participants in order to accomplish the formation of the sub-communities. This task is carried out by the Investigating Society of Participants (SIP).

21.3.1.2 Investigating Society of Participants (SIP)

The Investigating Society of Participants is characterized as the dynamical type, because in this kind of organization there is the need of social interaction. The option to use the Dependence-Based Coalition Model (DBC) for the dynamical organization is due to the fact that if the agent that integrates the society does not have the autonomy to carry out a certain activity, another member who can help is required. So, over the time the agents improve their knowledge about other agents.

The formation of coalitions on the DBC model occurs in the following way: (the steps of the model are written considering an example of a procedure that occurs in the society):

1. Choice of a goal: an agent Ag1 chooses a certain goal to be achieved. In case there is no longer a goal, agent Ag1 does not try to form coalitions anymore. The choice of a goal can be the search for participants that have interest in discussing issues about "Environment Pollution". The goal is always chosen based on the formed groups.
2. Choice of a plan: supposing that Ag1 chose the G1 goal, the next step is to choose a plan to accomplish it. As the agent can have more than one plan for the same goal, the choice of the plan is based on the notion of feasible plan. In case there are no more plans, step one is resumed. Based on the participants' profile, agent Ag1 can have several plans for this objective and this way he chooses one that can be used. If Ag1 finds the plan worthwhile, he executes the analysis of the plan actions.
3. Analysis of the plan actions: once a plan is chosen, Ag1 analyses its objective situations concerning G1, in case the situation is independent or dependent. If the situation is independent, Ag1 is considered independent to accomplish that

objective and this way does not need cooperation from any other agents. In the dependence situation, however, Ag1 cannot initiate the execution of his plan immediately, because he first needs to find an agent that accomplishes the action he does not know how to execute.

4. Choice of the partner: through the social resolution mechanism, Ag1 considers his relationships and dependence situations with the other agents related to G1 and through the pre-established criterion, Ag1 chooses the best possible partners. In case there are no possible partners for the actual action, Ag1 chooses a new plan to achieve G1 returning to step 2.
5. Coalition formation between the agents: once the best possible partner is chosen, here called Ag2, Ag1 will send it a coalition proposal, which can contain the following proposals:

   - Ag2 accepts the proposal and the coalition is formed. From this moment on, the works to solve G1 are started. At the end of this process, if the actions were accomplished correctly, G1 is considered concluded and an invitation is sent to the participants that have a profile similar to the group's to participate; and Ag1 can return to step 1;

   - Ag2 refuses the proposal and in this case Ag1 tries respectively to find another partner, returning to step 4. The proposal refusal by Ag2 can occur through the following factors:
     – Ag1 misunderstood Ag2, probably for having incorrect or incomplete information about Ag2. In this case, Ag2 informs such information to Ag1, and Ag1 can review his opinion about Ag2.
     – Ag2 did not find the proposal interesting for his goals.

The MAS uses the rules previously described to executes a classification method to look for sub-community participants.

Fig. 21.2 shows the algorithm that will calculate the total number of messages in a society with n agents to establish presentation communication and search for a partner, for a total number of cycles g. The algorithm is based on a previous analysis of the process within the CV-Muzar. At each cycle, all the communication between the agents takes place through the messages exchanges. The active agent sends messages of coalition proposal until he finds a partner or until there are no possible partners. The possible partner always responds to the coalition proposal sending a message of acceptance or review. When the active agent gets an acceptance message, he sends a coalition message establishing the agreement with the partner agent. If no partner is found the coalition message is not sent. Thus, considering a society with n agents, where m agents can accomplish the desired action, and coalition proposal messages are sent to k agents (means that k - 1 agents sent messages of refusal or review), the total number of sent messages in each cycle is: Scycle = 2m in case it didn't find any partner; Scycle = 2k + 1, where $0 > k \leq m$ and; Scycle = 0 if the agent is independent.

Now, considering a g cycles competition, the total number of exchanged messages between the agents after all the accomplished cycles (SDBC) is showed in

```
depint(n,g)
Sap = n * (n-1);        //calculates agents presentation
For i = 0 until i < g
    if Active.autonomous then start a new cycle;
    else
      m = Active.searchPartner(plan);
      //search and calculates the total os possible partners
      k = 0;
      //initiates the amount of proposals carried through
      findPartner = false;
      while (k <= m) or (not findPartner) do
            k = k + 1;
            If probably partner accepts proposal of collation
             formation then
               findPartner = true;
               k = k + 1;
       end while
       Sc = Sc + k; //total of messages during coalition
                    //formation
       end else
       Stot = Sap + Sc; //total of exchanges messages
end for
```

Fig. 21.2: Algorithm that calculates the total number of messages in a society

Fig. 21.3. where: Scycle = 2m in case it didn't find any partner; Scycle = 2k + 1, where $0 > k \leq m$; Scycle = 0, if the agent is independent.

These societies were implemented using the platform JADE and they were integrated to CV-Muzar, which is implemented in PHP. After the societies SIS-C and SIP execute, the system sends an invitation e-mail to the community participants that were chosen by the society to participate in a sub-community. The participants can accept the invitation or not.

$$S_{DBC} = S_{presentation} + \sum_{i=1}^{g} S_{cycle} = n\,(n-1) + \sum_{i=1}^{g} S_{cycle},$$

Fig. 21.3: Total number of exchanged messages between the agents after all the accomplished cycles (SDBC)

21.3.2 Initial Evaluation

In order to evaluate the group formation functionality, we have conduced two kinds of evaluation, a user's evaluation and a usability evaluation. For the user's evaluation, an experiment in the formation module of sub-communities was carried out. Fifteen people related to the museum were invited. The participants were divided into two different groups. The first group received a small description of how they should fill out the individual profile and the subject nominations for the groups formation. The second group didn't have any help to filling out the individual profile as well as for creating the groups. Over the two weeks' tests, the participants were invited, through messages sent by e-mail, to take part in the sub-communities created by the two groups. In all simulations carried out, the MAS (considering the data mining process) nominated correctly the sub-communities related to the participant's profile. However, it will be necessary to optimize the processing time of the information exchange between agents on the Dependence-Based Coalition Model, because it took a long time for sending the invitations.

The tool used for the usability evaluation was Ergolist. This tool provides a way to evaluate the facility of use of interactive software considering ergonomics aspects [15]. The tool is composed of 18 checklists. Each one of these checklists comprises from three to twenty-seven subjects. The usability inspection has been completed for all environment. The data obtained through the inspection tests shows the percentage of conformity, or not, with certain characteristics in the system. Most of the applicable approaches obtained percentage of conformity higher than 50%. The best categories are: Explicit Actions (100%), Minimum Actions (100%), Consistency (90%), Density of Information (88,8%) and Grouping for Location (88,8%). In contrast, Users Experience (25%) and Flexibility (33,3%) are the worst categories. This occurs due to certain characteristics are not present in CV-Muzar such as: existence of dialogues with the user's abilities and various forms of presenting the same information to different types of user.

21.4 Final Considerations and Future Work

The main contribution of this work is to present a multiagent system that uses data mining classification implemented using dependence-based coalition model in order to automate the sub-communities construction. The integration of multiaget systems and Dependence-Based Coalition Model within a museum virtual learning community is a new idea.

So, in this chapter we presented the background related to the proposed sub-communities formation sub-module. We have described how data mining classification is executed by the two agents' society (SIS-C) and SIP executing an agent-mining interaction that involves users' preferences and human intelligence. In order to evaluate the proposal we made two initial evaluations and users' evaluation and a usability evaluation.

To examine carefully the data mining generated by the MAS, considering the Dependence-Based Coalition model (DBC) and to obtain a clear analysis of the exchange flow when the dependence-based coalition process occurs, studies are being carried out to analyze the time that the society takes to process these data and send the invitation message to the participants. In this way, we choose to work with the Exchange Values theme, considering that the nature of the social relations depends, mostly, on the proper representation of the norms and social conventions. We are analyzing the work proposed by Dimuro and Costa [16].

Thanks to: Conselho Nacional de Desenvolvimento Científico e Tecnológico for the support through the edictal MCT/CNPq 15/2007 - Universal.

References

1. R. M. Palloff, K. Pratt. Construindo Comunidades de Aprendizagem no Ciberespao: Estrategias eficientes para salas de aula on-line, Artmed, Porto Alegre, 2002.
2. J. S. Sichman. Du Raisonnement Social chez les agents: Une Approche fondee sur la thorie de la dependence. Thse de doctorat de 1'NPG, Grenoble, Franca, 1995.
3. M. P. Silva. Ao Pedagógica: Uma questo a ser (re)pensada nos Museus de Arte. Porto Alegre: UFRGS, 2003. Master Thesis - Programa de Pos-Graduacão em Educacão, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.
4. H. Rheingold. The Virtual Community: Homesteading at the Electronic Frontier, [1993]. http://www.rheingold.com/vc/book, Dec. 2008.
5. L. Palazzo et al. Comunidades virtuais de formaão tecnológica: fundamentaão pedagógica e metodologia de construo. In: XIII Simpósio Brasileiro de Informática na Educacão, 2002, UNISINOS.
6. R. R. Souza. Aprendizagem Colaborativa em Comunidades Virtuais, Florianópolis: UFSC, 2000. Master Thesis - Universidade Federal de Santa Catarina, Florianópolis, 2000.
7. S. L. Vygotsky. The Collected Works of L.S.Vygotsky. v. 1. New York: Plenum Press, 1987.
8. J. Han and M. Kamber. Data Mining: Concepts and Techniques. San Diego: Morgan Kaufmann Publishers, 2001.
9. P-N. Tan et al. Introduction to Data Mining. Boston: Addison Wesley, 2006.
10. J. C. Silva et al. Distributed data mining and agents. In: Engineering Applications of Artificial Intelligence Volume 18, Issue 7, October 2005, Pages 791-807.
11. V. Gorodetskiy et al. Infrastructural Issues for Agent-Based Distributed Learning. In: Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology, 2006, Pages 3-6.
12. F. Lina and C. Hsuehb. Knowledge map creation and maintenance for virtual communities of practice.In: Information Processing & Management Volume 42, Issue 2, March 2006, Pages 551-568.
13. A. C. B., De Marchi. Um ambiente de suporte a comunidades virtuais baseados em repositório de objetos de aprendizagem informal em museus. 2005. Thesis, PGIE, UFRGS. Porto Alegre.
14. M. Wooldridge. Introduction to MuliAgent Systems. Wiley, 2002.
15. Ergolist Project (2000). Available at: http://www.labiutil.inf.ufsc.br/ergolist. Access in: 08/20/2008.
16. G. P. Dimuro, A. C. R. Costa. Exchange Values and Self-Regulation of Exchanges in Multi-Agent Systems: the provisory, centralized model. In: Engineering Self-Organising Systems: Revised Selected Papers of the Third International Workshop, ESOA 2005, Utrecht, The Netherland s, July 25, 2005 (Lecture Notes in Artificial Intellligence). Berlin: Springer, 2006, v. 3910, Pages 75-89.