Top-K Competitor Trust Mining and Customer Behavior Investigation Using Data Mining Technique

Nisha Joseph
M.Phil Scholar, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu, India

B. Senthil Kumar
Associate Professor, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu, India

Abstract – In the recent scenario web data and shopping domain is a rising domains. There are more and more applications and companies provide online services nowadays. However, the rapid growth of online shopping information imposes an increasing challenge for users who have to choose trusted sellers and effective seller choices from a large number of available records from ecommerce applications to improve the business. Several studies failed to perform the multiple attribute selection to mine the competitors. The proposed system performs an adaptive clustering system with competitor and trust management which combines customer buying similarity along with the interpersonal influence. The proposed system Identifies and analyzes customer behavior to investigate the value of stable analysis. To achieve this, the proposed CTMiner and utilized the data mining concepts to find the interpersonal and intra personal similarities of products and customers to mine the competitor from the sales log. Identifying customer's usage patterns and grouping the customers, who are all have been in a same cluster is also performed. The proposal applied a new CTMiner system on the ecommerce application with computational overhead reduction and adaptive grouping. The proposed system is experimented with both real time and synthetic datasets for dynamic clustering results analysis. The results show the proposed system finds the competitor by effectively grouping the customers and products by similarity detection. The Top-K competitors are generated with the trust value from the review datasets. The proposed work adopts the adaptive Top-K sub clustering to improve the detection accuracy with dynamic and unstructured dataset.

Index Terms – Competitor Mining, Customer Behavior Analysis, Clustering, Classification, Trust Management.

1. INTRODUCTION

Data mining is the method to handle huge datasets and retrieve useful patterns from those data's. In the current scenario like social networks and other applications are using data mining techniques to improve the data management. There are huge number of techniques and methods available in data mining for various applications. The proposed system is comes under business data mining, which helps to improve the ecommerce and business solutions by analyzing and finding the competitors. The most appropriate data mining techniques in the field of business intelligence and ecommerce like applications are: prediction, association mining and recommendation etc. The key significance of identifying and monitoring business competitors is an effective research area, which helps to handle several business challenges. Checking and distinguishing business competitors have not much considered in the previous studies. Data mining is the ideal method for dealing with such tremendous data's for mining competitors. The proposed system allows the business person to know their competitors and customer to get the trusted and best sellers. This will be very helpful to recommend and analyze the product and the appropriate sellers from the web. In any case, it is for the most part hard to see all reviews in various sites for focused items and get appropriate proposals easily. In the existing studies, many authors investigated such enormous client data analysis and competitor mining to improve the business [1] [2].

The proposal provides a study of developing to find optimal sellers and their competitors for ecommerce application and Retail applications. This utilizes the Retail market data and results from the different similarity features. The proposal introduces a new superfine algorithm which combines the Machine learning and Recommender system [3] to overcome the seller selection challenge in ecommerce applications such as mining top-k similarity measure problems. The system initially analyzed the drawbacks of the existing competitor mining and ecommerce applications and develops two different approaches based on machine learning approaches like clustering and the other one is trust alert System, which is for capturing and suggesting the trusted seller from large market dataset and reviews.

The proposed system aims to make competitor mining and seller trust analysis from the large unstructured database [4]. This aims to improve the detection accuracy and recommendation performance. This mainly aims to resolve the large unstructured data clustering problem. And finally helps to find accuracy in competitor mining. This aims to improve the recommendation system by selecting trusted seller based on review log. The proposed system helps to resolve the problem of providing an automatic analysis of competitor and competitor alert system. The followings are the research aim, which have been considered. Aims to improve the clustering

performance by applying adaptive methods, the proposed system aims to develop a Trust and competitor Miner CTMiner, which overcomes the earlier work of CMiner [5]. And finally finds the similarity between users, products and competitors iteratively to cluster. The proposed competitor and trust mining framework and recommendation system helps to identify the healthy and top-k competitors. There are many technical and domain challenges intrinsic in designing and implementing an effective competitor and rust mining framework for retail and ecommerce applications. This will create a multi feature generation problem and clustering problem. To overcome the challenges in the competitor mining, the current work proposed a superfine model named as CTMiner. This will be useful for the user to select trusted service from the large ecommerce business providers.

2. PROBLEM DEFINITION

Many researches was conducted the experiments on item feature extracting data and competitor analysis [6]. The problem of automatically extracting data records that are related to the user given may have two types of documents like structured and unstructured. Handling unstructured dataset in the web repository may always create many challenges. The recent survey on unstructured data handling techniques is summarized in [7]. This method performs a novel data extraction by means of identifying the data regions and merging followed by segmentation and query result set identification of the records. The extracted data should be converted into structured one and nested structures are identified. Even though the earlier work CMiner++ provides good result, it still limits in few cases like domain specification, data handling and dynamic data management issues.

3. PROPOSED SYSTEM

An alternative approach for CTMiner system with similarity computation of the customer buying behavior based on different factors has been introduced. The proposed scheme improves upon that in existing CMiner and CMiner++ in terms of reliability, efficiency and accuracy. In particular, the proposed scheme depends on the existing CTMiner systems, which significantly reduces communication and computational costs.

3.1 PROPOSED SYSTEM

Rapid growth of information generated by online ecommerce applicationss leads to increases in demand of effective competitor mining systems to improve the business. Traditional techniques become unqualified because they do not consider data of social relation in the ecommerce applications for giving topK competitors; existing competitor mining techniques consider only the feature similarity rather than the performance and other features. This is noteworthy and challenging to fuse social contextual factors which are derived from customers' motivation of social activities into social recommendation. With the introduction and popularity of ecommerce applications, ever more customers like to share their real life experiences, such as blogs, ratings, feedbacks and reviews. New latest aspects of ecommerce application like interpersonal influence and interest based on same feature and challenges for CTMiner system to resolve the problem of mining top k competitors process; the system introduces a new adaptive Competitive Miner system. The proposed work performs the following two processes.

- CTMiner system
- Adaptive clustering of customers with their buying behaviors
- Trust alert process

The trust alert system produces an alert for the customer regarding the trust seller information. Using this output, the user can know the trusted seller and the seller can know their competitor. The competitor mining process is performed using effective data mining and similarity calculation process.

3.2 Contributions:

This proposes adaptive clustering system with competitor and trust management which combining customer buying similarity along with the interpersonal influence. The thing of customer individual interest makes associations between customer and product with hidden features. The followings are the contributions of the proposed work.

- Identifying and analyzing customer behavior to investigate the value of stable analysis.
- The proposed CTMiner utilizes interpersonal and intra personal similarities from the sales log.
- Identifying customer's usage patterns and grouping the customers, who are all have been in a same cluster.
- Appling a new CTMiner system with computational overhead reduction and adaptive grouping.

The main contributions of this paper are summarized above. Such features of products and services make connections between customer and other customer using the same the feature vectors. So this proposes and utilizes sales log by enforcing customer personal interests. Identifying personal unique interest is modeled to get an accurate model for the above mentioned clustering problem, here the clustering customer and finding the competitors from their reviews and calculating the trust is important. The influence of the proposed system with different performance metrics has been modeled.

3.3 CTMINER Framework:

There are some potential problems with the existing Competitor mining System. One is the scalability, which is how

quickly a competitor mining system can generate accurate results second one is K size specification and also clustering Problem and better accuracy. The CTMiner framework overcomes the problems of the existing system. The following algorithm gives the steps to calculate the adaptive clustering and competitors from the huge dimensional datasets.

Algorithm 1 CTMiner

Input: Set of products *I*, Product of interest $i \in I$, feature space *F*, reviews F.

Collection $Q \in 2F$ of queries with non-zero weights, Adaptive subspace Cluster *DI*, int *k*

Output: Set of top-*k* competitors for *I and trust score T*

Step1: $ATOPK \leftarrow masters(i)$

Step 2: if ($k \leq |ATOPK|$) then

Step 3: return *ATOPK*

Step 4: end if

Step 5: $k \leftarrow k - |ATOPK|$

Step 6: $LB \leftarrow -1$

Step 7: $X \leftarrow$GETSIM($ATOPK;DI$) $\cup DI[0]$

Step 8: while ($|X|$!= 0) do

Step 9: X UPDATEATOPK($k;LB;X$)

Step 10: if ($|X|$!= 0) then

Step 11: *ATOPK* MERGE(*ATOPK;X*)

Step 12: if ($|ATOPK| = k$) then

Step 13: *LB* WORSTIN(*ATOPK*)

Step 14: end if

Step 15: *X* GETSIM(*X;DI*)

Step 16: end if

Step 17: end while

Step 18: return *ATOPK*

Step 19: Routine UPDATEATOPK*(k, LB, X)*

Step 20: $localATOPK \leftarrow \emptyset$

Step 21: $low(j) \leftarrow 0; \forall j \in X.$

Step 22: $up(j) \leftarrow \Sigma q2Qp(q) \times V\, qj;j\ ; \forall j \in X.$

Step 23: for every $q \in Q$ do

Step 24: $maxV \leftarrow p(q) \times V\, q\ i;i$

Step 25: for every product $j \in X$ do

Step 26: $up(j) \leftarrow up(j) - maxV + p(q) \times V\, q\ i;j$

Step 27: if ($up(j) < LB$) then

Step 28: $X \leftarrow X \setminus \{j\}$

Step 29: else

Step 30: $low(j) \leftarrow low(j) + p(q) \times V\, q\ i;j$

Step 31: *localATOPK:update*($j;\ low(j)$)

Step 32: if ($|localATOPK| \geq k$) then

Step 33: $LB \leftarrow$WORSTIN(*localATOPK*)

Step 34: end if

Step 35: end if

Step 36: end for

Step 37: if ($|X| \leq k$) then

Step 38: break

Step 39: end if

Step 40: end for

Step 41: for every product $j \in X$ do

Step 42: for every remaining $q \in Q$ do

Step 43: $low(j) \leftarrow low(j) + p(q) \times V\, q\ i;j$

Step 44: end for

Step 45: *localATOPK:update*($j;\ low(j)$)

Step 46: end for

Step 47: return ATOPK(*localATOPK*)

The CTMiner Algorithm: It is an exact algorithm for finding the top-k competitors of a given item by analyzing the different similarity calculations. The algorithm makes use of the adaptive subspace clustering in order to reduce the number of items that need to be considered. Given that it only cares about the top-k competitors, this can incrementally compute the score of each candidate and stop when it is guaranteed that the top-k has emerged. The pseudocode is given in Algorithm 1.

4. IMPLEMENTATION AND RESULTS

4.1 Dataset:

The system mainly deals with the problem of clustering and data scanning problem to mine the competitor. So this need a set of data related to the ecommerce or retail sales logs. The data collected from UCI repository websites, but some attributes are not available in the website, so the experiments additionally taken the synthetic dataset which is described below.

The below dataset shows the retail sales dataset collected from the UCI repository. This has been taken as a synthetic dataset for the current implementation. The implementation may

consist n number of products sales information's for experiments. CTMiner has taken 500 product details as initial dataset. This can be expanded up to 5 lacks tuples. As like the above dataset the system has another important dataset which creates more than 100 customers ecommerce purchase log. The second dataset has collected from the website developed with the ecommerce options, which saves the customer feedback for every product. Using these dataset, the competitor and trust sellers are gathered.

A	B	C	D	E	F	G	H	I
InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	Customer	Country	
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850	United Kingdom	
536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850	United Kingdom	
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850	United Kingdom	
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850	United Kingdom	
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850	United Kingdom	
536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12/1/2010 8:26	7.65	17850	United Kingdom	
536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/1/2010 8:26	4.25	17850	United Kingdom	
536366	22633	HAND WARMER UNION JACK	6	12/1/2010 8:28	1.85	17850	United Kingdom	
536366	22632	HAND WARMER RED POLKA DOT	6	12/1/2010 8:28	1.85	17850	United Kingdom	

Figure 1.0: Retail dataset with 5 lakh entries

productno	productname	dealername	b.	bdate	comp	compdet	companyname
2	camera	snapdeal	a..	1/7/2016 5:...	Not Delivered	not yet received	snap
1	camera	seattle	a..	12/12/2016 ...	Poor Service	poor service not good	Pooma
2	camera	snapdeal	a..	1/7/2016 6:...	Poor Service	poor	snap
1	camera	seattle	a..	1/7/2016 6:...	Damage	gh	Pooma
3	mobile	amazon	tt	4/5/2016 7:...	Not Delivered	not yet delivered	amazon

Figure 2.0: Ecommerce Synthetic data

The experiment uses the synthetic data sets for experiments. In particular, this creates synthetic data sets with customer code and ecommerce log detail with the reference from the literature. The system can have n number of tuples for experiments. After the synthetic data set is generated, and given the number of m local sites, each tuple from the synthetic uncertain database D is assigned to site Si chosen uniformly. Clearly, all local sites have the same data distribution. In particular, a local site server keeps a random sample set of the underlying data set, and the sample sets are mutually disjoint. In the experiments, every local server possesses an equal number of points, named the local cardinality.

4.2. Results:

In each category of e ecommerce dataset, we use 80% of data as the training set and the remaining 20% as the test set. This is a synthetic dataset where the count can be increased by running the application every time. This chapter shows the results of every process in CTMiner. As per experiments and analysis the following chart describes the performance difference between existing and proposed systems in terms of time, accuracy and scope along with ranking efficiency.

Table 1.0 Comparison table.

Model name	**Time delay**	**Accuracy**	**scope**	**Top_K ranking efficiency**
Existing(CMiner)	7.5	7	6.5	4.5
Proposed(CTMiner)	5.5	9	8.9	9.2

The table 1.0 represents the time delay, accuracy, scope and top k ranking efficiency comparison between existing Cminer and Proposed CTMiner.

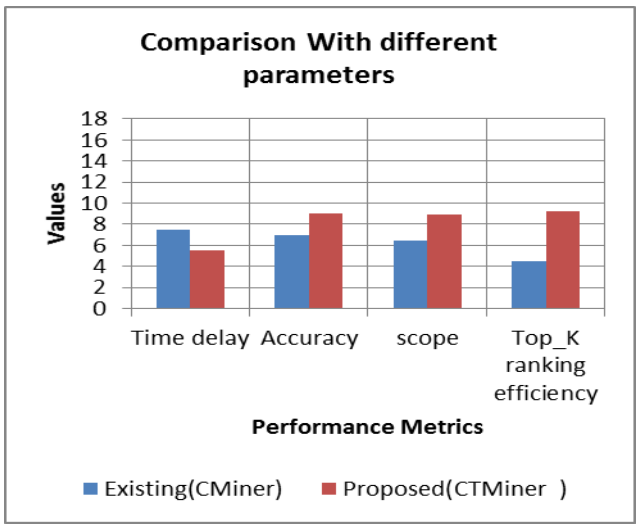

Figure 3.0 comparison chart

The above figure 3.0 shows the overall performance comparison between existing and proposed systems.

5. CONCLUSION

A personalized recommendation and alert approach has been proposed by combining several real time factors known as buying details and gathered from the dag. This overcomes the clustering customer details the problem, which is common in recommender system. To overcome the above issues, the system implemented with CMiner techniques. This approach is implemented in an e-ecommerce dataset with relevant similarity measurement phases. This measures individuality of rating items with the reference of experienced customers with various factors. At present the personalized recommendation model in the literature only takes interpersonal relationship and customer's historical rating records. In the proposed CTMiner

the system takes the area information and other attributes to recommend more personalized and real-time items to the customers.

REFERENCES

[1] Apte, Chidanand, Bing Liu, Edwin PD Pednault, and Padhraic Smyth. "Business applications of data mining." *Communications of the ACM* 45, no. 8 (2002): 49-53.

[2] Heinrichs, John H., and Jeen-Su Lim. "Integrating web-based data mining tools with business models for knowledge management." *Decision Support Systems* 35, no. 1 (2003): 103-112.

[3] Adomavicius, Gediminas, and Alexander Tuzhilin. "Context-aware recommender systems." In *Recommender systems handbook*, pp. 191-226. Springer US, 2015.

[4] Blumberg, Robert, and Shaku Atre. "The problem with unstructured data." *Dm Review* 13, no. 42-49 (2003): 62.

[5] Valkanas, George, Theodoros Lappas, and Dimitrios Gunopulos. "Mining Competitors from Large Unstructured Datasets." *IEEE Transactions on Knowledge and Data Engineering* (2017).

[6] Li, Rui, Shenghua Bao, Jin Wang, Yong Yu, and Yunbo Cao. "Cominer: An effective algorithm for mining competitors from the web." In *Data Mining, 2006. ICDM'06. Sixth International Conference on*, pp. 948-952. IEEE, 2006.

[7] Kumar, B. Senthil, and Nisha Joseph. "A Review on Competitor Mining and Unstructured Dataset Handling Techniques." *Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org* 7, no. 9 (2017).