



Evaluation and Ranking Online User Comments

Zahra HARİRİ^{1,*}, Mehran REZAİE²

¹ Master Student, Faculty of Graduate Studies, Safahan Institute of Higher Education, Isfahan, P.O.Box 817474319

² Professor, Faculty of Graduate Studies, Safahan Institute of Higher Education, Isfahan, P.O.Box 817474319

Received: 01.02.2015; Accepted: 05.05.2015

Abstract. From the viewpoint of online communities, the interactions between users and administrators is vitally important in order to start and maintain a community. Online newspaper is one of the online communities where the users are able to respond through their comments to the news they read. Indeed, users generally express purposeful emotions in their comments. In order to use the comments one is required to organize them, considering that they are often in conflict with each other and displayed disorganized. The different approaches in the field of text mining and sentiment analysis have come to help.

Keywords: Sentiment analysis, ranking comments, online users

1. INTRODUCTION

World Wide Web can be considered as a repository of ideas from users. The challenge that the manufacturers and web administrators are faced by is to analyze and organize their ideas. Analysis of emotions in online publications is a way of organizing user's ideas, which requires weighting of the words in comments. The weighting methods include genetic algorithms, artificial neural networks, regression equations and TF-IDF. In this paper we have used the TF-IDF method.

In the following we will demonstrate the weighting method and the way of using a dictionary for scoring the words in comments. The proposed method has been implemented in C#.Net.

2. TF-IDF

TF-IDF in data recovery works well. Also on weight of words, it has given satisfactory results. In fact, the TF-IDF weighting of word is one of the most common methods. It uses a combination of TF and IDF methods. (It should be noted that in conventional usage of TF-IDF, the worthless words like: an, that and some others which are repeated the most will be omitted in the beginning as a stop words.)

In this way, the weight of each word will be equal to the product of TF and IDF that is expressed in equation (I).

$$W_{t,d} = tf \times IDF = (1 + \log tf_{t,d}) * \log N / df_t \quad (I)$$

Where N: number of documents, df_t : documents which consist t on them.

*Corresponding author. Zahra HARİRİ

2.1. TF

The first time it was stated in [7]. If so, we consider the set of documents D as $d_i \in D$, TF means of repeated t_i in d_i that can be calculated in equation (II).

$$W_{t,d} = \begin{cases} 1 + \log tf_{t,d} & \text{if } tf_{t,d} > 0 \\ 0 & \text{else} \end{cases} \quad (\text{II})$$

2.2. IDF

This was proposed in [8] for the first time, it means that the number of d_i which contains t_i . In total it can be calculated by means of equation (III).

$$idf_t = \log N / df_t \quad (\text{III})$$

Where N: number of documents

3. SENTIMENT ANALYSIS

Ideas are central to all human activity and it can be said that anyone is willing to know other people's opinion before making any decision. From early 2000 onwards, sentiment analysis has been considered as one of the research areas of natural language processing, data mining, web mining and text mining. It can be defined as:

"A field study that states ideas, emotions, assessments, attitudes and feelings towards the existence of an organization, service, product, and some others like them."

Hence human had been looking for a way to classify and organize comments. With the passage of time it became a branch of analysis of emotions which is called "Document Sentiment Classification" and its purpose is document classification based on positive and negative ideas. This paper suggests that by the use of this method, all ideas will be classified in three groups of positive, negative and neutral.

Classification of emotion is essentially a problem of text classification. Therefore, any method of learning like Naïve Bayes classification, SVM, or any other method presented can be used. As mentioned earlier, the method used in this paper for the classification is TF-IDF method.

4. THE PROPOSED METHOD

In this section, the proposed method for ranking online user's comments has been described. For this purpose sentiment analysis and TF-IDF weighting method were used.

In general, the features of the code which is written in C#.Net can be stated in 4 cases.

1. The possibility of limiting the search to documents with pre-determined valued words.
2. Displaying the number of repetition of each word in each document, separately.
3. Displaying the number of documents containing each word, separately.

4. Displaying the final value for each document.(Equation (IV))

$$val = \{ \sum val_i | d_i \in Dic \} \tag{IV}$$

We calculated the weight of words in the document by using of TF-IDF method, then quantified the final weight of each word with the use of dictionary (Equation (V)).

$$val_i = d_{i,k} * tf * idf \tag{V}$$

Where $d_{i,k}$ is the value of t_i in dictionary and the final value of each word is val_i .

After calculating the final value of each document (means Val), by using equation (VI) the comments will be ranked.

$$if \begin{cases} val > 0 \\ val = 0 \\ val < 0 \end{cases} \quad \begin{matrix} d_i \text{ Positive} \\ d_i \text{ Neutral} \\ d_i \text{ Negative} \end{matrix} \tag{VI}$$

Figure 1 shows flow chart of the Visio drawing tool.

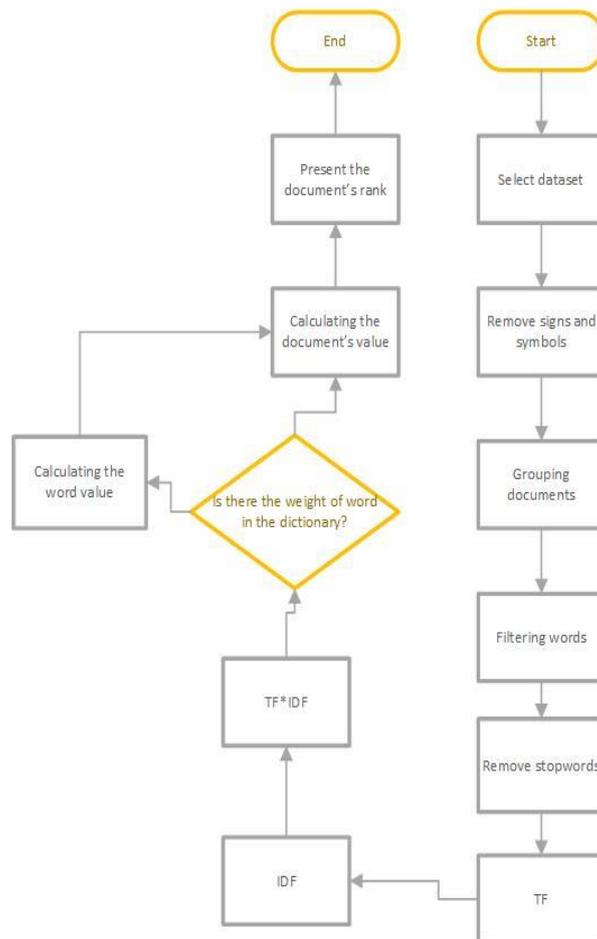


Figure 1. Flow chart

5. EVALUATION

We evaluated the efficiency of the proposed algorithm by using two datasets. Dataset 1 which is user's comments about Canon G3 and contains 639 records and dataset 2 that contains 380 records about NikonCoolpix 4300. Figure 2 shows the results of data set 1 and figure 3 shows the results of dataset 2.

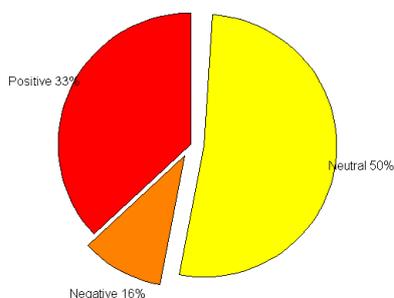


Figure 2. Result of dataset 1.

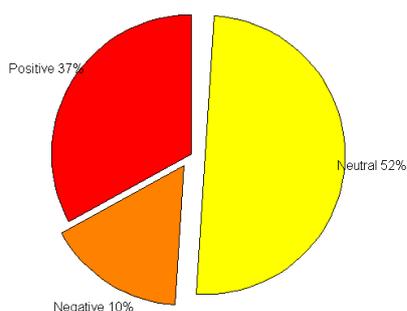


Figure 3. Result of dataset 2.

6. CONCLUSIONS AND FUTURE WORKS

In this paper, a method for ranking online user's comments has been provided. It was performed by using the practical concepts of emotion analysis and TF-IDF weighting method. The results demonstrated that the proposed algorithm has an acceptable performance.

In the future, it can also be used for ranking Persian comments or can be entered to any user's database and ranking comments on the recommendation systems.

Acknowledgment

This article is excerpted from graduate thesis at the Institute of Higher Education in Safahan institute.

REFERENCES

- [1] Sungho Shin, Sangkeun Park, Jinseop Shin, Sa-Kwang Song, " Analytics on Online Discussion and Commenting Services",Lecture Notes in Computer Science,IEEE, Vol 8016, 2013, pp 250-258.
- [2] Ying Zhang, Yi Fang, Xiaojun Quan," Emotion Tagging for Comments of Online News by Meta Classification with Heterogeneous Information Sources", 35th international ACM SIGIR conference on Research and development in information retrieval ,2012, pp 1059-1060.
- [3] Pooja Kherwa, Arjit Sachdeva, Dhruv Mahajan Nishtha Pande, Prashast Kumar, "An approach towards comprehensive sentimental data analysis and opinion mining", Advance Computing Conference (IACC), 2014 IEEE International, 21-22 Feb. 2014,ISBN: 978-1-4799-2571-1,pp:606-612.
- [4] Mina Maleki, Ahmadollah Zade Bar Forush , "New feature weighting method based on class information in classified documents", Twelfth International Conference Forum, Martyr Beheshti University, Tehran,Iran , February 20-22 , 2007.
- [5] Bing Liu,"Sentiment Analysis and Opinion Mining", Morgan &Claypool Publishers, May 2012.
- [6] Diksha Behl, Sahil Handa, Anuja Arora,"A Bug Mining Tool to Identify and Analyze Security Bugs using Naïve Bayes and TF-IDF",2014 International Conference on Reliability, Optimization and Information Technology -ICROIT 2014, India, Feb 6-8 2014.
- [7] Luhn, H.P, "A Statistical Approach to the Mechanized Encoding and Searching of Literary Information" ,IBM Journal of Research and Development, Vol. 1, No. 4, pp.309-317, 1957. 10]
- [8] Salton, G., Yang, C.S., "On the Specification of Term Values in Automatic Indexing", Journal of Documentation,Vol. 29, No. 4, pp. 351-357, 1973.
- [9] Udwadia, F.E. and Trifunac, M.D. (973), "Ambient Vibration Test of Full Scale Structures",Proc. Of the 5th World Conf. On Earthquake Engineering Rome.