Data Mining in E-Commerce for Electronics Products

Manpreet Kaur[1], Jyoti Arora[2]

[1,2]Desh Bhagat University, Mandi Gobindgarh

Abstract: *We are going to discuss improved mining strategies which are required to maintain optimized website structure which in turn is helpful for increase businesses, to keep check on competitor's websites, comparison of various brands, attracting new customers and to retain the old customers. Each product are classified. The classification of products and services enables reliable and efficient electronic exchanges of product data across organizations. Products are classified with Dataset, Category Score, Tokenization and Keyword Classification.*

Keywords: Data mining, E-commerce, Web mining, Web ontologies, Classification

1. Introduction

Information about the mining could be the technique of inspecting & gather info by different perspectives along with outlining the idea in beneficial details. The idea details helpful to improve income, reduces prices, or both equally. It really is used for gather the data by different- different web sites. Information about the mining could be

the technique of finding information within big relational listings. The item a good choice for end users to handle the information effortlessly. Process for finding files habits undetectable inside significant files units.

Digital trade, often called E-commerce/ecommerce, will be trading inside offerings using pc communities, such as Net. Digital trade attracts on technology such as cellular trade, automated funds transport, source cycle supervision, and Affiliate marketing, on-line deal finalizing, automated files interchange (EDI), stock supervision techniques and also automatic files series techniques.

2. Literature Survey

2001 Ed H. Chi, Peter Pirolli, Kim Chen, James Pitkow have proposed on the Web, users typically forage for information by navigating from page to page along Web links. Their surfing patterns or actions are guided by their information needs. Researchers need tools to explore the complex interactions between user needs, user actions, and the structures and contents of the Web. In this paper, they describe two computational methods for understanding the relationship between user needs and user actions. First, for a particular pattern of surfing, we seek to infer the associated information need. Second, given an information need, and some pages as starting points, they attempt to predict the expected surfing patterns. The algorithms use a concept called ─information scent", which is the subjective sense of value and cost of accessing a page based on perceptual cues. They present an empirical evaluation of these two algorithms, and show their effectiveness.

2015 D. Ramamonjisoa, R. Murakami, B. Chakraborty have proposed the number of user-contributed comments is increasing exponentially. Such comments are found widely in social media sites including internet discussion forums and news agency websites. In this paper, they summarize the current approaches to text analysis and the visualization tools which deal with opinion and topics mining of those comments. They then describe experiments for topic modeling on users' comments and examine the possible extensions of methods on visualization, sentiment analysis and opinion summarization systems.

2015 Yongyi Cheng, Yumian Yang, Jianhua Jiang and GaoChao Xu have proposed with the rapid development of E-Commerce, how to evaluate the E-Commerce sites accurately has become an important issue. However, to cluster E-Commerce sites correctly and accurately is not an easy thing based on characteristics of high dimensions and uneven density for E-Commerce sites. This leads to bad performance of the cluster result.

2012 Belsare Satish and Patil Sunil have proposed Data mining has matured as a field of basic and applied research in computer science. The objective of this dissertation is to evaluate, propose and improve the use of some of the recent approaches, architectures and Web mining techniques (collecting personal information from customers) are the means of utilizing data mining methods to induce and extract useful information from Web information and service where data mining has been applied in the fields of e-commerce and e-business (that means User's behavior). In the context of web mining, clustering could be used to cluster similar click-streams to determine learning behaviors in the case of e-learning or general site access behaviors in e-commerce. Most of the algorithms presented in the literature to deal with clustering web sessions treat sessions as sets of visited pages within a time period and do not consider the sequence of the click-stream visitation. This has a significant consequence when comparing similarities between web sessions. Wang and Zaiane propose an algorithm based on sequence alignment to measure similarities between web sessions where sessions are chronologically ordered sequences of page accesses.

2015 Prof. Vishal Parikh, Parth Shah have proposed analyses content based recommendation for e-commerce site. Recommendation system use to generate

Paper ID: SUB159265

1255

recommendation of the product that customer may want to buy. This system increase the sale of vendor and easy to find product from available product. Association rule mining and clustering technique use to make real time recommendation system. From the user's transaction data-set they can generate rules for customer buying tendency. Based on customer purchased product and customer profile, they can generate recommendation using association rule mining technique. Association rule mining is very time consuming process for large data-set. So, it is not feasible for real time recommendation system. To overcome this problem clustering technique is used. Using hierarchical clustering we can make partition of whole large data-set in to tree of clusters. It decrease the time for real time recommendation system.

3. Present Work

We propose the use of a rich set of Product Classification features like positive, negative, automatic product feature classification. The proposed feature selection method can improve product classification performance. The proposed Feature Relation Network is rule-based product classification method that finds the product features category and subcategory from the given dataset. The proposed algorithm consisted of four basic components: Dataset, Category Score, Tokenization, Keyword classification for each category.

Dataset: The first step is used to read the products saved in the excel file. The key features are read and classified.

Tokenization: In this module, the program reads the known words from the user comment based on the word list prepared on the basis of most common words. This word is list is obtained by matching the words in the user comments along with a valid word list file already saved as a text file. The text file is loaded into the memory and passed to the tokenization process for the further computations. The process of tokenization then extract all of the words from the user comment and filter them on the basis of a list of common words containing no features. The filtered list is obtained after deleting those matching words in the user comment. These common words are not given in the word weight file containing the rank/weight of each word being used in the common English language, which contains a neutral, positive or negative emotion.

Category Score: In this, the product's entire name match with the dictionary. The Dictionary contains all company names with their category. When the product matches to particular company name, then it will be categorized according to company's product category. If any company working with two categories, then it will be subcategorized further.

Keyword Classification: In this, every possible keyword for each category has own score. First all keywords of product features break down into separate. After breaking down each keyword matches with keyword score. After that total score will be calculated on the basis of keyword matching criteria. This way, product subcategory will be found.

Algorithm 1: Brief Design of **P**roduct **Classification for E-commerce Platforms (PCEP)** Algorithm

1. Obtain the data from the e-commerce websites.
2. Extract the list of products P from the e-commerce websites' products.
3. Extract N number of words from feature list M using dictionary based tokenization
4. Filter message content with STOPWORD list of common English words while Tokenization
5. Load product and feature classification (PFC) knowledge data
6. Classify the product after comparing it with the PFC data
7. Acquisition of the feature classification (SEC) knowledge data
8. Calculate the feature score after comparing it with SEC data
9. Classify the feature according to the score and increment the product or product feature index accordingly
   a. If score is more than zero
      i. Increment the positive index
   b. If score equals zero
      i. Increment the neutral index
   c. If score is less than zero
      i. Increment the negative index
10. Load product PFC knowledge data
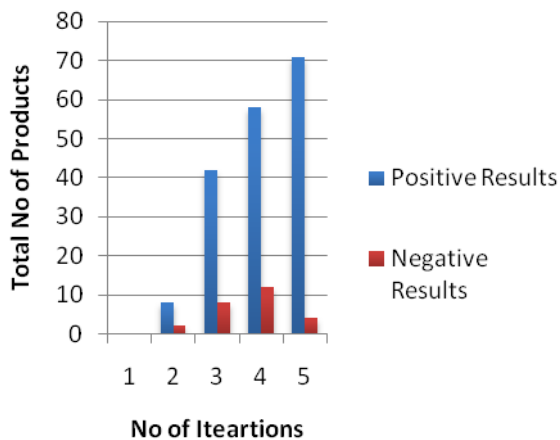11. Prepare the classification report according to product category and subcategory

Workflow Diagram
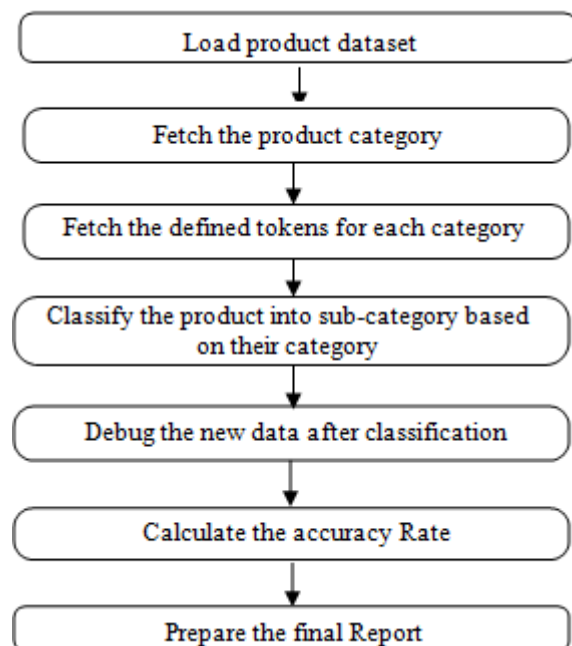
4. Graphical Analysis

The graphical analysis of the results has been given in the following figures:

Paper ID: SUB159265

1256

Comparison b/w Positive and Negative Results

Analysis of various numbers of products

Total Number of Products	Positive Results	Negative Results
10	8	2
50	42	8
70	58	12
75	71	4

5. Conclusion & Future Work

In the future, the proposed algorithm can be enhanced to calculate product features of more products or with a wider range. The above categories can be probably calculated using the dictionary based phrase specification methods. Also, the proposed algorithm can be improved on the basis of execution time and accuracy.

References

[1] K. R. Suneetha, Dr. R. Krishnamoorthi ―Identifying User Behavior by Analyzing Web Server Access Log File.‖ IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.4, April 2009.

[2] Duen-Ren Liu, Ya-Yueh Shih ―Integrating AHP and data mining for product recommendation based on customer lifetime value.‖Information & Management 42 (2005) 387–400.

[3] Ron Kohavi, Llew Mason, Rajesh Parekh, Zijian Zheng ―Lessons and Challenges from Mining Retail E-Commerce Data.‖Machine Learning, 57, 83–113, 2004.

[4] Minqing Hu, Bing Liu ―Mining Opinion Features in Customer Reviews.‖ 2014.

[5] Anol Bhattacherjee ―Acceptance of E-Commerce Services: The Case of Electronic Brokerages.‖ IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, Vol. 30, No. 4, July 2000.

[6] Bamshad Mobasher ―Data Mining for Web Personalization.‖The Adaptive Web, LNCS 4321, pp. 90–135, 2007.

[7] Ron Kohavi ―Mining E-Commerce Data: The Good, the Bad, and the Ugly.‖ KDD 2001's Industrial Track.

[8] Zan Huang, Wingyan Chung, and Hsinchun Chen ―A Graph Model for E-Commerce Recommender Systems.‖ Journal of the American Society for Information Science and Technology, 55(3):259-274, 2004.

[9] Rakesh Agrawal, Amit Somani, Yirong Xu ―Storage and Querying of E-Commerce Data.‖ Proceedings of the 27th VLDB Conference, Roma, Italy, 2001.

[10] Ed H. Chi, Peter Pirolli, Kim Chen, James Pitkow ―Using Information Scent to Model User Information Needs and Actions on the Web.‖ SIGCHI'01, March 31-April 4, 2001.

[11] D. Ramamonjisoa, R. Murakami, B. Chakraborty―Comments Analysis and Visualization Based on Topic Modeling and Topic Phrase Mining.‖ Proceedings of the Third International Conference on E- Technologies and Business on the Web,2015.

Paper ID: SUB159265