CommTrust by Mining Feedbacks, Tweets & Comments

Vijaya Chavan, Dr. M. Z. Shaikh

Lecturer, Dept of Computer Technology, Bharati Vidyapeeths Institute of Technology Navi Mumbai, India

Principal, Bharati Vidyapeeths College of Engineering, Navi Mumbai, India

ABSTRACT: With the rapid growth of information and telecommunication industry, most of the people start using internet for education, entertainment and shopping. Online shopping is proliferating with number of Smartphone users grows. People have started selling and buying online because of many benefits like easy of shopping and selling, a lot of products are available in discount price, and lot of option are available. The selection of product to buy with require quality is biggest issue to online shopper, so they always looks for rating for products. The reputation reporting systems is very essential for success of ecommerce business. The current reputation reporting system showing average rating by buyers, which is "all good reputation" because of almost all product have rating close to 4-5. Rating system hardly assist buyer to select product. By using ratting people cannot predict aspect what is good in product on particular site, whether quality of product is good or shipping is fast or buying process on website is good?. Even rating for particular product is less on customer are not able to predict exactly which aspect is not good in product on website. In order to assist people to buy product it is better to analyze comment on ecommerce site or social media post regarding products instead of showing just rating. To solve this problem we have proposed the multidimensional trust (MDT) model, which take buyers comment on ecommerce site and tweet about products to analyze buyers view on different aspect of product and transaction. We have proposed aspect level opinion mining on feedback comment and tweets using topic modeling algorithm and type dependency analysis on POS tagged comment. MDT model can effectively for the "all good reputation" issue and rate product effectively for different aspect in order to assist buyer in online shopping.
energy metric.

KEYWORDS: Natural Language Processing, Topic Modelling, Tweets, E-Commerce.

I. INTRODUCTION

There is a rapid growth in a world wide web from the last few years facilitated by evolution in information technology and telecommunication industry. Now days most of the population in urban and rural area have started using Smartphone, provide easy access to internet. Almost all work can be done with the help of Smartphone and internet, from simple mobile phone recharge to biggest business deals can be done online [7].

Buyer's gives positive feedback ratings, and also they express some positive aspect, disappointment and negativeness in free text feedback comments, often towards specific aspects of transactions. For example, comments like "The product was very good." Expresses positive opinion towards the product aspect, whereas the comment "Postage time was little extended but otherwise, nice service". This gives negative opinion towards the postage time but a positive opinion to the transaction. So comments in free text format about product provides good information to customer to select product base on opinion about different aspect of product and transaction [6].

In MDT, we have propose to use aspect level opinion mining on ecommerce feedback and tweets. We have propose an approach that combines Natural Language Processing (NLP) and Opinion mining on feedbacks. To extract the particular aspects from comments we have proposed the Latent Dirichlet Allocation (LDA) topic modeling algorithm. To find the opinion of different aspects we have proposed type dependency analysis. It can be seen that with the modifying relations generally the noun or verb expresses the target concept under consideration whereas the adjective

or adverb expresses opinion towards the target concept. So to find dependencies we have proposed to use POS tagging on sentence to extract different part of speech from sentence. So we have proposed MDT to reduce the strong positive bias in existing reputation systems, and solve the "all good reputation" problem and rank product effectively with respect to different aspect of product and transaction.

II. RELATED WORK

In E-commerce applications trust and reputation are decisive factors in order to rate the sellers or products. Reputation Systems used in e-commerce have the main aim is to guide the users in a proper way so that they can choose between the different available sellers on E-commerce. In reputation system the models are meant to project the different sellers as per their services and quality they provide to users. Different Models differ between each other by the methods they opt to calculate the trustworthiness measure for different sellers. Review mining or Feedback mining method of trust calculation will be a very good idea as the users are free to express themselves in textual feedback comments. So system given in paper have considers only feedback data on portal. They have ignored a lot of data available on social media as some time user does not comment on ecommerce portal, instead they share experience on social media [3].

The system given in the paper [4] intended to design framework, which utilizes the perception made by purchasers or buyer or user. Depending on criticism survey, remarks, feedback, comment, ratings are mined. In the given survey paper have studied a multidimensional trust model for registering notoriety scores from client input remarks. Social network has gained remarkable attention in the last decade. Accessing social network sites such as Twitter, Facebook, LinkedIn and Google+ through the internet and the web 2.0 technologies has become more affordable. People are becoming more interested in and relying on social network for information, news and opinion of other users on diverse subject matters. The heavy reliance on social network sites causes them to generate massive data characterized by three computational issues namely; size, noise and dynamism. These issues often make social network data very complex to analyze manually, resulting in the pertinent use of computational means of analyzing them. Data mining provides a wide range of techniques for detecting useful knowledge from massive datasets like trends, patterns and rules [5].

The paper given in system may have cold start problem to calculate rating for few product[7].The system in given article [8] presents a brief overview of opinion mining and its classifications and specifically focuses on the sub topic aspect-based opinion mining, its approaches, metrics used for evaluation and latest research challenges.

III. PROPOSED ALGORITHM

The feedback comments as a source where buyers express their opinions more honestly and openly. Aspect level opinion mining on feedback comments help or guide new buyer to select appropriate product on website with desire quality. Two major tasks in aspect based opinion mining are aspect extraction and aspect sentiment classification. Process of identifying the opinion words from the given sentence is called aspect extraction and categorizing the extracted opinion words into one of the polarity scales is called aspect sentiment classification.
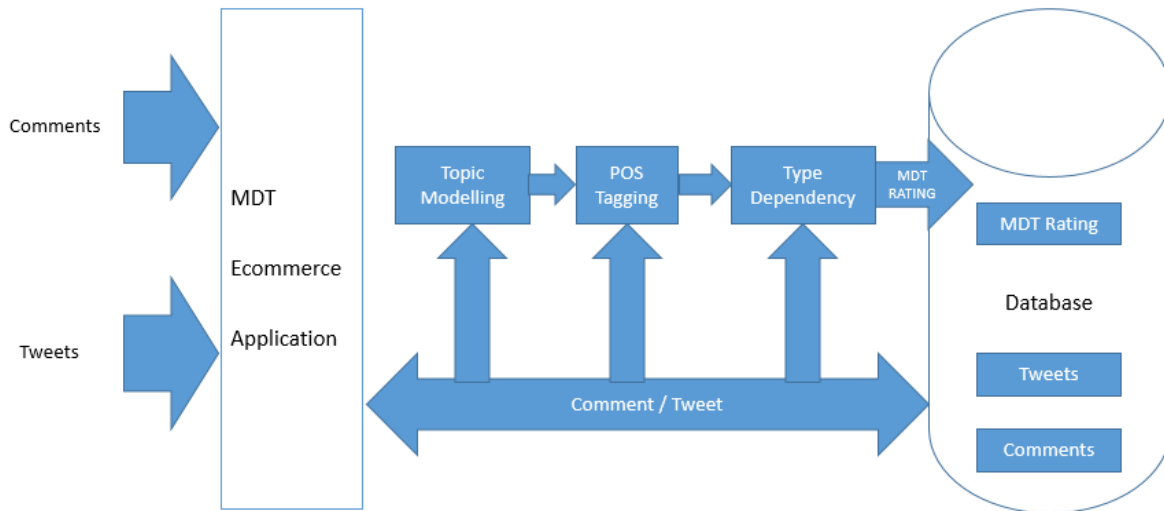
3.1 System Architecture:

Fig 3.1 System Architecture

To solve the problem statement we have proposed to implement custom ecommerce site and upload comments from amazon, flipkart and some comment on proposed ecommerce site. The propose ecommerce website also able to download tweets related different product using twitter OAUTH API. The Latent Dirichlet Allocation (LDA) topic modeling algorithm is use to extract the aspect from comment. To extract the opinion towards aspect we have proposed to find type dependency analysis, it can be seen that with the modifying relations generally the noun or verb expresses the target concept under consideration whereas the adjective or adverb expresses opinion towards the target concept. To find the type dependency like noun and adjective we have proposed to use POS tagging to find different part of speech in comment.

Topic Modelling : Latent Dirichlet Allocation
Topic models are a suite of algorithms that uncover the hidden thematic structure in document collections. These algorithms help us develop new ways to search, browse and summarize large archives of texts. Right now, we work with online information using two main tools—search and links. We type keywords into a search engine and find a set of documents related to them. We look at the documents in that set, possibly navigating to other linked documents. This is a powerful way of interacting with our online archive, but something is missing.

POS Tagging:
The process of assigning a part-of-speech to each word in a sentence. Parts of speech are useful features for finding named entities like people or organizations in text and other information extraction tasks. Parts-of-speech influence the possible morphological affixes and so can influence stemming for informational retrieval, and can help in summarization for improving the selection of nouns or other important words from a document.
Stochastic Tagging:
Based on probability of certain tag occurring, given various possibilities
By Bayes' Rule,

$$P(T|W = P(W|T) \, P(T)/P(W = \alpha P(W|T)P(T)$$

So find T that maximizes $P(T)P(W|T)$

Chain rule:

$$P(T) = P(t_1)P(t_2|t_1)P(t_3|t_1,t_2)P(t_3|t_1,t_2,t_3) \dots P(t_n|t_1,t_2,\dots,t_{n-1})$$

As an approximation, use

$$P(T) \approx P(t_1)P(t_2|t_1)P(t_3|t_2)\dots.P(t_n|t_{n-1})$$

Assume each word is dependent only on its own POS tag: given its POS tag, it is conditionally independent of the other words around it. Then

$$P(W|T) = P(w_1|t_1)P(w_2|t_2)\dots.P(w_n|t_n)$$

So

$$P(T)P(W|T) = P(t_1)P(t_2|t_1)\dots.P(t_n|t_{n-1})P(w_1|t_1)P(w_2|t_2)\dots.P(w_n|t_n)$$

Type Dependency Analysis:

The typed dependency relation representation is a recent NLP tool to help understand the grammatical relationships in sentences. With typed dependency relation parsing, a sentence is represented as a set of dependency relations between pairs of words in the form of (head, dependent), where content words are chosen as heads, and other related words depend on the heads. For example of analyzing the comment "Super quick shipping. Product was excellent. A great deal. ALL 5 STAR.". The comment comprises four sentences, and the sentence "Super quick shipping." is represented as three dependency relations. Shipping does not depend on any other words and is at the root level. The adjective modifier relations amod (shipping-3, super-1) and amod (shipping-3, quick- 2) indicate that super modifies shipping and quick modifies shipping. The number following each word (e.g. shipping-3) indicates the position of this word in a sentence. Words are also annotated with their POS tags such as noun (NN), verb (VB), adjective (JJ) and adverb (RB).Among the dependency relations expressing grammatical relationships, we select the relations that express the modifying relation between adjectives and nouns, and adverbs and verbs. It can be seen that with the modifying relations generally the noun or verb expresses the target concept under consideration whereas the adjective or adverb expresses opinion towards the target concept.
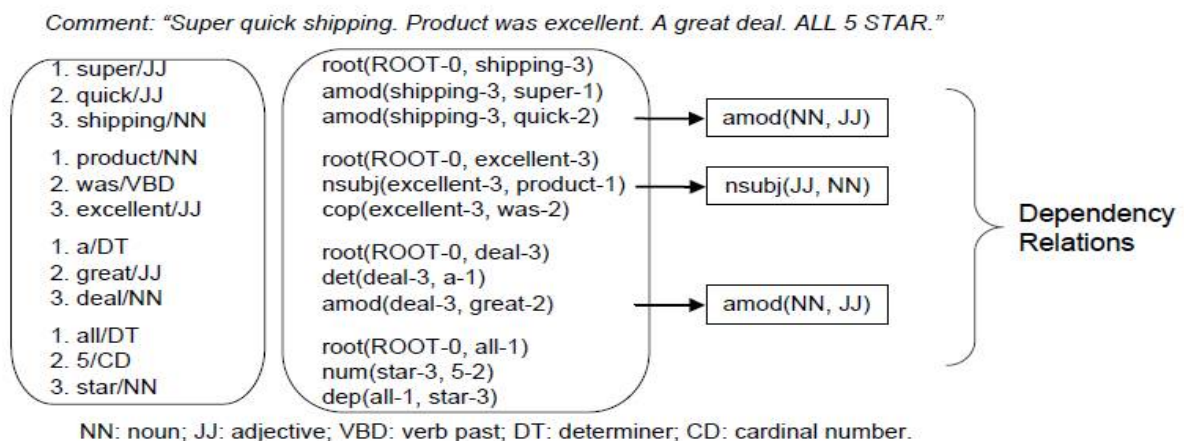

Fig 3.2 Type Dependency Relationship

If a comment expresses opinion towards dimensions then the dimension words and the opinion words should form some dependency relations. Ratings from dimension expressions towards the head terms are identified by identifying the prior polarity of the modifier terms. We have proposed the polarities on scale of 1-5 star as 1 star mean Very Negative, 2 star means negative, 3 means neutral, 4 means positive, and 5 means Very Positive.
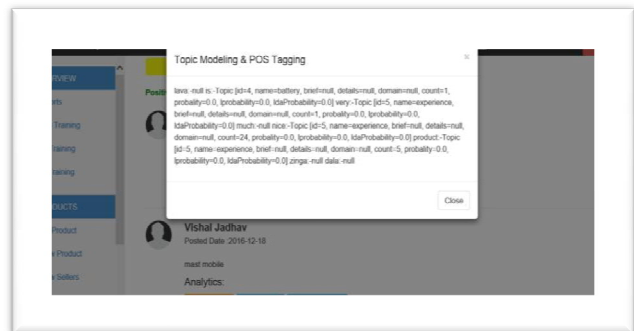
IV. SIMULATION RESULTS

Topic Modelling:
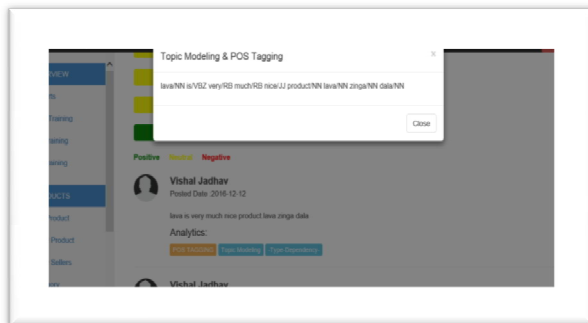
POS Tagging:

The above figures shows the analysis of the comment by admin. The admin can initiate the multidimensional trust analysis for any seller. The aspect base ratting for product and transaction will be updated in database in order to assist visitor to buy a product. The above page shows the topic modeling model trainer. The training sample documents will be given as an input and the document will be split into sentences and word. The each word will be tagged with appropriate topic and respective probabilities of word for the LDA will be updated with trained model.

Type Dependency:
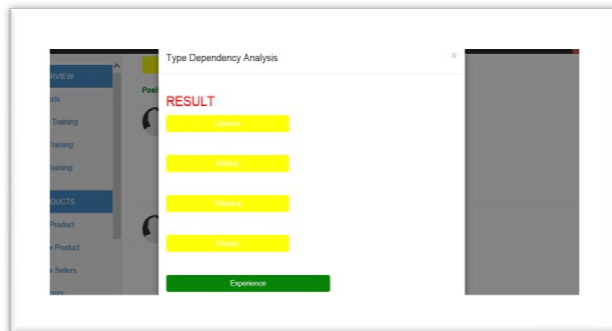
FeedBack Results

The above figures show the POS trainer page. The training document will be given as an input and stochastic POS tagger model will be trained or updated with the respective probabilities. Also this gives the result for the comment in terms of negative neutral and positive with the parameters camera, shipping, battery, display, experience.

V. CONCLUSION AND FUTURE WORK

Reputation reporting system is very essential in today's ecommerce scenario, as there are lot of product available on different site from different sellers. In order to assist buyer to choose appropriate product with desire quality ecommerce site should provide reputation of product. The problem with current reputation system is famously called as all good reputation, almost all product have rating close to 4-5. By viewing average rating buyer cannot predict aspect what is good in product on particular site, whether quality of product is good or shipping is fast or buying process on website is good?. Even rating for particular product is less on customer are not able to predict exactly which aspect is not good in product on website. Buyer often leave feedback in free text format on ecommerce portal or on social about different aspect of product and transaction. So we have propose the solution of aspect level opinion mining on comments and tweet using Topic modeling, POS tagging and type dependency analysis to assist buyer to select product with desire quality. We have also propose star rating on scale of 1-5 for different product and transaction using feedback on website and tweeter data.

REFERENCES

1. Xiuzhen Zhang, Lishan Cui, and Yan Wang, "Computing multi-dimensional trust by mining ecommerce feedback comment", IEEE, Vol 26, No 7, 2014
2. P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman, "Reputation Systems: Facilitating T "Computing Multi-Dimensional Trust by Mining E-Commerce Feedback Comments" Xiuzhen Zhang, Lishan Cui, and Yan Wang, Senior Member, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING VOL:26 NO:7 YEAR 2014
3. Priyanka Kumbhar, and Manjushi Mahajan, "Trust evaluation by mining ecommerce feedback mining", IJSR, Vol. 4, Issue 11, Nov 2015.
4. Sonali Ghare, "Survey on feedback comment analysis (Ecommerce)", IJIRCCE , Vol. 3, Issue 11, Nov 2015.
5. Mariam Adedoyin-Olowe, Mohamed Medhat Gaber, and Frederic Stahl, "Survey of Data Mining technique for social network analysis", School of computing science and digital media Rober Gordon University.
6. Ms. Pooja Chavan, and Prof. Rahul Kulkarni, "Mining Ecommerce Feedback Comment Using Multi dimensional Trust Computation," JMEST, vol. 2, Issue 7,July 2015.
7. Richa Sharma, Shweta Nigam, and Rekha Jain, "Mining Product review at aspect level", IJFCST, vol. 4. No.3 May 2014.
8. K. Vivekanandan, and J. Soonu Arvindan, "Aspect base opinion mining survey", IJCA, Vol. 106, No. 3, Nov 2014.
9. Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau, "Sentiment Analysis on Twitter Data", Department of Computer Science, Columbia University.
10. Nozomi Kobayashi ¤ Kentaro Inui, and Yuji Matsumoto , " Extracting Aspect-Evaluation and Aspect-of Relations in Opinion Mining", Graduate school of information science, Takayama, Japan.
11. Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, "Analysis of recommendation algorithm in ecommerce", Department of Computer Science, University of Minnesota, Oct. 2000.
12.

BIOGRAPHY

Mrs Vijaya Sayaji Chavan is a Lecturer in the Computer Technology Department, Bharati Vidyapeeths Institute Of Technology, Mumbai University. She received Bachlor of Computer Eengineering BE(Comp) degree in 2005 from BVCOE, Kolhapur,India and pursuing ME(Comp) from BVCOE Navi Mumbai

Dr.M.Z.Shaikh is Principal in Bharati Vidyapeeths College Of Engineering ,Mumbai University ,India.