

A literature survey to identify high profitable items from E-commerce data sets

Monica Makhija, Daljeet Choudhary, Aditya Tiwari, Jyothi Pillai

CSE Student: 6th sem CSE Student: 6th Sem Assistant Professor Associate Professor
Bhilai institute of Technology Raipur (CG)

Monicamakhija20@gmail.com, daljeet2510@gmail.com, adityarise0609@gmail.com
jyothipillai71@gmail.com

ABSTRACT

Highprofitable rare item sets, is an approach to extract out all the products that emerges high profit over selling. Data mining is a technique that collects data and figure out all relevant products which urges high profit over selling and having high margins. In this paper, keeping in mind customers' purchasing behaviour, profitable rare item sets which figure out all those products which can earn high margin profit, selling of these products through e-trade can lead to high profit. Rare item sets mining is a challenging task where the key issues are identifying interesting rare patterns and efficiently discovering them in large datasets. This data can contribute to have high access to all those products which are rarely purchased but consuming high margins, which in future can lead to high business utilities in e-trade.

Keywords: High profitable rare item sets, E-trade, high margin products.

1. INTRODUCTION

1.1 E-commerce

E-commerce is an innovative concept. It is, at present, heavily relying on the internet and mobile phone revolution to reshape the way businesses reach their customers. While in countries such as the US and China, E-commerce has taken notable steps to achieve high revenue, the industry in India is, still at its beginning. However over the past few years, the division has grown. E-trade has become the rapidly-growing segment. If this flourishing growth continues over the next few years, the size of the e-trade industry is estimated to be 20 to 30 billion USD by 2025. This increment is determined by increased customer purchases.

In this paper, sale of high profitable goods which are rare products having high margin, is taken under consideration and suggestions have been made to increase the sale of these products such as mobile-phone accessories, beauty products, footwears and many more.

1.2 Data Mining

Data mining is a technique of finding out the hidden details and information from the provided databases. It also analyses, explores and deducts some structured data that fulfils our needs based on some query or needs.

1.3 Association Rule Mining (ARM)

Association rule is a model that analysis or associates data mining task of uncovering relationships of data. It specifies types of data association, like it analyses the type of frequent products that are being purchased by the customers, it also predicts the future aspects of data.

1.4 Customer Purchasing Behaviour (CPB)

It basically describes the customers' behaviour of buying products, as data mining, association rules are the basis for this concept according to which the data is analysed and provided to the customer. Focusing on customers' need of buying high profitable rare item sets which earns them high margins of profit, this data analysis benefits them with structured data that are directly provided to the customers.

1.5 Rare Items

Rare item-sets are those item-sets which occur infrequently in the transactional dataset. In most of the realistic situations, rare item-sets having high profit provide very beneficial insights to the customer. Rare patterns may also indicate the occurrence of selling of high-margin products which is considered exceptional in the data. (Jyothi Pillai, 2013)

1.6 Profitable Items

The sale of rare items incorporates a substantial amount of high profit than other products, which is very beneficial to the business. Those rare items upon selling transform into "profitable items". The items having more margin yield more profit upon their sale.

1.7 Recommender Systems-

A recommender system or a recommendation system is a subcategory of processed-data filtering system that helps to predict the "rating" or "preference" that a customer would assign to a product. Recommender systems specifically generate a list of recommendations in one of the following two methods–

Collaborative filtering.

Content-based filtering.

2. Literature Survey

As in earlier research papers, this topic had been discussed under various terms referring to the same moot point of earning high profits over rare products. As in literature reporting of previous research, the labels "recommendation agents", "high utility products", "association rule mining", "frequent item sets", "adaptive interviewing" have been used.

Bo Xiao et al (2007), have discussed about recommendation agents, these agents depict about the interest of the customers for buying any product by analysing the pattern of their searching and selecting products online. That paper was designed to figure out the individual requirements of

particular customers. They gain users' trust to a great extent. The suggestions that are made to a particular customer is customized which is very crucial.

E-commerce provides more reach to customers, there is low operational cost, easy access to variety of products, more scope of comparing products. So, in this huge variety of products, on the basis of search history and customer profile these agents provide suggestions that encourage the user to look up for that product in various websites.

Venkateswari et al (2011), have focused about association rule mining there is "market-basket analysis" which performs a study on the habits of customers which is the source of motivation behind Association rule mining. The number of transactions made by the customer for the particular category of products is the basis for extracting information about the customers' preferences. The paper also suggests "Association rule mining in E-commerce data provide navigation and purchasing suggestions to customers."

Markus et al, have proposed the Utility – based decision TreeOptimization. This paper basically is about conducting an interview sessions for customers regarding their needs, all expected wants about that product, and "Minimizing the exit risk". As per their needs a decision tree is created and optimized result is generated, all the solutions that lead to an optimized result show the preferences of the sellers and buyers. This method actually deals to optimize the result by reducing the size of steps and finally suggest the buyer the only product to buy.

As E-commerce provide us with plenty of options to compare our products, this paper comes with the concept of "Minimizing the exit risk", which focuses on reducing the number of users exiting and turning over other sites. It wants its users to buy their products before exiting it. So they come up with reducing the number of questions necessary to draw a meaningful want.

Vikram et al, have suggested U-Hist tree – efficient data structure for high utility pattern mining from transaction databases. This paper focuses over digging the "high utility items" from data collected. The approach here used is about using data structure called "UP-tree", it maintains a graph at each node of the tree. This approach leads to better utility data, it actually defines the utility pattern of relative importance of particular items but it treats each utility with equal importance so transaction data is used as it creates copies of items which create a high utility mining pattern. This pattern eliminates the non-candidate items and evades the analysis of non-required results and computes better estimated results and improves the performance of the mining.

Sarwar et al (2000), have proposed a recommender system which analyses large scale purchase and producing valuable recommendations, it receives information about customers interest and accordingly suggest the products that fits their needs. The technique here used is "Collaboration filtering" which filters out the database and to discover customers' requirements. It focuses on improving the scalability of the collaborative filtering and improving the quality of recommendations for the customers.

Sawant et al (2014), (Ganesh Sawant¹, 2014) (Yussuf Abu Shaaban, 2009) have suggested techniques for mining high utility item sets from transactional database, the aim is to find out frequent pattern mining, high weighted pattern, high utility pattern to discover the products that have higher probability. The approach here applied is basically discussed about the technique of Discarding Global Unpromising items(DGU) and Decreasing Global node Utilities(DGN) to optimise the transactional data to figure out high utility itemsets.

Jane Hillston(2009), has solely focused over preserving the Quality of Service(QoS), as with changing characteristics of internet commerce, dynamic QoS needs to be enhanced. Poor QoS leads to customers' disapproval. So this paper comes out with the concept of cost-based admission

control(CBAC), which on high loads situation schedules the customers' request to more lightly loaded time period enhancing dynamicity of the system, which leads to better productivity of resources, endure throughput, reduces request failures.

Thorleuchter et al(2012), papers have come up with the idea over impact of textual information from existing E-commerce websites. Through web mining, (Dirk Thorleuchter, 2012) the information is gathered and are documented, and through website-matrix a training set is created and as per the pattern identification is done over the documents. These documents are further sub divided into training sets and test sets and are processed using text mining methods. The matrix training results are then analysed with the success factors of the e-website companies. The success factor shows the probability of the occurrence of semantic pattern, this approach led to focus over less successful e-websites.

Jyothi et al(2013), have proposed paper over transaction profitability which describes business intelligence, analysing over predicting future objectives, data mining's association rule being the whole some of the processing technique, these outcomes can be utilised for productive decision making like retail shelf management, customer segmentation. Association Rule Mining (ARM), helps to identify repeated items from the vast database and applying association rule over that data and creating conclusions about the data. This paper basically focused over the utility of the item, which is generalised by its contribution over the quantity of items sold.

The approach here described is about Transaction Profitability using HURI [TPHURI] is a modified version of HURI. TPHURI finds profitable transactions consisting of high utility rare item sets.

Jian et al(2011), has focused over the customers behaviour of purchasing a product having high marginal utility factor . This paper therefore explores the approach how to select a product with higher marginal utility, using Singular Value Demonstration (SVD) algorithm which outshines the users about the products that are new to customers

Akazue et al. (2016) has suggested that there is a lack of trust shown by the customers, towards a high valued product. Due to the lack of verification and inspection of the product before it is delivered, the customer cease to buy those high value products. This trust can be gained by providing an option to them to upload a live image of the item that is original to them and scrutinize with the image of the product they want to purchase and hence the similarity will be directly proportional to the authenticity of the product. If the product resemblance is very low, then the customer can reject it without further ado. This results in increasing the sale of high profitable rare products.

Bhosale et al. (2014) has focused on the point that the extraction of high beneficial products can be done by the method of maintaining a dedicated database for the transactions. However a large number of item-sets cause the system to take heavy storage space as well increased execution time. So to manage it, UP-tree and UP-tree mining algorithms (Utility Pattern Tree) are used that provide a compact representation of the long transactions occurring within a database.

Bipul et al. (2017) has proposed a matrix factorization method that will incorporate recommendations based on customer's ideal inclination towards "long tail products" which means insignificant number of hits for most of the products i.e. (TLT) phenomenon in the context of recommender systems. Here matrix factorization approach has been used to capture the impact of the included parameter to calculate the customers' tastes of the products. By training the customer's liking in such a way that would cause customer's attraction towards long tail items and hence promoting recommender system to recommend more of long tail items to specific customers.

Adam et al. suggested to construct a new methodology to recognize a dataset, such that the collection of reiterated item-sets represents an important diversion from expected arbitrary dataset, hence these item-sets can be distinguished as crucial. This methodology focuses on a “Poisson approximation” of the distribution of the quantity of reiterated item-sets of a proposed size, which is the main outcome of the paper. A significant feature of this method is that, it takes into consideration the complete dataset rather than individual discoveries, (Bipul Kumar, 2016) (Adam Kirsch, -)hence it is able to differentiate between important observations and arbitrary deviations in data, thus generating fewer fake discoveries.

3. CONCLUSION

In this paper, using data mining approach we tried to collect data of product that leads to high profit to the customers which can be applied to do e-trade. From all the discussed papers we have come across all the methods and techniques to earn out high profit over item sets by applying any algorithms required. After the analyses of these item sets customers can directly get their items.

REFERENCES

- Adam Kirsch, M. M. (-). A Rigorous Statistical Approach for. *Copyright 200X ACM XXXXXXXXX/*, 1-2.
- Akazue M. I., E. E. (2016). ENHANCING THE SUITABILITY OF HIGH VALUED COST ITEMS IN. *SAU Sci-Tech. J, 2016, 1(1) ISSN: 2536 - 6866*, 1-3.
- Badrul Sarwar, G. K. (2014). Analysis of Recommendation Algorithms for E-commerce. *University of Minnesota*, 1-2.
- Bhosale, S. P. (2014). Mining High Beneficial Itemsets from Transactional. (*IJCSIT*) *International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 6902-6904, 6902-6904*.
- Bipul Kumar, P. K. (2016). Fattening The Long Tail Items in E-commerce. *Received 7 June 2016; received in revised form 8 April 2017; accepted 27 May 2017*, 1-2.
- Bo Xiao, I. B. (2007). E-commerce product recommendation agents use, characteristics, and impact. *MIS Quarterly Vol. 31 No. 1*, 137-140.
- Dirk Thorleuchter, D. V. (2012). Predicting E-commerce company success by mining the text. *Expert Systems with Applications 39 (2012) 13026–13034, 13026-13027*.
- Ganesh Sawant¹, B. K. (2014). Efficient Techniques for Mining High Utility Itemsets from Transactional Databases:. *ISSN (Online): 2319-7064*, 1-3.
- Jian Wang, Y. Z. (2011). Utilizing Marginal Net Utility for Recommendation in. *University of California, Santa Cruz*, 1003-1004.
- Jyothi Pillai, O. (2013). TRANSACTION PROFITABILITY USING HURI. *International Journal of Business Information Systems Strategies (IJBISS) Volume 2, Number 1*, 1-3.

Markus Stolze, M. S. (-). Utility-based Decision Tree Optimization:. *IBM Research, Zurich Research Laboratory, Säumerstr. 4,, 1-2.*

Venkateswari S, S. R. (Apr 2011). ASSOCIATION RULE MINING IN ECOMMERCE:. *ISSN : 0975-5462 Vol. 3 No. 4, 4.*

Vikram Goyal, S. D. (-). *Indraprastha Institute of Information Technology, Delhi, 1-2.*

Yussuf Abu Shaaban, J. H. (2009). Cost-based admission control for Internet Commerce QoS enhancement. *Electronic Commerce Research and Applications 8 (2009) 142–159, 141-145.*