

Sampling based Multi-Agent Joint Learning for Association Rule Mining

(Extended Abstract)

Junyi Xu
Information System and
Management College
National University of Defense
Technology
Changsha, China
xujunyi0923@163.com

Le Li
Information System and
Management College
National University of Defense
Technology
Changsha, China
lile10@126.com

Li Yao
Information System and
Management College
National University of Defense
Technology
Changsha, China
liyao6522@gmail.com

Yifan Chen
Information System and
Management College
National University of Defense
Technology
Changsha, China
yifan_clifford@163.com

ABSTRACT

In order to achieve distributed data mining quickly and efficiently, this paper proposes SMAJL, a model for sampling based multi-agent joint learning which integrates sampling technology and multi-agent argumentation in the field of association rule mining. By sampling, this model can reduce the size of dataset and improve the efficiency of data mining; through joint learning from argumentation, it can effectively integrate inconsistent knowledge from different samples to improve the quality of distributed mining. We experimentally show that, in a variety of sampling strategies, SMAJL can almost achieve 90% accuracy using sample having a size of only 30% of that of original dataset.

Categories and Subject Descriptors

I.2.11 [Artificial Intelligence]: Multi Agent systems

Keywords

sampling; argumentation; joint learning; association rule

1. INTRODUCTION

In the field of traditional data mining, distributed data mining can quickly discover potential knowledge. While in the background of big data, people want to identify potential knowledge from massive data quickly and efficiently. However for massive data such as census and meteorological data, we cannot simply use distributed mining methods to analyze. Therefore, distributed data mining has its flaw in big

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*
Copyright © 2014, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

data applications. For big data, how to enhance the efficiency and quality of distributed data mining, and integrate inconsistent knowledge of distributed mining, has become the research hotspot.

Because sampling technique which uses samples to replace the original data, can effectively reduce the size of dataset and improve the efficiency of data analysis, sampling technique has been widely used in data mining and database fields.

Argumentation technology as new knowledge integration method has been used in data mining in recent years. Through argumentation, individual agent can analyze and evaluate knowledge of different samples between each other, and integrate inconsistent knowledge effectively. Therefor this article will propose a sampling-based multi-agent joint learning model called SMAJL in order to improve the efficiency of distributed association rules mining.

2. SAMPLING BASED MULTI-AGENT JOINT LEARNING MODEL

In order to achieve distributed data mining quickly and efficiently, this article will combine sampling technique with argumentation based multi-agent joint learning method, and propose sampling based joint learning model called SMAJL. By sampling, the model can reduce the data scale, and reduce the cost of data analysis to improve the efficiency of data mining; through learning from argumentation, agents can achieve the effective integration of distributed knowledge.

SMAJL method divides the learning process into three stages: sampling, individual mining and multi-agent joint learning. In SMAJL, first individual agent performs sampling in the overall dataset, and samples form the individual experience base of each agent. In local experience bases, agents perform independent association rule mining, thus forming individual knowledge bases. Each Agent uses its

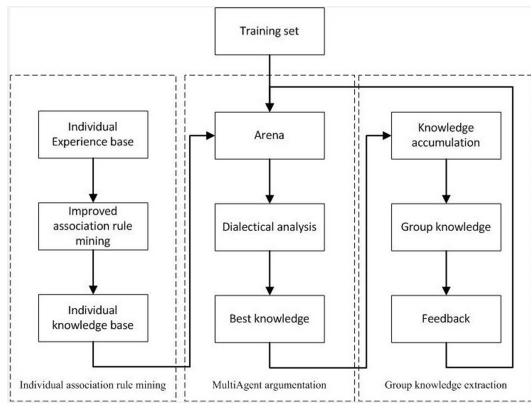


Figure 1: Argumentation based joint learning process

own local knowledge to argue with other Agent in the argumentation platform Arena. Through this kind of communication and learning process, knowledge of defects can be eliminated, and knowledge of high quality of multi-agent system can be extracted out, forming the sharing group knowledge and to guide the subsequent joint learning process. In this way, SMAJL can integrate local knowledge of different samples and improve the quality of distributed association rule mining.

3. MULTI-AGENT JOINT LEARNING FROM ARGUMENTATION

In SMAJL, argumentation is applied as a manner of interaction and learning in multi-agent system for knowledge integration of different samples. This section will describe the principle of the joint learning from argumentation method.

Guided by "Knowledge spiral" model in knowledge management area [3], we propose "joint learning from argumentation" (Multi-Agent Learning jointly from Argumentation, MALA). MALA method has three stages: the individual association rule mining, multi-agent argumentation and the group knowledge extraction, as shown in Figure 1.

The method of "joint learning from argumentation" can effectively merge the local experience knowledge of individual agent from different samples: Individual agents can realize the function of knowledge externalization; Through argumentation, individual agents realize the transformation of individual knowledge into organizational knowledge using the argumentation platform Arena; Ultimately, the shared knowledge of multi-agent system continue accumulation and refinement to form the spiral evolution process.

4. EXPERIMENTS

In sampling based joint learning method, we use four sampling strategies: sampling without replacement (SRSWOR), sampling with replacement (SRSWR), cluster sampling (C-S), stratified sampling (SS). We also compare SMAJL to joint learning method with non-sampling (MALA); simultaneously comparing with the association rule mining algorithm TFPC [1].

In order to empirically evaluate SMAJL, we use nursery dataset which contains 12960 examples and 4 agents to take part in SMAJL. Each agent produces association rules with

Table 1: Accuracy and rule number of different methods

	SRSWOR	SRSWR	SS	CS	MALA	TFPC
Accuracy	92%	90%	90%	88.7%	94%	63.3%
Number	65.8	63.2	67.7	67.4	79.5	343.7

the confidence level to 50% and the support level to 1% using Apriori-TFP algorithm [2].

During the sampling phase, each agent samples 1000 records from the original dataset as individual empirical data. While in MALA, four agents split the dataset equally. And we used TCV test on the same 300 records. In an experimental run, we use the training set to form the sharing knowledge base, which will be evaluated using the test set.

From experimental results in Table 1 we can see, accuracy of SMAJL can reach almost 90%. Compared with TFPC it can be increased by about 30%. Compared with TFPC, the number of final rules of SMAJL is more refined. Relative to MALA, SMAJL takes only about 30% of the input data and achieves accuracy of about 90%. For different strategies, SMAJL using SRSWOR gets the highest accuracy of 92%, which is closest to MALA ; while SMAJL using CS is less effective with 88.7%.

In short, all these strategies combined with SMAJL get relatively satisfactory results. With decrease of the input data, the accuracy of SMAJL can still be close to the original joint learning method, and the resulting knowledge size is also more refined.

5. CONCLUSION

This paper has presented SMAJL, which provides a new way to improve the quality and efficiency of distributed association rule mining. The key idea is joint learning from argumentation can be used to exchange and discuss local knowledge learnt by agents from individual samples using association rule mining. SMAJL is performed by three processes: sampling, individual mining and multi-agent joint learning. The results of experiments reveal SMAJL has an effective capability in learning from argumentation with small samples and reducing the cost of data analysis.

6. ACKNOWLEDGMENTS

This work is funded under National Natural Science and Foundation of China (No. 70971134) and Doctoral Research Fund (No.20124307110023).

7. REFERENCES

- [1] F. Coenen. The lucs-kdd tfpc classification association rule mining algorithm. *Department of Computer Science, University of Liverpool*, 2004.
- [2] F. Coenen, P. Leng, and S. Ahmed. Data structure for association rule mining: T-trees and p-trees. *Knowledge and Data Engineering, IEEE Transactions on*, 16(6):774-778, 2004.
- [3] I. Nonaka and H. Takeuchi. The knowledge-creating company: How japanese companies create the dynamics of innovation. *Long Range Planning*, 29(4):592, 1996.