

A Review on Automatic Product Classification for the E-Commerce Portals

Manpreet Kaur¹, Jyoti Arora²

^{1,2}Desh Bhagat University, Mandi Gobindgarh, India

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 products are classification. The classification of products and services enables reliable and efficient electronic exchanges of product data across organizations.

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

1. Information

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified.



Electronic commerce, commonly known as E-commerce/ecommerce, is trading in products or services using computer networks, such as Internet. Electronic commerce draws on technologies such as mobile commerce, electronic funds transfer, supply chain management, Internet marketing, online transaction processing, electronic data interchange (EDI), inventory management systems and automated data collection systems.

Now days E-Commerce sector is growing in various dimensions. It is considered as fastest growing sector across the globe. Several E-Commerce statistics shows that now India is moving form nascent stage to more refined and advanced stage. There are many reasons for growth of E-commerce such as busy lifestyle, increased buying power, high computer literacy rate, feedback availability of products, cash back and easy exchange policies and many discounts offers. Now day's money transaction over Internet is easy and reliable. It is right to say that some of the online

businesses are well liked because of their well-structured and user-friendly websites. To concentrate on an effective and optimized website is a real challenge for this industry.

One of the growing and associate research area is Web Mining. Web Mining is helpful to extract useful information from user's surfing pattern on the web. With the help of these patterns, we can make the website more easy to use. Improve availability of information for the products. Web mining can be categorized as follows:

Web Content Mining means extracting the knowledge from content of web pages. It can be used for web page content improvement, quality of search result improvement etc.

Web Usage Mining deals with usage records of web pages. The log records are carefully examined and the extracted knowledge is used in taking important organizational decisions.

Web Structure Mining is related with organization of various connected web pages with in a website or among the websites for easy retrieval of information to user. Web Structure Mining supports a strong base in website restructuring. [19][20]. In this paper we will discuss about how web mining can be deployed intelligently to get benefit in various dimensions of E-Commerce, which is not only useful for customer but also for data analysts to take various important decisions to better satisfy their organizational needs. Accordingly this paper is divided into various sections. Section II describes literature review, Section III explain about the research problem, Section IV illustrates the objectives of research. Section V describes the implementation architecture of web mining with E-Commerce. Section VI gives clear understanding of research methodology, proposed and innovative Apriori-IC algorithm, followed by graphical analysis.

2. Literature Survey

2014. Stolz, Alex et. al. has proposed a web system known as PCS2OWL to automatically classify the products online. The authors have proposed the generic approach based on deriving web ontologies from product classification systems. In this paper, the authors have described a generic, semi-automated method for deriving OWL ontologies from

product classification standards and proprietary category systems. Also, they have shown that our approach generates logically and semantically correct vocabularies, and have presented the practical benefit of our approach. The resulting product ontologies are compatible with the GoodRelations vocabulary for e-commerce and with schema.org and can be used to enrich product and offer descriptions on the Semantic Web with granular product type information from existing data sources.

2014. Nederstigt, Lennart J. et. al. has implemented the system names FLOPPIES for ontology population for product classification. The authors have built a framework for large-scale ontology population of product information from tabular data in e-commerce stores. This paper proposes FLOPPIES, a framework capable of semiautomatic ontology population of tabular product information from Web stores. By formalizing product information in an ontology, better product comparison or parametric search applications can be built, using the semantics of product attributes and their corresponding values. The framework employs both lexical and pattern matching for classifying products, mapping properties, and instantiating values

2014. Sun, Chong et. al. has proposed the system named as Chimera for the classification on Large-scale data using machine learning, predefined rules, and crowdsourcing. In this paper the authors have described Chimera, their proposed solution to classify tens of millions of products into 5000+ product types at WalmartLabs. They have also shown that at this scale, many conventional assumptions regarding learning and crowdsourcing break down, and that existing solutions cease to work. The authors have described how Chimera employs a combination of learning, rules (created by in-house analysts), and crowdsourcing to achieve accurate, continuously improving, and cost-effective classification.

2014 Bhargav, Anshul, and Munish Bhargav [1] have worked on pattern discovery and users classification through web usage mining. The proposed framework is based on three steps. In the first step, preprocessing is done to remove useless data from web log file so as to reduce its size. In the second step, this cleaned log file is used for discovering usage patterns. Finally, the discovered patterns lead to the classification of users: on the basis of countries; on the basis of direct entry to the site or referred by the other site; on the basis of time of access, i.e., either different seasons or different months or different days. This information can then be used by the website administrators for efficient administration and personalization of their websites and thus the specific needs of specific communities of users can be fulfilled and so the profit can be increased.

2013 Eltahir, Mirghani, and Anour FA Dafa-Alla [2] have worked on extracting knowledge from web server logs using web usage mining. The main problem that faces any website admin or any web application system is data increase per-second, which is stored in different types and formats in server log files about users, their future needs and maintains the structure and content of website or web services according to their previous data. Web usage mining aims at discovering useful information or knowledge from usage data registered in log files, based on primary kinds of data

used in the mining process. By using one of the web mining techniques, this paper uses a web usage mining technique to procure knowledge from web server log files where all user navigation history is registered.

2014 Gupta, Ashika, Rakhi Arora, Ranjana Sikarwar, and Neha Saxena [3] have proposed a method for web usage mining using improved Frequent Pattern Tree algorithms. Web usage mining itself can be categorised further depending on the kind of usage data considered they are web server, application server and application level data. This Research work focuses on web use mining and specifically keeps tabs on running across the web utilization examples of sites from the server log records. The bonding of memory and time usage is compared by means of Apriori algorithm and improved Frequent Pattern Tree algorithm.

2014 Sharma, Murli Manohar, and Anju Bala [4] have proposed an algorithm for frequent access pattern identification in web usage mining. In the field of web mining the analysis of the web logs is done to identify the user search patterns. In the usual approaches of finding the patterns, pattern tree is created and then the analysis is done, but in this proposed algorithm there is no need of tree creation and the analysis is done based on the website architecture, which will increase the efficiency of the other pattern matching algorithms and needs only one database scan.

3. Problem Formulation

Product classification is the technique to filtering the products according to their types. Products are further classified into their sub categories under main category. Benefits of the product classification in e-commerce websites are 1) To Automate listing the products 2) Easy to search for the particular product by the user 3) Easy to maintain the inventory.

The existing system is working based on ontology product classification. It fetches data from ecommerce websites and classifies the products according to their categories. There is a wide scope to improvise the existing model. It can be enhanced in such a way to classify the products into their category and further categories can also be categorized. Also, in the new system we can propose the accuracy rate for the classification.

4. Proposed Work

The proposed work will be based upon the classification of the products based upon the product's category. The proposed work will be using the large products based database for the purpose of tokenization, which is the process to fetch the product's category out of the database. The proposed model would be able to classify the product's category and their subcategory based upon the product's database. The proposed model will also offer the accuracy results and a human-software intelligent interaction, which can help the machines to learn the new rules for the higher accuracy than the existing system.

We are proposing the use of a rich set of product analysis like Electronic, Food, Books, Health care, Cosmetic etc. 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's category and subcategory. The proposed algorithm consisted of four basic components: Product Category, Tokenization, Product Model and sub-category Analysis.

5. Conclusion

In which products are divided into different- different category and subcategory. Through category classification, customers check easily about the products. We implement only on some product category. In future we can increase the speed for finding about the products and also search all world products easily.

References

- [1] Stolz, Alex, Bene Rodriguez-Castro, Andreas Radinger, and Martin Hepp. "PCS2OWL: A generic approach for deriving web ontologies from product classification systems." In *The Semantic Web: Trends and Challenges*, pp. 644-658. Springer International Publishing, 2014.
- [2] Nederstigt, Lennart J., Steven S. Aanen, Damir Vandic, and Flavius Frasincar. "FLOPPIES: A Framework for Large-Scale Ontology Population of Product Information from Tabular Data in E-commerce Stores." *Decision Support Systems* 59 (2014): 296-311.
- [3] Sun, Chong, Narasimhan Rampalli, Frank Yang, and AnHai Doan. "Chimera: Large-scale classification using machine learning, rules, and crowdsourcing." *Proceedings of the VLDB Endowment* 7, no. 13 (2014).
- [4] Bhargav, Anshul, and Munish Bhargav. "Pattern discovery and users classification through web usage mining." In *Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference on*, pp. 632-636. IEEE, 2014.
- [5] Eltahir, Mirghani, and Anour FA Dafa-Alla. "Extracting knowledge from web server logs using web usage mining." In *Computing, Electrical and Electronics Engineering (ICCEEE), 2013 International Conference on*, pp. 413-417. IEEE, 2013.
- [6] Gupta, Ashika, Rakhi Arora, Ranjana Sikarwar, and Neha Saxena. "Web usage mining using improved Frequent Pattern Tree algorithms." In *Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on*, pp. 573-578. IEEE, 2014.
- [7] Sharma, Murlu Manohar, and Anju Bala. "An approach for frequent access pattern identification in web usage mining." In *Advances in Computing, Communications and Informatics (ICACCI), 2014 International Conference on*, pp. 730-735. IEEE, 2014.
- [8] Neha Verma, Prof. (Dr.) Jatinder Singh "Improved Web Mining for E-Commerce Website Restructuring." *2015 IEEE International Conference on Computational Intelligence & Communication Technology*.