Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 4, Issue, 03, pp.117-120, March, 2012

INTERNATIONAL JOURNAL
OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

MINING ACTIONABLE PATTERNS USING COMBINED ASSOCIATION RULES

[1,*]Kavitha, K. and [2]Ramaraj, E.

[1]Department of Computer Science, Dr.Umayal Ramanathan College for Women, Karaikudi, Tamilnadu, India
[2]Directorate of Distance Education, Madurai Kamaraj University, Madurai, Tamilnadu, India.

ARTICLE INFO

Article History:
Received 14th December, 2011
Received in revised form
19th January, 2011
Accepted 24th February, 2011
Published online 31st March, 2012

Key words:

Actionable patterns,
Combined association rules,
Utility,
Composite items.

ABSTRACT

Data mining promises to discover actionable patterns to users. However, patterns generated from traditional association mining are often difficult to understand and put into action, thus cannot satisfy the needs of real world completely. Actionable patterns address that patterns are deemed actionable if the user can act upon them in her favor. Recent studies present combined association rule mining can help extract useful knowledge from learned single rules, but even in this case, there still exists some interesting rules which cannot be found out by combined association rule mining. We propose a generic framework that uses utility in decision making to drive the data mining process. We use concepts from meta-learning and build on earlier work by Elovici and Braha, that uses decision theory for formulating an utility measure, to specialize the framework for classification tasks,and use the combined association rule mining to extract further actionable pattern and perfect combined association rule mining. This study proposes a novel approach to discover actionable combined patterns with composite items.

INTRODUCTION

The primary motivation for the field of data mining is to provide support for decision making by detecting useful patterns in large volumes of data. The decisions that are made based on the mined patterns are often crucial to the functioning of an organization or enterprise. We call such patterns that support decision making as *actionable patterns*. Knowledge discovery in database (KDD) is an active area of research that resolves the non-trivial process of identifying valid, potentially useful, and ultimately understandable patterns in data. In other words, the origin of KDD, many researchers realized the need from 'data' to 'knowledge' for the business decision-making, such as. Recently, more efforts have shifted from 'valid' and 'understandable' knowledge to actionable knowledge especially for real world data mining applications. In simple terms, a pattern is actionable if the user can act upon it to her advantage. Furthermore, actionable patterns can not only afford important grounds to business decision-makers for performing appropriate actions, but also deliver, expected outcomes to business.

Most data mining algorithms and tools stop with mining and delivery of patterns satisfying intrinsic measures (such as accuracy, support) and ignore decision making with respect to the pattern. While researchers have looked at measuring the utility of discovered patterns in decision making, there has not been much work on using such utility measures for driving the mining process as such. In our work we propose a framework for "closing the loop", i.e., using utility in decision making to drive the mining process. We introduce extrinsic measure which evaluate patterns with respect to the decisions made. Our goal is thus to include *decision making* as a part of Knowledge Discovery in Databases (KDD) process and come up with a general architecture to measure pattern's usefulness with respect to the decision made. Association rule mining is a main method to produce patterns. However, as large numbers of association rules are often produced by association mining algorithm, sometimes it can be very difficult for decision makers to not only understand such rules, but also find them a useful source of knowledge to apply to the business processes. In other words, association rules can only provide limited knowledge for potential actions. Therefore, there is a strong and challenging need to mine for more informative and comprehensive knowledge for decision-making in the real world. In order to overcome the drawback of association rule mining, some researchers emphasized on combined association rule mining . Simply put, combined rules are further extracted from learned single rules (or simple rules). Zhao et al. presented that combined patterns would provide more interesting knowledge and more useful results than simple association rules. However, in some cases, combined association rules extracted from the single rules which may enable it cannot discover certain useful rules. For example, rules $A\rightarrow C$ and $B\rightarrow C$, combined association rules may find combined rule $A\wedge B\rightarrow C$ at most. Meanwhile, Wang et al. pointed out that association rules with composite items can help to discover some useful rules which cannot be discovered by other algorithms without composite items. Thus, it is obvious to notice that composite items $A\vee B$ are not emphasized in combined association rules.

***Corresponding author:** kavitamil79@yahoo.com

Framework

In order to present our framework for mining actionable patterns we look at mining and evaluation stages of KDD process. Data mining stage tries to optimize some intrinsic measure such as accuracy,while pattern evaluation evaluates pattern for its actionabilty based on some decision theoretic framework. The most natural way to develop an architecture for mining actionable pattern is to merge two stages into one. Toachieve this we need to develop a objective function that can drive mining of patterns based on their utility in data mining. Except in some cases this is not easy to achieve. In most formulations of the evaluation stage the optimal decision is related to patterns mined via a complex optimization procedure and no simple relationship exists to underlying intrinsic measures employed in data mining stage. Statistical machine learning algorithms used in data mining consist of two stages : model selection and parameter estimation. While model selection is typically done manually, many parameter estimation procedures operate by gradient descent on an error/objective/likelihood function. But given the complexity of the optimization problem in the decision stage it is not usually possible to define a differentiable objective function for the data mining stage to directly optimize. Thus it is not tractable to drive the parameter estimation stage to produce actionable patterns within the existing class of data mining algorithms. Our approach is to drive the model selection stage using the utility of patterns in decision making. For this we use ideas from the field of Meta-learning.

The field of meta-learning, or learning to learn refers to a class of algorithms, that choose approriate classes of models and hyper-parameters, in order to optimize some objective function. This function is typically an intrinsic one, often the same objective function as in the parameter estimation phase. Meta-learning allows us to adjust the bias of the mining algorithm. This in turn affects the performance of the algorithm, since the patterns the algorithm detects is heavily influenced by the bias of the algorithm. In our framework we look at the utility in decision making as the performance measure which is optimized using the meta-learning approach. We define the space of hyper-parameters, S that we will modify using meta-learning, say, the number of neurons in the hidden layer, and the learning rate. We then choose a neighborhood structure N, with N(S), S 2 S, denoting the neighbors of a state S. The particular neighborhood structure that we adopt is a manhattan structure, in that we allow only one hyper-parameter to change at a time, by a fixed step size. We use the best-improvement strategy to search the neighborhood. In order to speed up the process when the neighborhood is large, we can employ a first-improvement strategy, where in the first neighbor, in some arbitrary order, with a better evaluation than the current state is chosen. Since we only find local optima, we repeat the entire local search procedure several times by picking different start states for the hyper-parameters. We finally use the hyperparameter setting that gave us the best evaluation across all repetitions.

Related work

Usually, there are a lot of tasks of data mining, such as classification, clustering and rule mining. Among of these tasks, the mining of patterns for discerning relationships between data items in large databases is a well studied technique. In order to introduce the key research, more details of association rule mining, combined association rule mining and association rule mining with composite items will be presented in this section.

Association rule mining

Association rule mining, a widely used data mining technique, is used to reveal the nature and frequency of relationships or associations between entities. Support and confidence, are the two major indices, which have useful applications to evaluate the rules. For instance, consider rule *X*: if *E* then *F*. Suppose that *X* has 60% confidence and 40% support. It expresses that 40% of records contain *E* and *F*. In fact, this means that in 40% of total records, rule *X* is valid. Additionally, it expresses that 60% of records that contain *E*, contain *F* as well. However, conventional association rules, as discussed above, can only provide limited knowledge for potential actions.

Combined association rule mining

Strictly speaking, traditional association rule mining can only generate simple rules. However, the simple rules are often not useful, understandable and interesting from a business perspective. Thus, Zhao et al.and Zhang et al.proposed combined association rules mining, which generated through further extraction of the learned rules. In other words, to present associations in an effective way, and in order to discover actionable knowledge from resultant association rules, a novel idea of combined patterns is proposed.

Mining composite items

The issue of mining association rules with composite items was proposed a couple of years ago. The basic concept of composite item is introduced referred to.Let $W=\{v1, v2, ..., vm\}$ be the set of atomic items. A composite item is formed by combining several items. The general form of a composite item is $v1 \vee v2 \vee \ldots \vee vn$, *for* $j \in [1,n]$, where $vj \in W$. A database is consisted of transactions where each transaction contains at least one of the atomic items. A transaction contains a composite item if the transaction contains at least one of the atomic items which form the composite item. Atomic items and composite items will be referred as items generally. An item is large if the member of transactions containing the item exceeds the minimum support. In addition, an itemset is a set of items such that none of the items in the set has similar items. For instance, $\{E, F \vee G\}$ is a valid itemset. However, $\{E, E \vee G\}$ is not a valid itemset as *E* and $E \vee G$ have the similar item *E*.As such, in comparison with simple association rule, association rules with composite items while treating each composite item as an independent item like the atomic items are more interesting or useful in real life applications.

Mining actionable combined patterns with composite items

Under the environment of actionable mining, in order to fulfill the objective of integrating combined association rule mining and mining association rules among composite items as a whole, this section will discuss some metrics contributing the approach proposed first, and then present the novel method.

Metrics

Interestingness measure and constraint based mining are employed in data mining to filter out such redundancy and

	Utility 1		Utility 2		Utility 3	
	Spam	Non-spam	Spam	Non-spam	Spam	Non-spam
Accept mail	-1000	300	-300	300	-300	300
Reject mail	400	-500	400	-1000	400	-500

Table 1: Utility Matrix

	Utility 1		Utility 2		Utility 3	
	Experiment 1		Experiment 1		Experiment 1	
	Best utility	Avg utility	Best utility	Avg utility	Best utility	Avg utility
Before	263.30	204.53	264.85	181.49	273.15	269.31
After	266.19	264.02	266.76	234.26	275.10	272.97
	Experiment 2		Experiment 2		Experiment 2	
Before	266.42	214.01	265.82	217.32	276.06	241.142
After	266.42	262.87	265.82	245.72	276.77	273.66
	Experiment 3		Experiment 3		Experiment 3	
Before	263.43	259.85	245.47	236.58	275.03	260.39
After	269.09	266.34	245.47	242.64	275.03	261.82

Table 2: Best and Average utility

uninteresting patterns. The most effective way of reducing volume of discovered pattern is so called interestingness measures. There are two types of such measures namely, objective and subjective measures. Objective measures are those that depend only on the structure of a pattern such as the confidence/support measure and usually quantified by using statistical methods. Subjective measures are further divided into actionable, unexpected and novel. A pattern is actionable which means that decision makers can get benefit from discovered patterns. A pattern is novel if it is completely new to the user. Actionable pattern or knowledge can lead to deliver the right information at the right time as well as provide appropriate gateways to the information space. Kaur noted that if a pattern found to be actionable this implies directly/indirectly the existing of other subjective measures such as background knowledge about domain. However, the following factors contribute to the difficulties of actionable mining. First, there exists the dilemma that expert knowledge and domain knowledge should be acquired to define actionability, but the ability to provide such knowledge has the possibility to invalidate the need for mining actionable patterns. Second, it is hard to model the usefulness of a pattern. Since the information is not available until after patterns deployed in the real world. Therefore, in order to obtain the actionable knowledge for the decision making effectively, the balance between objective and subjective measures should be both involved in the mining process. To mining combined association rules with composites items, the key point is how to build the objective and subjective measures between combined association rule mining and mining association rules among composite items. Support and Confidence, are two major objective indices, which have useful applications to evaluate the association rules. Since combined association rule mining is further extracted from the simple learned rules. Support and confidence are two major metrics of combined association rules. Subjective metrics such as domain knowledge and expert knowledge are not involved in combined association rules obviously. Furthermore, association rule mining with composite items can discover interesting rules when all atomic items which form the composite item may not large. When the composite item exceeds the minimum support, every atomic item in the composite item may be less than the threshold. Thus, modified confidence is introduced as follows:

For a rule $\{v1 \vee v2 \vee \ldots \vee vm,\ u\} \rightarrow R$, modified confidence is defined to be minimum Support= Support ($vi \cup u \cup R$) / Support ($vi \cup u$) , where $i \in [1,n]$. The reliability of rules should be evaluated. If a rule contains composite item, the larger its modified confidence is, and the stronger the intensity of substitutability of the atomic items in a composite is, the more reliable the rule is. In addition, as discussed above, expert knowledge and domain knowledge play a significant role for any success in business. However, most researchers emphasize the importance of finding rules but without well-defined strategy of discovering them. Since the knowledge of expert and domain is introduced, actionable patterns which aim to develop patterns those are actionable in the business world. Thus, general profit can be employed as the measure criteria in real life applications. To this end, in order to generate actionable patterns, interesting measures including objective measure such as support and confidence, and subjective measure such as expert knowledge and domain knowledge should be introduced in actionable combined association rule mining with composite items. In particular, in the mining process, expert knowledge and domain knowledge should be involved at first. Subsequently, the patterns generated should be evaluated.

Proposed approach

The proposed approach in this study consists of there steps. The first step is to employ expert knowledge and domain knowledge to ascertain the composite items generated from itemset. Then use association rules algorithms to find the rules with composite items. Finally, use combined association rule integrating the rules generated. The outline of actionable combined association rule mining with composite items is shown as follows:

1. Ascertain the items of set *S* according to the knowledge of users are interested in.
2. Mine the association rules with composite items *CI* which belongs to *S*.
3. Generate association rules *ARC* with algorithm such as FP-tree .
4. Use combined association rules to discover the association rules *AR*.
5. Integrate *AR* and *ARC* with the help of objective measures to find the final pattern *PA*.
6. Check the reliability of *PA*.

Conclusions

This study discusses a generic framework that uses utility in decision making to drive the data mining process and we use concepts from meta-learning that uses decision theory for formulating an utility measure, to specialize the framework for classification tasks, and mine actionable combined patterns with composite items. Firstly, association rule, combined association rule and association rules with composite items are presented respectively. Then in order to implement actionable pattern mining in real world applications, metrics such as objective and subjective measures are discussed. Finally, this paper proposes a novel approach of mining actionable association rules with composite items integrating combined association rules and association rules with composite items. Domain application such as supplier optimization and customer retention will be discussed in future research.

REFERENCES

[1] H. Xu, G. L. Liu, J. Guo and Z. L. Lou, "Grain Price Impact Analysis System on Data Mining", JCIT: Journal of Convergence Information Technology, Vol. 6, No. 1, pp. 207 -211, 2011.

[2] W. S. Pan and P. W. Chen, "A Study on the Logistic Service Satisfaction for Internet Marketing Enterprise Using Data Mining Technology ", AISS: Advances in Information Sciences and Service Sciences, Vol. 3, No. 2, pp. 114-120, 2011.

[3] P. S. Wang, "Survey on Privacy Preserving Data Mining", JDCTA: Journal of Digital Content Technology and its Applications, Vol. 4, No. 9, pp. 1 - 7, 2010.

[4] L. B. Cao and Y. C. Zhao "The Evolution of KDD: Towards Domain-Driven Data Mining", International Journal of Pattern Recognition and Artificial Intelligence, Vol. 21, No. 4, pp. 677- 692, 2007.

[5] Y. C. Zhao, H. F. Zhang and L. B. Cao, "Combined Pattern Mining: From Learned Rules to Actionable Knowledge", Lecture Notes on Artificial Intelligence, Vol. 5360, No. 1, pp.393-403, 2008.

[6] Y. C. Zhao, H. F. Zhang, F. Fiqueiredo, L. B. Cao and C. Q. Zhang, "Mining for Combined Association Rules on Multiple datasets", In Proceedings of ACM SIGKDD Conference on Knowledge Discovery in Data- International Workshop on Domain Driven Data Mining, pp. 18-23, 2007.

[7] H. F. Zhang, Y. C. Zhao, L. B. Cao and C. Q. Zhang, "Combined Association Rule Mining", In Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 1069- 1074, 2008.

[8] K. Wang, J. N. K. Liu and W. M. Ma, "Mining the most Reliable Association Rules with Composite Items", In Proceedings of IEEE International Conference on Data Mining-Workshops, pp. 749-754, 2006.

[9] Z. Zhang and H. W. Wang, "An Efficient Association Rule Mining Algorithm and Business Application", In Proceedings of International Conference on Communications, Circuits and Systems, pp. 959-965, 2007.

[10] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules", In Proceedings of 20th International Conference on Very Large Databases, pp. 487-499, 1994.

[11] L. B. Cao, P. S. Yu, C. Q. Zhang and Y. C. Zhao, "Domain Driven Data Mining", Springer, New York, USA, 2010.

[12] K. Wang, Y. L. Jiang and A. Tuzhilin, "Mining Actionable Patterns by Role Models", In Proceedings of International Conference on Data Engineering, pp. 16-25, 2006.

[13] SwapnaRaj, P., and Ravindran, B. 2008. Personalized webpage rendering system. In *14th International Conference on Management of Data*.

****