AN ENHANCED SEMANTIC MEASURE IN DATA MINING

U.VENKANNA
PG student
Laqshya college of Computer Studies, Kakatiya University, AP, INDIA

Abstract - The important concept in the field of data mining is Utility mining. It is an enhancement to frequent pattern mining where it consists in determining the frequent patterns on transactional databases. The main objective of frequent item-set mining is it must satisfy minimum support threshold.To solve the problems exists in frequent item-set mining another concept known as Utility mining is introduced where in this the transaction database contains transactions where purchase quantities are taken into account as well as the unit profit of each item. The main purpose of data mining and analytics is to find novel, potentially useful patterns that can be utilized in real-world applications to derive beneficial knowledge. In recent years, there is a more demand for utility-oriented pattern mining. Utility pattern mining is a vital task, with numerous high-impact applications, including cross-marketing, e-commerce, finance, medical, and biomedical applications. In this paper we are going to introduce the methods involved in Utility pattern mining. First, we introduce an in-depth understanding of UPM, including concepts, examples, and comparisons with related concepts. A taxonomy of the most common and state-of-the-art approaches for mining different kinds of high-utility patterns is presented, including Aprioirity-based, tree-based, projection-based, vertical-/horizontal-data-format-based, and other hybrid approaches**.**

Keywords: Utility pattern mining, Frequent Item set mining, Rare item sets, Association rule mining

I. INTRODUCTION

Data mining is the way toward discovering patterns designs from large datasets including techniques at the crossing point of machine learning, measurements, and database systems. Data mining is an interdisciplinary subfield of software engineering and insights with a general objective to separate data (with wise strategies) from an informational collection and change the data into an intelligible structure for further use. Data mining is the examination venture of the "learning revelation in databases" process, or KDD. Beside the crude examination step, it likewise includes database and information the board viewpoints, information pre-handling, model and induction contemplations, intriguing quality measurements, unpredictability contemplations, post-preparing of found structures, perception, and online updating.[1] The distinction between information investigation and Data mining is that information investigation is to outline the history, for example, breaking down the adequacy of a showcasing effort, conversely, Data mining centers around utilizing explicit machine learning and factual models to foresee the future and find the examples among information.

The expression "information mining" is in actuality a misnomer, on the grounds that the objective is the extraction of examples and learning from a lot of information, not the extraction (mining) of information itself.[7] It additionally is a buzzword[8] and is habitually connected to any type of extensive scale information or data preparing (accumulation, extraction, warehousing, examination, and measurements) just as any utilization of PC choice emotionally supportive network, including man-made consciousness (e.g., machine learning) and business knowledge. The book Data mining: Practical machine learning instruments and procedures with Java was initially to be named simply Practical machine learning, and the term information digging was included for promoting reasons. Frequently the more broad terms (extensive scale) information investigation and examination – or, when alluding to real strategies, man-made brainpower and machine learning – are progressively fitting.

The genuine Data mining assignment is the self-loader or programmed investigation of vast amounts of information to remove beforehand obscure, fascinating examples, for example, gatherings of information records (bunch examination), unordinary records (peculiarity discovery), and conditions (affiliation rule mining, consecutive example mining). This generally includes utilizing database methods, for example, spatial records. These examples would then be able to be viewed as a sort of synopsis of the info information, and might be utilized in further investigation or, for instance, in machine learning and prescient examination. For instance, the Data mining step may distinguish various gatherings in the information, which would then be able to be utilized to acquire increasingly precise expectation results by a choice emotionally supportive network.

II. ASSOCIATION RULE MINING

Mining association rule mining is one of the researches in data mining. Association rule mining[5] is a procedure which is meant to discover frequent patterns, correlations, associations,

or causal structures from data sets found in various kinds of databases such as relational databases, transactional databases, and other forms of data repositories. There are three main applications of Association rule mining.

Basket data analysis - is to analyze the association of purchased items in a single basket or single purchase as per the examples given above.

Cross marketing - is to work with other businesses that complement your own, not competitors. For example, vehicle dealerships and manufacturers have cross marketing campaigns with oil and gas companies for obvious reasons.

Catalog design - the selection of items in a business' catalog are often designed to complement each other so that buying one item will lead to buying of another. So these items are often complements or very related.

III.   UTILITY MINING

High Utility Pattern Mining (HUPM) has become emerging research topic in data mining. In HUPM the *utility* of each item can be predefined based on users background knowledge. Utility is a subjective measure where it indicated a subjective value is associated with a specific value of utility is assigned by the user according to his interpretation of domain specific knowledge measured by a specific value, such as cost, interestingness measures can be categorized in objective, subjective[11] and semantic measures. Objective measures such as support or confidence for pattern mining. Subjective measures such as unexpectedness by considering users domain knowledge. For semantic measures such as utility consideration of data long with the users expectation will be done. Hence utility is the quantitative measure for user preferences and the usefulness of the itemset can be quantified in terms of utility value. How useful an item set is can be determined by the utility as it must satify a specific utility constraint.

To resolve the above issues utility pattern mining has become an emerging topic in data mining where in this each object or item set has a unit utility also known as unit profit and can appear more than once in each transaction or event. The utility must satisfy specified utility constraint.

The ARM[9] approaches think about the utility of the things by its essence in the exchange set. The recurrence of itemset isn't adequate to mirror the genuine utility of an itemset. For instance, the deals director may not be keen on incessant itemsets that don't produce noteworthy benefit. As of late, a standout amongst the most difficult data mining assignments is the mining of high utility itemsets proficiently. Distinguishing proof of the itemsets with high utilities is called as Utility Mining. The utility can be estimated as far as cost, benefit or different articulations of client inclinations. For instance, a PC framework might be more productive than

a phone as far as benefit. Utility mining model was proposed in [19] to characterize the utility of itemset. The utility is a proportion of how valuable or productive an itemset X is. The utility of an itemset X, i.e., u(X), is the entirety of the utilities of itemset X in every one of the exchanges containing X. An itemset X is known as a high utility itemset if and just if u(X) >= min_utility, where min_utility is a userdefined least utility edge [11]. The fundamental target of high-utility itemset mining is to discover each one of those itemsets having utility more prominent or equivalent to client characterized least utility edge.

3.1 Methodologies

There are four different methodologies used in utility pattern mining

3.1.1 Data Structure

A compact tree structure such as UP tree used to facilitate the mining performance and void scanning of original database repeatedly.

3.1.2 UP growth

It is more compressed utility pattern tree and proposed well known utility pattern growth algorithm which is used efficiently to mine HUIs.UP growth is enhancement for FP growth where it consists of four novel strategies name DLU(Discarding Local Unpromising Items),DLN(Decreasing Local Node utilities),DGU(Discarding Global Unpromising items during the construction of global UP tree).After two scans of original database the UP tree can be constructed. In the first scan the utility of each transaction can be calculated. In the second scan transactions are inserted into UP tree by using DGU and DGN strategies.

3.1.3UP Growth+

UP-Growth achieves better performance than FP-Growth by using DLU and DLN to decrease overestimated utilities of item sets. However, the overestimated utilities can be closer to their actual utilities by eliminating the estimated utilities that are closer to actual utilities of unpromising items and descendant nodes. We propose an improved method, named UP-Growth+, for reducing overestimated utilities more effectively. In UP-Growth, minimum item utility table is used to reduce the overestimated utilities. In UP-Growth+, minimal node utilities in each path are used to make the estimated pruning values closer to real utility values of the pruned items in database.

IV.   CONCLUSION

In the era of Data Mining, Association Rule Mining is one of the most essential assignments. Countless calculations are accessible for affiliation rule mining, which considers mining of frequent itemsets. Be that as it may, a rising point in Data Mining is Utility Mining, which focuses on utility contemplations amid itemset mining. Utility Mining covers all

parts of monetary utility in information mining and aides in identification of uncommon itemset having high utility. Uncommon High Utility itemset mining is extremely helpful in a few reality applications. In this paper, we have displayed a brief diagram of different calculations for high utility uncommon item set mining.

V. REFERENCES

[1] R. Agrawal, T. Imielinski and A. Swami, 1993,

"Mining association rules between sets of items in large databases", in Proceedings of the ACM SIGMOD International Conference on

Management of data, pp 207-216.

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

[3]Attila Gyenesei, "Mining Weighted Association

Rules for Fuzzy Quantitative Items", Lecture notes in Computer Science, Springer, Vol. 1910/2000, pages 187-219, TUCS Technical Report No.346, ISBN 952-12-659-4,ISSN 1239- 1891, May 2000.

[4] R. Chan, Q. Yang, Y. D. Shen, "Mining High

utility Itemsets", In Proc. of the 3rd IEEE Intel. Conf. on Data Mining (ICDM), 2003.

[5] H. Yun, D. Ha, B. Hwang, and K. Ryu. "Mining

association rules on significant rare data using relative support". Journal of Systems and Software, 67(3):181–191, 2003.

[6] H.Yao, H. J. Hamilton, and C. J. Butz, "A

Foundational Approach to Mining Itemset Utilities from Databases", Proceedings of the Third SIAM International Conference on Data Mining, Orlando, Florida, pp. 482-486, 2004.

[7] G. Weiss. "Mining with rarity: a unifying framework",.SIGKDD Explor. Newsl., 6(1):7– 19, 2004.

[8] Liu, Y., Liao, W., and A. Choudhary, A., "A Fast High Utility Itemsets Mining Algorithm", In

Proceedings of the Utility- Based Data Mining

Workshop, August 2005.

[9] Lu, S., Hu, H. and Li, F. 2005. "Mining weighted association rules. Intelligent Data Analysis",5(3):211–225.

[10]V. S. Tseng, C.J. Chu, T. Liang, "Efficient

Mining of Temporal High Utility Itemsets from

Data streams", Proceedings of Second International Workshop on Utility-Based Data Mining, August 20, 2006

[11]H. Yao, H. Hamilton and L. Geng, "A Unified

Framework for Utilty-Based Measures for Mining Itemsets", In Proc. of the ACM Intel. Conf. on Utility-Based Data Mining Workshop (UBDM), pp. 28-37, 2006.