MINING AND SUMMARIZING ONLINE ECOMMERCE REVIEWS WITH WORD ALLIGNMENT MODEL

Chetna Gyneshwari[1], Claneta D'cruz[2], Karishma Kale[3], Prof. Rinku A.Badgujar[4]

[1] *Student, Computer, JSPM's BSIOTR, Maharashtra, India*
[2] *Student, Computer, JSPM's BSIOTR, Maharashtra, India*
[3] *Student, Computer, JSPM's BSIOTR, Maharashtra, India*
[4] *Assistant Professor, Computer, JSPM's BSIOTR, Maharashtra, India*

ABSTRACT

The mining opinion targets and opinion words from online reviews are important tasks for fine-grained opinion mining, the key component of which detecting opinion relations among the words. In this his paper proposes a novel approach based on the partially-supervised alignment model, which regards identifying opinion relations as an alignment process. A graph-based co-ranking algorithm is exploited to estimate the confidence of each candidate. Finally, the candidates with higher confidence are extracted as opinion words or opinion targets. Compared to the previous methods based on the nearest-neighbor rules, our model is captures opinion relations more precisely, especially for the long-span relations. Compared to syntax-based methods, our word alignment model effectively alleviates negative effects of parsing errors when dealing with informal online texts. Compared to the traditional alignment model, the proposed model obtains better precision because of the usage of partial supervision. In addition to, when estimating the candidate confidence, we penalize higher-degree vertices in our graph-based co-ranking algorithm to decreases the probability of error generation. Our experimental results on three corpora with different sizes and languages show that our approach effectively outperforms state-of-the-art methods.

Keyword: - *Feedback Mining, Text mining, Reputation System, Electronic commerce, Trust Score, Positive Bias, Opinion Extraction.*

---

1. INTRODUCTION

The reputation-based trust models are widely used in the e-commerce applications, and feedback ratings are aggregated to compute sellers' reputation trust scores. All good reputation" problem however is prevalent in current reputation systems reputation scores are universally high for sellers and it is difficult for potential buyers to select trustworthy sellers. In this project work, based on the observation that buyers often express opinions openly in free text feedback comments, we have proposed Comm Trust, a multi-dimensional trust evaluation model, for computing comprehensive trust profile for sellers in E-commerce applications. Different from existing multi-dimensional trust models, we compute dimension trust scores and dimension weights automatically via extracting dimension ratings from feedback comments. Using LDA algorithm approach to mine feedback comments for dimension rating profiles. Both approaches Achieve significantly higher accuracy for extracting dimension ratings from feedback comments than a commonly used opinion mining approach.

1.1 PROBLEM STATEMENT

Online reviews usually have informal writing styles, including grammatical errors, typographical errors, and punctuation errors. This makes the existing parsing tools, which are usually trained on formal texts such as news reports, prone to generating errors. Accordingly, these syntax-based methods, which heavily depend on parsing performance, suffer from parsing errors and often do not work well. To improve the performance of these methods, we can specially design exquisite, high-precision patterns. However, with an increase in corpus size, this strategy is likely to miss more items and has lower recall. Therefore, how to precisely detect the opinion relations among words is a considerable challenge in this task. The collective extraction adopted by most previous methods was usually based on a bootstrapping framework, which has the problem of error propagation. If some errors are extracted by an iteration, they would not be filtered out in subsequent iterations. As a result, more errors are accumulated iteratively. Therefore, how to alleviate, or even avoid, error propagation is another challenge in this task.

2. LITERATURE REVIEW

There are various researches that have been done related to reputation calculation. Some of the works are presented below.
Xiuzhen Zhang and Lishan Cui have presented In CommTrust, an approach that combines dependency relation analysis, a tool recently developed in natural language processing (NLP) and lexicon-based opinion mining techniques to extract aspect opinion expressions from feedback comments and identify their opinion orientations. We further propose an algorithm based on dependency relation analysis and Latent Dirichlet Allocation (LDA) topic modelling technique to cluster aspect expressions into dimensions and com

Authors X. Wang, L. Liu In [12] has presented open environments trust relationship which is build using ratings. Customer Ratings are also be called as recommendations or feedbacks provided by uses. There are many Rating aggregation algorithms are used to build up trust relationship for sellers by using rating aggregation algorithms. As other Complex methods and algorithms are not always cost effective and resistant to fake ratings provided by buyers. One of the system named Review aggregator is one such system using rating aggregation algorithm given by P. Thomas and D. Hawking, in [3]. First it stores different reviews and makes use of these to support websites where the users can read this reviews. They have assign each review a numeric value based on the positive polarity expressed in that particular review and based on that an average assessment is made.

Authors H. Zhang, Y. Wang presented [5] a PeerTrust framework used in peer to peer systems. Here they have used contextual factors for computing trust scores and weights for different peers. The contextual factors includes transaction item details, item transaction amount and transaction time. The first term Transaction item refers to the product in traded in a transaction second the properties of the item like product qualities, product categories of which determine the nature of the transaction. Third term Transaction amount refers to sum of prices of all products in a transaction done by a particular user at a specific time. Higher the transaction amount more is the chance for fraud to happen. Here the term Transaction time refers to time when a transaction happens. While calculating trust Transaction time has a specific feature. Here the consideration is that any query on temporal dimension should start from a previous point (e.g. one week ago) and end at present time.
The main disadvantage of this work is that it uses a bit large amount of data space as well as computation time. Another limitation is flexibility while considering the contextual factors because the factors are chosen while the system is designed. So the output of this system is that, the exact ranking of the sellers cannot be ensured.

EigenTrust algorithm proposed by S. D. Kamvar and M. T. Schlosser [6] uses rating matrix representation for local trust scores and calculates the global ratings for peers from the rating matrix. This system is a reputation management algorithm for peer to peer Systems. In this work each peer in the network assigned a unique global trust value based on peer history of uploads and reduce the number of unauthenticated files in peer to peer network. They have consideration here that if peer a trust peer b, then all peers trusted by b can also be trusted a. Here they compute the local trust value for all peers that have provided it with authenticate or fake downloads based on the satisfactory or unsatisfactory transactions that it had.
The main limitation of this system is that it assumes that feedback ratings are given already and the aggregation algorithms are given priority.

Another concept come into the picture such as Opinion mining is also called sentiment analysis which is a part of natural language processing and computational linguistics which identify and extract subjective information [7] from the source like comments, reviews. The main part of these systems is text analysis. Here the concept opin ion mining targets to determine the polarity of a document with respect to some context of the text. Further it determines the different opinions expressed by different authors about some topics. The methods used for Accuracy measurement are precision or recall functions.

Aspect opinion extraction is a concepts presented in various papers. Several existing work on aspect mining on product reviews, apps reviews and many reviews [8]–[10]. Out of these M. Hu and B. Liu in [8] used frequent nouns and noun phrases as aspects for product reviews and further opinion lexicon is developed to identify opinion aspect orientations. Authors G. Qiu and B. Liu [9] have further improved the previous methods to apply lexical different patterns to improve the aspect mining accuracy. And next in [10] dependency relation main is parsing and that is used for mine aspect opinions for movie related reviews. Here they have not group the aspect opinion expressions into clusters.

Unsupervised topic modeling approaches are presented by some authors which is used to jointly model opinions and aspects . They have considered here the probabilistic Latent Semantic Analysis model [9]. There are various models presented related to this which differ in granularities and how aspects and opinions interactions with each other [14], [15].

3.IMPLEMENTATIONS DETAILS

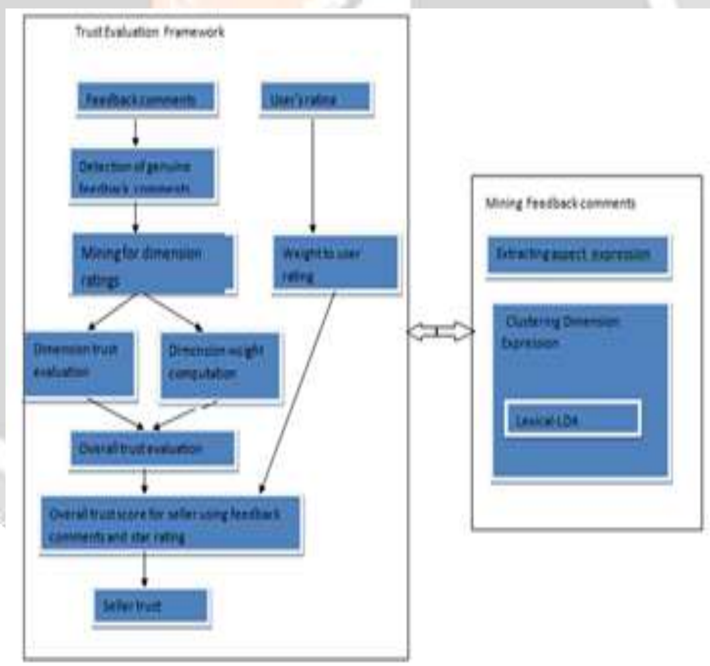

Fig -1: System Architecture

3.1 ALGORITHM

LDA Algorithm Pseudo code: -

Input: - User comment

Output: -

Step1:divided the user comments into the sentences based on user identifiers such as but, and

Step 2:Stored each sentences into the array .

Step 3:Reapete step 1 and step 2 until all comment convert in To the sentences.

Step 4: This sentence will be tokenized into individual words To analyse them.

Step 5: The tokenized words are now compared with words In the database to decide the dimension whether it is

Shipping, quality etc.

Step 6: These words are then compared with words in another database to decide upon the dimension

(positive ,negative)

Step 7 : These two databases are used to know the direction of dimension whether it is positive

(ex:"good delivery") or negative (ex: "slow shipping").

Step 8: Once the direction of dimension and dimension weights are computed, rating will be given accordingly which is stored in another database. This process continues until all the sentences are given rating.

Step9;Those value will be taken and giving the final rating to all comment.

Step 10:add all positive rating in to the positive score,negative rating to the negative score .

Step11:final reting=[Positive score + negative score/number of row in the table].

4. MORPHISM

Morphism refers to a structure-preserving mapping from one mathematical structure to another. The notion of morphism recurs in much of contemporary mathematics. In set theory, morphisms are functions in linear algebra, linear transformations; in group theory, group homomorphisms ; in topology, continuous functions, and so on. **Morphism** is an abstraction derived from structure-preserving mapping between two mathematical structures.There are two operations which are defined on every morphism, the domain (or source) and the codomain (or target).

If a morphism$f$ has domain X and codomain Y, we write $f:X \rightarrow Y$. Thus a morphism is represented by an *arrow* from its domain to its codomain. The collection of all morphisms from X to Y is denoted hom(X, Y) and called the **hom-set** between X and Y.For every three objects X, Y, and Z, there exists a binary operation hom(X, Y) × hom(Y, Z) → hom(X, Z) called composition . The composite of $f:X \rightarrow Y$ and $g : Y \rightarrow Z$ is written $g \circ f$ or gf. The composition of morphisms is often represented by a commutative diagram.

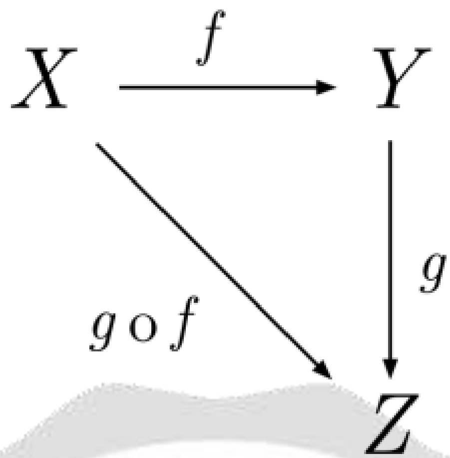

Fig -2: Commutative diagram

Morphisms satisfy two axioms:

- *Identity:* for every object X, there exists a morphism $\mathrm{id}_X : X \to X$ called the **identity morphism** on X, such that for every morphism $f : A \to B$ we have $\mathrm{id}_B \circ f = f = f \circ \mathrm{id}_A$.
- *Associativity:* $h \circ (g \circ f) = (h \circ g) \circ f$ whenever the operations are defined.

Describing the morphism in our system:

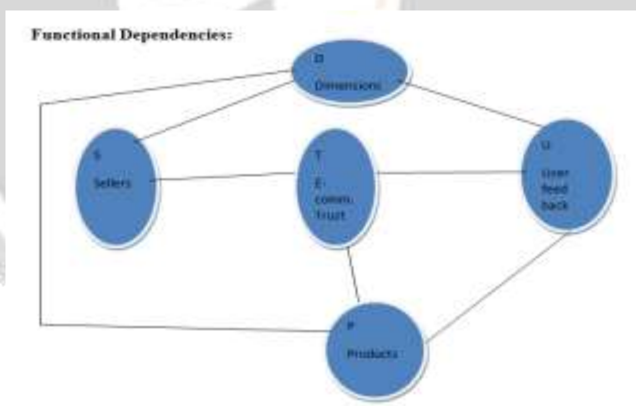

Fig -3: Functional Dependencies between user, product, Sellers and e-commerce review.

E(C)=E-Commerce WAM Model

Now consider another set S, a set of sellers selling their products through e-commerce application. Which can be represented as:

F(S) :S = {S1,S2,S3….Sn}

Now various users can buy products through this shopping portal , these products posted by various sellers. After getting a product users give feedback to the seller related to various dimensions such as quality, shipping.

F(D):D={D1,D2,D3….Dn} which is a set of dimensions which we are using to evaluate the trust score of each seller.

F(U) :U={U1,u2, U3…..Un}

Consider a set U as a set of users registering with our system which is represented as

F(C): C = {C1,C2,C3….Cn}

A set C as feedback comment set we are using to analyse multidimensional trust of a seller which can be represented as

F(T):T = (∑ TD1+TD2+....+TDn) /(n)

This score T represents the overall trust score of seller ,So we are here showing the multidimensional as well as the total overall score of a seller to the user of a system. Same score we calculated for the Product.

5. FUTURE WORK

In future work, we can improve mining methods to identify terms more accurately and the comments with more word count by storing them in database and reviews can be multi languages so that can more efficient to users and which would improve the overall accuracy of the rating system

6. CONCLUSION

In this paper we presents an alignment-based approach with graph co-ranking to collectively extract opinion targets and opinion words. Our main contributions can be summarized as follows: To precisely mine the opinion relations among words, we propose a method based on a monolingual word alignment model (WAM). An opinion target can find its corresponding modifier through word alignment.We further notice that standard word alignment models are often trained in a completely unsupervised manner, which results in alignment quality that may be unsatisfactory. We certainly can improve alignment quality by using supervision. Thus, we further employ a partially-supervised word alignment model (PSWAM). We believe that we can easily obtain a portion of the links of the full alignment in a sentence. These can be used to constrain the alignment model and obtain better alignment results. To obtain partial alignments, we resort to syntactic parsing. To alleviate the problem of error propagation, we resort to graph co-ranking. Extracting opinion targets/ words is regarded as a co-ranking process. Specifically, a graph, named as Opinion Relation Graph, is constructed to model all opinion target/word candidates and the opinion relations among them.

7. ACKNOWLEDGEMENT

We expressed a deep sense of gratitude to honorable Prof. Rinku A.Badgujar our guide and head of the department of computer engineering. She has been a consistent source of motivation and inspiration to us.We thank her for giving us a friendly space, providing laboratory facilities and believing on us that we have potential to develop this model of our project.

8. REFERENCES

[1]. H. Zhang, Y. Wang, and X. Zhang, "Efficient contextual transaction trust computation in e-commerce environments," in *Proc. the 11th IEEE Int. Conf. on Trust, Security and Privacy in* Computing and Communications (TrustCom 2012), 2012.

[2]. "A trust vector approach to transaction context-aware trust evaluation in e-commerce and e-service environments," in *Proc. the fifth IEEE Int. Conf. on Service-oriented Computing and Applications (SOCA 2012)*, 2012.

[3]. M. Gamon, "Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis," in *Proc. the 20th Int. Conf. on Computational Linguistics*, 2004.

[4]. Y. Hijikata, H. Ohno, Y. Kusumura, and S. Nishida, "Social summarization of text feedback for online auctions and interactive presentation of the summary," *Knowledge-Based Systems*, vol. 20, no. 6, pp. 527–541, 2007.

[5]. M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. the fourth Int.Conf. on KDD*, 2004, pp. 168–177.

[6]. M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. the fourth Int.Conf. on KDD*, 2004, pp. 168–177.

BIOGRAPHIES

	Chetna Gyneshwari is a student of Computer, JSPM's BHIVARABAI SAWANT INSTITUTE OF TECHNOLOGY AND RESEARCH, PUNE,01.working on Mining And Summarizing Online Ecommerce Reviews With Word Allignment Model
	Claneta D'cruz is a student of Computer, JSPM's BHIVARABAI SAWANT INSTITUTE OF TECHNOLOGY AND RESEARCH, PUNE,02.working on Mining And Summarizing Online Ecommerce Reviews With Word Allignment Model
	Karsihma Kale is a student of Computer, JSPM's BHIVARABAI SAWANT INSTITUTE OF TECHNOLOGY AND RESEARCH, PUNE,03.working on Mining And Summarizing Online Ecommerce Reviews With Word Allignment Model
	Prof. Rinku A.Badgujar ,JSPM's BHIVARABAI SAWANT INSTITUTE OF TECHNOLOGY AND RESEARCH, PUNE.