International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 03 | Mar -2017 www.irjet.net p-ISSN: 2395-0072

Sentiment Analysis of Product Reviews and Trustworthiness Evaluation using TRS

Vivek Panchal[1], Zaineb Penwala[2], Sneha Prabhu[3], Rhea Shetty[4], Prof. Reena Mahe[5]

[1,2,3,4] *Student, Dept. of Information Technology, Atharva College of Engineering, Maharashtra, India*

[5] *Asst. Professor, Dept. of Information Technology, Atharva College of Engineering, Maharashtra, India*

Abstract— *There has been a rapid growth in the E-commerce industry which market and sell products as well as allow users to express their opinions about products. Buyers generally refer to these reviews and opinions before making a buying decision and to obtain first-hand experience of the product from certified buyers. However, users tend to adapt modern writing styles such as abbreviations, misspelled words, phrases instead of sentences and emoticons. Hence, automatic summarization of product reviews using Sentiment Analysis has great significance as it helps the company know what the user liked and disliked about the product as well as helps buyers make an online purchase decision. The basic task of sentiment analysis is sentiment classification which classifies a user review as positive, negative, neutral. Also, it is important to calculate the degree of trust of the user who posts a review, the review's trustworthiness and generation of a global reputation score of the product.*

***Keywords*— Abbreviations, Opinions, Review, Sentiment Analysis, Summarization, Trustworthiness**

1. INTRODUCTION

Sentiment Analysis or Opinion Mining is the study of people's opinions, attitudes and emotions toward an entity. An entity represents events or topics. These topics are most likely to be covered by reviews. Sentiment Analysis identifies the sentiment expressed in a text and then analyzes it. Therefore, the target of Sentiment Analysis is to find opinions, identify the sentiments expressed, and then classify their polarity. The data sets used in SA are important. The main sources of data are from the product reviews. These reviews are important to the business holders as they can take business decisions according to the analysis results of users' opinions about their products. The review sources are mainly review sites [1]. Consumers make a better choice, if they have access to reviews and experiences from other consumers who have made similar choices. Mistakes that other consumers made can be avoided and confusion about products or services can be cleared.

Presently, most of the leading e-commerce websites, tend to focus on the product or its features and give very little emphasis to what is being said about the product. This paper focuses on semantically analyzing and evaluating product reviews. It does not compare products feature-wise but detects sentiments in a review and gives an overall rating. It also determines the degree of trust of the user providing the review by assigning it a global reputation score [3]. This is done by providing the user with critic reviews and asking user to like or dislike a particular critic review. Thus, sentiment of review provided can be evaluated and degree of trust of review determined through feedback system. However, few of them such as [6, 7, and 8] have been devoted to the semantic analysis of textual feedbacks in order to generate a most trustful trust degree of the user.

2. HISTORY

The accuracy of a sentiment analysis system is, in principle, how well it agrees with human judgments. This is usually measured by precision and recall. However, according to research human raters typically agree 79% of the time [4]. Thus, a 70% accurate program is doing nearly as well as humans, even though such accuracy may not sound impressive. If a program were "right" 100% of the time, humans would still disagree with it about 20% of the time, since they disagree that much about any answer. For sentiment analysis tasks returning a scale rather than a binary judgment, correlation is a better measure than precision because it considers how close the predicted value is to the target value [4]. The system design of the opinion summarizer cum classifier has been implemented by L. Zhao and C. Li in this paper [5]. An opinion review database was generated by crawling some popular websites that categorically post product reviews by actual users. The product opinion summarizer consisted of three main phases that are: (1) Preprocessing phase (2) Feature extraction phase, and (3) Opinion summarization and classification phase [6].

Trust is an important factor in any social relationship, especially in commerce transactions. In the e-commerce

context, there is a lack of direct trust assessment. Although cryptography, electronic signatures and certificates assist users in making the transaction more secure, they remain insufficient to construct a trustful reputation about a product or a service. Thus, users are not able to conceive a reputation for the product without any additional help. In such circumstances, Trust Reputation Systems (TRS) are solicited in e-commerce applications so as to create trustworthiness, among a group of participants, toward transactions, product's characteristics and toward users' past experiences. In fact, e-commerce users prefer to focus on users' opinions about a product, in order to conceive their own trust and reputation experience. Users believe in their common interest which is to know about the trustworthiness of the transaction and product. Therefore, feedbacks or reviews, scores, recommendations and any other information given by users are very important for the assessment of trust reputation. TRS are indeed essential mechanisms that aim to detect malicious interventions of users whose intention is to falsify the reputation score of a product positively or negatively. There are many works that propose algorithms for calculating a reputation or defining a specific set of possible reputations or ratings. However, few of them have been devoted to the semantic analysis of textual feedbacks in order to generate a most trustful trust degree of the user [9].

3. METHODOLOGY

The reviews posted by the users are stored in a database. Since the reviews are posted by non-experts, it may contain spelling mistakes, white spaces and incorrect punctuation, so here cleaning tasks like spell correction and sentence boundary detection is performed. The objective content is removed from product reviews and the subjective content is extracted for future analysis. Subjective content consists of sentiment sentences which has at least one positive or negative word. Next step is feature extraction where frequently occurring nouns (N) and noun phrases (NP) are treated as possible opinion features followed by Parts of Speech Tagging (POS) tagging. A Part-of-speech-tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each such as noun, verb, and adjective. Opinion words are thus extracted and then its polarity is identified. Here SENTIWORDNET is used which gives result in notions of 'positivity' or 'negativity' [2]. Further, trustworthiness of user is calculated using Trust Reputation System (TRS). In e-commerce users are obliged to trust information from unknown and unreliable sources and anonymous users.TRS ensures that the product gets perfect rating according to the corresponding review for a product using concordance algorithm and helps the customer to make right choice about the product. Using existing TRS architecture along with different algorithms the reputation score is calculated.

4. IMPLEMENTATION

The user starts by giving an appreciation (rating) and a textual feedback about a specific product. When he clicks on submit, in order to validate the given information, user is redirected to another interface showing this message for example: "please give us your opinion about the following feedbacks before validating the information you gave:" In fact, in this interface we will find chosen feedbacks from the database from different types. Those feedbacks can be fabricated in order to summarize numerous users" feedbacks stored in the database. The generated feedbacks can be stored in another knowledge base. So as many feedbacks we add in the ordinary data base, we will fill the knowledge data base with prefabricated feedbacks using text mining algorithms and tools [3]. However, some users can give already summarized feedbacks that can directly be included in the knowledge data base. Before sending the users feedback and appreciation about the product to the trust reputation system, we have to verify the concordance between them in order to avoid and eliminate contradiction or malicious programs attacking the system. In the redirected interface, several feedbacks from different types will be displayed. Through this redirection, we are trying to detect and analyze the user intention behind his intervention on the e-commerce application. Hence, we examine and evaluate his intention using other pre-fabricated feedbacks with different types. The trustworthiness of each feedback is already present. Consequently, we use TRS to generate the user trust degree which plays the role of a coefficient and then rectify his appreciation according to his trust degree and generates the score of the feedback. Each feedback has trustworthiness in a threshold [-5,5]. The closest is the trustworthiness to 5, the most trustworthy the feedback is. The closest is the trustworthiness to -5, the very untrustworthy is the feedback. If the feedback is trustworthy its score would be included in [0,5] else it would be included in [-5,0].

Pseudo-code for the values standardization:

```
// to respect the threshold [-5;5]
If (Degree_trust_user<-5)
Degree_trust_user=-5;
Else if (Degree_trust_user>5)
Degree_trust_user=5;
Return degree_trust_user;
//the end of function
/*then we affect the trust degree of the user to the degree of trustworthiness of the user's feedback*/
Ufeedtrustworth=Degree_trust_user;
```

After that, we have to generate the global trust reputation score of the product using the user's appreciation (rating) and his trust degree. To calculate the global trust score of the product, we sum all the appreciation values multiplied by their respective coefficient and then divide the result of the summation on the summation of all coefficients:

$$\frac{X + b * y}{a + b} = \frac{\sum \text{user's appreciation} * \text{user's trust degree}}{\sum \text{user's trust degree}}$$

Where 'X' represents the summation of all users' appreciation.
'y' represents the new appreciation given by the user.
'b' represents the new coefficient to be added.
'a' represents the summation of all users' trust degrees.

In the formula above, the summation of a user's appreciation multiplied by each user's trust degree is calculated respectively. We divide the result by the summation of all users' trust degree from the first until the last one for whom we have just calculated the trust degree or updated it. We can store the 'X' and the 'a' in different areas in the data base, so we can get them separately and then calculate easily:

$$\frac{X + b * y}{a + b}$$

5. RESULT AND OUTPUT

We have implemented the proposed system and provided below the screenshots which show the outputs of the system. User can add a product to cart and buy it. In order to review a product, it is mandatory for a user to first buy it, failing which the user is prompted to buy it first. Fig. 1 shows a page where user can write a review and rate the product that he/she has purchased.

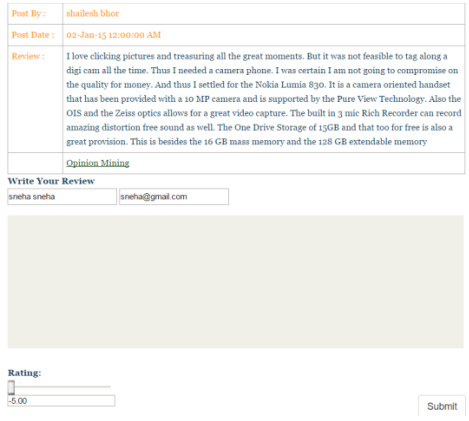

Fig -1: Review Page

Fig. 2 shows the sentiment polarity that pops when the user submits the review and rating.

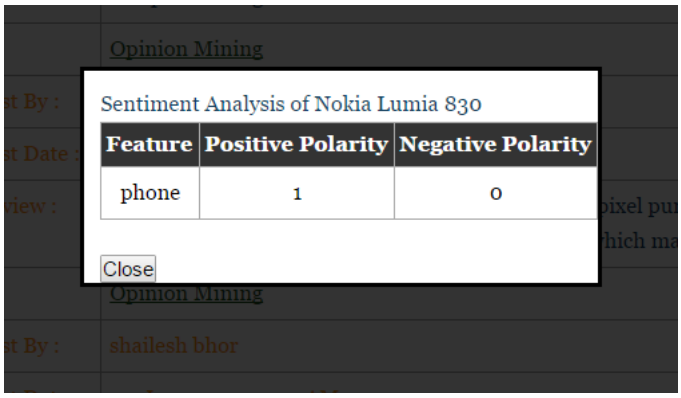

Fig -2: Sentiment Analysis Popup Message

Fig. 3 shows the trust reputation score of the user after they have provided feedback in the form of 'like' or 'dislike' for the prefabricated feedbacks whose trustworthiness was already known.

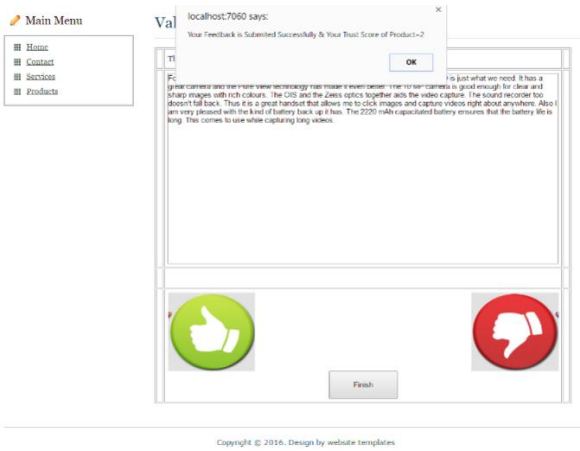

Fig -3: Page requesting feedback of the user in the form of 'likes' and 'dislikes'

Fig. 4 shows the trust score that pops up as a dialogue box to the user. Thus, we have determined the degree of trust of a user.

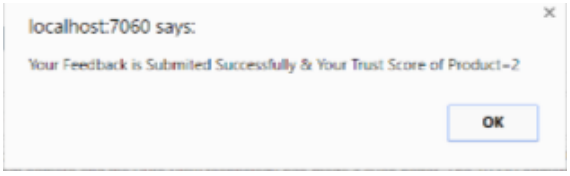

Fig -4: Trust Reputation Score of the User

6. CONCLUSION

In this paper, we have put forth a working system to identify sentiment of reviews on e-commerce websites and classify them as positive or negative. This is done while also determining the degree of trust of the user. The system makes sure that a certified buyer is only allowed to post review so as to obtain reviews as genuine as possible and only store them in the database. Thus, after posting a review, the trustworthiness of the user is determined through a feedback system for which the user's opinion is asked. These reviews are prefabricated and already present in the knowledge base. The sentiment polarity and trustworthiness of these are already known.

REFERENCES

[1] Walaa Medhat, Ahmed Hassan and Hoda Korashy, "Sentiment Analysis Algorithms and Applications: A survey", Ain Shams Engineering Journal, December 2014.

[2] Alexandra Cernian, Valentin Sgarciu, Bogdan Martin "Sentiment Analysis from Product Reviews using SentiWordNet as Lexical Resource", 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Year: 2015

[3] Hasnae Rahimi, El bakkalihanan, "Toward a New Design of Trust Reputation System in e-commerce", in the proceedings of ICMCS (International conference on Multimedia computing and systems, Tangier, Morocco, IEEE 2012).

[4] L.Dey and S. M.Haque, "Opinion Mining From Noisy Text Data," International Journal on Document Analysis and Recognition, vol. 12, no. 3, pp. 205–226, 2009.

[5] L. Zhao and C. Li, "Ontology Based Opinion Mining for Movie Reviews," in Proceedings of the 3rd International Conference on Knowledge Science, Engineering and Management, pp. 204–214, 2009.

[6] Audun Josang and Jennifer Golbeck, "Challenges for Robust Trust and Reputation Systems", 5th International Workshop on Security and Trust Management (STM 2009), Saint Malo, France, September 2009.

[7] Félix Gómez Mármol, Joao Girao and Gregorio Martínez Pérez, "TRIMS,a Privacy-Aware Trust and Reputation Model for Identity Management Systems", in the proceedings of Computer Networks, 54(16):2899-2912, September 2010.

[8] Jennifer Golbeck and James Hendler, "Inferring Reputation on the Semantic Web", in the proceedings of WWW 2004, May 17-22,2004, New York, NY USA. ACM.

[9] Prof. Reena Mahe, Rahul Jadhav, Pratik Gaikwad, Rahul Gadekar, Kiran Bhise, "Trustworthiness in E-Commerce Context using TRS Algorithm", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 10, October 2015.