MULTI-AGENT SYSTEMS IN DATA IMPUTATION OF COLLABORATIVE FILTERING
In Case of e-WeddingThailand

Kunyanuth Kularbphettong
Science and Technology Faculty, Suan Sunandha Rajabhat University, Dusit Bangkok, Thailand

Phayung Meesad, Gareth Clayton
Information Technology Faculty, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

Keywords: Collaborative filtering, Multi-agent system, e-WeddingThailand, Web services, Ontology, Data imputation, Data preprocessing techniques.

Abstract: Multi-agent system is rapidly emerging as a new paradigm to develop complex and intelligent commerce application systems in e-Business. In this paper, we present the findings on the techniques used for data imputation techniques in Collaborative filtering based on multi-agent systems (MAS) of the on-going project, e-WeddingThailand. The aim of our project is to implement MAS combined with various techniques, like Web Services, Ontology, Web Semantic and Data Mining techniques. However, the present paper focuses on the data imputation technique in collaborative filtering utilized in order to treat missing values of customer behavioral patterns for Wedding business. As a result, a model obtained is therefore used as a benchmark for testing potential patterns so that they are used to strengthen the derived model in enhancing the overall system performance.

1 INTRODUCTION

Collaborative Filtering is a well known technique used in recommendation systems to suggest products and services for customers on e-Commerce systems because it is highly rated by some other customers with similar tastes. This technique generally analyzes relationships between users and products or services to identify the user product/service associations (Yifan, et al., 2008). However, the collaborative filtering technique has some disadvantages, like the inaccuracy of prediction and the lack of the transparency in the predictions (Janusz, 2006), when the number of similar users is small and the collected data from customers contain many missing values.

Generally, missing values can generate a bias impacting on the quality of the performance of data mining/machine learning techniques. The more incorrect and irrelevant data presented, the more unreliable the results. Especially in collaborative filtering technique, missing data impacts on the difficulties in estimation and inference (Schafer, 1997). Therefore, Missing Data Imputation is one of the challenging issues that impacts on improving the performance of Data Mining and Machine Learning approaches.

Also, with the growth and success of the Internet, there are tremendous numbers of customers using recommendation systems to rapidly perform adequate and reliable results in order to meet the demands of customers. With widely developed advanced mechanisms, Multi-agent System is one of the promising approaches utilized to enhance the performance of the recommendation systems. A multi agent system is a computational system, or a loosely coupled network in which two or more agents interact or work together to perform a set of tasks or to satisfy a set of goals. Each agent is considered as a locus of a problem-solving activity which operates asynchronously with respect to the other agents (Sandholm and Lesser, 1996). Although there has been an enormous amount of research done recently applied MAS in a wide range of problem in Data Mining and Machine Learning techniques,

Kularbphettong K., Meesad P. and Clayton G..
MULTI-AGENT SYSTEMS IN DATA IMPUTATION OF COLLABORATIVE FILTERING - In Case of e-WeddingThailand .
DOI: 10.5220/0003186304570461
In *Proceedings of the 3rd International Conference on Agents and Artificial Intelligence* (ICAART-2011), pages 457-461
ISBN: 978-989-8425-41-6

there is little of research focused on using MAS in the data preprocessing step. Therefore, in this paper we propose and implement the MAS framework of an on-going project, called e-WeddingThailand, which merely emphasizes using MAS to handle the problems of missing values in Collaborative filtering techniques.

The remainder of this paper is organized as follows. Section 2 reviews related literature and research works. Section 3 presents the experimental results based on the purposed framework based on multi-agent systems. This prototype demonstrates how to succeed in adapting multi-agent systems in data imputation in collaborative filtering techniques. Finally, we conclude the paper with future research issues in section 4g.

2 RELATED WORKS

A literature search shows that most of the related research has been deployed multi-agent to develop e-Commerce in collaborative filtering techniques and the data imputation techniques by following this:

According to Wei et al, 2003, they present a market –based recommendation system based on MAS where an agent acts on behalf of its users and sells the recommended slide bar space. Also, Gulden Uchyigit and Keith Clark, 2003, applied an agent based approach to collaborative filtering and used a process of pre-clustering to form shared interest based on their interest profiles. MANET (Collins et al, 2002) a mobile agent based system, enables users to compare products from different sellers. However, it has a problem of system performance when the number of sellers increases. Furthermore, according to Zili Zhang and research group, 2003, they applied agent technology to collect and integrate data distributed over various computing platforms to facilitate statistical data analysis in replacing the missing values by using either Approximate Bayesian Bootstrap (ABB) or Ratio Imputation and using MAS improves the execution time by focusing on spatial knowledge in order to extract knowledge in Predictive Modeling Markup Language (PMML) format (Baazaou et al, 2005). Moreover, other research works show that agent technologies representing in various ways that are related with data mining techniques (Huang et al, 2007), (Zili et al, 2003) and (Chien et al, 2009). Hence, from previous literature, it presents that there are many research studies exploiting multi-agent technology in collaborative filtering and data mining techniques blended together. Consequently, in order to succeed on their frameworks, agents should have abilities to perform as a behalf of user to handle with given tasks such as planning, reasoning and learning. Also, data mining techniques are an important way to make a reason for agent under uncertainty and in incomplete information situations. Notwithstanding, data imputation technique acts as a crucial task in enhancing performance of the system when mining incomplete data.

3 THE PROPOSED FRAMEWORK AND EXPERIMENTAL RESULTS

This section displayed the experimental results in data imputation of collaborative filtering of the e-WeddingThailand project and also compares the results between the original data set and the imputed data set.

3.1 The Proposed Framework

For illustration of framework as figure 1 (Kobkul, 2007), and (Kobkul et al, 2010), we select the wedding businesses and its environment. Nowadays, the wedding business is increasingly becoming growth. Booming wedding businesses affect other businesses like Hotels, wedding studios, car hiring companies, flower shops, music shops, travel agencies, and even media businesses. Due to the explosion of internet, couples use the wedding portal sites as a medium to search for needed information. However, it takes timing because of the overwhelming information available on the internet. Couples must take time to looking for wedding packages and related services, and rely on information of wedding business agencies. Further, couples make their decision based on comparing various wedding packages. The wedding package is composed of hotel wedding package, wedding studio package, music, floral decoration, card & gift agencies and etc.

Also, the proposed architecture of the e-WeddingThailand, this framework is concerned about four main aspects: multi-agent system, Web services, ontology and data mining techniques. In the multi-agent system, each agent is autonomous to be able to make decisions and act proactively. Agents can communicate, exchange knowledge, collaborate or negotiate with each other, to efficiently achieve the common goal. They receive

and process users' requirements, make use of user's preference, perform as a mediator of services business. They select and contact with appropriate wedding business agencies like hotels, wedding studio, and so on through Web services and give couples the optimal result. Web services define, provide related services, and interact with negotiator agents and ontology is the meaningful data which can be directly accessed by agents or people through the Web.

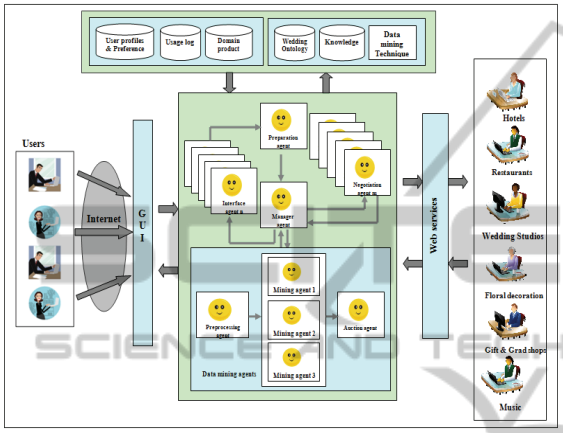

Figure 1: The purposed architecture of the e-WeddingThailand system.

Figure 2: The rating score web page in e-WeddingThailand System.

Figure 2 shows the rating score web page of e-WeddingThailand and figure 3 presents the operation of JADE in this project. However, in this paper we present merely a multi-agent system dealing with data imputation of collaborative filtering techniques.

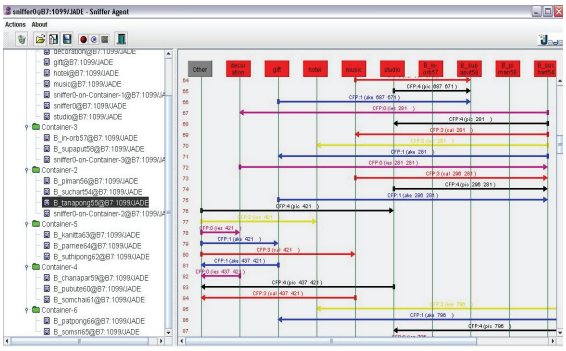

Figure 3: The operation of JADE in e-WeddingThailand System.

3.2 The Experimental Result

We worked on the expected customer data set from this project where users entry and rated the relevant wedding products and services and we were able to collect data on about 315 users. All data was collected in accordance with appropriate end user agreements and privacy policies. The analysis was done with fully anonymous data. It means no personally identifiable information collected in connection with this research. We collected some relevant wedding products and services. The training data contains R_{ui} values, for each user u and item i, which represent how user u rated, item i. In addition, we use a similarly constructed test set. Our system is trained in order to generate predictions of what users will prefer the related wedding products and services and we evaluate the quality of the different algorithms for computing the predictions using the mean absolute error (MAE). It is commonly used to measure the performance of prediction.

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N} \quad (1)$$

where p_i is the user-specified score
and q_i is the calculated score from the system

In our experiment, we filled missing data in the rating table by implementing two different algorithms KNN Imputation and the composite imputation method between hot deck and nearest neighbour methods based on mean substitution as described in section 3 and we tested them on our data sets.

Table 1: MAE Performance of CF without imputed and CF algorithms with KNN Imputations.

	MSE of CF without Imputed	MAE of CF with KNN Imputation				
		n=10	n=20	n=30	n=40	n=50
D_1	0.768	0.763	0.761	0.758	0.755	0.753
D_2	0.776	0.773	0.771	0.769	0.763	0.759
D_3	0.753	0.751	0.749	0.743	0.741	0.738
D_4	0.759	0.754	0.751	0.748	0.744	0.741
D_5	0.761	0.759	0.756	0.753	0.749	0.747

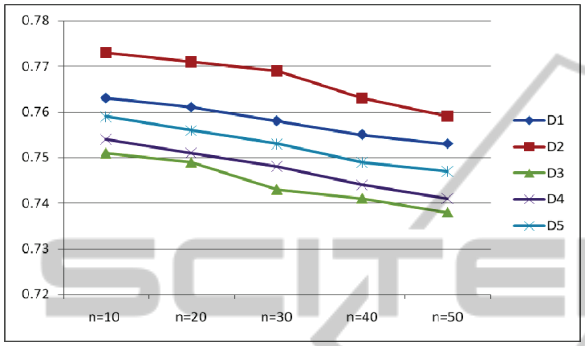

Figure 4: MAE Performance of CF algorithms with KNN Imputations.

Table 2: MAE Performance of CF without imputed and CF algorithms with the composite imputation method between hot deck and nearest neighbour methods.

	MAE of CF without Imputed	MAE of CF with the composite imputation method between hot deck and nearest neighbor methods				
		n=10	n=20	n=30	n=40	n=50
D_1	0.768	0.761	0.758	0.753	0.751	0.747
D_2	0.776	0.767	0.763	0.759	0.753	0.749
D_3	0.753	0.750	0.747	0.744	0.741	0.738
D_4	0.759	0.755	0.751	0.747	0.742	0.739
D_5	0.761	0.758	0.755	0.751	0.748	0.745

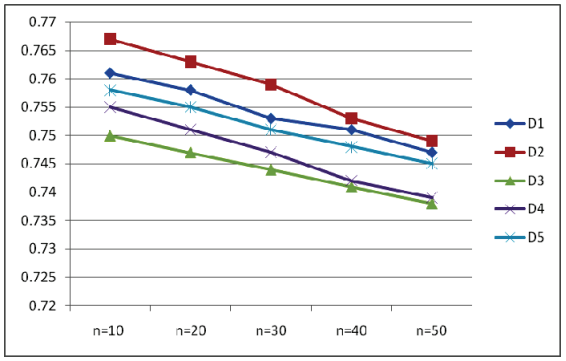

Figure 5: MAE Performance of CF algorithms with the composite imputation method between hot deck and nearest neighbour methods.

After computing for each imputation algorithm, we performed them with the collaborative filtering algorithm with the neighbourhood and used weight to generate the prediction and tested these experiments with MAE. Table 1 and 2 show the experimental results. There are five data sets used in this experimental in which 80 % of data set was of the training and 20% was of the tested set. And figure 4 and 5 present the MAE performance of CF algorithm with KNN Imputations and with the composite imputation method between hot deck and nearest neighbour methods.

Nevertheless, from the experiment, we found that the value of MAE performance of CF algorithms with the composite imputation method between hot deck and nearest neighbour methods was lower than the value of MAE with other approaches. It is clear that using imputation approaches helps to reduce the error of the prediction results. However, it takes time consuming to run the experiment which is shown in figure 6.

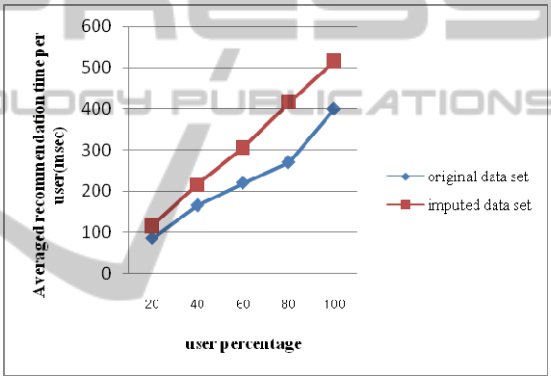

Figure 6: Recommendation Time to user.

4 CONCLUSIONS

In this paper we presented our preliminary ideas of building multi-agent systems in data imputation of collaborative filtering based on e-WeddingThailand system. In the part of MAS, we have implemented this prototype by using JADE platform. JADE is quite easy to learn and use. Moreover, it supports many agent approaches such as agent communication, protocol, behaviour and ontology. As for the future work, we need to explore more reasonable data mining technologies.

ACKNOWLEDGEMENTS

The authors would like to thank Suan Sunandha Rajabhat University for scholarship support this project.

REFERENCES

Chien-Ho Wu, Yuehjen E. Shao, Jeng-Fu Liu, and Tsair-Yuan Chang, 2009. "On Supporting Cross-Platform Statistical Data Analysis Using JADE", Book Series Studies in Computational Intelligence, Springer Berlin Heidelberg, *issn1860-949X (Print) 1860-9503 (Online)*, Volume 214/2009.

Gulden Uchyigit and Keith Clark, 2003. "A multi-agent Architecture for Dynamic Collaborative Filtering", Proceedings of the 5th International Conference on Enterprise Information Systems, Angers, France, April 22-26, pp363-368.

Huang Xin Li; Chosler, R., 2007. "Application of Multilayered Multi-Agent Data Mining Architecture to Bank Domain", Wireless Communications, Networking and Mobile Computing .International Conferenceon Volume, Issue , 21-25 Sept. pp 6721 – 6724.

H. Baazaoui Zghal, S. Faiz, and H. Ben Ghezala, 2005. "A Framework for Data Mining Based Multi-Agent: An Application to Spatial Data", World Academy of Science, Engineering and Technology.

Janusz Sobecki, 2006. "Implementations of Web-Based Recommender Systems Using Hybrid Methods", International Journal of Computer Science & Applications, 2006 Vol.3 issue 3 pp 52-64.

J. Collins, W. Ketter and M. Gini, 2002. *A Multi-agent Negotiation Tested for Contracting Tasks With Temporal and Precedence Constraints*, Int'l Journal of Electronic Commerce, pp 35-57.

Kobkul (Kunyanuth) Kularbphettong, 2007. "*e-Negotiation based on Multi-agent ssyetm*",JCSSE 2007 –The international Joint Conference on Computer Science and software Engineer,Thailand.

Kobkul (Kunyanuth) Kularbphettong, Gareth Clayton, and Phayung Meesad, 2010. "*e-Wedding System based on Multi-System*", Advances in Intelligent and Soft-Computing ,series of Springer.

Kobkul (Kunyanuth) Kularbphettong, Gareth Clayton, and Phayung Meesad, 2010. "*A Hybrid System based on Multi-agent System in Data Preprocessing stage: e-Wedding*", International Journal of Computer Science and Information Security.

Sandholm, T. and Lesser, 1996. "*Advantages of a Leveled Commitment Contracting Protocol*." Thirteenth National Conference on Artificial Intelligence (AAAI-96), pp. 126-133, Portland, OR.

Schafer, J. L., 1997. Analysis of Incomplete Multivariate Data, *New York: Chapman and Hall*.

Wei Y., Moreau L., Jennings N., 2003. *Recommender Systems: A Market-Based Design*, Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-agent Systems, ACM Press, 600-607.

Yifan Hu, Yehuda Koren,and Chris Volinsky, 2008. "Collaborative Filtering for Implicit Feedback Datasets", Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pages: 263-272.

Zili Zhang, Chengqi Zhang and Shichaozhang, 2003. "An Agent-Based Hybrid Framework for Database Mining", Applied Artificial Intelligence, 17:383–398.