International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:5, 2007

Actionable Rules: Issues and New Directions

Harleen Kaur

***Abstract*—** Knowledge Discovery in Databases (KDD) is the process of extracting previously unknown, hidden and interesting patterns from a huge amount of data stored in databases. Data mining is a stage of the KDD process that aims at selecting and applying a particular data mining algorithm to extract an interesting and useful knowledge. It is highly expected that data mining methods will find interesting patterns according to some measures, from databases. It is of vital importance to define good measures of interestingness that would allow the system to discover only the useful patterns. Measures of interestingness are divided into objective and subjective measures. Objective measures are those that depend only on the structure of a pattern and which can be quantified by using statistical methods. While, subjective measures depend only on the subjectivity and understandability of the user who examine the patterns. These subjective measures are further divided into actionable, unexpected and novel. The key issues that faces data mining community is how to make actions on the basis of discovered knowledge. For a pattern to be actionable, the user subjectivity is captured by providing his/her background knowledge about domain. Here, we consider the actionability of the discovered knowledge as a measure of interestingness and raise important issues which need to be addressed to discover actionable knowledge.

***Keywords*—** Data Mining Community, Knowledge Discovery in Databases (KDD), Interestingness, Subjective Measures, Actionability.

I. Introduction

IN today's information Society, without appropriate theories, technologies, and tools allow us to access enormous amount of data which leads to poor information and hence decision making more inefficient and therefore the risk of information fatigue unless we adapt our selves to new technologies to solve this complexity of today's streaming of data.

Knowledge Discovery in Databases (KDD) is a active area of research that resolves the complexity mentioned above. Knowledge Discovery in Databases is the effort to understand, analyze, and eventually make use of the huge volume of data available. Through the extraction of knowledge in databases, large databases will serve as a rich, reliable source for knowledge generation. It combines many algorithms and techniques used in Artificial Intelligence, statistics, databases, machine learning, etc. KDD is the process of extracting previously unknown, not obvious, new, and interesting information from huge amount of data. KDD is the extraction of interesting patterns in large database. It has been recognized that a discovery system can generate a plenty of patterns which may be no interest. This is one of the central problem in the field of knowledge discovery in the development of good measures of interestingness of the discovered patterns.

Harleen Kaur is Ph.D scholar in the Department of Mathematics, Jamia Millia Islamia, New Delhi, INDIA; (email : harleen_k1@rediffmail.com).

Even after using data mining strategies such as pruning and optimization techniques, the volume of discovered patterns remain another mining problem, which leads to discovery of unreliable and uninteresting knowledge.

Interestingness measure and constraint based mining are used in data mining to filter out such redundancy and uninteresting patterns [4] example of such constrains are knowledge type constraints, data constraints, rule constrains, etc. The constraints based mining are not studied further since they are beyond the scope of our paper.

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 and usually quantified by using statistical methods. On the other hand, subjective measures are based on the subjectivity and understandability of the user who examine the patterns. These subjective measures are further divided into actionable, unexpected and novel. a pattern is unexpected if it contradicts and hence surprise the user expectations [13, 8, 9, 15]. A pattern is novel if it is completely new to the user. Actionability is defined as the extent to which a user can get benefit from the discovered patterns [1,3, 6, 7].

The most important subjective measure that has not been attacked in the data mining literature is actionability, which we believe is crucial toward making an appropriate decision. in addition, we assume that if a pattern found to be actionable this implies directly/indirectly the existing of other subjective measures.

Actionability is the key concept of any success either in business or academia. The proposed strategies are aimed to filter out the redundant patterns that may mislead decision makers, that is, relevant patterns to the decision maker should not be redundant and therefore actionable. Usually the users intended to make actions towards their interest in order to enable them to perform their jobs perfectly that lead to better and more efficient decision-making. Actionable knowledge can lead to deliver the right information at the right time as well as provide appropriate gateways to the information space.

There is an automated information technology that can capture and analyze not just information but actionable information. It is also called as actionable information based intelligent system. Actionable knowledge is information that one can be acted upon, something that leads to an action, and something that makes things happen.

Always there is a need to find a measurement, which would help to find relevant, and actionable patterns in the vast amount of information stored the databases or data warehousing. This will give better and efficient analytical tools, which can help in decision making.

II. Motivation

One of the central issues in data mining community is to make the mined patterns actionable. Actionable patterns are those patterns, which are interesting to the user that is user can do something about knowledge for his/her interest. Actionability is an important aspect of interestingness that has to be quantified based on subjectivity of the user to facilitate decision-making process [14].

Most the work that have been done on actionability focused on measuring this measure based on unexpectedness [13]. In addition, it is assumed that novelty is a good measure for actionability. The main difficult problem is to capture all the different actions from a specific domain, that is, if the problem is solved, then the next step will be to list all the possible action into different groups of pattern. The resultant patterns examined against user subjectivity to report actionable patterns.

Since the data keep on changing over time, subsequently, the actions should be changed accordingly. The optimal action is the one that change current status of profit to higher ones. This lead us to change the current state to another with a higher probability to get more profit. From the above stated literature, the action rule can be defined as the rule, which can be changed according to user. These rules are to be flexible to the user according to their requirements.

Since the data keep on changing over time, subsequently, the actions should be changed accordingly. The optimal action is the one that change current status of profit to higher ones. This lead us to change the current state to another with a higher probability to get more profit. From the above stated literature, the action rule can be defined as the rule, which can be changed according to user. These rules are to be flexible to the user according to their requirements. Final Stage.

III. Related Work

There are many approaches that studied the subjective measure of interestingness. Most of these approaches are proposed to discover unexpected patterns [8, 9, 10]. This important measure is further used to determine the actionable pattern explicitly. In [10] it is argued that the actionability is a good measure for unexpectedness and unexpectedness is a good measure for actionability. The patterns are categorized on the basis of these two subjective measure as patterns that are both unexpected and actionable, patterns that are unexpected and not actionable, and rules that are expected and actionable.

In fact, the above categorization scheme does not solve the problem of determining how actionability affects interestingness. The actionability and unexpectedness measures must be addressed individually and each represented separately in the interestingness measure.

In [5], the authors mine the actionable knowledge from the viewpoint of data mining tasks and algorithms. The tasks, such as clustering, association, outliers detection etc are explained along with the actionability techniques.

[19] proposed an approach to find out the best action rules. The best k-action rules are selected on the basis of maximizing the profit of moving from one decision to another. The technique used as post analysis to the rules extracted from decision tree induction algorithm. A novel algorithm is presented that suggest action to change customer status from an undesired status to desired one. In order to maximize profit-based, an objective function is used to extract action rules.

Furthermore, [2] has proposed a framework to quantify novelty of the discovered rules. They show that, the novelty is a good measure for actionability. Novelty of currently discovered rules is quantified with respect to domain knowledge and previously discovered knowledge based on newly discovered knowledge.

In [12, 13] they define actionablity in term of defining two types of attributes, namely, stable and flexible attributes. an action is takes when a change in flexible attribute is encountered. The approach takes into consideration the changes of attributes value as well and gives suggestion of, to which attributes value should a an attribute be changed in order to get some action. For a given two rules, there was no constraints values of flexible attributes listed in both rules or in one of these rules. The proposed constraints make use of confidence of new rules, which are called as Extended action rules, much higher than the confidence of corresponding to the action rules. DEAR2 is a system which is proposed by [16] as an extension to [17] and is based on a tree structure that partitions the set of rules, having the same decision value, into equivalence classes each labeled by values of stable attributes.

IV. Profit Mining as a measure for Action Rules

Data mining is viewed as the process of turning data into information, information into action, and action into value or profit [13]. However, the task of finding actionable rules is not trivial. Most papers emphasize the importance of finding actionable rules but without well-defined strategy of discovering them. Actionability is seen as an elusive concept

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:5, 2007

because it is difficult to know the space of all rules and the actions to be attached to them. Products should not be selected based on their individual profitability, but rather on the total profitability, they generate.

The profits due to cross-selling effects with other products in the assortment, therefore, to evaluate product profitability, it is essential to look at frequent sets rather than at individual product items. To improve customer relation, the enterprise must know what actions to take to change customers from undesired status such as attritors to a desired one such as a loyal customer [19].

Unlike distributed knowledge, in order to extract actionable knowledge from output to other data mining algorithms, one must take into account resource constraints. Actions, such as direct mailing and sales promotion, cost money to the enterprise. At the same time enterprises are increasingly constrained by cost cutting. There is a strong limitation on the number of customer segments that the company can take on, or in the number of actions the company can exploit.

The attribute value changes will incur costs. Domain knowledge and domain experts can only determine these costs. For each attribute used in the decision tree, a cost matrix is used to represent such costs. If some values of some attributes cannot be changed with any reasonable amount of money then these attributes are called hard attributes but if the value changes are possible with reasonable costs, then those attributes are called soft attributes. So in order to discover action rules it is required that the set of conditions (attributes) is partitioned into stable attributes and flexible attributes. For example, *Age and Sex* is a stable attribute, and *interest rate* on any customer account is a flexible attribute (dependable on bank).

Profit mining is to find the best probable recommendations using any prediction model and modify their rank based on probability and profit.

V. Comparative Analysis

For the sake of understandability of the approaches proposed in the literature, we will perform a comparative analysis between two techniques of finding action rules, the first approach uses the notion of cost and feasibility of an action rules was introduced [11], the technique was based on calculating the cost of changing the attribute value from v_1 to v_2.

Let $b \in A_i$ is a flexible attribute and b_1, b_2, b_3, and $b_4 \in V$ are its values, then if the cost of the changing of the attribute value from b_1 to b_2 is less than the cost of the changing of the attribute value from b_3 to b_4 then we say that the change of values from b_1 to b_2 is more feasible than the change from b_3 to b_4 [18].

The second approach uses a technique which is based on providing a set of pattern transactions and pre-selected target items and recommending target items and promotion strategies to new customers and maximizing the net profit.

Suppose a rule has the form $\{g_1,\ldots, g_k\} \rightarrow \langle I,P \rangle$ where $g_1,\ldots, g_k$ are generalized non-target sales such that no g_i is a generalized sale of other g_j and $\langle I,P \rangle$ is a generalized target sale. Consider a customer represented by a set on non-target sales $\{s_1, \ldots, s_p\}$. A rule $\{g_1,\ldots, g_k\} \rightarrow \langle I,P \rangle$ matches the customer if $\{g_1,\ldots, g_k\}$ generalizes $\{s_1, \ldots, s_p\}$. If a matching rule $\{g_1,\ldots, g_k\} \rightarrow \langle I,P \rangle$ is selected to make recommendation and if the customer buys some quantity Q of I under the recommended promotion code, the profit generated is (Price(P)-Cost(P)) $\times$ Q [18].

The table given below shows the subjective measures and their correlation with each other. It summarizes the most influential measures of subjectivity. Note that the most important measures for decision making are the pattern found to be actionable - novel and actionable-unexpected which considered to be the most interesting patterns. They are overlapping with each other.

TABLE I
SUBJECTIVE INTERESTINGNESS MEASURES CATEGORIES

	Unexpected Rule	Expected Rule	Novel Rule	Non-novel Rule
Actionable Rule	Most Interesting	Not Interesting	Most Interesting	Less Interesting
Non-Actionable Rule	Less Interesting	Not Interesting	More Interesting	Not Interesting

VI. Conclusion

This paper studies the issues of Actionability based on the discovered knowledge. This scheme is extracting patterns on profit basis and combining them into a global actionable plan for decision makers and customer relationship from the data about the customers. New techniques for discovering action rules will be investigated. In future work, I will construct algorithm for flexible attributes and evaluate the effectiveness of algorithms in the real world deployment of the action-oriented community.

ACKNOWLEDGMENT

I would like to thank Prof. S K Wasan, Department of Mathematics, Jamia Millia Islamia, New Delhi, India and Dr. Vasudha Bhatnagar, Deptt. of Computer Science, Delhi University, New Delhi, India for their encouragement and support. In addition, great thank to Mr. Ahmed Sultan Al-Hegami, Lecturer, Sana University, Yemen for fruitful discussion.

REFERENCES

[1] Adomavicius, G. & Tuzhilin, A. 1997. *Discovery of Actionable Patterns in Databases: The Action Hierarchy Approach*. In Proceedings of the Third International Conference of Knowledge Discovery & Data Mining. The AAAI Press.

[2] A. S. Al-Hegami, V. Bhatnagar, and N. Kumar. *Novelty Framework for Knowledge Discovery in Databases* . In Proceedings of 6th International Conference on Data Warehousing and Knowledge Discovery.(DaWaK 2004), Spain, 2004.

[3] A. S. Al-Hegami. *Subjective Measures and their Role in Data Mining Process*. In Proceedings of the 6th International Conference on Cognitive System. New Delhi, India, 2004.

[4] Han, J. and Kamber, M. 2001. *Data Mining: Concepts and Techniques.* 1st Edition, Harcourt India Private Limited.

[5] He., Zenyou. & Xu., Xiaofei., & deng., Shengchum. *Data Mining for Actionable Knowledge: A Survey*. Supported by the High Technology-gyResearch & Development Program of China (No. 2003 AA4Z2170, No. 2003 AA413021) and the IBM SUR Research Fund.

[6] Liu, B. & Hsu, W. & Lee, H-Y. & Mum, L-F 1996. *Tuple Level Anaysis for Identification of Interesting Rules*. In Technical Report TRA5/95, SoC, National University of Singapore, Singapore.

[7] Liu, B. & Hsu, W. & Chen, S. & Ma, Y. 2000. *Analyzing the Subjective Interestingness of Association Rules*. IEEE Intelligent Systems.

[8] Padmanabhan, B. & Tuzhilin, A. 1997. *Unexpectedness as a Measure of Interestingness in Knowledge Discovery*. Working paper # IS-97- Dept. of Information Systems. Stern School of Business, NYU.

[9] Padmanabhan, B. & Tuzhilin, A. 1998. *A Belief-Driven Method for Discovering Unexpected Patterns*. KDD-98.

[10] Padmanabhan, B. & Tuzhilin, A. 2000. *Small is Beautiful: Discovering the Minimal Set of Unexpected Patterns*. KDD-2000.

[11] Ras, Z.W. & Tzacheva, A., In search for action rules of the lowest cost. In "*Monitoring, Security and Rescue Techniques in Multiagent Systems*". Advances in Soft Computing , Springer-Verlag, 2005, 261-272

[12] Ras, Z.W. & Wieczorkowska, A.. Action Rules : *How to increase profit of a company*. In Principles of Data Mining and Knowledge Discovery, (Eds. D. A. Zighed, J. Komorowski, J. Zytkow) Proceedings of PKDD'00, Lyon, France, LNCS/LNAI, No. 1910, Springer-Verlag 2000, 587-592.

[13] Silberschatz, A. & Tuzihilin, A. 1995. *On Subjective Measures of Interestingness in Knowledge Discovery*. In Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining.

[14] Silberschatz, A. & Tuzihilin, A.1996. *What makes patterns Interesting in Knowledge Discovery Systems*. IEEE Trans. And Data Engineering. Vol.5, no. 6.

[15] Suzuki, E. 1997. *Autonomous Discovery of Reliable Exception Rules*. In Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining. Newport Beach, CA, USA.

[16] Tsay, L.-S., & Ras, Z. W. (2004). *Action Rules Discovery System DEAR, Method and Experiments. In the Special Issue on Knowledge Discovery, (Ed. Z. W. Ras)*, Journal of Experimental and Theoretical Artificial Intelligence, Taylor & Francis.

[17] Tsay, L.-S., & Ras, Z. W. (2005). *Action Rules Discovery System DEAR2, Method and Experiments*. In the Special Issue on Knowledge Discovery, (Ed. Z. W. Ras), Journal of Experimental and Theoretical Artificial Intelligence, Taylor & Francis. Vol. 17, No.1-2, January-June 2005, 119-128.

[18] Wang, K., & Zhou, S., & Han. J. (1999). *Profit Mining: From Patterns to Actions*. National Science and Engineering Research Council of Canada.

[19] Yang, Q., & Yin, J. and Ling, X. C., & Chen, *T. Postprocessing Decision Trees to Extract Actionable Knowledge* . In IEEE International Conference on Data Mining (ICDM'03).