Abstract

In today’s world, micro-blogging sites has become a platform for individuals or organizations across the world to express their opinions, sentiment and experience in the form of tweets, status updates, blog posts, etc. This platform has no political and economic restrictions. This paper discusses an approach where a published stream of tweets on electronic products from the twitter micro-blogging site are then subjected to preprocessing and classified based on their emotional content as positive, negative and neutral. The performance of the unsupervised algorithm is then analyzed. The paper concludes with the comparison of the existing system with the proposed systems and applications of the research.

EXISTING SYSTEM

The existing system works only on the dataset which is constrained to a particular topic. The existing systems also do not determinethe measure of impact the results determined can have on the particular field taken into consideration and it does not allow retrievalof data based on the query entered by the user i.e. it has constrained scope. In simple words, it works on static data rather thandynamic data. Unsupervised algorithms like Vector Quantization, are used for data compression, pattern recognition, facial andspeech recognition, etc and therefore cannot be used in determining sentiment in twitter data. Apriori algorithm fails to handlelarge datasets and as a result can generate faulty results.

PROPOSED SYSTEM

In the proposed system, we will retrieve tweets from twitter using twitter API based on the query. The collected tweets will besubjected to preprocessing. We will then apply the supervised algorithm on the stored data. The supervised algorithm used in oursystem is Support Vector Machine (SVM). The results of the algorithms i.e. the sentiment will be represented in graphical manner(pie charts/bar charts). The proposed system is more effective than the existing one. This is because we will be able to know howthe statistics determined from the representation of the result can have an impact in a particular field.

COMPARISON

|  |  |  |
| --- | --- | --- |
| S.NO | Existing System | Proposed System |
| 1. | Existing system takes a stored dataset on a particular topic into consideration. | Proposed system will gives you the freedom to choose the data of any topic |
| 2. | It fails to determine the impact the results might or will have in the respective field | Here, it gives you the impact the results and statistics will have on the respective field. |
| 3. | Existing system does not allow the retrieval of data based on the query entered by user | Proposed system allows retrieval of data based on the query entered by the user |
| 4. | Existing system does not provide accurate feature selection | Proposed system will provide accurate feature selection. |

CONCLUSION

Sentiment analysis has become an important factor in decision making process in a particular field. In this paper we discussed techniques for preprocessing and information retrieval of tweets through twitter. Also we studied about the supervised learning technique: Support Vector Machine for text categorization which can be used to find out the polarity of textual tweet. From study we can conclude that SVM acknowledges some properties of text like High Dimensional feature space, few irrelevant feature, sparse instance vector. The performance of SVM can be evaluated using precision and recall. Different results show that SVM gives good performance on text categorization as compared with ANN. With ability to generalize high dimensional feature space, SVM eliminates need of feature selection.

REFERENCES

[1] Geetika Gautam, Divakar Yadav. (2014). Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis. IEEE 2014.

[2] Neethu M S, Rajasree R. Sentiment Analysis in Twitter using Machine Learning Techniques. IEEE 2013.

[3] B. Gokulkrishnan, P. Priyanthan, T. Ragavan, N. Prasath and A. Perera,. Opinion Mining and Sentiment Analysis on a Twitter Data Stream. IEEE 2012.

[4] <http://twitter4j.org/en/>

[5] <https://en.wikipedia.org/wiki/Vector_quantization>

[6] <https://en.wikipedia.org/wiki/Apriori_algorithm>

[7] Phillip Tichaona Sumbureru. Analysis of Tweets for Prediction of Indian Stock Markets. IJSR 2013.

[8] Xing Fang, Justin Zhan. Sentiment analysis using product review data. Journal of Big Data 2015.

[9] Gilad Mishne. Experiments with Mood Classification in Blog Posts. Live Journal 2005.

[10] S. A Kanade, S. Shibu and Abhishek Chauhan. Review of Aspect Based Opinion Polling. IJREST 2014.