International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 3 Issue: 2
ISSN: 2321-8169
481 - 483

Analyzing Public Sentiments: A Review

Pankaj M. Bhongade[1]
PG Student , Dept. of Comp. Tech.
Yeshwantrao Chavan College of Engg.
Nagpur, India
e-mail: pankaj.bhongade@yahoo.com

Dr. M. M. Kshirsagar[2]
Professor , Dept. of Comp. Tech.
Yeshwantrao Chavan College of Engg.
Nagpur, India

Abstract:- Large number of users share their opinions on Social networking sites. it can be useful for analyzing sentiments of different peoples about different domains/products. so that this analysis can be beneficial for making various decisions in various fields. For instance a company can analyze sentiments about products while a politician can view comments about them to improve their position in the society. in previous studies only track of sentiments can be taken but in this we are trying to analyze the public sentiments and trying to find out the possible reasons behind variation about comments based on that we tries to propose a system and tries to improve performance of system.

Keywords:-*Data preprocessing, public sentiment, topic mining, sentiment analysis,*

*****

I. Introduction

With the large number of users using the social networking sites, it has become a media where millions of users share their views about various domains. This can be useful for taking decisions in various fields such as a company tries to analyze the feedback about their products, While a politician tries to improve his position according to the comments. Large number of previous research has been done on opinion mining, it is important concept where opinions about peoples can be taken to mine possible reasons. The sentiment analysis can also be very useful to know about product reviews. Previous studies only taken track of public sentiment but in this we are moving one step further to analyze the public sentiments and tries to taken out predictions from that.

Our paper is divide into four sections, first section gives idea about the proposed system, in section 2 related work which is done with respect public sentiment on social sites then in section 3 we discussed the methodology and techniques of proposed system like short text classifier, text representation, machine learning based classification and prediction using fuzzy algorithm, and in the last section conclusion about the proposed work and future development scope .

II. Related work

Yuheng hu [5] described a joint statistical model ETLDA that characterizes topical influences between an event and its associated Twitter feeds (tweets). Their model enables the topic modeling of the event/tweets and the segmentation of the event in one unified framework. Evaluation of topic modeling done in both ways quantitative and qualitative by ET-LDA. Based on the results improvement has been done. They believe this paper presents the first step towards understanding complex interactions between events and social sites.

Deepayan Chakrabarti[7] tackled the problem of constructing real-time summaries of events from twitter tweets. They proposed an approach based on learning an underlying hidden state representation of an event. Through experiments on large scale data on American Football games they showed that SUMMHMM clearly outperforms strong baselines on the play-by-play summary construction task, and learns a underlying structure of the sport that makes intuitive sense. They have not yet evaluated the summaries generated by their approach in real-time on search engine users; this is something hope to do in the future.

Brendan O'Connor [10] find that a relatively simple sentiment detector based on Twitter data replicates consumer confidence and presidential job polls. come without caution, it is encouraging that expensive and with the simple-to-gather text data that is generated from online social sites. The paper suggest advanced NLP techniques to improve opinion estimation may be very useful. The textual analysis could be substantially improved. Besides the clear need for a well-suited lexicon, the modes of communication should be considered. Modeling traditional survey data is a useful application of sentiment analysis. But it is also a stepping stone toward larger and more sophisticated applications.

Hila Becker[9] proposed a general framework for identifying events in social media documents via clustering, and used similarity metric learning approaches in this framework, to produce high quality clustering results. They discussed and experimented with ensemble based and techniques of classification combining a set of similarity metrics to predict when social media documents correspond to the same event. The experiments suggest that similarity metric learning techniques yield better performance than the baselines on which we build. In particular, their classification based techniques show significant improvement over traditional approaches that use text-based similarity but can-not distinguish between events and non events documents.

Johan Bollen [12] proposed profile of mood state approach for tweets .for doing this, they express six dimensional vector representing moods. They aggregate mood components on a daily scale and compare our results to the timeline of cultural, social, economic, and political events that took place in that time period. They speculate that collective emotive trends can be modeled and predicted using large scale analysis of user generated content.

Jaewon Yang[16] explored temporal patterns arising in the popularity of online content. First they formulated a time series clustering problem and motivated a measure of time series similarity. then developed K-SC, a novel algorithm for time series clustering that efficiently computes the cluster centroid under our distance metric. Finally, they improved the scalability of K-SC by using a wavelet-based incremental approach. We investigated the dynamics of attention in two domains. A massive dataset of 170 million news documents and a set of 580 million Twitter posts. The proposed K-SC achieves better clustering than K-means in terms of intra-cluster homogeneity and intercluster diversity. They also found that there are six different shapes that popularity of online content exhibits. Interestingly, the shapes are consistent across the two very different domains of study, namely, the short textual phrases arising in news media and the hash tags on Twitter. They showed how different participants in online media space shape the dynamics of attention the content receives. And perhaps surprisingly based on observing a small number of adopters of online content we can reliably predict the overall dynamics of content popularity over time. All in all, the work provides means to study common temporal patterns in popularity and the attention of online content, by identifying the patterns from massive amounts of real world data. The results have direct application to the optimal placement of online content. Another application of the work is the discovery of the roles of websites which can improve the identification of influential websites or Twitter users.

III. Methodology

3.1 Short text Classifier

3.1.1 Sentimental word dictionary:

The dictionary for the positive sentiments and negative sentiments can be maintained.

3.1.2 Preprocessing:

Dataset of product comments can be taken as for preprocessing. This preprocessing technique can have two methods stemming [14] algorithms and stop word removal algorithm [13]

3.1.3 Stemming:

In this word stem to original word removing unnecessary part eg. Running, runly, runned which stem to run.

3.1.4 Stop word removal:

In this stop words such as the, of, from, at, which, on etc can be removed.

3.2 Text Representation

In this text can be filtered according to the sentimental text and user comments about products. And part of speech Tagging can be done to that.

3.3 Machine learning based Classification

In this association rule mining algorithm [15] ie. a priory algorithm is used to find which comments are positive comments and which are negative comments by using sentimental score applied to each word .

3.4 Prediction using Fuzzy algorithm

After the classification as positive and negative comments the predictions on the different types of product can be done and graph can be maintained to show positive, negative and neutral comments about each product for predicting which product got better response and which are need to be replaced.

IV. Conclusion

The present paper reviews and summarizes methodology for analyzing public sentiments and review paper suggests that the proposed algorithm can be work better for public sentiment analysis and somewhat proper prediction can be taken about products. The proposed system can work better but there is some improvement needed in future as reasons about the sentiments can be mined for further analysis.

REFERENCES

[1] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Jiajun Bu Chun Chen, and Xiaofei He, Member, IEEE “Interpreting the Public Sentiment Variations on Twitter” IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 5, May 2014.

[2] Xin Chen, Mihaela Vorvoreanu, and Krishna Madhavan “Mining Social Media Data for Understanding Students’ Learning Experiences” IEEE Transactions On Learning Technologies, 2013

[3] Balakrishnan Gokulakrishnan, Pavalanathan Priyanthan, Thiruchittampalam Ragavan, Nadarajah Prasath, AShehan Perera “Opinion Mining and Sentiment Analysis on a Twitter Data Stream” , The International Conference on Advances in ICT for Emerging Regions - ICTer ,2012 .

[4] Malhar Anjaria, Ram Mahana Reddy Guddeti "Influence Factor based Opinion Mining of Twitter Data Using Supervised Learning", IEEE Conference ,2014
[5] Y. Hu, A. John, F. Wang, and D. D. Seligmann, "Et-lda: Joint topic modeling for aligning events and their twitter feedback," in Proc.26th AAAI Conf. Artif. Intell, Vancouver, BC, Canada, 2012.
[6] J. Weng and B.-S. Lee, "Event detection in twitter," in Proc. 5thInt. AAAI Conf. Weblogs Social Media, Barcelona, Spain, 2011.
[7] D. Chakrabarti and K. Punera, "Event summarization using tweets," in Proc. 5th Int. AAAI Conf. Weblogs Social Media, Barcelona, Spain, 2011.
[8] T. Minka and J. Lafferty, "Expectation-propagation for the generative aspect model," in Proc. 18th Conf. UAI, San Francisco, CA, USA, 2002.
[9] H. Becker, M. Naaman, and L. Gravano, "Learning similarity metrics for event identification in social media," in Proc. 3rd ACM WSDM, Macau, China, 2010.
[10] B.O'Connor, R. Balasubramanyan ,B.R. Routledge , and N. A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series," in Proc. 4th Int. AAAI Conf. Weblogs Social Media, Washington, DC, USA, 2010.
[11] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in Proc. 15th ACM SIGKDD, Paris, France, 2009.
[12] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in Proc. 5th Int. AAAI Conf. Weblogs Social Media, Barcelona, Spain, 2011.
[13] Chong Tze Yuang, Rafael E. Banchs, Chng Eng Siong, "An Empirical Evaluation of Stop Word Removal in Statistical Machine Translation" Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 30–37, Avignon, France, April 23 - 27 2012. c 2012
[14] Julie Beth Lovins, "Development of a Stemming Algorithm" Mechanical Translation and Computational Linguistics, vol.11.
[15] Seyed Hamid Ghorashi, Roliana Ibrahim, Shirin Noekhah and Niloufar Salehi Dastjerdi, "A Frequent Pattern Mining Algorithm for Feature Extraction of Customer Reviews" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012.
[16] Chong Tze Yuang, Rafael E. Banchs, Chng Eng Siong, "An Empirical Evaluation of Stop Word Removal in Statistical Machine Translation" Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 30–37, Avignon, France, April 23 - 27 2012. c 2012.
[17] Seyed Hamid Ghorashi, Roliana Ibrahim, Shirin Noekhah and Niloufar Salehi Dastjerdi, "A Frequent Pattern Mining Algorithm for Feature Extraction of Customer Reviews" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012.
[18] Chong Tze Yuang, Rafael E. Banchs, Chng Eng Siong, "An Empirical Evaluation of Stop Word Removal in Statistical Machine Translation" Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 30–37, Avignon, France, April 23 - 27 2012. c 2012.