ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 109- 112 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp22.pdf

FINDING RATING OF FEEDBACK COMMENTS USING EXTENDED LLDA ALGORITHM

G.Umakiran[1], A.Nageswara Rao[2]

PG Student (CSE), Department of CSE, S.V.College of Engineering, Tirupati, AP, India.
Professor & HOD, Dept. of CSE, S.V. College of Engineering, Tirupati, AP, India.
umaleelavathi@gmail.com[1]
hod_cse_svce@svcolleges.edu.in[2]

ABSTRACT: For e-commerce systems accurate trust evaluation is crucial for the success. The eBay reputation management system reported "all good-reputation" problem is an issue. The systems reporting reputation have been implemented in e-commerce systems such as eBay and Amazon (for third-party vendors). The over-all reputation scores for sellers are computed by aggregating feedback ratings and comment's. To choose a trusted e-commerce web site, here we proposed extended LLDA Lexical Latent Dirichlet Allocation algorithm to find the rating of feedback comments (free text), a fine grained multi-dimensional trust evaluation model by mining e-commerce feedback comments. To the best of our knowledge, the algorithm computes fine-grained multi-dimension trust profiles automatically by mining feedback comments.
Keywords: E-Commerce; LLDA; Trust; Reputation; Feedback comments

1 INTRODUCTION

The process of transforming data into useful information is known as data mining. In other words it is mining of knowledge from data. A large amount of data is available these days due to increasing computerization and digitization related to all areas like business, industry, science, finance, banking, healthcare, etc. Data mining is about finding insights which are reliable in statistics, previously unknown from data. Data mining an interdisciplinary sub-field of computer science is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, statistics, database systems and machine learning. The ultimate goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

Data mining, or Knowledge discovery, is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data, data mining derives its name from the similarities between searching for valuable information in a large database and mining a mountain for a vein of valuable ore, Both processes require either sifting through an immense amount of material, or intelligently probing it to find where the value resides.

As search engine optimization (SEO) relies heavily on keyword targeting, mining keyword plays a crucial role in the success or failure of an optimization campaign. According to SEO strategy proper keyword data mining is evidently essential. Finding the right data mining for effective and efficient keyword research, however, time taking and expertise. Keyword data which is to be mined is constantly changing, so a source of data a six months ago may not be as reliable today.

The data mining includes applications in following areas: Medical and health care, Retail, Marketing, Customer Management, Fraud detection, Search engine optimization, Banking and finance, Telecom industry, Computer Security, Education.

Data mining techniques are used to discover hidden patterns and relationships in the data, which is very most helpful in decision-making. In e-commerce systems accurate trust evaluation is crucial for the success. The eBay reputation management system reported "all good-reputation" problem[1] is an issue. The systems reporting reputation have been implemented in e-commerce systems such as eBay and Amazon. The total reputation scores for sellers are computed by aggregating feedback ratings and comment's. To choose a trusted e-commerce web site, here we proposed extended LLDA algorithm to find the ratings of feedback comments in the form of free text, it is a trust evaluation model [7] by mining e-commerce feedback comments. As per the extent of our knowledge, the algorithm computes fine trust profiles automatically by mining the feedback comments. In later sections, we use the terms reputation score and trust score interchangeably.

2 LITERATURE SURVEY

According to [2] Trust is an essential component in any relationship: interpersonal, in social structures, as well as in business relationships. As an interpersonal relationship, trust is the willingness of one person to increase his or her vulnerability to the actions of another person whose behaviour he or she could not control. As a structural relationship between people in a social system , trust is a collective and institutional attribute that can be drawn on to achieve certain societal goals. Trust in business relationships or economic transactions encourages exchange partners (agents) to work at preserving relationships through cooperative transactions. The transactions may occur on an individual-to-individual, individual-to-firm, or firm-to-firm level

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 109- 112 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp22.pdf

The purpose of this research is to build a model of multidimensional trust formation for online exchanges in B-to-C E-commerce. Further, to study the relative importance of the dimensions between two expert groups i.e., academics and practitioners, two semantic networks and content analysis are conducted: one for academician's perspectives and another for practitioners' perspectives of trust in B-to-C electronic commerce. The results shown in two perspectives which are divergent in some ways and complementary in other ways. Our belief is that the two are needed to be combined to represent meaningful trust-building mechanisms in websites.

Drawbacks:

- Here, modeling formulation , representation and implementation is for textual documents but not for dependency relation expressions
- Clustering is not performed on the dependency relation representations of aspect opinion expressions.

According to [3] In this research, they tried to consider trust among the members while they select an item based on the opinion of fellow beings. The public reputation is calculated of that item based on the general opinion given by previous users or customers. Then we combine this reputation with the trust among the opinion giver and the member who is going to select the item. As the recommendation comes from a trusted friend and it also includes the general opinions of public, so that quality of the opinion may improve. Currently, none of the web-based social network is considering combining the public reputation of an item with the trust among the members of the network to suggest or recommend an item. In general, people like to express their opinion and are interested about others opinion regarding the items they have concern. One of the popular ways to obtain customer feedback is collecting ratings about the product or services by the end users. In addition to the customer ratings, there is also great online customer feedback information available over the Internet as customer reviews in the form of free text, newsgroups post, comments, discussion forums or blogs about the product or services. This vital information also can be used to generate the public reputation of the sellers. To achieve this, techniques of data mining, specially recently emerged opinion mining (Hu & Liu, 2004a), (Popescu & Etzioni, 2005), (Ku, Liang, & Chen, 2006) could be a trust worthy tool. Mining and organizing the opinions from the feedback of the customer or user of an item could be useful for the person or organization that is going to use the item in future.

Drawbacks:

- Here system dimension weights are computed directly by aggregating aspect opinion expressions.
- But in proposed system dimension weights are as regression from overall ratings

According to [4] The rapid development of Web technology has resulted in an increasing number of hotel customers sharing their opinions on the services of hotel. The need of effective visual analysis of online customer opinions, as it has a significant impact on building a successful business. In this paper, we present Opinion Seer; it is an interactive visualization system that could visually analyze a large collection of online reviews of hotel customers. The system is built on a new visualization-centric opinion mining technique that considers uncertainty for faithfully modeling and customer opinion analysis. A new visual representation is developed to convey customer opinions by augmenting well-established scatter plots and radial visualization. Thus, to provide multiple-level exploration, by introducing subjective logic to handle and organize subjective opinions with degrees of uncertainty. we illustrated from several case studies the effectiveness and usefulness of Opinion Seer on analyzing relationships among multiple data dimensions and comparing opinions of different groups. Apart from data on hotel customer feedback, Opinion Seer could also be applied to visually analyze customer opinions on other products or services.

Drawbacks:

- This paper outlines the retaliatory feedback problem. Data gathered from eBay is used to show that users are worried about the possibility of retaliation. Finally, a simple solution, involving the rating of feedback.
- The analysis performed was very limited. Future research should further detail the problem through various research methodologies. While the proposed solution is simple and seems to resolve most of the issues addressed herein, future research into the solution needs to be conducted.

From the above references we proposed a system on opinion mining using Lexical Latent Dirichlet Allocation technique.

3 EXTENDED LLDA

Our work is related to opinion mining, or also called as sentiment analysis on free text documents There has been computing aspect ratings from overall ratings in e-commerce feedback comments or reviews(positive or negative) their aspect ratings and weights are computed based on regression from overall ratings and the positive bias in overall ratings is the focus.

In e-commerce systems although buyers leave positive feedback ratings, they express some disappointment and negativeness in free text feedback comments often towards specific aspects of transactions.

For example, a comment like "The products were as described." this phrase expresses positive opinion towards the product aspect, whereas the comment expresses negative opinion about delivery and a positive opinion about the transaction in general is "Delivery was a little slow but otherwise, great service highly recommended."

By analyzing the wealth of information in feedback comments we can uncover buyers' detailed embedded opinions towards different aspects of transactions, and compute comprehensive reputation profiles for sellers.

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 109- 112 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp22.pdf

To give rating for e-commerce systems following steps involved as shown in Figure 1:

3.1. Extracting user comments: The typed dependency relation representation is a recent NLP (Natural Language Processing) tool to help understand the grammatical relationships in sentences. For example analyzing the comment “Super quick shipping, Excellent product, its a great deal, ALL five STAR." using the Stanford typed dependency relation parser. The comment comprises four sentences, “Super quick shipping," this comment is represented as three dependency relations. Shipping does not depend on any other words and is at the root level. The adjective modifier relations amod (shipping - 3, super - 1) and amod (shipping - 3, quick - 2) indicate that super modifies and also indicates the quick modifies shipping with same meaning.

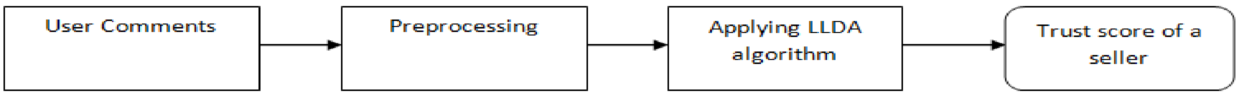

Figure 1: Steps involved in mining Feedback Comments

Table 1. Head Term Cluster Dimensions

Dim	Manual Clustering	Extended LLDA (α=0.3)	Lexical-LDA
1	Item, bag, product, dress, ear rings, outfit, top, boots, zip, scarf leggings, ring, jacket, necklace, backpack, suit, material	Item: 435, bag: 124, dress: 89, outfit:45, coat:10, ring:25, backpack:12 scarf:3, received: 57	Item: 346, leggings:67, suit:9, materials: 23, outfit:15, zip: 24
2	Quality, condition, look, size, color, description, fit, described	Look: 20, condition: 12, size: 10, color: 8, fit: 10, curtains: 7	Size: 16, refund: 9, zip: 10, design: 12, business: 6
3	Delivery, shipping, postage, dispatch, time, arrived, received, post, shipment, arrival, came	Delivery:789, payment: 78, shipping: 45, deal: 60, came: 20, arrival: 5, shipment: 45	Delivery: 678, shipping: 456, response: 50, deal: 78, arrived: 78, shipment: 8
4	Seller, ebayer	Seller: 345, ebayer: 321, described: 8, leggings: 5	Seller: 567, service: 231, communication: 121, value: 23, buy: 12, backpack: 4
5	Service, response, track communication	Communication: 112, service: 97, product: 78, value: 67, buy: 34, time: 12	Goods: 45, response:23
6	Transaction, buy, deal, purchase, order, business	Transaction: 123, order: 45	Transaction: 90
7	Payment, price, value, refund	Price: 45, refund: 12, suit: 4, shipment: 56	Payment: 45, mode: 34
8	Exchange, return, cash back, redeem	Exchange: 45, return: 23	Redeem: 21, delivery: 11

3.2. Preprocessing: We propose the Extended Lexical-LDA algorithm to cluster aspect expressions into semantically coherent categories, which we call dimensions. Informal expressions like A+++ is replaced with AAA/A and thankx were replaced with thanks. The Stanford dependency relation parser [5] was then applied to produce the dependency relation representation of comments and dimension expressions were extracted.

The dimension expressions were then clustered to dimensions by the Extended Lexical-LDA algorithm. After clustering the comments, the following steps are calculated the feedback score is the total number of positive ratings for a seller from past transactions. The Detailed seller ratings of a seller are five-star ratings on the following four aspects: Item as described (Item), Communication (Comm), Shipping time (Shipping) and Shipping and handling charges (Cost).

From last 12 months the positive feedback percentage is calculated based on the total number of positive and negative feedback ratings for transactions.

$$FP = \frac{\#positive\ ratings}{\#postive\ ratings + \#negative\ ratings}$$

3.3. Applying LLDA Algorithm: Our proposed approach is the Extended LLDA which clusters the expressions in to semantic categories, we say dimensions. In this technique, takes the document as input data, Extended LLDA uses of Shallow Lexical Knowledge to achieve the effective clustering of dependency relations.

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 109- 112 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp22.pdf

We are using lexical knowledge in two types to supervise the clustering expressions into dimensions which produces the clusters as meaningful.

- In comments the co-occurrence of terms namely head terms is not very informative. Instead of we use occurrence of dimension expression w.r.t a modifier which potentially increases the meaning of the context in the dimension expression.
- In our observation it is very rare and difficult to find the different aspect of e-commerce transactions and the dimension expressions are extracted from the same comment.

In manual clustering dimension expressions are in the form of set (modifier, head) which are used to remove noise from comments and those set's will be considered to support the head terms. Through this we can reduce the 0.1% of noise from the comments. We can combine all the dimension expressions which exist in all categories by overlapping the head terms. As a result of manual clustering we obtained totally 8 clusters for each seller. As shown in the Table 1.
Extended LLDA was implemented in [6] on the Mallet Topic modeling toolkit.

3.4 Trust score of a seller: The Extended LLDA parameter settings [8] were prior pseudo counts for topics as $\alpha_k = 0.1$ and for terms as $\beta_t = 0.01$. And the number of topics k = 3, 5, 8 for evaluating the algorithm and the iterations was set to 500.
Now we evaluate the Extended LLDA against the LLDA for clustering and also against the human clustering. As we got 8 categories as a result of human clustering i.e., k=8 for Extended LLDA. Our experimental results shows the differences between LLDA and Extended LLDA algorithm as 0.73 ~ 0.76. This proved that Extended LLDA algorithm has accurate performance in mining. As shown in the Table 1 clusters we got the difference ($\alpha = 0.3$).

Table 2. Precision of Identifying Different Ratings

	Positive	Negative	Neutral	Average
eBay	0.67±0.01	0.60±0.02	0.87±0.02	0.71±0.01
Amazon	0.76±0.02	0.70±0.01	0.72±0.01	0.72±0.01

The tool SentiWordNet is used to show the prior difference of modifier terms. The Table 2 shows the difference in positive, negative, neutral ratings on eBay and Amazon. It is observed that our approach achieved average precision for all ratings as 0.80 ± 0.18 as per the eBay and Amazon datasets.

4. CONCLUSION:

The problem "high reputation ratings for sellers cannot rank the sellers effectively so therefore the buyers are misguided to select the genuine and trustable sellers. Also observed that the buyers give their feedback positively and negatively directly the user comments in the form of free text. In this paper we have proposed the new technique which gives more accurate and also effective scores to rank the sellers. By combining the NLP (natural language Processing) with opinion mining we can evaluate the trustworthy sellers in the e-commerce application.

5. REFERENCES:

[1] Xiuzhen Zhang, Lishan Cui, and Yang Wang, "CommTrust: Computing Multi-Dimensional Trust by Mining E-Commerce Feedback Comments" IEEE VOL. 26, NO. 7, pp. 1631-1643, JULY 2014.
[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., Vol. 3, Issue No:01, pp. 993–1022, Jan. 2003.
[3] T. Hofmann, "Probabilistic latent semantic indexing," in Proc. 22nd ACM SIGIR, Vol:01, Issu No: 01, pp. 50–57. New York, NY, USA, 1999.
[4] Ross A. Malaga," The Retaliatory Feedback Problem: Evidence from eBay and a Proposed Solution," emerging trends and challenges in information technology management, Volume: 1, issue No: 01, PP: 5-10, 2006.
[5] M. De Marneffe and C.Manning, "The Standford typed dependencies representation," in Proc. CrossParser, Stroudsburg, PA, USA, 2008.
[6] A. K. McCallum. (2002). MALLET: A Machine Learning for Language Toolkit [Online]. Available: http://mallet.cs.umass.edu
[7] Y. Zhang and Y. Fang, "A fine-grained reputation system for reliable service selection in peer-to-peer networks," IEEE Trans. Parallel Distrib. Syst., vol. 18, no. 8, pp. 1134–1145, Aug. 2007.
[8] G. Heinrich, "Parameter estimation for text analysis," Univ. Leipzig, Leipzig, Germany, Tech. Rep., 2005.

6. BIOGRAPHY:

Ms.G.Umakiran M.Tech. She received Under Graduation (IT) degree in 2013 from JNTUA, Anantapur and Post Graduation (Computer Science and Engineering) degree in 2015 from JNTUA, Ananthapur. His research interests are Opinion Mining, Web Mining, Big Data.

Prof .A.Nageswara Rao M.Tech (Ph.D). He received Bachelor of Engineering (CSE) degree in 1998 from CBIT - Hyderabad, Osmania University and Master of Technology (Computer Science) degree in 2000 from Hyderabad Central University HCU). He has 15.6 Yrs Teaching Experience in various reputed Engineering Colleges. At present he is working as Professor & HOD in the department of CSE in SV College of Engineering (SVCE), Tirupati from may 2013. His research interests are Data Mining and Big Data.