International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Product Reputation Analysis System Based on Partial Supervised Word Alignment Model

Nikita N. Khose[1], Prof. V. V. Dakhode[2]

[1]Dept. of Computer Engineering SKNCOE,
Savitribai Phule Pune University, Pune, India

[2]Assistant Professor, Dept. of Computer Engineering, SKNCOE,
Savitribai Phule Pune University, Pune, India

Abstract: ***As increase in web network there is also increase in social, commercial product services and user interest their review for the product. When user wants to buy some products from the web they put their reviews and rating to the products. For another user who buy the products for which the rating and review system is very helpful for both user and organization. The review contains product feature and rating because of this product popularity decides. Reputation system has ratings for the product analysis. If large number of rating is present then there may be unsure review of product. Hence proposed system used good reputation system having precise sentiment and opinion word and opinion target.***

Keywords: Reputation system, Opinion mining, Opinion target, Opinion words

1. Introduction

The opinion about the product is express by the user using the websites. Rating review and reputation score is more important factor about decision making purpose. Reputation score is depends on positive rating be cause‟s user easily shop the product and negative rating causes user reject the products. The ratings are given by the user to calculate the reputation score of the product. Reputation system consists of positive, negative ratings. Consider an example Users U1 buys product M and gives 4 out of 5 rating to the product as this rating is greater than 2.5 it is considered as positive. User U1 deducted one point because M is not having good quality camera. If another U2 wants to buy mobile with good quality camera (8MP), suppose M is the product which has required specification (8MP) for U2. Practically U1 experienced that M is not performing like 8MP and hence he reduced one point out of five. When U2 considers the Reputation score of the product it does not reflects negative opinion about the camera given by the U1 because his rating is considered as positive in reputation score calculation.

If U2 writes a review about product, his review will cover that camera is not performing as per specification. If U1 reads review given by the U2 then U2 can make his decision wisely. It is not easy for U2 to read all reviews in detail for decision making. Reading many reviews may result into confusion for U2, so there is need to summarize and reflect the overall opinions of the users for decision making purpose.

Every review is combined that‟s why each review is not effect on the sentiment of the user about the product. For example take one review about Camera “Great quality of pictures but camera has battery problem. Display on camera is average”, in this review we can‟t conclude overall sentiment of user about camera. Review has three dimensions, Quality of pictures, Battery, Display and sentiment of each different. To analyze these reviews it is needed to extract words about which user is writing and these words are called as Opinion target (Quality of pictures, Battery, Display). There is need to extract words which describe opinion of users about the opinion target called as Opinion word.

2. Literature Review

The target extraction using Centering Theory method express by this paper [1]. According to centering theory center always represents the focus of attention and analogues to Targets in the news comments. Paper does considered the need labeled data for training.

When the available number of labeled data is low then model is not trained properly and gives inefficient results Uses the supervised approach for opinion target extraction discussed in this paper [1]. This problem can be solved by, proposed the domain adaption method for supervised learning, in which model of target extraction in different domain is used for target domain in which there is lack of labeled data consider in this paper[2]. Opinion word and opinion target showed by the results given proposed system.

Opinion target and opinion word has strong association between them is discovered by many method. Scheme in this paper [3] also uses this fact and extracts Opinion target and Opinion words jointly.

Double propagation [4] [11] [12] means it propagates the target from words here opinion word and opinion target is extracted based on double propagation. Dependency parser is used to identify syntactic pattern. In this grammar is not followed by the user and this is biggest disadvantage of this. Results shows that method is efficient for medium size corpus but not for small and large size corpus. At that time

this method fails because dependency parser can't handle such reviews.

Scalability problem is resolved in this paper [5]. "Part-whole" and "no" patterns are added to increase the recall. Precision of the extracted features is improved using Ranking of the features on the basis of feature importance [6]. Feature importance is depends on ranking. Feature importance depends on two factors: 1) Feature relevancy and 2) Feature frequency.

To extract the relation between target and word this paper [9] [14] focuses on word alignment model. Then graph based method is used to extract the opinion target and opinion word from relation of target and word. For the quality of alignment the WAM model introduced for the analysis. The span relation in sentence is depends on the nearest neighbor method. Without parsing sentences this method is more robust than the syntax pattern.

The machine learning framework [7] which uses lexicalized HMMs (Hidden Markov Models). This approach integrates linguistic features, such as part-of speech, phrases, internal formation patterns and surrounding contextual clues of words into automatic learning. The unsupervised knowledge-lean topic modeling approach [8] automatically identify aspects from reviews and their representative words and it extended previous topic models to identify opinion words and aspects jointly.

The word-based translation model [10] extract opinion targets which uses the monolingual scenario to mine the associations between targets and opinion words. Also graph based algorithm is exploited to extract opinion targets. To identify implicit topics and sentiment words topic modeling [13] is used which aims to extract target list or word lexicon from reviews.

3. System Architecture

The system can consists of 5 steps described as follows

- **Preprocessing**

In this step reviews are preprocessed using standard methods.

- Reiews are fetched online (e.g. from a website)
- These reviews are preprocessed to be given to the next step.

- **Extraction of the Target and Opinion Words**

Target and opinion words are extracted using method proposed. Detecting the relation between the potential target and potential opinion words Object relation graph is generated. Here, confidence of each vertex in ORG calculated using random walking with restart algorithm to propagate confidence among candidates. The actual target and opinion words are depends on confidence value and their threshold value.

- **Weight Calculation of the Target**

Every target word cannot treat equally for reputation score of the product. For example, Mobile M is one product then color of the product is less important than the processor of the mobile, therefore considering the color of mobile and processor equal in process of the reputation score will be wrong. For the importance of target there is needed to calculate weight of the words for the analysis. If we consider the review or feedback of the mobile then processor is frequently discussed than color of the mobile.

$$W(T) = (\sum_{i=1}^{n} f(t)) / F(T) \quad (1)$$

Where
N is number of reviews
T is target word
F (t) frequency of the t in ith review
W (T) is weight of the target.
F (T) summation of frequency of each target word

- **Sentiment Score of the Opinion Word**

To calculate the Sentiment score system will use Sentiwordnet .There is need to find the sentiment of the opinion word. SentiWordNet assigns 3 scores word's positive score, negative score and neutral score, to the given word. Sum of these three words will be 1. For example when we pass „Good" word to the sentiwordnet it give positive score as 0.75, negative score 0.25 and neutral score 0. In our system we will consider the two values, positive score and negative score. In our case word will be either positive denoted by 1 or negative denoted by 0. Step 4 and Step 5 can be run parallel manner.

- **Reputation Score Calculation**

We have extracted each target word with its associated words and stored in vector.

$$V(T) = \{ow1, ow2, ow3 \ldots.. owN\} ;$$

Where
V (T) is vector of the associated words of target word T in all reviews.
Owi is the ith opinion word about target in the vector V(T).

In third step, each ow word's sentiment score is calculated, if positive score is more than negative score then ow is considered as positive and replace by 1 in V (T) vector to form a Sentiment vector of T

$$SV(T) = \{0, 1, 0, .0\}$$

Where
SV (T) is sentiment vector of T, 0 values at first index shows that first word in V (T) is negative about this Target T.

For each target, target score is calculated as product of the summation of values in SV (T) and weight of the target W (T)

$$TS(T) = W(T) * \frac{(\sum_{i=1}^{n} SVi(T))}{Size\ of\ SV\ (T)} \quad (2)$$

Where SVi(T) is ith value in the sentiment vector.

To calculate overall Reputation score of product P mean of the target score taken as follows.

$$R(P) = \left(\sum_{i=1}^{n} TS(Ti)\right) / n \qquad (3)$$

Where

n is number of the targets

TS (Ti) is target score of the ith target and R(P) is reputation score of the product.

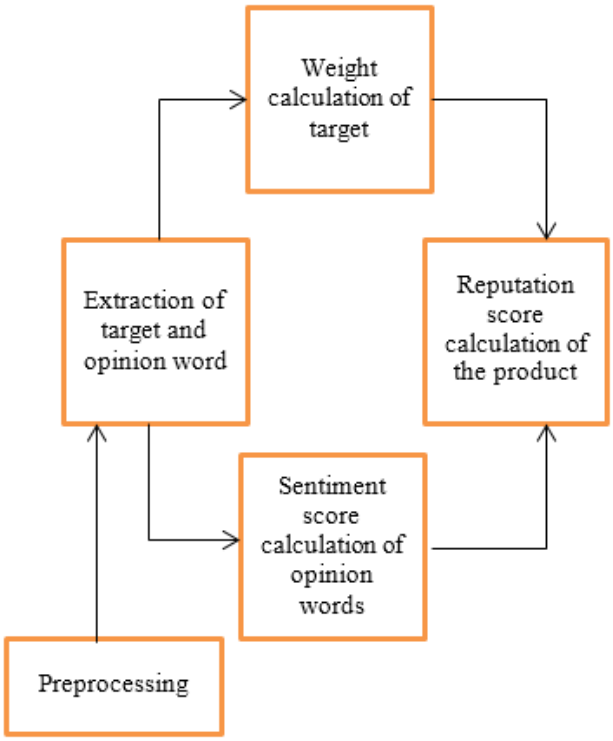

Figure 1: Architecture of Proposed System

4. Mathematical Model

Let, S be the reputation system having Input, Processes and Output. It can be represented as,

S = {I, P, O}

Where, I is a set of all inputs given to the System, O is a set of all outputs given by the System, P is a set of all processes in the System

- I = {I1, I2, I3,…..I7}
  - I1 is set of all text reviews of product P
  - I2 is set of preprocessed dataset
  - I3 is list of tagged words.
  - I4 is alignments from the dataset.
  - I5 is list of potential opinion target and opinion words.
  - I6 Final opinion target and opinion words.
  - I7 List of target nodes and attached opinion words in node links

- P= {P1, P2, P3….P10}

P1- is preprocessing step which takes I1 and outputs the preprocessed text.

P2 - This process extracts nouns, noun phrases a, verb and adjectives, this process done by part of speech tagger

P3 - Potential opinion targets and opinion words are extracted using alignments from Giza and Stanford parser.

P4 - Process takes input as O2 and forms Object relation graph (ORG)

G: (V, E, W)

Where

V is set of vertices V = Vt U Vo

Vt is set of potential target words

Vo is set of potential opinion words

E is set of Edges Eij

W is set of weight of the edges.

P5- Alignment probabilities between a potential target Wt and Potential opinion word Wo is calculated using

P (Wt|Wo) = Count (Wt,Wo) / Count(Wo)

And Opinion association between Wt and Wo calculated as follows

OA (Wt,Wo) = (a * P(Wt|Wo)) + 1 / ((1-a) P(Wo|Wt))

where a is harmonic factor used to combine alignment probabilities.

P6- Find confidence of the target / Opinion word

Conf (Wt, K+1) = (1-u) X OA (Wt,Wo) X Conf(Wo,k) + u X It

Conf (Wo, K+1)= (1-u) X OA (Wt,Wo) X Conf(Wt,k) + u X Io

Where

Conf(Wt,K+1) is Confidence of Potential target word Wt as target word at K+1 round

Conf(Wo,K+1) is Confidence of Potential opinion word Wt as opinion word at K+1 round

Conf(Wo,k) is Confidence of Potential target word Wt as target word at Kth round

Conf(Wt,k) is Confidence of Potential opinion word Wt as opinion word at Kth round

It and Io denote prior knowledge of candidates being opinion targets and opinion words

u is impact of the prior knowledge u ϵ (0,1)

P7 – Vt and Vo are arranged in decrement order using their confidence and words above certain confidence threshold are considered as extracted target and opinion words.

P8 – weight of each extracted word T is calculated using

$$W(T) = \left(\sum_{i=1}^{n} f(t)\right) / F(T)$$

Where n is number of reviews

f (t) frequency of the t in ith review

W (T) is weight of the target

F (T) summation of frequency of each target word

P9 – Sentiment score of each opinion word calculated

Sscore = SentiWordNet(Wo)

Sscore is set of Positivity (P) , negativity (Ng) and neutrality (Nu) score of word

If P > Ng
Then Wo is positive
Else Wo is negative
Output will be O7

P10- reputation score of the product is calculated
For each target, target score is calculated as product of the summation of values in SV (T) and weight of the target W (T)

$$TS(T) = W(T) * \frac{(\sum_{i=1}^{n} SVi(T))}{Size\ of\ SV\ (T)}$$

Where SVi(T) is ith value in the sentiment vector. To calculate overall Reputation score of product P mean of the target score taken as follows.

$$R(P) = (\sum_{i=1}^{n} TS(Ti)) / n$$

Where n is number of the targets
RP(P) will range from 0 to 1.

TS (Ti) is target score of the ith target and R (P) is reputation score of the product.

- Output will be Reputation score of the product

O = {O1, O2, O3, O4, O5, O6, O7, O8}
Where,
O1 is preprocessed text
O2 is set of noun, noun phrase, verb and adjective
O3 is set of Alignment probabilities among all target and opinion words.
O4 is set of confidence of target and opinion words
O5 set of target and opinion words
O6 weight of target word
O7 Vector of sentiment of each extracted opinion word
O8 Reputation score of the product

5. Experimental Results

For experiments, the reviews and rating of the iPhone 6s and Samsung Galaxy J7 are collected from Amazon e-commerce website. Each review is stored in text file. The Reputation score of the each product is calculated using proposed system and existing rating based reputation system. Reputation system of the Amazon e-commerce takes average of the all rating. Reputation score by both systems for each product is compared.

TABLE 1: Reputation Score

Product Name	*Reputation Score by Proposed Reputation System*	*Reputation Score by Existing reputation System*
iPhone 6	3.10	2.4
Samsung Galaxy J7	3.65	3.9

Reputation score using proposed system is calculated using opinion targets and words therefore it covers more sentiments of the users about product than reputation score calculated based on rating of the product.

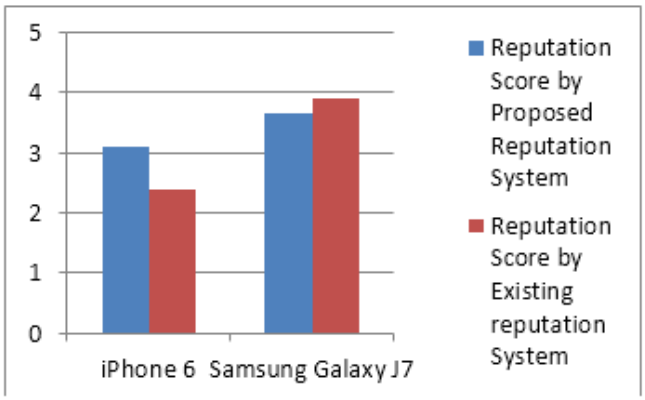

Figure 2: Reputation Score

6. Conclusions

Overall reputation score is reflected using rating given by users to the product. Trust Or Reputation score of the product is very important factor for decision of the user. Rating given by user does not reflect the various dimensions of user opinion and it is very important to consider the dimensions for Reputation of the product. This paper proposed reputation system which considers the dimensions of the user opinion.

References

[1] T. Ma and X. Wan, "Opinion target extraction in chinese news comments." in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 782–790.
[2] F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, "Cross-domain co-extraction of sentiment and topic lexicons," in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Jeju, Korea, 2012, pp. 410–419
[3] M. Hu and B. Liu, "Mining opinion features in customer reviews," in Proc. 19th Nat. Conf. Artif. Intell, San Jose, CA, USA, 2004, pp. 755–760
[4] G. Qiu, L. Bing, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," Comput. Linguistics, vol. 37, no. 1, pp. 9–27, 2011
[5] L. Zhang, B. Liu, S. H. Lim, and E. O'Brien-Strain, "Extracting and ranking product features in opinion documents," in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010
[6] Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in Proc. Conf. Empirical Methods Natural Lang. Process., Singapore, 2009, pp. 1533–1541.
[7] W. Jin and H. H. Huang, "A novel lexicalized HMM-based learning framework for web opinion mining," in Proc. Int. Conf. Mach. Learn., Montreal, QC, Canada, 2009, pp. 465–472.
[8] W. X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid," in Proc. Conf. Empirical Methods Natural Lang. Process., Cambridge, MA, USA, 2010, pp. 56–65.
[9] Kang Liu, Liheng Xu, and Jun Zhao, Co-Extracting Opinion Targets and Opinion Words from Online

Reviews Based on the Word Alignment Model , IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 3, MARCH 2015

[10] K. Liu, L. Xu, and J. Zhao, "Opinion target extraction using wordbased translation model," in Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn., Jeju, Korea, Jul. 2012, pp. 1346–1356.

[11] B. Wang and H. Wang, "Bootstrapping both product features and opinion words from chinese customer reviews with crossinducing," in Proc. 3rd Int. Joint Conf. Natural Lang. Process., Hyderabad, India, 2008, pp. 289–295.

[12] G. Qiu, B. Liu, J. Bu, and C. Che, "Expanding domain sentiment lexicon through double propagation," in Proc. 21st Int. Jont Conf. Artif. Intell., Pasadena, CA, USA, 2009, pp. 1199–1204.

[13] A. Mukherjee and B. Liu, "Modeling review comments," in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Jeju, Korea, Jul. 2012, pp. 320–329.

[14] K. Liu, H. L. Xu, Y. Liu, and J. Zhao, "Opinion target extraction using partially-supervised word alignment model," in Proc. 23rd Int. Joint Conf. Artif. Intell., Beijing, China, 2013, pp. 2134–2140.

[15] Bing Liu., "Sentiment analysis and subjectivity. Handbook of Natural Language Processing", Second Edition,2010.

Author Profile

Nikita N. Khose pursing ME degree in Computer Engineering from Smt. Kashibai Nawale College of Engineering, Pune-411041, India and BE degree in Computer Engineering from Pune University, India, in 2014. Her current research area is opinion mining, information extraction and data mining.

Prof. Virendra Dakhode received ME degree in Computer Engineering from Rajarshi Shahu College of Engineering, University of Pune, India in 2013 and BE degree in Information Technology from Amravati University, India, in 2007. His current research area is database and data mining. Currently he is working as Assistant Professor in Department of Computer Engineering, Smt. Kashibai Nawale College of Engineering, Pune-411041, India. He is having 7 years of teaching experience. He is a member of ISTE, CSI and presented three papers in International journal.