Reputation System: Evaluating Reputation among All Good Sellers

Vandana Jha, Savitha R, P Deepa Shenoy and **Venugopal K R**
Department of Computer Science and Engineering,
University Visvesvaraya College of Engineering,
Bangalore University, Bangalore, India
Email: vjvandanajha@gmail.com

Abstract

A reputation system assists people selecting whom to trust. The "all good reputation" problem is common in e-commerce domain, making it difficult for buyers to choose credible sellers. Observing high growth of online data in Hindi language, in this paper, we propose a reputation system in this language. The functions of this system include 1) review mining for different criteria of online transactions 2) calculation of reputation rating and reputation weight for each criteria from user reviews and 3) ranking sellers based on computed reputation score. Extensive simulations conducted on eBay dataset show its effectiveness in solving "all good reputation" problem. So far as our knowledge is concerned, this is the first work in Hindi language on reputation system.

1 Introduction

Reputation is an estimation of the trust the community has built in you. It is a complex and context-dependent opinion of the community about any entity in question. It is highly influencial in e-commerce applications. In traditional transactions, the product is physically available for inspection whereas in online transactions, the people transact among strangers, without any physical manifestation. The consumer is forced to pay for the goods and services before receiving/trying them. Reputation is crucial for the success of e-commerce systems. Reputation system calculates and reports reputation score for an entity based on opinions from other members of the community having direct interaction with the entity in question. These opinions are collected in the form of ratings and/or free text reviews. Reputation scores are publicly available to all the members of the community so that participants can decide about future transactions. Reputation systems are used for third-party sellers by various e-commerce sites, such as, eBay and Amazon. For example on eBay, the reputation score is calculated on the total number of positive and negative feedback ratings for transactions using the formula: $\frac{\#Positive\_Ratings}{\#Positive\_Ratings+\#Negative\_Ratings}$ for feedbacks left in the last 1 month, 6 months and 12 months[1].

There is an overview of studies and experiments on eBay reputation system in (Resnick et al., 2006). Surprisingly, most of the feedback ratings on eBay are positives (99% on average) (Resnick et al., 2001) which lead to an issue known as "all good reputation" issue (Resnick et al., 2001), (Resnick and Zeckhauser, 2002). Even though detailed seller ratings are available on eBay site based on four criteria, *item as described*, *communication*, *shipping time* and *shipping and handling charges* (as shown in Figure 1) but these are also positively biased. This is misleading and does not help buyer in taking decision for the transaction. Although the ratings or scores are positively biased but the feedback comments (from now onwards called as reviews) expressed in the form of free text give the clear picture of the disappointments for some criteria during transaction (O'Donovan et al., 2007). For example, *smooth transaction, great product* but *slow delivery*, with a positive reputation score. We can extract buyers' experience about all the criteria of transactions by mining the knowledge embedded in their reviews

---

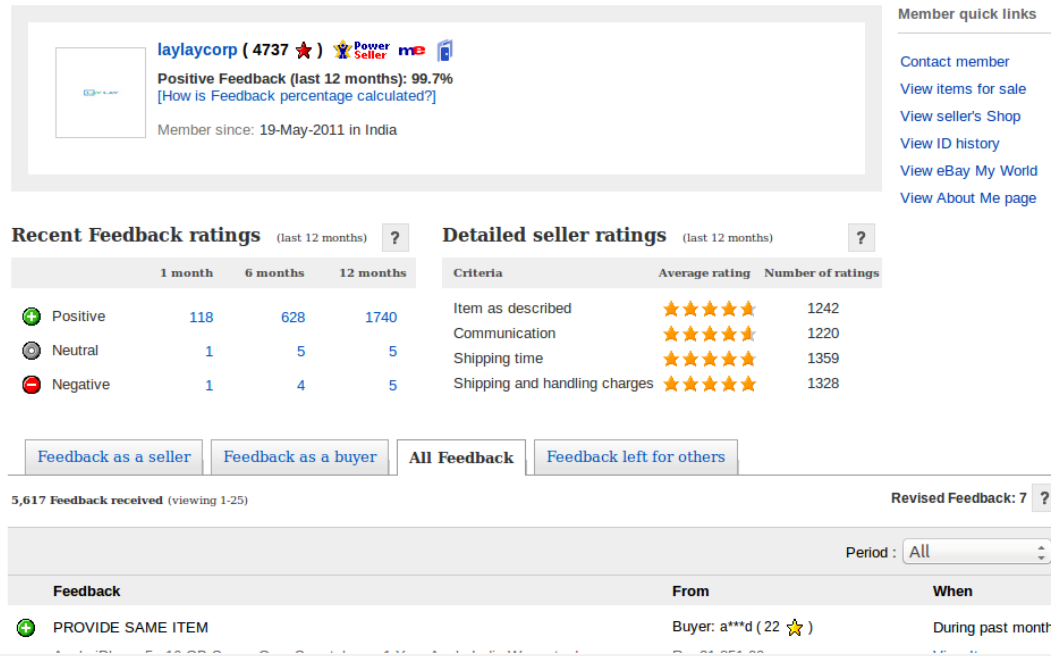[1] http://pages.ebay.in/help/feedback/allaboutfeedback.html

Figure 1: Feedback profile of a seller (laylaycorp) from `www.ebay.in` with 99.7% positive feedback score, Detailed seller ratings and Feeback comments

and calculate broad and complete reputation score for sellers.

1.1 Motivation

Online shopping through these e-commerce sites (`www.ebay.in`, `www.amazon.in` etc.) are gaining popularity in India. This popularity is increasing more because of their mobile applications accessed on smart phone. Smart phone has made one more thing possible, that is, writing in someone's own language using language selector. Both the benefits of smart phone get combined together and give the advantage of buying products from e-commerce sites and leaving reviews in Hindi, Bengali, Tamil etc. The proposed reputation system is for Hindi language. Hindi, the 4th largest spoken language, makes 4.7% of the world population[2]. It is the official language of India. English is understandable only by 25.9% of Internet users[3] so research in other languages is the need of the hour.

---

[2]http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers
[3]http://www.internetworldstats.com/stats7.htm

1.2 Contribution

A reputation system is proposed in this paper, which can evaluate reputation and rank sellers by review mining. Here "all good reputation" sellers, with more than 90% reputation score, are considered. Reputation system calculates a comprehensive reputation profile for sellers which include criteria based scores and weights, as well as total reputation scores by adding all the criteria scores. This approach combines opinion mining techniques (Pang and Lee, 2008), (Liu, 2012) with Natural Language Processing techniques. Our reputation weights are calculated by using criteria based opinion expressions, unlike other methods (Lu et al., 2009), (Wang et al., 2011), (Wang et al., 2010), to reduce positive biasing in ranking of sellers. The simulations conducted on eBay dataset show its effectiveness in solving "all good reputation" issue.

The organization of the paper is as follows: A brief overview of the related work is in section 2. Section 3 describes the proposed reputation system. Simulation runs on eBay datasets and all the results concerned are discussed in section 4. Conclusions are given in section 5.

2 Related Work

So far our knowledge is concerned, this is the first work on reputation system in Hindi language so we are unable to provide related work specifically in this language. Following is the state of art about reputation system in other languages and the methods applied to compute it.

Related Work can be branched into two main parts: 1) Review analysis and mining and 2) Review mining for computing reputation and trust. Reviews can be movie reviews, product reviews or other forms of free text.

2.1 Review analysis and mining

There have been works focussing on sentiment classification. (O'Donovan et al., 2007) and (Gamon, 2004) showed that reviews, in the form of free text, are noisy and mining knowledge out of it, is a challenging task. (Hijikata et al., 2007) focuses on summarizing the reviews and deleted 80.8% of courteous comments saying that these comments contain almost no information. (Lu et al., 2009) focuses on summarising short comments, each associated with an overall rating. (Hu and Liu, 2004) focuses on developing opinion lexicon for identifying opinion orientation for product reviews. (Qiu et al., 2011) focuses on applying syntactic relations for improving the accuracy of aspect extraction. (Zhuang et al., 2006) focuses on a multi-knowledge based approach, which integrates WordNet, statistical analysis and movie knowledge. But these works are neither in Hindi language nor use the concept of grouping opinion expressions.

Review mining in Hindi Language for polarity detection is the center point for (Jha et al., 2015c), (Jha et al., 2015a), (Jha et al., 2015b), (Jha et al., 2016), (Bakliwal et al., 2012), (Narayan et al., 2002).

2.2 Review mining for computing reputation and trust

Literatures (Resnick et al., 2001), (Resnick and Zeckhauser, 2002) and (O'Donovan et al., 2007) are available which focus on reputation systems and shown strong biasing towards positive rating. Even though sufficient solution to this problem has not been suggested, paper (O'Donovan et al., 2007) has proposed to analyse review comments. Paper

(Jøsang et al., 2007) is a survey on all the systems which can be used to find trust and reputation in e-commerce domain. Various statistical methods like the Beta reputation (Jsang and Ismail, 2002), Rating aggregation algorithm (Resnick et al., 2006) and Kalman inference (Wang et al., 2012) are also proposed for computing trust.

3 Reputation System

We have considered buyers' reviews from *eBay* dataset for eight sellers ($s = 8$). The reputation score is the weighted summation of criteria based reputation ratings for each seller and calculated by the following formula:

$$ReputationScore = \sum_{i=1}^{c} R_i * W_i, \qquad (1)$$

where R_i and W_i are reputation rating and weight respectively for criteria c. Here, $c = 4$, i.e., *Item as described (I)*, *Communication (C)*, *Shipping time (S)* and *Shipping and handling charges (Cost)*.

3.1 Criteria Based Reputation Rating

We calculated criteria based reputation rating using Bayesian method given in (Jøsang et al., 2007). Reputation rating for each criteria can be computed from the count of positive and negative ratings for that criteria. According to Bayes rule, the updated (posteriori) reputation rating is estimated from previous (priori) reputation ratings (Jsang and Ismail, 2002). The reputation rating can be defined by the beta probability density functions parameter tuple (α, β) (where α and β are the count of positive and negative ratings respectively). It is denoted by $beta(p|\alpha, \beta)$ and can be declared using the gamma function Γ as:

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

The probability expectation value of the beta distribution, $\alpha/(\alpha + \beta)$, is linearly combined with the mean, y/n, to compute reputation rating (Heinrich, 2008) i.e., $R_i = \frac{y+\alpha}{n+\alpha+\beta}$, where y is the number of positive ratings and n is the total number of ratings. When there is no prior ratings, then $\alpha = \beta$. Let's assume, $\alpha + \beta = c$, then $\alpha = \beta = 1/2 * c$ and R_i can be defined as:

$$R_i = \frac{y + 1/2 * c}{n + c} \qquad (2)$$

S	SellerName	ProductType	#Reviews	#P	#N	Pos (%)	DSR (Number of ratings)			
							I	C	S	Cost
1	uniqcorp	Mob.Acc.	47	44	3	93.6170	4.5 (40)	4.4 (41)	4.2 (40)	5(44)
2	Jeelus-com	Mob.Acc.	58	54	4	93.1034	4.5 (51)	4.4 (51)	4.3 (53)	5(58)
3	exclusiveretail	Mob.Acc.	70	69	1	98.5714	4.8 (66)	4.8 (63)	4.9 (63)	5(67)
4	aaa999acessoriesshop	Mob.Acc.	63	57	6	90.4762	4.5 (55)	4.3 (51)	4.2 (54)	5(64)
5	rkaquafreshindia2015	HomApp.	84	79	5	94.0476	4.3 (79)	4.3 (76)	4.1 (74)	5(98)
6	xiting.deals	HomApp.	49	47	2	95.9183	4.5 (30)	4.4 (28)	4.2 (29)	5(30)
7	Stallions-elex	HomApp.	49	47	2	95.9183	4.6 (51)	4.5 (50)	4.6 (50)	5(47)
8	trinitronestore	HomApp.	49	47	2	95.9183	4.4 (41)	3.9 (40)	3.8 (40)	5(42)

Table 1: Seller Dataset from `www.ebay.in`

3.2 Criteria Based Weight

In this step, the reviews are tokenized and Part-Of-Speech (POS) tagging is performed using hindi-postagger[4]. The typed dependency relation representation (De Marneffe and Manning, 2008) concept is used for extracting opinion expressions from the reviews. These expressions are clustered for each criteria and criteria based weight is computed.

The typed dependency relation representation (De Marneffe and Manning, 2008) provides a straightforward description of grammatical relations in sentences. A sentence can be depicted by a group of dependency relations between pairs of words using (head, dependent), where criteria or criteria words are represented by heads and related words or modifiers become dependent on heads. This can be accomplished using hindi-dependency-parser-2.0[5]. After parsing, words with adjective, noun, verb and adverb POS tags are separated as these only express subjectivity (Turney, 2002). The pairs like adjectives and nouns, and adverbs and verbs, express opinion expression where nouns or verbs represent criteria words and adjectives or adverbs represent opinion towards these criteria. Next, we generate clusters for each criteria. These clusters are formed by matching words in the reviews with criteria words (cword). If the match occurs then one word before and after criteria word is tested for being a modifier or not being a criteria word. The words which satisfy this condition, are clustered together for that criteria.

The dependency relations are used to compute

criteria based weight based on LDA topic modelling technique (Blei et al., 2003). We are taking the results from Gibbs sampler for LDA (Heinrich, 2008), (Griffiths and Steyvers, 2004) for determining weight of the words. W_i can be computed as follows: $CriteriaWeight = \sum_{i=1}^{t} \frac{(n-m)}{(n+m)},$ where t is the number of cluster words in each criteria, n= count of number of words which are *otherthanWordOtherthanCriteria*, m=count of number of words which are *otherthanWordSameCriteria*.

$$W_i = \frac{CriteriaWeight}{Wordcount} \quad (3)$$

4 Performance Evaluation

4.1 Datasets

We have crawled 477 buyer's reviews for eight eBay sellers from `www.ebay.in`, where four sellers are randomly selected for each of two categories namely Mobile Accessories (Mob.Acc.) and Home Appliances (HomApp.) from "Shop by category" list. These reviews are converted into Hindi language using translator[6]. Further machine translation is corrected manually by one native speaker. The preprocessed reviews are stored as Reviews file. We have also extracted the feedback profile for each seller for the evaluation of our reputation system[7] (Figure 1). It consist of following informations for a seller:

- The *Feedback score (#P)*, which is the total number of positive ratings for transactions in the past.

---

[4]http://sivareddy.in/downloads#hindi_tools

[5]http://sivareddy.in/downloads#hindi-dependency-parser

[6]https://translate.google.com/

[7]http://pages.ebay.in/services/forum/feedback.html

118

Seller	CriteriaReputationRating			
	I	C	S	Cost
1	0.8636	0.8462	0.8091	0.9583
2	0.8709	0.8524	0.8347	0.9677
3	0.9337	0.9325	0.9513	0.9718
4	0.8729	0.8338	0.8166	0.9706
5	0.8427	0.8420	0.8036	0.9804
6	0.8529	0.8325	0.7988	0.9412
7	0.8895	0.8704	0.8889	0.9608
8	0.8462	0.7545	0.7364	0.9565

(a) Criteria based ReputationRating

Seller	CriteriaWeight			
	I	C	S	Cost
1	0.1530	0.1917	0.1601	0.3707
2	0.0924	0.3241	0.1060	0.4076
3	0.1787	0.3479	0.0446	0.2737
4	0.3578	0.2570	0.2689	0.2262
5	0.0898	0.3403	0.1285	0.2667
6	0.2586	0.2835	0.1545	0.4393
7	0.1928	0.4433	0.1524	0.3146
8	0.1453	0.3530	0.2083	0.3362

(b) Criteria based Weights

Table 2: The computed results of CriteriaReputationRating and CriteriaWeight for eight sellers

- The *Positive Feedback percentage (Pos (%))*, which is calculated as $\frac{\#Positive\_Ratings}{\#Positive\_Ratings + \#Negative\_Ratings}$ for transactions in the past 12 months.

- The *Detailed seller ratings (DSR)*, which consists of four criteria for the transaction: *I, C, S* and *Cost* together with Average rating (in the form of five-star ratings) and Number of ratings for each criteria.

Table 1 shows the details of this dataset.

4.2 Evaluation Results

Table 2 shows the results of computing CriteriaReputationRating, R_i and CriteriaWeight, W_i for sellers. R_i is calculated using DSR, these are the star ratings given by the users, without considering opinions expressed in the reviews. W_i is computed using opinion expressions present in the reviews. W_i is normalized to keep reputation score between 0 and 1. This is required for its comparison with *Pos (%)* to evaluate ranking among sellers. Our reputation system uses both the informations i.e., DSR star ratings as well as opinion expressions from the reviews and calculate $ReputationScore$ (equation 1).

Table 3 shows seller 6, seller 7 and seller 8 have same value for *Pos (%)* which is 0.9592 but each has different value for DSR. In this case, if *Pos (%)* is considered alone then these sellers (seller 6, seller 7 and seller 8) should be at the same rank, which is inaccurate as they have different DSR values. If only DSR value is considered for ranking then three sellers (seller 1, seller 4 and seller 6) should be at the same rank, as shown by underlined equal values (=

Seller	Pos (%)	DSR	ReputationScore	Rank
1	0.9362	4.5	0.7790	7
2	0.9310	4.6	0.8396	5
3	0.9857	4.9	0.7997	6
4	0.9048	4.5	0.9657	3
5	0.9405	4.4	0.7269	8
6	0.9592	4.5	0.9935	2
7	0.9592	4.7	0.9951	1
8	0.9592	4.3	0.8643	4

Table 3: Comparision between existing Reputation Score from `www.ebay.in` and computed Reputation Score by our system

4.5). Our reputation system rank them at different levels by considering their DSR values as well as opinions expressed in their reviews.

5 Conclusion

Reputation systems are commonly used in online transactions but they are suffering from "all good reputation" problem. The high reputation score of every seller makes it difficult for the buyers' to choose amongst them for the transaction. We have proposed a reputation system combining the concepts of Natural Language Processing, Opinion Mining and Summarisation Methods to compute reputation score and rank the sellers effectively. Our reputation score is the weighted summation of criteria based reputation ratings for each seller. This combines the knowledge out of user ratings given in the form of five-star ratings and the opinions expressed in the form of reviews. The effectiveness of the system is shown by the simulations conducted on eBay dataset. This is able to solve "all good reputation" problem.

119

References

Akshat Bakliwal, Piyush Arora, and Vasudeva Varma. 2012. Hindi subjective lexicon: A lexical resource for hindi polarity classification. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC)*.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Marie-Catherine De Marneffe and Christopher D Manning. 2008. The stanford typed dependencies representation. In *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*, pages 1–8. Association for Computational Linguistics.

Michael Gamon. 2004. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In *Proceedings of the 20th international conference on Computational Linguistics*, page 841. Association for Computational Linguistics.

Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5228–5235.

Gregor Heinrich. 2008. Parameter estimation for text analysis. *University of Leipzig, Tech. Rep*.

Yoshinori Hijikata, Hanako Ohno, Yukitaka Kusumura, and Shogo Nishida. 2007. Social summarization of text feedback for online auctions and interactive presentation of the summary. *Knowledge-Based Systems*, 20(6):527–541.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM.

V Jha, R Savitha, SS Hebbar, PD Shenoy, and KR Venugopal. 2015a. Hmdsad: Hindi multi-domain sentiment aware dictionary. In *Computing and Network Communications (CoCoNet), 2015 International Conference on*, pages 241–247. IEEE.

Vandana Jha, N Manjunath, P Deepa Shenoy, and KR Venugopal. 2015b. Hsas: Hindi subjectivity analysis system. In *2015 Annual IEEE India Conference (INDICON), (IEEE INDICON 2015)*, pages 312–317. IEEE.

Vandana Jha, N Manjunath, P Deepa Shenoy, KR Venugopal, and LM Patnaik. 2015c. Homs: Hindi opinion mining system. In *Recent Trends in Information Systems (ReTIS), 2015 IEEE 2nd International Conference on*, pages 366–371. IEEE.

Vandana Jha, N Manjunath, P Deepa Shenoy, and KR Venugopal. 2016. Hsra: Hindi stopword removal algorithm. In *2016 IEEE International Conference on Microelectronics, Computing and Communications (MicroCom 2016)*, National Institute of Technology Durgapur, India. IEEE.

Audun Jøsang, Roslan Ismail, and Colin Boyd. 2007. A survey of trust and reputation systems for online service provision. *Decision support systems*, 43(2):618–644.

Audun Jsang and Roslan Ismail. 2002. The beta reputation system. In *Proceedings of the 15th bled electronic commerce conference*, volume 5, pages 2502–2511.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.

Yue Lu, ChengXiang Zhai, and Neel Sundaresan. 2009. Rated aspect summarization of short comments. In *Proceedings of the 18th international conference on World wide web*, pages 131–140. ACM.

Dipak Narayan, Debasri Chakrabarti, Prabhakar Pande, and Pushpak Bhattacharyya. 2002. An experience in building the indo wordnet-a wordnet for hindi. In *First International Conference on Global WordNet, Mysore, India*.

John O'Donovan, Barry Smyth, Vesile Evrim, and Dennis McLeod. 2007. Extracting and visualizing trust relationships from online auction feedback comments. In *IJCAI*, pages 2826–2831.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27.

Paul Resnick and Richard Zeckhauser. 2002. Trust among strangers in internet transactions: Empirical analysis of ebays reputation system. *The Economics of the Internet and E-commerce*, 11(2):23–25.

P Resnick, R Zeckhauser, E Friedman, and K Kuwabara. 2001. Reputation systems: Facilitating trust in internet interactions. working paper, mimeo.

Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood. 2006. The value of reputation on ebay: A controlled experiment. *Experimental Economics*, 9(2):79–101.

Peter D Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics.

Hongning Wang, Yue Lu, and Chengxiang Zhai. 2010. Latent aspect rating analysis on review text data: a rating regression approach. In *Proceedings of the 16th*

ACM SIGKDD international conference on Knowledge discovery and data mining, pages 783–792. ACM.

Hongning Wang, Yue Lu, and ChengXiang Zhai. 2011. Latent aspect rating analysis without aspect keyword supervision. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 618–626. ACM.

Xiaofeng Wang, Ling Liu, and Jinshu Su. 2012. Rlm: A general model for trust representation and aggregation. *Services Computing, IEEE Transactions on*, 5(1):131–143.

Li Zhuang, Feng Jing, and Xiao-Yan Zhu. 2006. Movie review mining and summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 43–50. ACM.