OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Mining E-Commerce Feedback Comments to Evaluate Multi Dimension Trust

Ayinavalli Ventaka Ramana[1], N. Venkateswara Rao[2]
*(Department of Information Technology, GMRIT/JNTUK,India
** (Department of Information Technology, GMRIT/JNTUK,India

ABSTRACT: *In E-Commerce application, the Reputation based trust models are widely used. To allocate ranks to the sellers feedback ratings and comments gathered together. We proposed a system namely CommTrust, it uses observations of buyers on products often express their opinions in feedback about products in text format. These reviews are mined. In CommTrust we propose K-Means algorithm is proposed for mining feedback comments which are used for weights and ratings of product. And also we propose aggregation method for ranking the sellers. I want to conduct experiments on various websites like Amazon and eBay, the CommTrust proved to be very effective.*

Keywords: *E-Commerce, text mining, k-means, FSM.*

I. Introduction

Many of the Reputation systems have been implemented in e-commerce systems such as eBay, Amazon and etc. where the ranking score for sellers calculated based on the feedback ratings given by buyers. For example on eBay, the reputation score for a seller is the positive percentage score, as the percentage of positive ratings out of the total number of positive ratings and negative ratings in last few months. A well consider issue with the eBay reputation management system is the "all good reputation" problem. In eBay all seller got 99% positive feedback on Average. This strong positive average hardly guides and forces the buyers to select one seller among many sellers are available of that product in in eBay. We take DSRs are aggregated ratings on best-worst. Still the strong positive rating is presented mostly best or good. There one possible negative rating is the chance for the lack of negative ratings at e-commerce sites, it attract the users who gives the negative feedback about the products it damages their own reputation score in purchasing sites like eBay.

Although the buyers give positive feedback ratings and negative ratings like some disappointment about the products and its transactions. For example "the product is very excellent" treated as a positive review towards the product aspect, where the comment "delivery charge is too expensive, otherwise it's good" treated as a negative review comment towards the price aspect but positive opinion to the transaction in general.

We propose Comment-based Multi-dimensional trust (CommTrust), a well multi-dimensional trust evaluation model by mining e-commerce feedback comments. In CommTrust, we propose K-Means algorithm for cluster feedback comments. K-Means is a numerical, unsupervised iterative method. It is Original and very fast, so in many practical applications this method is proved to be effective way that can produce very good results in clustering. But the complexity of k- Means is very high, particularly for large datasets. And also we propose one of the aggregation methods in ranking namely Feature selection method (FSM) based on feedback comments. Ranking to sellers is necessary task for in any purchasing site like eBay, Amazon and etc. It is very useful for buyers to select best and Trust worthy sellers based on ranking. By conducting experiments on purchasing web sites like eBay and Amazon reputation systems, we have to solve all reputation problems and allocate ranks to the sellers very effectively.

II. Commtrust

A well noticed multi-dimensional trust evaluation model by mining e-commerce feedback comments is called as Comment-based Multi-dimensional trust (CommTrust). A complete trust profile of sellers is computed by CommTrust. It is the first system which calculates fine-grained multidimensional trust profiles automatically by mining feedback comments is CommTrust.

Related Work: Related work divided into three main areas: 1) computational approaches to trust, especially reputation based trust evaluation and recent developments in well noticed trust evaluation; 2) e-commerce

feedback comments analysis and 3) aspect opinion extraction and summarization on movie reviews, product reviews on free text feedback comments.

2.1 Trust Evaluation:

In literature [5]-[7], the good positive feedback rating particularly eBay reputation system is well acknowledged. As proposed in [7], to study feedback comments to get seller status scores down to a balanced scale. There comments that do not determine explicit positive ratings are judged negative ratings on transactions.

Similar to that buyers and sellers are considered as individuals in e-commerce applications. Peers and agents are terms used to specify the individuals in open systems in various applications in the trust evaluation literature. The complete overview of trust model is provided in [11]. Individual level trust models aims to compute the trustworthiness of peers and assist buyers in their work of decision making [15]-[17].

Rating aggregation algorithms for calculating individual reputation scores include simple positive feedback percentage or average of ratings as in the eBay and Amazon reputation systems in [18]. Reputation based on statistical sharing assumption for ratings, as well as more advanced models calculates trust score variance and confidence level in e-commerce.

2.2 Feedback Comment Analysis:

After examined these 4([7], [12], [13], [14]) we analyzing feedback comments in e-commerce purchasing web sites like eBay and Amazon. It says that their focus was not albeit the complete trust evaluation. The main focus of [7] and [13] was sentiment classification of feedback comments. It is verifies that feedback comments are noisy and later analyzing them is a challenge. [7] States that missing aspect comments are judged negative and also models built from aspect ratings are used to sort the feedback comments into positive or negative. [14] Proposed a technique for summarizing feedback is mentioned. It aims at to filter out considerate comments that do not provide real feedback. Lu.Et al. [12] elaborates on producing "rated aspect summary" from eBay feedback comments. Its statistical generative model has basis on regression on the overall transaction ratings.

2.3 Aspect Opinion Extraction And Summarization:

Opinion mining or sentiment analysis on free text documents is related to our work. Aspect opinion mining on product reviews and movie reviews [19]–[21] are the existing work. In [19] for product reviews noun phrases and frequent nouns are considered as aspects. To mine aspect opinions for movie reviews dependency relation parsing is used in [21]. Some work group's aspects into clusters, assuming Aspect opinion expressions are given [22]. To extract aspects a semi supervised algorithm [23] was proposed and group them into meaningful clusters as managed by user input words. Some recent work on calculating aspect ratings from overall ratings in e-commerce feedback comments or reviews [12], [24], [25]. Based on regression from overall ratings and the positive partiality in overall ratings by using this to calculate aspect ratings and weights.

III. Commtrust: Comments-Based Multi-Dimensional Trust Evaluation

Buyers express their opinions more fairly and openly through comments in feedback, we take this feedback as a source. Our experiments on eBay and Amazon, we analysis review comments it reveals that even if a buyer gives a positive rating for a transaction. Buyer leaves the comments of varied opinions about different features of transactions in feedback comments. For example, a buyer wants to give negative feedback about a transaction, but unfortunately s/he leave the following comment: "price is good and also fast shipping". Obviously the buyer has positive opinion towards the price and delivery aspects of the transaction, even though an overall positive feedback rating towards the transaction. We call these aspects as dimensions of e-commerce transactions. Comments-based trust evaluation is therefore multi-dimensional.

IV. Mining Feedback Comments For Ranking

We propose K-means algorithm for clustering expressions in feedback, based on the ratings (best-bad) we rank the sellers by using FSM. We simply show the seller rankings in bar chart format. This graph contains vertical bars and each bar represents a specific seller ratings.

4.1 Algorithm:

We propose k-means algorithm for clustering review comments in feedback.

K-means clustering is a method of vector quantization, it's from signal processing, that is popular for clustering analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation be connected to the cluster with the nearest mean, serving as a prototype of the cluster.

K-Means algorithm:

K means clustering is a partition-based cluster analysis method. According to this algorithm we firstly select k data value as initial cluster centers, then calculate the distance between each data value and each cluster center and assign it to the closest cluster. K means clustering aims to partition data into k clusters in which each data value belongs to the cluster with the nearest mean.

The most common algorithm, described below, uses an iterative refinement approach, following these steps:

1. Define the initial groups' centroids. This step can be done using different strategies. A very common one is to assign random values for the centroids of all groups. Another approach is to use the values of K different entities as being the centroids.
2. Assign each entity to the cluster that has the closest centroid. In order to find the cluster with the most similar centroid, the algorithm must calculate the distance between all the entities and each centroid.
3. Recalculate the values of the centroids. The values of the centroid's fields are updated, taken as the average of the values of the entities' attributes that are part of the cluster.
4. Repeat steps 2 and 3 iteratively until entities can no longer change groups.

The K-Means is computationally efficient technique, being the most popular representative-based clustering algorithm. The pseudo code of the K-Means algorithm is shown below

Algorithm 1: K-Means Algorithm

Input: $E = \{e_1, e_2, \ldots, e_n\}$ (set of entities to be clustered)
k (number of clusters)
$MaxIters$ (limit of iterations)

Output: $C = \{c_1, c_2, \ldots, c_k\}$ (set of cluster centroids)
$L = \{l(e) \mid e = 1, 2, \ldots, n\}$ (set of cluster labels of E)

foreach $c_i \in C$ **do**
  $c_i \leftarrow e_j \in E$ (e.g. random selection)
end
foreach $e_i \in E$ **do**
  $l(e_i) \leftarrow argminDistance(e_i, c_j) j \in \{1 \ldots k\}$
end

$changed \leftarrow false$;
$iter \leftarrow 0$;
repeat
  foreach $c_i \in C$ **do**
    $UpdateCluster(c_i)$;
  end
  foreach $e_i \in E$ **do**
    $minDist \leftarrow argminDistance(e_i, c_j)\ j \in \{1 \ldots k\}$;
    if $minDist \neq l(e_i)$ **then**
      $l(e_i) \leftarrow minDist$;
      $changed \leftarrow true$;
    end
  end
  $iter++$;
until $changed = true\ and\ iter \leq MaxIters$;

Feature Selection Method:

To select features from the entire feedback dataset set, in our method we first define the importance score of each comment, and define the similarity between any two review comments. Then we employ an efficient algorithm to minimize the total similarity scores and maximize the total importance scores of a set of comments. Feature selection is a term commonly used in data mining to describe the tools and techniques available for reducing inputs to a manageable size for processing and analysis.

The capability to apply feature selection is critical for effective analysis, because frequently more information contained by datasets than is needed to build the model. For example, a dataset might contain 500 columns that describe the characteristics of sellers, but if the data in some of the columns is very sparse. You would gain very little benefit from adding them to the model. You could use feature selection techniques to automatically discover the best features and to exclude values that are statistically insignificant.

The feature selection method for ranking (FSM Rank) algorithm is shown in below.

Algorithm 1 FSMRank

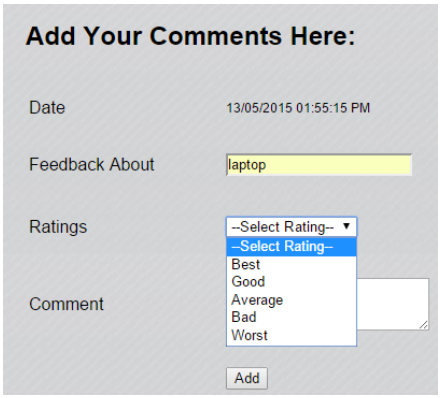
Input: $S = \{X^{(q_k)}, Y^{(q_k)}\}_{k=1}^{n}$, λ_1, λ_2, T and L_0; Output: w_T
Initialize: $w_0 = z_1 = \mathbf{0}$, $\alpha_1 = 1$, $\gamma = 2$, $L = L_0/\gamma$ [10]
For $t = 1, \ldots, T$
  Let $g = \ell'(z_t)$.
  While (true)
    $w_t = argmin_{w_t : w_t \in R_+^{2d}} \; Q(w_t, z_t)$ **(projection step)**
    If $\ell(w_t) \leq \ell(z_t) + \langle \ell'(z_t), w_t - z_t \rangle + \frac{L}{2}||w_t - z_t||^2$
      break;
    End If
    $L = \gamma L$;
  End While
  IF $\frac{|F(w_t) - F(w_{t-1})|}{|F(w_{t-1})|} \leq \epsilon_s$
    break;
  $\alpha_{t+1} = \frac{1+\sqrt{1+4\alpha_t^2}}{2}$
  $z_{t+1} = w_t + \frac{\alpha_t - 1}{\alpha_{t+1}}(w_t - w_{t-1})$
End For

V. Experiment Results

General experiments on review datasets were conducted to evaluate various aspects of CommTrust, including the trust model and the K-Means algorithm for clustering feedback comments.

5.1 Datasets:

We take some large number of feedback comments on products like laptops, mobiles, etc. and we randomly consider the ratings from best too bad for ranking the sellers. Based on these ratings the buyer can see how many user give positive feedback between good and best ratings. Here best and good ratings are considering as 5to 4 scoring rates. The dataset contains different review comments including ratings about particular product as shown in below sample pic.



The sample pic shows how the feedback comments are given by user or buyers about product laptop. The feedback comments ratings treated as feedback score for sellers. These ratings stored in the dataset and based on these ratings we provide rankings to the sellers by using FSM .

Here the sample dataset for feedback ratings about products and the count of the ratings shown below.

Product	**Feedback rating**	**Count**
Laptops	Best	124
Laptops	Good	90
Laptops	Average	150
Laptops	Bad	35
Mobiles	Good	200
Mobiles	Best	150
Mobiles	Average	201

By using these dataset ratings and based on these ratings the we can analyses and search which seller got these ratings. Based on this, we give ranking to the sellers using FSM.

VI. Conclusion

The "all good reputation" problem is well known for the reputation management systems of popular e-commerce web sites like eBay and Amazon. The sellers good rankings score cannot force the buyers to select one of the trustworthy seller among the many sellers. In this paper we propose feedback ratings in text format from best too worst and also take many review comments. We cluster these feedback comments by using K-means and by applying FSM algorithm we provide rankings to the sellers and also allows the buyers to see the sellers rankings related to that particular products in e-commerce application.

REFERENCES

[1]. Li Xiong, Ling Liu," A Reputation-Based T rust Model for Peer-to-Peer ecommerce Communities"
[2]. P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation systems. Communications of the ACM, 43(12), 2000.
[3]. Paul Resnick, Richard Zeckhauser," Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System"
[4]. D. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures.Boca Raton, FL, USA: CRC Press, 2003.
[5]. P. Resnick, K. Kuwabara, R. Zeckhauser, and E.Friedman, "Reputation systems: Facilitating trust in internet interactions," Commun. ACM, vol. 43, no. 12, pp. 45–48, 2000.
[6]. P. Resnick and R. Zeckhauser, "Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system," Econ. Internet E-Commerce, vol.11, no. 11, pp. 127–157, Nov. 2002.
[7]. J. O'Donovan, B. Smyth, V. Evrim, and D. McLeod, "Extracting and visualizing trust relationships from online auction feedback comments," in Proc. IJCAI, San Francisco, CA, USA, 2007, pp.2826–2831.
[8]. P. Thomas and D. Hawking, "Evaluation by comparing result sets in context," in Proc. 15th ACM CIKM, Arlington, VA, USA, 2006, pp. 94–101.
[9]. Xiuzhen Zhang, Lishan Cui, and Yan Wang,"CommTrust: Computing Multi-Dimensional Trust by Mining E-Commerce Feedback Comments", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 7, JULY 2014.
[10]. D. R. Thomas, "A general inductive approach for analyzing qualitative evaluation data," Amer. J. val., vol. 27, no. 2, pp. 237–246, 2006.
[11].S. Ramchurn, D. Huynh, and N. Jennings, "Trust in multi-agent systems,"Knowl. Eng. Rev., Vol. 19, no. 1, pp. 1–25, 2004.
[12].]Y. Lu, C. Zhai, and N. Sundaresan, "Rated aspect summarization of short comments," inProc. 18th Int. Conf. WWW,NewYork,NY, USA, 2009.
[13].M. Gamon, "Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis," inProc. 20th Int. Conf. COLING, Stroudsburg, PA, USA, 2004.
[14].Y. Hijikata, H. Ohno, Y. Kusumura, and S. Nishida, "Social summarization of text feedback for online auctions and interactive presentation of the summary,"Knowl. Based Syst.,vol.20,no.6, pp. 527–541, 2007.
[15].B. Yu and M. P. Singh, "Distributed reputation management for electronic commerce,"Comput. Intell., vol. 18, no. 4, pp. 535–549, Nov. 2002.
[16]. M. Schillo, P. Funk, and M. Rovatsos, "Using trust for detecting deceptive agents in artificial societies,"Appl. Artif. Intell., vol. 14, no. 8, pp. 825–848, 2000.
[17]. J. Sabater and C. Sierra, "Regret: Reputation in gregarious societies," inProc. 5th Int. Conf. AGENTS, New York, NY, USA, 2001, pp. 194–195.
[18].P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood, "The value of reputation on eBay: A controlled experiment," Exp. Econ., vol. 9, no. 2, pp. 79–101, 2006.
[19].M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. 4th Int. Conf. KDD, Washington, DC, USA, 2004, pp. 168–177.
[20]. G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," Comput. Linguist., vol. 37, no. 1, pp. 9–27, 2011.
[21]. L. Zhuang, F. Jing, X. Zhu, and L. Zhang, "Movie review mining and summarization," in Proc. 15th ACM CIKM, Arlington, VA, USA, 2006, pp. 43–50.
[22]. Z. Zhai, B. Liu, H. Xu, and P. Jia, "Constrained LDA for grouping product features in opinion mining," in Proc. 15th PAKDD, Shenzhen, China, 2011, pp. 448–459.
[23].A. Mukherjee and B. Liu, "Aspect extraction through semisupervised modeling," in Proc. 50th ACL, vol. 1. Stroudsburg, PA, USA, 2012, pp. 339–348.
[24]. H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis without aspect keyword supervision," in Proc. 17th ACM SIGKDD Int.Conf. KDD, San Diego, CA, USA, 2011, pp. 618–626.
[25]. H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis on review text data: A rating regression approach," in Proc. 16th ACM SIGKDD Int. Conf. KDD, New York, NY, USA, 2010, pp. 783–792.

[26].P. Li, C. Burges, Q. Wu, J. Platt, D. Koller, Y. Singer, and S. Roweis, "McRank: Learning to rank using multiple classification and gradient boosting," in Advances in Neural Information Processing Systems, vol. 20. Cambridge, MA, USA: MIT Press, 2007, pp. 897–904.
[27].O. Chapelle and S. Keerthi, "Efficient algorithms for ranking with SVMs," Inf. Retr., vol. 13, no. 3, pp. 201–215, 2010.
[28].Y. Freund, R. Iyer, R. Schapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," J. Mach. Learn. Res., vol. 4, pp. 933–969, Nov. 2003.
[29]. Y. Yue, T. Finley, F. Radlinski, and T. Joachims, "A support vector method for optimizing average precision," in Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2007, pp. 271–278.
[30]. J. Xu and H. Li, "AdaRank: A boosting algorithm for information retrieval," in Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2007, pp. 391–398.
[31]. Z. Cao, T. Qin, T. Liu, M. Tsai, and H. Li, "Learning to rank: From pairwise approach to listwise approach," in Proc. 24th Int. Conf. Mach. Learn., 2007, pp. 129–136.
[32]. Ran Vijay Singh and M.P.S Bhatia , "Data Clustering with Modified K-means Algorithm", IEEE International Conference on Recent Trends in Information Technology, ICRTIT 2011, pp 717-721.
[33].D. Napoleon and P. Ganga lakshmi, "An Efficient K-Means Clustering Algorithm for Reducing Time Complexity using Uniform Distribution Data Points", IEEE 2010.
[34].Tajunisha and Saravanan, "Performance Analysis of k-means with different initialization methods for high dimensional data" International Journal of Artificial Intelligence & Applications (IJAIA), Vol.1, No.4, October 2010.
[35].Neha Aggarwal and Kriti Aggarwal,"A Mid- point based k –mean Clustering Algorithm for Data Mining". International Journal on Computer Science and Engineering (IJCSE) 2012.
[36].Barileé Barisi Baridam," More work on k-means Clustering algortithm: The Dimensionality Problem". International Journal of Computer Applications (0975 – 8887)Volume 44– No.2, April 2012.