Reputational Key Word Search by E-Commerce Feedback

K. Kumar[1],Bullarao Domathoti[2], NageswaraRao Putta[3]
[1, 2, 3]Department of CSE
[1, 2, 3]Swetha Institute of Technology &Science, Tirupati , AP, INDIA

Abstract*-Popularity-based trust models are widely utilized in e-commerce purposes, and feedback scores are aggregated to compute sellers' fame believe ratings. The all just right popularity" problem nonetheless is standard in current repute systems popularity scores are universally excessive for agents and it is complex for advantage purchasers to prefer nontoxic sellers. On this thesis, established on the statement that purchasers more commonly express opinions openly in free text feedback feedback, now we have proposed CommTrust, a multi-dimensional trust evaluation model, for computing comprehensive believe profiles for sellers in e-commerce functions. Exceptional from present multi-dimensional believe units, we compute dimension trust scores and dimension weights mechanically by way of extracting dimension rankings from feedback comments.*

Situated on the dependency relation parsing process, we've proposed Lexical-LDA (Lexical subject Modeling based method) and DR-mining (Lexical competencies based approach) techniques to mine suggestions comments for dimension ranking profiles. Each methods reap significantly better accuracy for extracting dimension scores from feedback feedback than a most likely used opinion mining procedure. Large experiments on eBay and Amazon data show that CommTrust can effortlessly deal with the all just right status" challenge and rank marketers easily. To the quality of our talents, our research demonstrates the radical software of combining average language processing with opinion mining and summarisation tactics in believe analysis for e-commerce purposes.

I. INTRODUCTION

There has been a tremendous growth in e-commerce applications, where buyers and sellers conduct transactions on the web. Users are attracted to online-shopping not only due to the convenience in accessing the information of items on-sold, but also because of the availability of other buyers feedback on their purchasing experience, item-related and/or seller-related. All major online-shopping websites encourage buyers to provide feedback, often in the form of ratings along with some textual comments, to facilitate potential transactions. Reputation reporting systems [Resnick et al., 2000; Xiong and Liu, 2004; Zacharia and Maes, 2000] have been implemented in e-commerce systems such as eBay and Amazon (for third-party sellers), where overall reputation trust scores for sellers are computed by aggregating feedback ratings. In e-commerce environments, reputation mechanisms are related to the ratings that a seller received from buyers. The ratings indicate the ability of the seller to provide satisfactory transactions in the future, which is bene_cial to new buyers. For example on eBay, the reputation score for a seller is computed by aggregating buyer feedback ratings in the past 12 months, such as either the total number of positive ratings minus the total number of negative ratings or the percentage of positive ratings out of the total number of positive ratings and negative ratings.1 A well-reported issue with the eBay reputation management system is the \all good sellers" problem [Resnick et al., 2000; Resnick and Zeckhauser, 2002] where feedback ratings are over 99% positive on average [Resnick et al., 2000]. Such strong positive bias can hardly guide buyers to select sellers to transact with them. At eBay detailed seller ratings for

1http://pages.ebay.com/help/feedback/allaboutfeedback.html

sellers (DSRs) on four aspects of transactions, namely item as described, communication, postage time, and postage and handling charges are also reported. DSRs are aggregated rating scores on a 1- to 5-star scale. Still the strong positive bias is present { our analysis on sample eBay data shows that on average over 60% of aspect ratings are 4.9 stars. One possible reason for the lack of negative ratings at e-commerce web sites is that users who leave negative feedback ratings can attract retaliatory negative ratings and thus damage their own reputation [Resnick et al., 2000]. Note also that DSRs are not used to compute the overall trust scores for sellers.

The textual comments can provide specific understanding that's not available in ratings. Although purchasers depart positive feedback ratings, they still categorical some disappointment and negativeness in free textual content feedback comments, traditionally closer to speci_c facets, or dimensions of transactions. For instance, a comment just like the merchandise had been as I expected." expresses optimistic opinion towards the Product dimension, whereas the remark delivery was a little slow but or else, nice service. Recommend particularly." expresses bad opinion in the direction of the supply dimension but a optimistic rating to the transaction most likely. There are a number of factors why feedback provide extra safe know-how. First, ordinal scores are interpreted di_erently by di_erent users. Some users tend to expense greater while others are inclined to cost lessen. Secondly, most online looking web pages additionally allow sellers to fee the buyers to counter-steadiness the affect of malicious purchasers. Considering that the normal ranking would affect the sales broadly, sellers could use this mechanism as a weapon to shield their trade, score down shoppers who provide low scores on their buy. As such, the mechanism effectively leads to pseudo excessive scores than what comments are reecting. From the customer's standpoint, whilst the natural ranking is probably not a thoroughly trustworthy measure, it's the handiest without problems obtainable measure. Searching by means of tens of pages of comments can also be time consuming, and to digest the expertise is a frightening task, as well. This calls for a greater measure to represent the popularity of seller safely. Such fame is many times referred to as believe, which is de_ned by way of Wang and Lim [2008] as the extent to which one celebration measures the opposite social gathering's willingness and capability to behave within the measuring get together's interest". By way of analysing the wealth of information in suggestions comments we are able to find customers' embedded opinions in the direction of different features of transactions, and compute complete fame profiles for retailers. Specifically utilizing the positive and bad subjectivity of opinions in the direction of aspects of transactions as dimension rankings, we advocate comment-foundedMulti-dimensional trust (CommTrust), anegrained multi-dimension believe analysis model for e-commerce purposes.

1.2 Research Problems

Di_erent from existing work of computing trust from user ratings, we propose a multidimensional trust model based on feedback comments. The trust is decomposed into multiple aspects to represent di_erent dimensions of a transaction, including such as the quality of products or the delivery status of orders. We derive trust dimensions from textual feedback comments and combine customer preferences on each dimension to highlight customer concerns. There are four main research questions:

1. How can multi-dimensional trust from extracted dimensions and the associated opinion polarity be computed?

   In e-commerce environments different transactions may have different contexts. The trustworthiness of a seller should be related to forthcoming transactions. How to efficiently compute the trust level for a seller from the sentiment expressed in buyer feedback comments and represent it effectively is our first task.

2. How can dimensions from online feedback comments that customers have expressed their opinions on be more accurately identified?

   In e-commerce, sellers provide products and services, and buyers pay for them. During the process of _nishing these transactions, the quality of products, communication of sellers (whether the seller has friendly communication with buyers), delivery time (whether the seller delivers items on time) and shipping charges (whether the charges are reasonable) might be some of the dimensions which buyers are interested in. In online feedback comments, di_erent customers describe di_erent aspects of dimensions. How to accurately identify these dimensions expressed in natural language textual comments is our second task.

3. How can weights for each dimension that extracted from feedback comments be evaluated?

   In e-commerce, the feedback comments and ratings leaved by buyers are highly noisy. There are many comments are writen from the same buyer and therefore are highly correlated. Some buyers are lenient or harsh raters and therefore their ratings should be taken with a grain of salt. How to e_ciently evaluate the weights of each dimensionis our third task.

4. How can sentiment from textual feedback for each dimension be more accurately classified?

   Sentiment classi_cation aims to identify view-point from information expressed in text. Whether a piece of text is expressing positive or negative attitude towards associated dimensions of comments need to be identi_ed. How to accurately classify sentiment is our fourth task.

1.3 Thesis objectives and scope

Our work aims to provide a comprehesive trust pro_les for sellers that allows buyers to conduct their online shopping based on past experience. Our focus is on extracting dimension ratings from feedback comments and further aggregating these dimension ratings to compute dimension trust scores. The motivation of our research is that online feedback comments contain disdinctinformatnion for users to rank sellers, therefore content of comments can be used to reliably evaluate the trustworthiness of sellers.

The contribution of this thesis are:

We recommend to make use of remark-centered Multi-dimensional trust (CommTrust), an grained multi-dimension evaluation model, to calculate the believe for e-commerce functions.

Whilst the mannequin is probably extensible to goal item-specific trust, in this be trained we focal point on computing complete trust pro_le for sellers. We advocate an algorithm to establish dimension ranking expresses from suggestions feedback by way of applying lexicon-based opinion mining methods [Pang and Lee, 2008] in combo with dependency relation analysis, a tool lately developed in natural language processing (NLP) [De Marne_e et al., 2006; De Marne_e and Manning, 2008]. We deal with the 4 study questions by way of two tactics:

1. The topic modelling approach is applied to develop the Lexical-LDA algorithm for grouping dimension rating extraction and trust computation. Lexical LDA makes use of two types of lexical knowledge based on dependency relations for clustering dimension expressions into dimensions so as to produce meaningful cluster. The _rst lexical knowledge is that the co-occurrence of dimension expressions with respect to a same modi_er across comments can provide more meaningful contexts for dimension expressions, compare to add on counts of dimension expressions by comments. The second knowledge is that the dimension expressions extracted from the same comment are very unlikely about the same topic. Based on these two types of lexical knowledge, we revised Latent Dirichlet Allocation (LDA) [Blei et al., 2003] to develop the Lexical-LDA algorithm.
2. With the seed dimension words we propose Dimension Rating mining (DR-mining), a knowledge-based approach that incorporates domain knowledge, meta-data, and general grammatical patterns to accurately identifying dimension rating expressions from feedback comments. The matrix factorisation technique applied to automatically compute trust weights.

To the best of our knowledge, CommTrust is the _rst piece of work that computes _negrained multidimensional trust pro_les automatically by mining feedback comments. The rest of this thesis is organized as follows. In Chapter 2, the necessary background knowledge of the trust evaluation, sementic analysis, and text comments mining related works is introduced. In Chapter 3, we propose the comment-based multi-dimensional trust (CommTrust) model to identify trustworthy and reliable sellers.. In Chapter 4, we present topic modelling approach to mining feedback comments for dimension rating pro_les. We propose the Lexical-LDA algorithm to conduct dimension rating extraction and trust computation. In Chapter 5, we propose a knowledge-based approach that incorporates domain knowledge, meta-data, and general grammatical patterns to mining feedback comments for dimension rating pro_les. We formulate the problem of computing dimension weights from ratings as a factor analytic problem and propose a matrix factorisation technique to automatically compute weights for dimensions from the sparse and noisy dimension rating matrix. We conclude out study in Chapter 6, where the work of the thesis is summarised, particularly in relation to the research questions. Furth more the future research problems are discussed.

II. LITERATURE REVIEW

Related work for our research falls into four main areas: 1) computational approaches to trust, especially reputation-based trust evaluation and recent developments in _ne-grained trust evaluation; 2) e-commerce feedback comments analysis and more generally mining opinions on movie reviews, product reviews and other forms of free text documents; 3) aspect opinion extraction and summarisation on movie reviews, product reviews and other forms of free text; and 4) applications of the matrix factorisation technique for recommender systems and other data mining tasks.

2.1 Computational trust evaluation

The strong positive rating bias in the eBay reputation system has been well documented in literature [Resnick et al., 2000; Resnick and Zeckhauser, 2002; ODonovan et al., 2007], although no e_ective solutions have been reported. Notably in [ODonovan et al., 2007] it is proposed to examine feedback comments to bring seller reputation scores down to a reasonable scale, where comments that do not demonstrate explicit positive ratings are deemed negative ratings on transactions. Ratings on transactions are further aggregated as the overall trust scores for sellers. In this study on the other hand, our focus is on extracting dimension ratings from feedback comments and further aggregating these dimension ratings to compute dimension trust scores.

2.2 Feedback comment analysis

The success of e-commerce applications, such as eBay and Amazon, depends highly on the availability of user interaction. Usually, reputable sellers attract a large user population to transact with them and leave comments afterwards. Intuitively, one can use a reputation score to quantify how good a seller is at providing good services. However, the strong positive rating bias in reputation system has been noted in literature [Resnick et al., 2000; Resnick and Zeckhauser, 2002; ODonovan et al., 2007]. There have been studies on analyzing feedback comments in e-commerce applications to capture negativity information to provide reasonable reputation range score for sellers [ODonovan et al., 2007; Gamon, 2004; Hijikata et al., 2007; Lu et al., 2009].

2.3 Aspect opinion extraction and summarisation

More generally our work is related to opinion mining and sentiment analysis on free text documents, especially opinion mining in product reviews and movie reviews. In these studies, product or movie features and the opinions towards them are extracted. Summaries are produced by selecting and re-organising sentences according to the extracted features. Review mining and summarization is the task of producing a sentiment summary, which consists of sentences from reviews that capture the authors opinion. Review summarization is interested in features or aspects on which customers have opinions. It also determines whether the opinions are positive or negative. This makes it di_er from traditional text summarization. A comprehensive overview of the _eld is presented in [Pang and Lee, 2008; Liu, 2012]. Most existing works on review mining and summarization mainly focus on product reviews. For example, [Hu and Liu, 2004b; Popescu and Etzioni, 2005; Shi and Chang, 2006] concentrated on mining and summarizing reviews by extracting opinion sentences regarding product aspects.

III. COMMTRUST: COMMENT-BASED MULTI-DIMENSIONAL TRUST EVALUATION

We view feedback comments as a source where buyers express their opinions more honestly and openly. Our analysis of feedback comments on eBay and Amazon reveals that even if a buyer gives a positive rating for a transaction, s/he still leaves comments of mixed opinions regarding different aspects of transactions in feedback comments. Table 3.1 lists some sample comments, together with their rating from eBay. For example for comment c2, a buyer gave a positive feedback rating for a transaction, but left the following comment: \bad communication, will not buy from again. super slow ship(ping). item as described.". Obviously the buyer has negative opinion towards the communication and delivery aspects of the transaction, despite an overall positive feedback rating towards the transaction. We call these salient aspects dimensions of e-commerce transactions. Comment-based trust evaluation is therefore multi-dimensional. Hereafter we will use the terms opinion and rating interchangeably to express the positive, negative and neutral polarities toward entities that expressed in natural language text.

3.1 The CommTrust model

Figure 3.1 depicts the CommTrust framework. Unlike existing trust models (including the one used on eBay) where explicit transaction feedback ratings (positive or negative) are used to compute overall trust scores for sellers. Aspect opinion expressions, and their associated ratings (positive or negative) are trust extracted from feedback comments. Dimension trust scores together with their weights are further computed by aggregating dimension ratings.

Table 3.1:

No	Comment	eBay rating
c_1	beautiful item! highly recommend using this seller!	1
c_2	bad communication, will not buy from again. super slow ship(ping). item as described.	1
c_3	quick response	1
c_4	looks good, nice product, slow delivery though.	1
c_5	top seller. many thanks. A+	1
c_6	great price and awesome service! thank you!	1
c_7	product arrived swiftly! great seller.	1
c_8	great item. best seller of ebay	1
c_9	slow postage, didn't have the product asked for, but seller was friendly.	1
c_{10}	wrong color was sent, item was damaged, did not even fit phone.	1

Note: 1 = Positive, 0 = Neutral, -1 = Negative

Definition 3.1.1. The overall trust score T for a seller is the weighted aggregation of dimension trust scores for the seller,

$$T = \sum_{d=1}^{m} t_d * w_d, \qquad (3.1)$$

where td and wd represent respectively the trust score and weight for dimension d (d = 1::m). The trust score for a dimension is the degree or probability that buyers express positive opinion towards the dimension, and roughly is positively correlated with the proportion of positive ratings towards the dimension. However, buyers only express limited positive or negative opinions towards some dimensions in feedback comments. Computing the trust score from a limited number of samples has a high chance of over estimate. For example, out of 1,956 feedback comments for ten eBay sellers of the transactions from 31 January to 18 March 2012, only 73 comments contain ratings towards the communication dimension, where 72 are positive. The percentage of positive ratings is thus 98.6% (72 out of 73). However, this estimation is made from a limited sample of only 73 ratings. We propose to apply Bayesian adjustment to compute the trust scores for dimensions from a limited number of ratings.

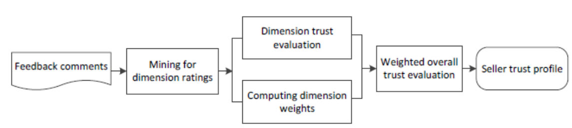

Figure 3.1: The CommTrust framework

Dimension satisfactorily. The trust score for a dimension can be estimated from the number of observed positive and negative ratings towards the dimension. Let

S={X1, ...,Xn} be n observations of binary positive and negative ratings, where y observations are positive ratings. S follows binomial distribution B(n; p). Following the Bayes rule, p can be estimated from observations and some prior probability assumption. Assuming the Beta distribution for the prior,

$$Beta(p|\alpha,\beta) == \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}$$

where _ and _ are hyper-parameters expressing prior beliefs, the Bayes estimate of p is formed by linearly combining the mean α/(α+β) from prior distribution and the mean y/n, as below [Casella and Berger, 1990; Heinrich, 2005]

$$\hat{p} = \frac{y+\alpha}{n+\alpha+\beta}. \tag{3.2}$$

It has been shown in the Beta reputation system [Josang and Ismail, 2002] that the assumption of Beta distribution for the prior belief leads to reasonable trust evaluation. The Beta reputation system adopts constant settings of α= β= 1 for Equation 3.2. We develop the approach further by introducing hyper-parameter settings for αand βto suit for a varying number of observed positive and negative ratings. It is preferable to have only one parameter for trust evaluation [Josang and Ismail, 2002]. With the prior belief of neutral tendency for trust, it can be assumed that α= β. Let α +β= m, then α = β = 1=2 *m. The trust score for a dimension is thus defined as follows:

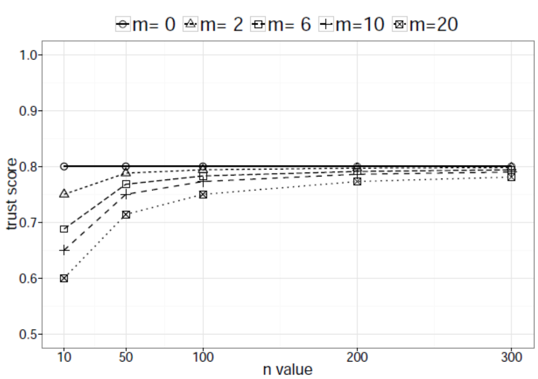

Figure 3.2: The dimension trust score model

$n = |\{v_d|v_d = +1 \vee v_d = -1\}|$, the trust score for d is:

$$t_d = \frac{|\{v_d|v_d = +1\}| + 1/2 * m}{n+m}. \tag{3.3}$$

Equation 3.3 is also called m-estimate [Karplus, 1995]. According to De_nition 3.1.2, td is in the range of [0..1], and 0.5 represents the neutral tendency for trust. In Equation 3.3, m is a hyper-parameter and can be seen as pseudo counts { 1=2 _m counts for the positive and negative classes respectively. The higher value of m, the more actual observations are needed to revise the natural neutral trust score of 0.5. More importantly by introducing the prior distribution using the super-parameter m, the adjustment can reduce the positive bias in ratings, especially when there are a limited number of positive and negative ratings [Resnick et al., 2000; Resnick and Zeckhauser, 2002].

Based on our experiment datasets (1,956 feedback comments for ten eBay sellers of the transactions from 31 January to 18 March 2012) refer to Section 5.3, Figure 3.2 plots trust score td by Equation 3.3 in relation to different settings of total number of ratings n and pseudo counts m. The figure is plotted for y/n = 0:8, and similar trends are observed forother values of y/n. It shows that when the total number of

No	Product	Delivery	Comm.	Cost	Tran.	eBay	Comment
c_1	1	0	0	0	1	1	beautiful item! highly recommend using this seller!
c_2	0	0	0	0	1	1	great service
c_3	0	0	1	0	0	1	quick response
c_4	1	-1	0	0	0	1	looks good, nice product, slow delivery though.
c_5	0	0	0	0	1	1	top seller. many thanks. A+
c_6	0	0	0	1	1	1	great price and awesome service! thank you!
c_7	0	1	0	0	1	1	product arrived swiftly! great seller.
c_8	1	0	0	0	1	1	great item. best seller of ebay
c_9	-1	-1	0	0	1	0	slow postage, didn't have the product asked for, but seller was friendly.
c_{10}	-1	0	0	0	0	-1	wrong color was sent, item was damaged, did not even fit phone.

Comm.: Communication, Tran.: Transaction

observed ratings n is large (n _ 300), td is not very sensitive to the settings of m and converges to the observed positive rating frequency of 0.8. When there is a limited number of observed ratings, that is n < 300, an observed high positive rating frequency y/n is very likely an overestimation, and so m is set to regulate the estimated value for td. With m = 2, when n _ 50 td _ 0:8. On the other hand, with m = 20, only when n < 300 td > 0:8. From our experiments, settings of m = 6::20 typically give stable results. By default, we set m = 6.

IV. CONCLUSION

In this chapter we have proposed comment-based multi-dimensional trust model, computing overall and dimensional turst scores from feedback comments. Based on an example of dimension ratings for an eBay seller, the CommTrust can signi_cantly reduce the strong positive bias in eBay reputation systems and solve the \all good sellers" problem. We have performed a user study. In this user study, the users are given the summary ofcomments from 45 pairs of ten different sellers, and they had to choose which seller is more trustworthy. The _ndings showed that the ranking from the user study is different from the ranking from eBay and Amazon for the same seller.

Since the user feedback are very rich and diverse, now we have proposed CommTrust to establish nontoxic and riskless sellers. In the subsequent chapter, more designated discussion shall be offered on the best way to mine the feedback to extract the dimension and dimension scores from the feedback, and the Ends of CommTrust Lexical-LDA.

REFERENCES

[1] L. Azzopardi, M. Girolami, and K. van Risjbergen. Investigating the relationship between language model perplexity and ir precision-recall measures. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, pages 369{370. ACM, 2003.

[2] S. Baccianella, A. Esuli, and F. Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proc. 7th Int. Conf. on Language Resources and Evaluation, 2010.

[3] R. Baeza-Yates, B. Ribeiro-Neto, et al. Modern information retrieval, volume 463. ACM press New York, 1999.

[4] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. the Journal of machine Learning research, 3:993{1022, 2003.

[5] J. Blitzer, M. Dredze, and F. Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In ACL, volume 7, pages 440{447, 2007.

[6] S. Brody and N. Elhadad. An unsupervised aspect-sentiment model for online reviews. In Proc. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 804{812, 2010.

[7] G. Carenini, R. Ng, and E. Zwart. Extracting knowledge from evaluative text. In Proceedings of the 3rd international conference on Knowledge capture, pages 11{18. ACM, 2005.

[8] G. Casella and R. L. Berger. Statistical inference. Duxbury Press, 1990.