CommTrust: Computing Multi-Dimensional Trust by Mining E-Commerce Feedback Comments

Aswini U \# and Akila Devi T.R \^2
\#M.Tech, CSE, B.S Abdur Rahman University, Vandalur 48, Chennai
\^Assistant Professor, CSE, B.S Abdur Rahman University, Vandalur 48, Chennai

Abstract - E-commerce applications are widely used by customers nowadays and they purchase products from the listed sellers based on their feedback ratings. Reputation-based trust models are used in these applications to rate the sellers. The “all good reputation” problem that exists in the current reputation systems show high reputation scores for sellers and potential buyers find it difficult to choose a seller who is trustworthy. Hence, based on the opinions that buyers express in the text feedback comments, a trust evaluation model “CommTrust” is proposed. This model includes 1) a multi-dimensional trust model for computing reputation scores from feedback comments 2) an algorithm for mining feedback comments for dimension rating and weights, combing natural language processing, opinion mining and topic modeling. Extensive experiments on eBay and Amazon data demonstrate that “CommTrust can effectively address the “all good reputation” problem and rank sellers effectively.

Index Terms— natural language processing, feedback comments, topic modeling, Reputation-based trust models.

I. INTRODUCTION

E-Commerce websites are popular nowadays and pave the way for customers to purchase products easily and conveniently, thanks to the online retailers Amazon, EBay, Flipkart etc. It is not uncommon for a product in a website listed, be sold by multiple sellers. So, how does a customer choose to purchase the product from a particular seller comes the question. The above mentioned retailers have built a reputation reporting system [1] that rates sellers (third party sellers for Amazon) with their scores. The higher positive rating a seller has, the higher is the reputation for that seller. Although, this system provides rating based on the customer input, there is an issue reported with the reputation management system of EBay, known as the “all good reputation problem” [1] [2]. The feedback ratings averages over 99% on a positive scale [1], thus creating a bias that makes difficult for a buyer to choose which seller to transact with. There may be two reasons for the strong positive bias. Firstly, the reason for the customer becoming unsatisfied may not be specified with the feedback rating.

For example, a customer may have liked a product, the packaging and the overall transaction, but the delivery would have been delayed. In this case, the customer may tend to score more 4.5 on a 5 star scale and comment about the delay in the text field. Secondly, customers fear they damage their own reputation by giving negative feedback ratings, as is the system in sites like EBay (http://pages.ebay.com/help/feedback/allaboutfeedback.html). In order to overcome the above mentioned “all good reputation” problem, a Comment Based, Multi Dimensional Trust (CommTrust) approach is proposed. CommTrust is a fine-grained multi-dimensional trust evaluation model achieved by mining the e-commerce feedback comments. With this approach, comprehensive trust profiles are computed for sellers that includes dimension reputation scores and weights and overall trust scores. Thus trust profiles for sellers are created by mining feedback comments.

In CommTrust, an approach that combines dependency relation analysis [3], [4] and lexicon based opinion mining techniques [5], [6] is proposed to extract aspect opinion expressions from feedback comments. Furthermore, based on the dependency relation analysis and Latent Dirichlet Allocation (LDA) topic modeling technique [7], an algorithm is proposed to cluster aspect expressions into dimensions and compute aggregated dimension ratings and weights, called the Lexical-LDA. Thus, the reputation profiles in CommTrust comprise dimension reputation scores and weights and overall trust scores for ranking sellers.

II. RELATED WORK

J. O’Donovan [8] Proposed to address the problem of unnaturally high trust ratings on Amazon, look to the free text comments and apply a classification algorithm tailored for capturing subtle indications of negativity in those comments. The situation arises frequently where users are afraid to leave a negative comment for fear of retaliatory feedback comments which could damage their own reputation. In many of these cases, a positive feedback rating is made, but the commenter still voices some grievance in the free text comment.

Pooja A. Rangari [9] Discussed that opinion mining often referred to as Sentiment Analysis is the computational field of study that analyzes people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards particular entities such as products, services, organizations, individuals, issues, events, topics, and their attributes.
The major rich source of review feedback comments are blogs, forums, social networking sites, online shopping website, etc. People mostly comment on these sites from which users and buyers Machine learning approaches are both supervised as well as unsupervised. But, opinion mining is basically a Supervised Approach where one needs to train a classifier on the training set before it is to be applied on a test set. It combines the techniques of natural language processing, information retrieval, text analytics and computational linguistics.

Yan Wang [10] Discussed that trust is a critical issue in e-commerce and e-service environments. In some applications (such as eBay), the trust management mechanism has been introduced to provide valuable information to buyers prior to placing orders and making payments. However, most studies and applications focus on determining the general trustworthiness of individuals but not providing transaction specific trust information that involves factors associated with forthcoming transactions, a new concept situational transaction trust, a novel approach to evaluate it, which binds existing trust data with a new transaction. This can deliver more accurate trust information to buyers and prevent some typical attacks.

S. Brody [11] Discusses about online review sites continue to grow in popularity as more people seek the advice of fellow users regarding services and products. Unfortunately, users are often forced to wade through large quantities of written data in order to find the information they want. This has led to an increase in research in the areas of opinion mining and sentiment analysis, with the aim of providing systems that can automatically analyze user reviews and extract the information most relevant to the user.

## III. COMMTRUST

In every E-Commerce website, feedback comments are sources in which users express their assessment, speculation, and conclusion openly. Analysis of those feedback comments in various websites reveals that, even if a buyer gives a positive rating for a product, they still might still leave comments of mixed opinions about different aspects of purchase. For example, buyer left the comment “Bad communication, will not buy again” in the text field and gave an overall positive rating for the transaction. Therefore, the comments based trust evaluation is multi dimensional.

**Definition 1** The overall trust score for a seller is the weighted aggregation of dimension trust scores for that seller.

\[
T = \sum_{d=1}^{m} t_d * w_d. \tag{1}
\]

where \( t_d \) and \( w_d \) represent respectively the trust score and weight for dimension \( d \) as in (1).

**Definition 2** Given the positive (+1) and (-1) negative ratings towards dimension \( d \), \( n = |\{vd|vd = +1 \lor vd = -1\}| \]

\[
t_d = \left[\frac{|\{vd|vd = +1\}| + 1/2 * m}{n + m}\right] \tag{2}
\]

The above equation is also called m-estimate (2) [12].
B. Clustering Dimension Expressions into Dimensions

The Lexical-LDA algorithm is proposed to cluster aspect expressions into semantically coherent categories, which we called dimensions. This algorithm makes use of two types of lexical knowledge to “supervise” clustering dimension expressions into dimensions so as to produce meaningful clusters.

- Comments are short and therefore co-occurrence of head terms in comments is not very informative. Instead, the co-occurrence of dimension expressions with respect to a same modifier across comments is used, and it potentially can provide more meaningful contexts for dimension expressions.

- It is observed in rare cases that the same aspect of e-commerce transactions is commented more than once in the feedback comment.

The clustering problem under topic modeling is formulated as: The dimension expressions for a same modifier term or negation of a modifier term are generated by a distribution of topics. Every topic is generated successively by a distribution of head terms. This allows making use of structured dependency relation representations from dependency relation parser for clustering. Hence, the input to Lexical-LDA will be in the form of (modifier, head) pairs, or their negations like (fast, shipping) or (not-good, seller).

V. EXPERIMENTS

We conducted experiments on e-commerce datasets to evaluate various aspects of CommTrust, including the trust model and the Lexical-LDA algorithm for clustering dimension expressions. Feedback comments from e-commerce websites for multiple sellers were obtained. The figures below show the trust ratings for sellers which were generated through CommTrust.

The Figure 4 shows the feedback comment analysis for a particular product “Tablet”. Color red represents positive reviews and yellow represents negative reviews.

The Figure 5 shows the analysis for the different models of Television. The x axis represents different product Ids and the y axis represents the total number of positive and negative comments.

A. Evaluation of Lexical-LDA

There was lot of informal language expressions used in the feedback comments. Hence, some pre-processing was
performed first. Spelling correction was applied. Informal expressions, for example “thankx” were replaced by “thanks”. Then, the Stanford dependency parser was applied to produce the dependency relation representation of comments and dimension expressions were extracted.

To evaluate the algorithm, the ground truth for clustering was created. Dimension expressions are (modifier, head) pairs, and to remove noise only those pairs with support for head terms of at least 0.1% or three comments were considered for manual clustering.

When we compared the obtained results with the actual feedback data in the respective e-commerce websites, accuracies hover over 70%.

VI. CONCLUSION

The “all good reputation” is a well known problem in popular websites like eBay and Amazon. The high reputation scores cannot rank sellers effectively. Therefore, this cannot guide potential buyers to choose trustworthy sellers to purchase products. It is observed that buyers express negative opinions in the free text feedback comment fields, although they provide higher ratings. We proposed to compute comprehensive multi-dimensional trust profiles for sellers from dimension ratings found in feedback comments.

Effective algorithms were proposed to compute dimension trust scores and dimension weights through extracting aspect opinion expressions from feedback comments and clustering them into dimensions. This approach demonstrates an application based on combining natural language processing, opinion mining and summarization techniques used for trust evaluation in e-commerce websites.

ACKNOWLEDGMENT

The authors are grateful to the Faculty of B. S.Abdur Rahman University for their valuable support in deploying the system and in providing experimental data and technical documentation concerning the “CommTrust: Computing Multi-Dimensional Trust by Mining E-Commerce Feedback Comments” project.

REFERENCES


Akila Devi T R holder of B. Tech & M. Tech in Computer Science & Engineering. Now she is Assistant Professor in B S Abdur Rahman University, vandalur, chennai.