
PROGRESS OF UTILITY MINER IN DATA MINING

G.Maragatham * & Prof. Dr. M. Lakshmi #

*Research Scholar, Computer Science & Engg , # Prof & HOD Computer Science & Engg
Sathyabama University – Chennai 600 119

ABSTRACT

As data mining deals with voluminous data. Knowledge extraction needs attention especially in decision making areas. The concept of utility can be integrated in mining process to draw useful patterns. This factor of course helps the researchers to bring out useful patterns in mining processes. Because of its significance it finds its application in mining of patterns in databases. Based on the type of application the utility factor varies. For example : With respect to the retail databases the utility factor can state about the profit of the items. In such cases the retailer can concentrate on profitable items for improving the business. With respect to the medical domain the utility factor can provide weightage on severity issues of a particular disease. This article gives an insight of Utility miner with respect to various aspects in the area of data mining.

Keywords: Utility miner, Data mining, Databases, Mining.

INTRODUCTION

Various researchers have proposed different interesting measures for analyzing the patterns from the data. The term utility is meant for specifying the usefulness of the item sets. The researcher Shen et al., has stated that the “interestingness of a pattern = probability with utility “. Based on the user’s specific objectives and the utility of the mined patterns, utility - based approaches may be more useful in decision making problems in real time. Utility Miner detects all item sets from a transaction database for utility values above than the minimum utility threshold. Utility mining helps a user to conveniently express the perspectives concerning the significance of patterns [5]. To achieve a user’s goal two categories of utilities are existing (i) transaction utility and (ii) external utility. Earlier utility , of an item can be directly obtained from the information stored in the transaction data set. Whereas , the later one - external utility reflects user preference and can be represented by means of a utility/preference table. Then, by using both transaction database and utility table together, data mining can be guided by the utilities of item sets. Hence, the inferred

knowledge is constructive for maximizing a user's goal. This is specified by Hong Yao et al., [2]. Hence, the issues related to utility based mining is discussed by various researches are discussed briefly.

A GLIMPSE ON UTILITY MINER CONCEPT IN DATA MINING AREA

Utility based mining adds new dimension to the association rule mining concept. It helps the users to express their perspectives towards the usefulness of the patterns to be mined. Based on this concept two types of utilities could be more useful (1) transaction utility (2) External utility. A transaction based utility of an item is directly obtained from the database whereas an external utility gives preferences of the users about the items. Yi - Dong Shen et al.(2002), has stated that the "interestingness of a pattern = probability + utility ". Based on the usefulness in real applications the rules are used in decision making processes. The notion to generate association rules that are both statistically and semantically related to the users objective is attempted by Yi - Dong Shen et al., (2002)., To achieve this target a new approach named OOA (Objective-Oriented utility based Association mining is developed. The method is able to generate Objective Oriented utility based association rules. Still for better rule set, Pruning based on utility constraint is introduced.

The fundamental concepts related to Utility is described by Yao, H., Hamilton et.al., (2004). The concept of Internal and external utilities are discussed along with the theoretical model of utility mining . The theoretical model of utility mining is presented with Lemmas and theorems for Utility Bound property and Support Bound property

Sven F. Crone et al., (2005) have presented a method for utility based mining with time series data. Commonly data mining applications with respect to corporate purpose use predictive classification algorithms. Data mining methods for regression and time series analysis generally ignore economic utility and use only simple accuracy measures. Techniques from disciplines such as computational intelligence and statistics alike minimize a symmetric statistical error, such as the sum of squared errors, for modeling ordinary least squares predictors. On the other hand, business applications elucidate that real forecasting problems that may contain non-symmetric errors. The costs emerging from over - versus under prediction are dissimilar for errors of identical enormity, requiring a correction phase in the prediction process to derive valid decisions. A asymmetric cost function is developed reflecting this issue which tends to do superior forecasts and a cost effective decision. A trained multilayer – perceptron using various objective functions, finds the performance in opposition to statistical forecasting methods. Ying Liu, et al.,(2005)., In this paper, the authors have devised a Two-Phase algorithm for efficiently pruning the number of candidates generated to obtain a set comprising of high utility item sets. In the first phase, a model is devised that applies the "transaction-weighted downward closure property" on the search space for the recognition of candidates. During the second phase, one additional database scan is done for the identification of the high utility item sets. The algorithm is tested in both synthetic as well as real databases. The algorithm performs very efficiently in terms of speed and memory cost even on large databases .

Hong Yao, Howard J. Hamilton and Liqiang Geng (2006), The authors have explored utility based measures for item set mining and proposed a unified framework with the utility

based measures such as utility value of an item, utility value of an item set for mining process using a unified utility function. Finally, mathematical properties of the utility based measures are discussed with respect to time and space complexity of the algorithm. Also, mining utilities from transaction databases is discussed by Yao, H., and Hamilton, H.J (2006).

Alva Erwin et al.,(2007) has discussed a method of Mining High Utility item sets from transactional databases. The authors have designed a novel method called CTU-PRO for mining high utility itemsets. The approach uses bottom up traversal on CUP – Compressed Utility tree. The efficiency of the algorithm is tested with various constraint to produce quality output.

Unil and Yun (2007) have proposed a different approach known as weighted interesting pattern mining (WIP) which recommend weight-confidence, a measure to mine correlated patterns with weight affinity. Weight boundaries have been determined using a weight range and support affinity patterns have been identified aided by an h-confidence. For performance improvement, instead of the upper bound of h-confidence, original h-confidence has been used without additional computation cost in WIP. WIP has been capable of detecting more correlated patterns by taking into account weight affinity and/or support affinity between items within patterns in addition to providing a balance between the two measures of weight and support. WIP has been shown to be efficient and scalable for determining affinity patterns by an all-inclusive performance study. In addition, it produced less but more useful patterns with the correlation. The selective use of w-confidence, h-confidence and weighted support consistent with requirement of applications has been capable of decreasing the number of thresholds.

Vid Podpecan et al.,(2007) has discussed utility-based data mining. The authors have presented a novel technique named FUFM – Fast Utility Frequent Mining for finding all utility – frequent itemsets with constraints such as utility and support factors. The performance shows that this technique is able to mine high utility items sets on large databases. Yeh J.S et., al.,(2007). From a retailer point of view to generate high profitable items from the transactional databases. The concept of Utility miner is used in conjunction with frequent pattern items. The authors have used a bottom up approach for frequently mining patterns of profitable items. It is a 2 phase algorithm which uses a novel concept called quasi-utility-frequency for mining lattice of item sets.

Jieh-shah (2008). In this paper the authors have applied incremental mining, which also identifies utility and temporal based itemsets. The authors have introduced two algorithms namely IUM – Incremental Utility mining and FIUM – Fast Incremental Utility Mining. The efficiency of the algorithms are studied with respect to various datasets. The concept of Weighted Utility ARM (WUARM) is introduced by Sulaiman (2008)., They have considered importance and the weight of the item sets for mining process. The efficiency of this algorithm is verified with respect to both real and synthetic datasets.

Chowdhury Farhan Ahmed et al. (2008) have proposed an AWFPM (adaptive weighted frequent pattern mining) algorithm by introducing the concept of adaptive weight for each item. Their algorithm has been capable of handling circumstances where the weight

(cost or importance) of an item changes with time. Their algorithm using adaptive weights have shown very efficient and scalable WFP mining by extensive performance analyses.

M. Sulaiman Khan et al. (2008) have proposed a framework for association rule mining known as weighted utility ARM (WUARM), which takes into account the varied significance and different frequency values of individual items as their weights and utilities. Thus, weighted utility mining has concentrated on discovering those item sets for which the weighted utility is greater than the weighted utility threshold specified by the user. They have conducted experiments using standard ARM, weighted ARM and weighted utility ARM (WUARM) on synthetic and real data sets and an analysis of the results has been presented. Shankar.S and T.Purusothaman (2009) have proposed a method to mine interesting association patterns from transaction data items of an organization to progress its business utility. The method has given consideration to significance, utility and subjective interestingness of the users while mining interesting association patterns. The interesting patterns that were mined using the proposed approach have been capable of giving useful suggestions to improve the business of the enterprise.

CONCLUSION

This article presents an overview on utility miner concept with respect to various aspects of data mining. The author concludes that the use of utility miner really helps the decision making process to a better standard. Hope this article, brings a good direction for the researchers in data mining area using utilities.

REFERENCES

1. Yi - Dong Shen , Qiang Yang and Zhong Zhang (2002).,“Objective - oriented utility-based association mining ”, Proceedings of the 2002 IEEE International conference on Data Mining.
2. Hong Yao, Howard J. Hamilton and Cory J. Butz (2004), “A foundational Approach to mining itemset utilities form Databases ”, Proceedings of 4th SIAM conference on Data Mining.
3. Sven F. Crone, Stefan Lessmann , Robert Stahlbock (2005) ,“ Utility based data mining for time series analysis: cost-sensitive learning for neural network predictors ”, UBDM 2005, Proceedings of the 1st International workshop on utility-based data mining , ACM Newyork, NY.
4. Ying Liu, W. K. Liao, and A. Choudhary (2005), “A Fast High Utility Item sets Mining Algorithm”, Proceedings of the First international workshop on Utility-based data mining, pp. 90 – 99, Illinois., Chicago.
5. Hong Yao, Howard J. Hamilton and Liqiang Geng (2006), “A unified framework for Utility based measures for mining itemsets ”, Philadelphia, Pennsylvania, USA.
6. Yao, H., & Hamilton, H.J (2006), “Mining itemset utilities from transaction databases. Data & Knowledge Engineering” , Vol : 59, Issue :3, Pp:603-626.
7. Alva Erwin, Raj P. Gopalan, N. R. Achuthan.(2007), “A Bottom – Up Projection Based Algorithm for Mining High Utility Item sets ”, - II workshop on Integrating AI and Data Mining -AIDM Gold Coast, vol : 84, kok-leong Ong, Junbin Gao, Wenyuan aLi, Ed.

8. Unil, Yun, (2007), “Efficient mining of weighted interesting patterns with a strong weight and/or support affinity”, An international journal of Information Sciences:, Vol. 177, No. 17, pp. 3477-3499.
9. Vid Podpecan, Nada Lavrac and Igor Kononenko,(2007), “A Fast Algorithm for Mining Utility-Frequent Itemsets”, In Proceedings of the International Workshop on Constraint-Based Mining and Learning - CMILE, 2007
10. Yeh J. S., Li, Y. C., Chang C. C., (2007) , “ Two-phase algorithms for a novel utility-frequent mining model”, LNCS - Proceedings of the international conference on Emerging technologies in knowledge discovery and data mining, pp. 433-444 .
11. Chowdhury Farhan Ahmed, Syed Khairuzzaman Tanbeer , Byeong-Soo Jeong and Young-Koo Lee, (2008), “Mining Weighted Frequent Patterns Using Adaptive Weights”, Lecture Notes in Computer Science, Springer Berlin / Heidelberg ,Vol.5326, pp. 258-265.
12. Jieh-shan Yeh, Chih-Yang Chang, Yao – Te Wang (2008), “Efficient Algorithms for Incremental Utility Mining”, ICUIMC 2008, January 31 - February 01, Suwon, Korea.
13. M. Sulaiman Khan, Maybin Muyebe, Frans Coenen(2008), “A Weighted Utility Framework for Mining Association Rules ”, Proceedings of 2008 Second UKSIM European Symposium on Computer Modeling and Simulation, pp. 87-92.
14. Shankar.S and T.Purusothaman,(2009), “A Novel Utility Sentient Approach for Mining Interesting Association Rules”, International Journal of Engineering and Technology Vol.1, No.5, December, 2009.