

Mining Online Reviews to Improve Sales Performance

Rashmi Ghuge
Department of Computer Engg.
JSPM RSSOER Narhe
Technical Campus Pune
411041

Supriya Pawar
Department of Computer Engg.
JSPM RSSOER Narhe
Technical Campus Pune
411041

Pooja Balakrishna
Department of Computer Engg.
JSPM RSSOER Narhe
Technical Campus Pune
411041

Nilam Chavan
Department of Computer Engg.
JSPM RSSOER Narhe
Technical Campus ,Pune
411041

S.S Shinde
Department of Computer Engg.
JSPM RSSOER Narhe
Technical Campus, Pune
411041

ABSTRACT

In Electronic commerce applications, the feedback ratings given by the customers are combined to calculate and assign seller's rating or their reputation trust score. The most widely used method in e-commerce applications to do this is via reputation trust models. But in such models the all good reputation problem is very commonly occurring, resulting in high trust scores or rating for all the sellers and this in turn makes it highly tricky for possible customers to select trustworthy and reliable sellers. In this paper, we propose a system for trust score calculation by mining feedback comments, on the basis of observation that the customers tend to express their views and opinions freely in the feedback comments section. The main objective is to implement a multidimensional trust model [1] for calculating the seller's reputation score from the information in the customer feedback comments and an algorithm for mining the text comments for dimension ratings and weights [1], using some techniques from natural language processing and from opinion mining [1] along with topic modelling. Comment based trust evaluation can successfully tackle the all good reputation problem and rank sellers efficiently, this has been revealed by previous experiments on eBay and Amazon data.

General Terms

Commerce, Transactions, Opinion, Feedback.

Keywords

Electronic Commerce, Text Mining, Natural Language Processing, Opinion Mining

1. INTRODUCTION

The conventional star rating system will give users the rating which can be biased which means even if the user gives a high star rating but in the comments if he mentions anything negative regarding the delivery of the product, then the rating becomes unfair.

Normally this type of rating system does not allow the buyer to rate the product in all aspects. So in order to conquer this issue, it is important to consider the ratings for comments as well. By taking the comments into consideration while rating the product or service will make the buyers to trust the sellers.

Comment rating serves the principle of ranking the sellers in another way. Buyers can deceive by giving false or bogus feedbacks and there can be challenge of trust context change during the dealings.

This type of rating makes the buyer more suitable to trust the seller thus making the site more trustworthy.

By analysing content of the feedback comments, the e-commerce site can comprehend the opinion of the buyer in various points like regarding the transactions, quality of product and service provided by the sellers.

2. EXISTING SYSTEM

2.1 Comment Based Multidimensional Trust Evaluation (CommTrust)

Comment-based Multi-dimensional trust (CommTrust) [1] is a trust evaluation model created from mining feedback comments from electronic commerce sites.

Figure 1 depicts the CommTrust framework. Unlike from existing trust evaluation models (including the one used on eBay) where accurate transaction feedback ratings (positive or negative) are used to calculate the seller's overall trust ratings, the CommTrust only consider the feedback comments. Aspect opinion expressions and their associated ratings (positive or negative) are first extracted from feedback comments then the dimension trust scores together with their weights are calculated by aggregating dimension ratings [1].

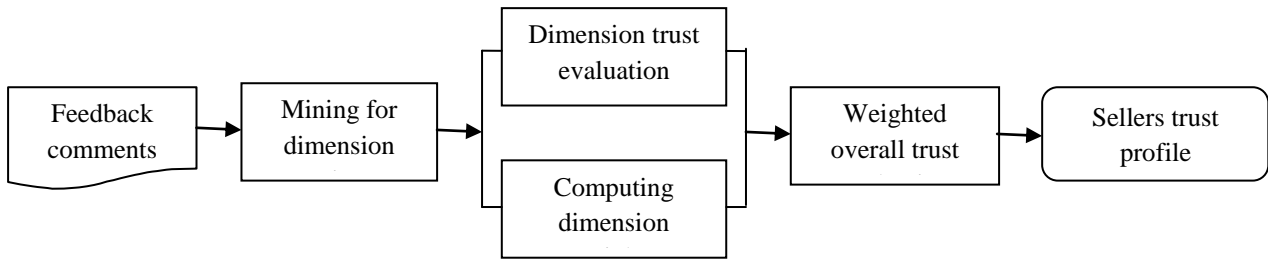


Fig 1: System Framework

In the existing systems semi-supervised algorithm is used to extract aspects. Unigram Representation is used for textual documents. Unsupervised topic modeling based techniques are also used. In all these only positive partiality in rating taken as whole is the center of attention.

3. PROPOSED SYSTEM

This research is associated to opinion mining, and sentiment analysis on text documents [1]. In recent times a semi-supervised algorithm was planned to uncover aspects and group them into significant clusters [5].

Using either the probabilistic Latent Semantic Analysis (pLSA) [3] or the Latent Dirichlet Allocation (LDA) [3] various unsupervised topic modeling-based techniques have been developed to mutually model opinions and topics [1].

The models vary in granularities and how aspects and opinions relate. None of these existing systems makes use of any lexical knowledge they all are based on the unigram depiction of documents [1].

From the ratings taken as whole in electronic commerce text feedback comments (positive or negative) some of the recent work focuses on computing aspect ratings [1]. Here, the positive bias in overall ratings is not the focus and aspect ratings and weights are calculated based on regression from the overall ratings [3].

3.1 System Architecture

Figure 2 shows the presented system architecture. The input to the system is in the form of feedback comments.

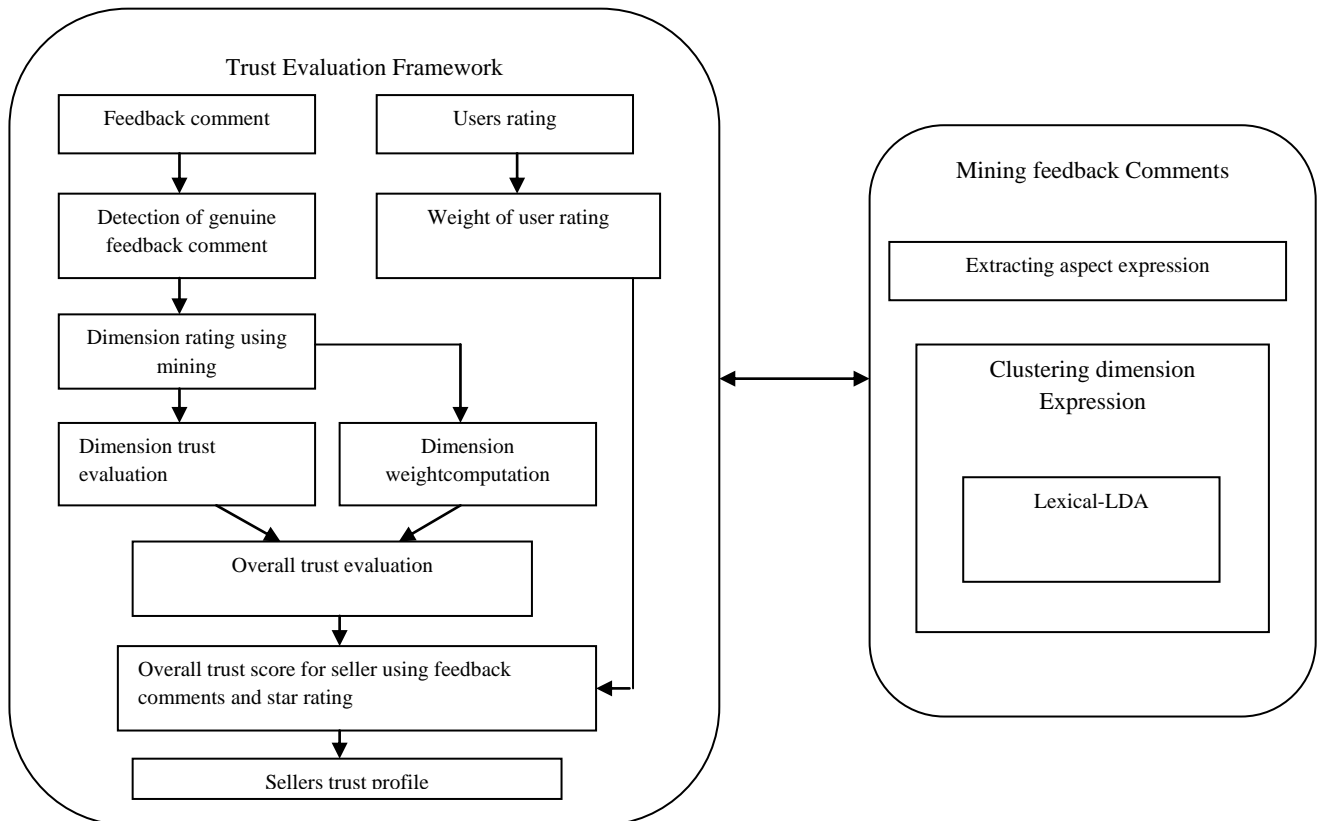


Fig2: System Architecture

3.2 Algorithm

Lexical LDA Algorithm:

a) Latent Dirichlet allocation

Latent Dirichlet allocation (LDA) is a useful representation that which can explain why several elements of the data are alike in natural language processing by allowing sets of observations to be explained by [4] overlooked groups. For instance, if explanations are words together in a document, it postulates that every document is an arrangement of a small number of subjects and that each word's formation is traceable to one of the document's subjects.

Every document can be seen as a blend of various subjects or topics. It is identical to probabilistic latent semantic analysis (pLSA), except for the fact that the topic allotment is assumed to have a Dirichlet prior in LDA. When it comes down to it, this results in more practical combinations of topics in a document. However, It has been observed, that the pLSA model is comparable to the LDA model under a uniform Dirichlet prior distribution.

LDA is a procedure that can find out topics contained by these documents automatically.

4. MATHEMATICAL MODEL

A. Feedback Comments

$$F = \{f_1, f_2, f_3 \dots f_n\}$$

Where, F is shows as a set of Feedback Comments and $f_1, f_2, f_3 \dots f_n$ are the number of feedback of sellers

B. User Ratings

$$U = \{u_1, u_2, u_3 \dots u_n\}$$

Where, U is represented as a set of user ratings i.e. star ratings and $u_1, u_2, u_3 \dots u_n$ are the number of user ratings.

C. Comments Mining Without Fake Comments

$$J = \{j_1, j_2, j_3 \dots j_n\}$$

Where, J is represented as a set of Feedback Comments after deletion of fake comments from input and $j_1, j_2, j_3, \dots, j_n$ are the number of real feedback comments for the seller.

D. Dimensions Ratings

$$D = \{d_1, d_2, d_3 \dots d_n\}$$

Where, D stands for as a set of Dimensions Ratings and $d_1, d_2, d_3 \dots d_n$ are number of ratings of sellers.

E. Dimensions Trust

$$T = \{t_1, t_2, t_3, \dots, t_n\}$$

Where, T stands for as a set of Dimensions trusts and $t_1, t_2, t_3, \dots, t_n$ is number of trusts of sellers.

F. Dimensions Weight:

$$W = \{w_1, w_2, w_3, \dots, w_n\}$$

Where, W is representing as a set of Dimensions Weights and w_1, w_2, w_3, w_n are number of weights of a sellers.

G. User Ratings Dimension Weight

$$Y = \{y_1, y_2, y_3 \dots y_n\}$$

Where, Y represents the set of User Ratings Dimension Weight and $y_1, y_2, y_3 \dots y_n$ are the number of weight of overall user ratings.

H. Overall Trust Evaluation by Feedback Comments

$$C = \sum_{d=1}^m td * wd$$

Where, C - Overall Trust Score

td - trust score for dimension $d = (1m)$

wd - weight for dimension $d = (1m)$

I. Overall Sellers Trust Score

$$Os = C + Y/2$$

Where, Os = Overall sellers trust score

C = Overall Trust Score

Y = User Ratings Dimension Weight

5. CONCLUSIONS

In the reputation management systems of popular e-commerce web sites like eBay and Amazon [1] the "all good reputation" problem is well recognized. Even though buyers or customers may give high feedback ratings on transactions, they express direct negative opinions on different aspects of transactions in the free text feedback comments [1].

In this paper, a system for trust score calculation by mining feedback comments is proposed, on the basis of observation that the customers tend to express their views and opinions freely in the feedback comments section. The main objectives are to implement a multidimensional trust model [1] for calculating the seller's reputation score from the information in the customer feedback comments and an algorithm for mining the text comments for dimension ratings and weights [1], using some techniques from natural language processing and from opinion mining [1] along with topic modelling.

CommTrust can be used to reliably calculate the trustworthiness of sellers. However, it will still need some improvement to mine more detailed information from feedback comments. In online feedback comments, casual language is commonly used to express users' opinion. In future work mining techniques can be improved to identify terms more accurately. Future work can explore the possibility of understanding the contents more in-depth.

6. ACKNOWLEDGEMENTS

We would like to express our gratitude to all those who gave me the possibility to complete this project. We would like to thank our guides Dr.Sulochana Sonkamble and Professor S.S Shinde, for their support, and encouragement in completing this paper. They have been great supervisors in motivating us on doing high quality work and keeping us on track of progress. The other members of JSPM faculty were equally helpful and their insights were also beneficial in our project, allowing us to engage in frequent discussions related to the topic, which helped to improve the ideas and presentation found in this paper

7. REFERENCES

- [1] Zhang, Xiuzhen, Lishan Cui, and Yan Wang. "Commtrust: computing multi-dimensional trust by mining e-commerce feedback comments." *Knowledge and Data Engineering, IEEE Transactions on* 26.7 (2014): 1631-1643.
- [2] Resnick, Paul, and Richard Zeckhauser. "Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system." *The Economics of the Internet and E-commerce* 11.2 (2002): 23-25.

- [3] O'Donovan, J., Smyth, B., Evrim, V., & McLeod, D. (2007, January). "Extracting and Visualizing Trust Relationships from Online Auction Feedback Comments". In *IJCAI* (pp. 2826-2831).
- [4] Wang, Hongning, Yue Lu, and Chengxiang Zhai. "Latent aspect rating analysis on review text data: a rating regression approach." *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010.
- [5] Wang, Hongning, Yue Lu, and Chengxiang Zhai. "Latent aspect rating analysis without aspect keyword supervision." *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011.
- [6] Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining." *LREC*. Vol. 10. 2010.
- [7] Hijikata, Yoshinori, et al. "Social summarization of text feedback for online auctions and interactive presentation of the summary." *Knowledge-Based Systems* 20.6 (2007): 527-541.
- [8] Jøsang, Audun, Roslan Ismail, and Colin Boyd. "A survey of trust and reputation systems for online service provision." *Decision support systems* 43.2 (2007): 618-644.
- [9] Koren, Yehuda. "Collaborative filtering with temporal dynamics." *Communications of the ACM* 53.4 (2010): 89-97.
- [10] Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis Lectures on Human Language Technologies* 5.1 (2012): 1-167.
- [11] S. Ramchurn, D. Huynh, and N. Jennings, "Trust in multiagentsystems," *The Knowledge Engineering Review*, 2004.