

Recognizing the fake reputation information, fair and healthy reputation system in E-Commerce

DesettiSrinuvasa Rao¹, Kesavarao Seerapu²

¹M.Tech Scholar(CSE) in Department of Computer Science and Engineering,

²Assistant Professor, Department of Computer Science and Engineering, Avanthi Institute of Engineering and Technology, Vizianagaram (Dist), Andhra Pradesh, India.

Received 26 November 2020; Accepted 10 December 2020

Abstract: E-commerce is all about carrying out business on the Web. It is about carrying out transactions, essentially buying and selling products and services by consumers and businesses respectively, on the web. Web is one of the largest sources of information, collection of many files stored in different web servers and its size is also growing rapidly. E-Commerce not only keep your business up and running but also make it tough to stand in the competitive world of e-business. E-commerce has provided a cost efficient and effective way of doing business in the web. Web mining is the application of data mining techniques to discover and extract useful and interesting information from the Web. User behavior on the web is based on web mining, security and e-commerce. Customer behavior pattern is analyzed to improve e-commerce websites. Different web mining algorithm and security algorithm provide security on e-commerce websites. Web mining algorithms such as like pagerank and trust rank are used to develop web mining framework in e-commerce website. Generally web mining framework is only based on the web content mining or web usage mining. In this proposed system web mining consist of web structure mining, web content mining, decision analysis and security analysis. Now a days, digitalization gaining popularity because of seamless, easy and convenience use of ecommerce. People choose online payment and eshopping; due to time convenience, transport convenience, etc. Because of that the frauds are increasing day by day. It is very important to detect such frauds and to find the solution to avoid these frauds.

Keywords: products, fraud, ecommerce.

I. INTRODUCTION

WUM deals with understanding user behavior in interacting with the web or with a website. One of the aims is to obtain information that may assist web site reorganization or assist site adaptation to better suit the user. Web usage mining model is a kind of mining to server logs and its aim is getting useful users' access information in logs to make sites can perfect themselves with appropriate users' requirements, serve users better and get more economy benefit. Web mining is the application of data mining techniques to extract knowledge from web data including web documents, hyperlinks between documents, usage logs of websites, etc. Web Usage Mining is a part of Web Mining which in turn, is a part of Data Mining. As data mining is the process of extracting meaningful and valuable information from large volume of data. Web usage mining is the process of mining useful information from server logs. Web usage mining is the process of finding out what users are looking for on internet. This information can then be used in a variety of ways such as, improvement of websites, e-commerce, website personalization, user future request prediction etc. The use of this type of web mining helps to gather the important information from customers visiting the site. This work focused on the web usage mining and identification of user's behavior on the web. The behavior of users on the web can be analyzed by extracting useful information from web log data. Web log file is automatically created and manipulated by every hit to the website. Log files usually contain noisy and irrelevant data. Preprocessing is done to remove unnecessary data from log file. After then pattern discovery and pattern analysis can be performed for extracting useful patterns. Such interested patterns can be generated using several techniques like classification, clustering, association rule mining. In this paper we deal with classification algorithms for studying the user/client behavior and for the generation of interested user patterns. Consideration of interested web users can be done on the basis of probability of relevant and irrelevant links. Relevant links are the most visited links that can be identified on the basis of time spend on a webpage or number of hits done to a particular link.

II. RELATED WORK

NachiketaSahoo et al. present a method to make personalized recommendations when user preferences change over time. When the user is observed over a long period of time a dataset on employees' blog reading behaviour is collected, which show that users' product selection behaviours change over time. So proposed a

hidden Markov model to correctly interpret the users' product selection behaviours and make personalized recommendations. He evaluates the model using three real world datasets that includes data on employee blog reading behaviour inside a firm, users' music listening behaviour collected through last.fm and users' movie rating behaviour at Netflix with different characteristics. Then comparison of the recommendation performance of the proposed model with that of a number of collaborative filtering algorithms and proposed prediction algorithm is done. He found that the proposed HMM based collaborative filter performs as well as the best among the alternative algorithms when the data is sparse or static. [3] Paula Cristina Vaz and David Martins De Matos presented a Hybrid Recommendation System for improving literary book recommendation system through author ranking. Here books are recommended for the book readers to decide which book to read next. It uses two item based collaborative filtering algorithms to predict books and authors that users will like. Author prediction is used to recommend the books.[11] Wahidah Hussain have given a Personalized Location Based Traveller Recommendation System that gives information to the travellers about the local area attraction such as local food, shopping spots etc.[12] A real Time Route Recommendation system is presented by Henan Wang and Guoliang Li. This system provides real time traffic aware routes that overcome the problem of recommending same routes for all users and do not take full advantage of real time traffic.[13]

Author Castellano et al.[10] proposed a new approach for data preprocessing of web logs. Here the author proves the pre-processing is the first stage of web usage mining. In particular, the author presents LODAP (LOGDATA Preprocessor), which is a software tool which they designed and implemented in order to perform preprocessing of log data. This tools allows creating reports containing the results obtained in each step and information summaries mined from the analysis of the considered log files. But the main limitation of this paper is, this only applicable for a particular website rather than the whole analysis. Author Ankit et al [11] has introduces many data pre-processing techniques helps tp identify unique users and user data session; this can be used to improve the performance features in web usage mining. The analysis of log files which can offer valuable insight into web site usage, those logs should be gathered properly. The given preprocessing techniques help to generate an useful log format from raw web log. Reddy, K. Sudheer, M. Kantha Reddy, and V. Sitaramulu[12] analyzed and summarized about various details about data preprocessing activities that are necessary to perform Web Usage Mining (WUM). The author shows the importance of data preprocessing in web mining applications. The experiment of the authors helps to estimate data preprocessing importance and finally they provided a new preprocessing approach. The approach increases the quality of the data available and reduces the size. But the authors not concentrated on the pattern discovery process from the preprocessed data.

III. RESEARCH METHODOLOGY

Selected e-commerce websites

We aimed to test a concise yet representational set of successful e-commerce portals. Ten websites were selected; five based on ranked popularity, and five on ranked sales conversion rates. The five most popular e-commerce websites, as sourced from the web service called Alexa, were (in order): Amazon, Ebay, Netflix, Walmart, and Ikea. The top conversion e-commerce websites, as defined by Nielson Company (Nicholls, 2010), were: Schwan's (41.7%), ProFlowers (26.5%), Vitacost (24.0%), Woman Within (22.4%), and Blair (20.5%).

Analysis method

This paper aims to explore, identify and understand the design of successful e-commerce websites by classifying their persuasive features. It also represents a new opportunity to learn from practice and understand how different persuasive features have been implemented by many successful ecommerce websites and how such implementation could be enhanced. Expert evaluation was used where one or more specialists evaluate the website against a checklist of list of persuasive design features. The checklist, presented in Table 1, was based on the 28 persuasion features defined by the PSD model (Oinas-Kukkonen and Harjumaa, 2009). Expert empirical evaluation was used to walk through regular tasks within specific websites, with functionalities checked against given definitions in the checklist. One evaluator made notes on an Excel sheet and recorded relevant comments. The evaluator sought other expert's opinions regarding the evolution of functionalities that are not clear and might be related to more than one persuasive feature. When the evaluator concluded the evaluation, another expert reviewed the recorded notes and comments against the utilised checklist and rigorous discussion and iterations took place to ensure that notes and comments were recorded against the right persuasive feature.

Table 1: Checklist for recognising persuasive features, adapted from Oinas-Kukkonen and Harjumaa (2009)

Persuasion Principle	Description and Characteristic
Social Role	Website adopts a role of social actor e.g. embedded conversational agent offering advice and assistance.

Trustworthiness	<i>Provides the customers with truthful, fair and unbiased information, e.g. website clearly states privacy, shipping and returning policies.</i>
Expertise	<i>Offers information showing knowledge, experience, and competence, e.g. website regularly updated, no dangling links or out of date information.</i>
Surface credibility	<i>Website has a competent look and feel, e.g. the design reflects the context.</i>
Real-world feel	<i>Provides information about organisation providing the services, e.g. provides clear customer service contact details like phone, post, or online chat.</i>
Authority	<i>Website refers to organisation or people in the role of authority, e.g. quoting authority statements from authorised organisation or people.</i>
Third-party endorsements	<i>Provides endorsements from respected sources e.g. security certificate logos from well-known security companies.</i>
Verifiability	<i>Provides means to verify the accuracy of the website content via outside sources.</i>
Social Learning	<i>Allows users to observe other customers' behaviour and outcomes, e.g. customer review board, showing what other users are experiencing, etc.</i>
Social Comparison	<i>Allows customer to compare own performance with other customers, e.g. customer review board.</i>
Normative Influence	<i>Provides means for gathering together customers who share the same goal or leverage, normative influence or peer pressure to increase the likelihood that a person will adopt target behaviour.</i>
Social facilitation	<i>Allows users to discern other users performing the same behaviour along with them, e.g. what others watch and buy right now.</i>
Cooperation	<i>Allows customers to cooperate, e.g. customer review board.</i>
Competition	<i>Provides means of competing with other customer, e.g. online auction.</i>
Recognition	<i>Provides public recognition for users who perform their target behaviour, e.g. name of awarded people are published on the website.</i>

The strategy: Designers identify the route and message that will promote persuasion. The routes may be direct (using logical argument), indirect (using cues) or a mix of both, depending on the user characteristics and ability to assess the meaning of persuasive messages. The analyses of the websites indicated compact presentation of messages, and both direct and indirect routes were used in almost all websites. As all e-commerce websites uses unobtrusive persuasive technologies, they neither state the persuasion intent nor the type of change being targeted. We assume they are targeting both change in behaviour and attitude, i.e. they aim to persuade the user to buy online and to be loyal to their website. Although some of the websites attempt to consider user background, all failed to include such differences in persuasive message and route types.

Identified Persuasive Features

The use of the PSD model assisted us in identifying precise persuasive features for system qualities that promote persuasive interactions. Results show that dialogue support (i.e. 85% feature coverage across the 10 sites) is the most common persuasive features with credibility support, primary task support, and social support gaining scores of 74%, 71%, and 51% coverage respectively.

Dialogue support

Dialogue support features were the most highly utilised features, with all websites incorporating features that support dialogue between the user and online trader. 90% of the websites used praise,

i.e. giving positive verbal feedback after registration or purchasing products/services from the website; e.g. 'Congratulations, you have successfully completed your order'. Also, all of the websites rewarded the users with online vouchers if they have performed particular actions, such as successfully inviting friends to buy from the website, or offering points for future purchase. This can help motivate users to i) purchase online and ii) be loyal to the online trader. Reminders were employed in different forms, such as emails regarding specific events. Additionally, websites regularly reminded users to check out when they have added products to the shopping basket, i.e. to ensure that users are reminded to perform the behaviour of purchasing from the website. Suggestions came in a form of recommender systems, i.e. where the website suggests other products that other users bought when buying a particular product or suggesting items that the user may like based on their previous search interest. Seven websites adopted a social role, i.e. automated online assistance, with one of them offering

an online community centre where users can chat with each other to discuss related issues. Similarity was the least used dialogue support feature; with only half of the websites adapting this feature.

Credibility support

Credibility is commonly defined as believability (Fogg, 2003). Credibility has significant influence on persuasion and website persuasiveness (Fogg, 2003, Cugelman et al., 2009). Based on results, it appears that all websites have a high credibility. Trustworthiness, expertise, surface credibility and real-world feel have been implemented in all of the websites. Interestingly, the five top websites, i.e. in terms of conversion rate, do not have physical stores. Despite a pure online presence, all employ real-world features to increase their credibility by providing information about their organisations and their contact details with four of them providing online chat. Surface credibility, as opposed to earned credibility, is related to website look and feel. The influence of surface credibility is assumed to be useful in the initial stages of interaction, i.e. when earned credibility has not yet been established (Harjumaa et al., 2009). Users, for example, often link website appearance with website quality and credibility. Surface credibility was measured in relation with two particular aspects: i) the website should reflect the context, and ii) the website does not have an unreasonable number of advertisements on it. Expert evaluation concerning this feature showed that the top five websites, in terms of conversion, did not present any advertisements for other websites, i.e. they only presented their own products/services, avoiding passing customers over to other websites. It seems particularly important to keep the users on their websites, and to avoid distracting them from the main target behaviour. Third party endorsements were used in eight websites, i.e. where websites provide endorsements from respectable external sources, e.g. PayPal, McAfee SECURE™ Trustmark, etc. Authority, i.e. providing a means to verify the accuracy of the site content via external sources, were seldom employed; with only two websites employing these features i.e. Amazon and EBay where they verify the credibility of small sellers by allowing them to sell their products under the umbrella of Amazon or EBay.

Primary Task Support

The primary task is the main task or activity that the website is designed to support, i.e. support the user's primary task of purchasing online. The analysis indicated that all primary task features were utilized, with the exception of simulation and rehearsal features. Reduction, tunnelling, tailoring, and self-monitoring (see Table 1) appeared to be the most common features as they were employed by all selected websites. Personalisation was applied in 60% of websites, which offered personalised recommendations based on users' interests and search history.

Social support

Social support was the least common category utilised in the websites. Social support influences the persuasiveness of the website as users' attitude or behaviour is usually influenced by the attitude and behaviour of other people (OinasKukkonen and Harjumaa, 2009). When users see others using the same website, and when the website allows the users to compare their behaviour with the outcomes of others, user attitude is often influenced either positively or negatively. In this study, normative influence was the most common social support feature; adopted in 9 websites. Customers review boards and social media links are common implementation examples of this feature. Users were allowed to email items or share questions through social media to get others comments. Social learning, social comparison and cooperation were adapted in 8, 7 and 7 websites respectively. These features are extremely important to improve user motivation to perform the target behaviour. Social facilitation was applied in 30% of the websites; increasing users' motivation and website persuasiveness by allowing users to discern the behaviour of other users who use the website at the same time i.e. 'what other users watch and buy right now'. Competition and recognition was only used by Ebay; an online marketplace website. This is due, in part, to the non-competitive nature of e-commerce business.

IV. PROPOSED METHODOLOGY

Businesses that integrate social media into their marketing strategy – from customer acquisition, to sales, to re-engagement campaigns – will benefit. Marketers can see in real-time what your audience cares about most, their interests, the conversations they're having and what they like. Use your social networks to better segment audience and understand your target demographics. This will help you optimize your campaigns and deliver more targeted messaging. Immediacy is big in social media; we want information and we want it now. That's why social networks are so great for customer service. They enable businesses to quickly respond to customer inquiries. Plus, social media makes it easier to spot and respond to unpleasant customer experiences. Develop a strategy for responding to customer inquiries via social media. Gathering the details from the user's social network profile and creating product recommendations. The profile details will match with the news server for gathering recommended news information. According to the news three news channels will be

consolidated for getting the confident information. Recommended product information can be changed according to the user's public chat and user's updated profile details. News details available in search mode also, news can be search area wise, city wise and state wise.

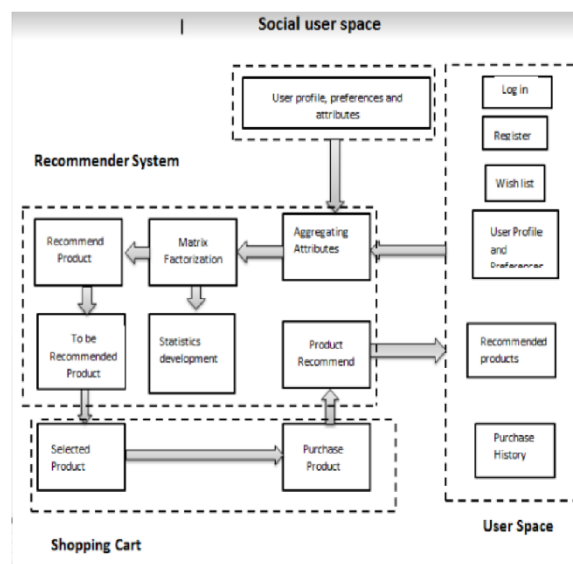


Fig -1: Proposed Architecture

In above architecture there are mainly five layers of control are present in fraud detection system. First layer that is terminal layer checks the security for all the transactions. This layer is used when any transaction is initialized. Security checks like correct PIN code, number of attempts, valid user name available balance, validity of are performed by this layer. After checking all valid checks the transaction will proceed further otherwise it will denied. Then transaction blocking rules are the rules which are defined for secure transaction. These rules use the few information available when the payment is requested, without analysing historical records or cardholder profile. If there is internet transactions initialized Fraud Detection using Machine Learning on a website which is unsecured then deny the transaction request. Transactions blocking rules are designed so that it should guarantee realtime operations and avoid blocking many genuine transactions. Scoring rules are also expert-driven models that are expressed as if-then statements. An example of scoring rule can be IF previous transaction in a different continent AND less than one hour from the previous transaction THEN fraud score =0.95. Scoring rules can be subjective as they can be designed differently. It is expected that fraudulent patterns should be detected from this layer. Only a limited number of alerted transactions are reported to the investigators, which represent the final layer of control. Investigators are the professionals experienced in analyzing transactions and are responsible of expert-driven layers of fraud detection system. Any card that is found victim of a fraud is immediately blocked, to prevent further fraudulent activities and this task is performed by investigators. Using this system architecture, transaction is detected as fraud or normal. Also, PCA for dimensionality reduction and for classification is used to train the data. And successfully applied on the dataset which contains details about past transaction.

V. CONCLUSION

The performances of all this techniques are examined based on accuracy, precision and specificity metrics. We have selected supervised learning technique Random Forest to classify the alert as fraudulent or authorized. This classifier will be trained using feedback and delayed supervised sample. Next it will aggregate each probability to detect alerts. Further we proposed learning to rank approach where alert will be ranked based on priority. The suggested method will be able to solve the class imbalance and concept drift problem. Future work will include applying semi-supervised learning methods for classification of alert in FDS.

REFERENCES

- [1]. Herna'ndez Sergio, Pedro A' lvarez, Javier Fabra and Joaqu'ınEzpeleta, Analysis of users' behaviour in structured e-commerce websites, 2169-3536 (c) 2016 IEEE Translations and content mining.
- [2]. S.E. Kim, T. Shaw and H. Schneider, "Web Site Design Benchmarking Within Industry Groups", Internet Research: Electronic Networking Applications and Policy, vol. 13, 2003, pp. 17- 26, doi: 10.1108/10662240310458341.

- [3]. M. Zhao and R.R. Dholakia, "A Multi-Attribute Model of Web Site Interactivity and Customer Satisfaction – An Application of the Kano Model", *Managing Service Quality*, vol. 19, 2009, pp. 286-307, doi:10.1108/09604520910955311.
- [4]. M.H. Huang, "Designing Website Attributes to Induce Experiential Encounters", *Computers in Human Behavior*, vol. 19, July 2003, pp. 425-442, doi: 10.1016/S0747-5632(02)00080-8.
- [5]. S.W. Lee and R.J. Koubek, "Understanding User Preferences Based on Usability and Aesthetics Before and After Actual Use", *Interacting with Computers*, June 2010, doi:10.1016/j.intcom.2010.05.002.
- [6]. S.J. Simon, "The Impact of Culture and Gender on Web Sites: An Empirical Study", *The DATA BASE for Advances in Information Systems*, vol. 32, 2001, pp. 18- 37, doi: 10.1145/506740.506744.
- [7]. B.J. Parker and R.E. Plank, "A Uses and Gratifications perspective on the Internet as a new information source. *American Business Review*, vol. 18, pp. 43-49.
- [8]. K.C.C. Yang, "Factors Influencing Internet Users' Perceived Credibility of News-Related Blogs in Taiwan", *Telematics and Informatics*, vol. 24, May 2007, pp. 69-85, doi:10.1016/j.tele.2006.04.001.
- [9]. Internet World Stats, "Internet Usage Statistics - The Internet Big Picture", 2010, <http://www.internetworldstats.com/stats.htm>.
- [10]. A. Angehrn, "Designing Mature Internet Business Strategies: The ICDT Model", *European Management Journal*, vol. 15, Aug. 1997, pp. 361-369, doi:10.1016/S0263-2373(97)00016-9.

Authors

DESETTI SRINUVASA RAO Holds a B.Tech Degree in Computer Science & Engineering from Sri Venkateswara College of Engineering and Technology, Srikakulam Affiliated to JNTU Kakinada. He presently Pursuing M.Tech (CSE) in Department of Computer Science and Engineering from Avanthi Institute of Engineering and Technology Cherukupalli(V), Near Tagarapuvalsa Bridge, Vizianagaram (Dist) - 531162, Andhra Pradesh. Area of interest include Web Technologies, Web Programming and Python programming, Machine learning and Data science.



Mr. Kesavarao Seerapu Working as Asst. Professor, Department Of Computer Science and Engineering in Avanthi Institute of Engineering and Technology Cherukupalli(V), Near Tagarapuvalsa Bridge, Vizianagaram (Dist) - 531162, Andhra Pradesh. His Area of interests include Machine Learning, Artificial intelligence, Network security and Sentiment Analysis.



Desetti Srinuvasa Rao, et. al. "Recognizing the fake reputation information, fair and healthy reputation system in E-Commerce." *IOSR Journal of Engineering (IOSRJEN)*, 10(12), 2020, pp. 33-38.