MINING E-COMMERCE FEEDBACK COMMENTS FOR A TRUSTED SELLER'S PROFILE USING SUPERVISED LEARNING ALGORITHMS

Sruthi sathyanandan s[1, 2] Dhanya Sreedharan

[1]*Sree Buddha college of engineering, Alappuzha, India,*

[2]*Sree Buddha college of engineering, Alappuzha, India*

---***---

Abstract-There is a tremendous growth in E-commerce applications where buyers and sellers conduct transactions through the web. Today users are attracted to online-shopping due to the convenience in accessing reviews given by post purchase regarding item-related as well as seller-related. Almost all online-shopping websites encourage its buyers to provide feedback in the form of text feedback comments as well as star ratings in order .The aim of system lies in finding trustworthy sellers from E-commerce feedback comments. From the overall feedback comments obtained we generate a trusted seller's profile. For this purpose firstly we mine all of the feedback comments. Feedback comments given in the form of star ratings are also considered. Then the corresponding weights for each stars are calculated. Then we compute dimensions weights and trust for the seller is computed.

*Keywords—*Aspect mining; LDA; NLP; LEXICAL-LDA

I.INTRODUCTION

When we make a purchase from an E-commerce site we browse through all the reviews posted by post purchase users. Most of the reviews we find in an E-commerce site play a major role to help customers in deciding whether to buy a product or not. Today there are many Reputation-based trust models are commonly used in E-commerce applications, and their feedback ratings are computed to find the sellers reputation trust scores. The "all good reputation" problem is a very common issue in E-commerce sites. The reputation scores for sellers in an E-commerce site is very high and it is difficult for buyers to select trustworthy sellers. In this paper we consider users reviews in the form of text as well as in the form of star ratings. System design consists of five parts. They are the following:(i)analysing feedback comments(ii)Mining of the total e-commerce feedback comments,(iii)Aspect opinion extraction(iv)classifying fake and genuine comments and v)seller's trust profile.
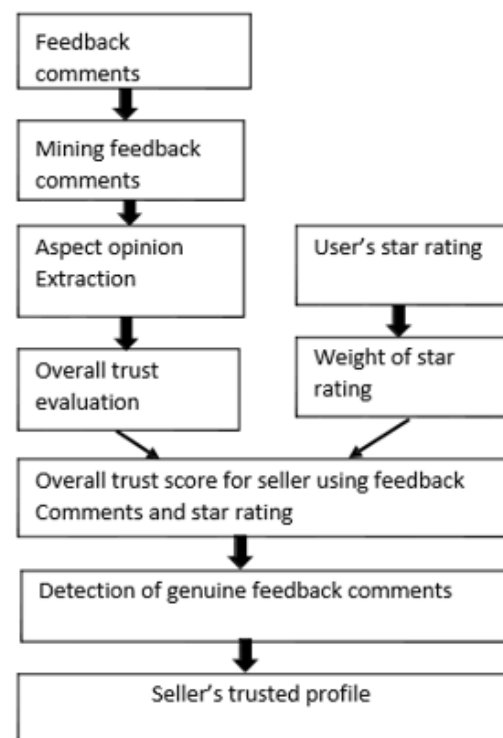

Figure 1: Architecture of proposed system

The system deals with the following modules.

A. Feedback comment analysis

The success of E-commerce sites depends highly on the availability of customer interaction. Usually reputed sellers attract their customers to transact with them and leave comments. So a customer purchasing from the site can check up with these reviews and can check how good a seller is in providing good services to their customers. An issue in the reputed sites like amazon and eBay is their strong positive rating bias. There have been lot of studies in analysing feedback comments in e- commerce sites in order to identify fake information to provide a genuine reputation range score for each sellers. Many studies have been conducted in the past

for analysing feedback comments in E-commerce applications. The comments and feedbacks in an E-commerce site are noisy and thus analysing them is a challenging problem.

B. Mining feedback comments

Mining feedback comments refers to techniques in natural language processing (NLP) and linguistics features helps in identifying and extracting subjective information. This process involves methods like text review analysis and feature extraction from the text. The text review mining means to determine actual polarity of the document with respect to their context. It identifies the different opinions expressed by customers about certain topics whether it is negative or positive. There will be different types of sentiments expressed for the same topic. The paper called "A Joint Model of Text and Aspect Ratings for Sentiment Summarization" discusses on a statistical model that discovers the relevant topics included in a text. This paper calculates aspect summarization based on two problems. First problem is aspect identification and the second problem is mention extraction. The goal behind this is to achieve a set of relevant aspects.

C.Opinion extraction

The system is more generally related feedback analysis on free text documents. Aspect based extraction is done by selecting and re-organising sentences in order to extract the relevant features. Review mining and summarization is the task of producing a sentiment summary, which consists of sentences from reviews.

The paper called "Computing Product Rating Using Real-Time Feedback Comments from E-Commerce Portal "discusses about the rating of products in real time from an e-commerce portal. Here the system calculates the feedback scores and classify each seller based on the scores obtained. The system discussed here uses a multi-dimensional trust based algorithm for mining the feedback comments. Here feedback comments are mined in order to get the dimension ratings and weights. For this LDA algorithm is proposed here.It order to achieve an efficient result we use shallow lexical analysis also. It is done by constructing a dependency relation analysing model.

D.Classifying fake and authentic comments

The paper called "Supervised Learning for Classification of Authentic and Fake Online Reviews "introduces various supervised learning algorithms to classify online reviews to fake and genuine. The fake reviews are identified based on four features like writing style, level of details, name, able to understand and certain cognition indicators.

2. LITERATURE SURVEY

2.1 Trust relationships from online feedback comments

[1] Discusses E-commerce customer satisfaction based on trust. The system uses a B2C (business to consumer) application called business to customer using various e-commerce application. The Theory of Reasoned Action (TRA) used in the paper explains about the attitudes and behaviours of the customers. This paper examines the overall text data based on a business to customer satisfaction. Trust and loyalty are the main factors considered. Other main factors while analysing the feedback comments are the quality of the user interface, the information quality, privacy and security. The interface quality is completely related to customer trust.

[2] Discusses the rating of products in real time from e-commerce portal. Here the system calculates the feedback scores and classify each seller. The paper uses a multi-dimensional trust based algorithm for mining the feedback comments. Here feedback comments are mined to get dimension ratings and weights. For this we use LDA algorithm.Inorder to get an efficient result we use shallow lexical analysis. It is done by constructing a dependency relation. Firstly the text data is constructed to a matrix form. Then the text data is represented as dimensions and the dimensions are clustered to get meaningful clusters. The dimensions are generated by various topics in the text data and each topic is got from the head terms in a text data.

2.2 Classification of fake and authentic reviews

[3] Introduces supervised learning algorithms to classify online reviews to fake and authentic. Here the fake reviews are identified based on four features like writing style, level of details, name, able to understand and certain cognition indicators. The algorithms used are classification algorithms like random forest, support vector machine etc. For supervised learning machine learning algorithms are also used here.For this support vector machines are used. While considering a feedback review eight kinds of POS tags are used. The POS tags use are based on nouns, adjectives, prepositions, names, verbs, adverbs, pronouns, conjunctions. The structure of the feedback text is calculated based on the number of characters in a single word, total number of words and the number of words contained in a single review.

[4]Discusses developing a theory for checking the diagnosticity of online reviews. This paper considers the user's review, the reviewer's past track review record is also checked. Diagnosticity considers various factors in a review like its depth, rating, profile, product type and readability. Here a theory is constructed for diagnosticity of the reviews.

Here the review ratings are collected based on star ratings. Usually for all the E-commerce feedback reviews we can see a question at the bottom of each feedback like whether the review was useful or not. So the users who read the feedback usually gives a "yes" or "no" option for such feedbacks. So this can also be used as a factor while considering the diagnosticity of the review.

[5] Discusses the online trust in e-commerce transactions. This paper explains that trust is a major factor the determines the success of a business in e-commerce transactions. This paper uses a technology acceptance model to find the overall trust in e-commerce transactions. For this two categories of trust is considered. They are cognition based trust and personality based trust.

[6] Discusses trust computation in e-commerce. Here a trust vector approach is identified based on important dimensions including transaction time, amount, product category etc. First a new index called CMK-TREE is computed to support efficient computation. The trust vector considers the various features of a seller. It checks the trustworthiness of a seller in the product hierarchy category, the trustworthiness of a seller in his product's price range. And the higher the value of a trust vector the more the trustworthiness of the seller.

2.3 Social summarization of e-commerce feedback comments

[7]Discusses the social summarization in the e-commerce feedback comments. The system explained here gives appropriate weightages to each sentence based on some factors and extract sentence that have higher weight from other documents.

[8] Discusses the summarization in e-commerce feedback comments. The system explained here classifies and summarize information into seven types. They are based on the frequency of the keyword, its location, title, structure of sentence, description of key, lexical relationship between words, distance in vector space model.

[9] Discusses the sentimental analysis in e-commerce feedback comments using natural language processing. The system explained here finds how sentiments are expressed in a sentence and checks out the positive and negative aspects in them.For this a lexicon based dictionary is constructed to find the semantic relatonships.Then the system concludes that the statements that express sentiments is more better when compared to overall opinion.

[10] Discusses lexicon based approach to find the positive and negative aspects in a feedback and thus make a classification of the whole review. Here sentiment analysis is done for the classification of reviews. Sentiment analysis is done to understand the customer opinions and determine the emotions of a user. The system uses text from small documents to huge documents. A lexicon based dictionary constructed for the system can identify common or default sentiment words, blind negation words, negative words and split words.

[11] Discusses the extraction of aspects and aspect relations from opinion. The system uses opinion extraction in two steps. First step is the extraction of aspect evaluation and the second step is the extraction of aspect relations. The methods used for opinion extraction in this paper are of aspect evaluation relation extraction, opinion-hood determination, aspect of relation extraction.

3. CONCLUSION

In the proposed system we analysed the feedback comments. Stop word removal is applied to each text feedback comments. Then the expressions got after applying stop word removal is clustered to make dimensions from the expressions. Then a dependency analysis is done. In dependency analysis Natural Language Processing is applied (NLP).NLP is applied to find sentences, names, split sentence, tokenize sentence, and POS tagging is done to each tokens. Then dimension ratings are found out based on the sum of positive and negative feedback Comments. Based on this a trust score is calculated. Then based on linguistic clues we classify each feedback comments to fake or genuine. For classification we used SVM.

Acknowledgements

REFERENCES

[1]Identifying honest recommenders in reputation systems frag afzelin king Fahd University of petroleum and minerals, information and computer science department, Saudi Arabia.

[2]International journal of scientific & engineering research, volume 6, issue 11, november-2015 137 ISSNl 2229-5518 ijser © 2015 http://www.ijser.org

[3]Using supervised learning to classify authentic and fake online reviews, wee school of communication & information, Nanyang technological university, Singapore.

[4]Developing a theory of diagnosticity for online reviews Alton y. k. Chua, and Snehasish Banerjee proceedings of the international multiconference of engineers and computer scientists 2014.

[5] A literature review of online trust in business to consumer e-commerce transactions, 2001-2006 shan-yan hang, national dong haw university, Taiwan, jie18.roye28@msa.hinet.net

[6]Reputationpro: the efficient approaches to contextual transaction trust computation in e-commerce environments hailing Zhan and yarn wag, Macquarie University, Sydney.

[7] Social summarization of text feedback for online auctions and interactive presentation of the summary yoshinori graduate school of engineering science, Osaka University.

[8]Constructing literature abstracts by computer: techniques and prospects, department of computer, university of Lancaster, Chris d pace, Germany 2014

[9] Sentiment analysis: capturing favourability using natural language processing testy nasukawa IBM research, Tokyo research laboratory 1623-14 shimotsuruma, yamato-shi, Kawakawa-ken, 242-8502, japan

[10] Serendib: simple and practical lexicon based approach to sentiment analysis prabu palanisamy, vineet Yadav and harsh elchuri serendio software pvt ltd guindy, Chennai, harsha}@serendio.com

[11] Extracting aspect-evaluation and aspect-of relations in opinion mining nozomi kobayashi kentaro inui, and Yuji matsumoto graduate school of information science, japan, 2010