

SENTIMENT ANALYSIS

Diksha Saxena¹

Shubham Gupta²

Joy Joseph³

Prof.Richa Mehra⁴

ABSTRACT: *Text mining and sentiment analysis have recently received huge attention in recent year. Sentiment analysis is one of the major tasks of NLP. The data is usually unstructured and contains noise; therefore the task of gaining information is complex and expensive. There is a growing need for developing different methodologies and models for efficiently processing the texts and extracting information. One way to extract information is text mining and sentiment analysis. This paper provides an overview of different techniques used in text mining and sentiment analysis elaborating on all tasks. It also categorized at sentence level sentiment analysis and document level sentiment analysis.*

KEYWORDS: *Sentiment Analysis, Supervised Learning, Unsupervised Learning, Text Mining, Feature Extraction, Feature Representation.*

INTRODUCTION: Sentiment analysis, also known as text mining, in essence, is the process of quantifying the emotional value in a series of words or text, to gain an understanding of the attitudes, opinions and emotions expressed. Sentiment analysis can be applied to various sectors such as ecommerce, banking, mining social media websites like Facebook, Twitter and so on. One of the applications of sentiment analysis is recommendations systems, for instance YouTube recommends on the basis of consumers likes, dislikes and comments provided by the user. In this paper, we study various text mining and sentiment analysis

techniques applied to different areas in multi lingual format and from different resources. A sentiment analysis and text mining frameworks typically includes following subtasks acquiring text data, data cleaning and preprocessing, conversion of text to machine readable vectors, features selection, and finally applying NLP and machine learning algorithms. In this paper, we present a literature review on recent trend in text-mining and sentiment analysis. Topic modeling is successfully combined with sentiment priors to generate topics and sentiment classes simultaneously. Emoji's and emoticon sentiment are included in many of the studies to improve accuracy of results and so on.

LITERATURE REVIEW

In ^[1]Duwairiet. al mentioned that sentiment analysis determine the polarity of given text either using machine learning approach or using lexicon based approach. The classifiers applied on the datasets were Naïve Bayes, Support Vector Machine (SVM) and K-Nearest Neighbors (k=10)) where SVM gave highest precision and KNN gave the highest recall. Also to test the data sets 10-fold cross validation was used which demonstrated

that the precision got by SVM i.e. 75.25 was the best precision and the recall got by KNN i.e. 69.04 was the best recall. Therefore, to get better classification results, bigger data sets were required and to label them crowd sourcing was considered followed by semi supervised learning. In ^[2]Kouloumpiset. al demonstrated the usefulness of linguistic features and existing lexical resources used in micro-blogging to detect the sentiments of twitter messages. From this paper the researchers concluded that micro blogging features were more useful as compared to POS(Part-of-Speech) features and features from existing sentiment lexicon.^[3] consists of a new method formed by combination of rule based classification, supervised learning and machine learning which showed the improvement in micro and macro averaged F1. To get better effect, Prabowoet. al considered semi-automatic approach. From this paper they concluded that hybrid classification was better than the classification by any individual classifier. They also concluded that reduction of rules will produce less effect on F1.From ^[4]Mudinaset. al concluded that concept level sentiment analysis system was better as compared to pure lexicon based system and pure learning

based system due to more precision in polarity classification and well structured, readable results. On experimenting, they confirmed that hybrid approach was better than sentiment strength. From their paper, they concluded that system obtained high precision than pure lexicon based system but near to pure learning based system. It also gave well structured, readable result and more resistance to writing style of text. In short, the proposed hybrid approach was capable in combining a carefully designed lexicon and a powerful supervised learning algorithm. In [5], Lin et. al identified subjective information using automated tools and a novel probabilistic modeling framework called joint sentiment/topic model, which detects sentiment and topic together from text. They concluded that the proposed JST model was fully apart as compared to other machine learning approaches. Basically, they proposed this model on movie dataset to classify the sentiment polarity and to improve the sentiment classification accuracy. In this paper, a joint sentiment/topic (JST) model had been proposed with the help of which document level sentiment classification could be depicted and mixture of topics from text simultaneously could be extracted. On the

other hand, existing approaches in sentiment classification were based on supervised learning, while the proposed JST model was fully unsupervised, hence comes up with more flexibility and could be easily combined with other applications. When the results were compared with existing supervised approaches then they found out that this model gave a competitive performance in document level sentiment classification. On other side it also had one limitation of classifying each document as a bag of words which results in ignoring the word ordering for example predicting sentiment of "not good movie" being positive and of "not bad movie" being negative. This leads them to include bigrams and trigrams in their model. Another step which would be included in future was to detect the polarity of text at several granularity levels, e.g. detecting sentiment labels for more fine-grained topics. Model performance on datasets from different domains was also evaluated. In their paper, Li et. al [6] studied online forums hotspot and forecast using sentiment analysis and text mining approaches. First of all, to inspect the sentiment polarity for each piece of text, an algorithm was created. Afterwards to develop unsupervised text

mining approach the algorithm was joined with k-means clustering and support vector machine (SVM). Described text mining approach had been used to group forums into various clusters, whose center represent a hotspot forum within the current time span. The datasets had been taken from SINA sports forum. Experimental results showed that SVM forecasting gets high consistent results with k-means clustering. The top 10 hotspot forums given by SVM forecasting resemble 80% of k-means clustering results. Both SVM and k-means achieved the same results for the top 4 hotspot forums of the year. In this paper they had created an algorithm that automatically analyzes the sentiment polarity of a text, with the help of which text values were obtained. Influential power of text was represented by absolute value and sentiment polarity by the sign of text. Previously created algorithm was then combined with k-means clustering and SVM classification to integrated approach for online sports forums cluster analysis. Unsupervised algorithm had been applied to group the forums into various clusters, whose center represent hotspot forum with the current time span. In addition to clustering the forums based on data from

the current time window, forecasting for the next window was also done by them. Proof for existence of correlations between post text sentiment and hotspot distribution was given by empirical studies. Results showed that both SVM and k-means produce consistent natural groupings. Several companies could be benefited from these hotspot predicting approaches in different ways. These companies could also combine results for market basket analysis to yield comprehensive decision support information. A firm in financial sector or the financial department of a giant company might get profit from such a sentimental and text mining process. In financial market, right before a security market opens and trading begins, analysts people on sales and trading desks usually try to get an overall fix on market sentiment and for particular investments.

FINDINGS

- Algorithm design could be improved to yield a more accurate calculation of sentiment. Even for supervised learning, algorithms other than SVM, or variations of SVM, could be joined as well.

- They had incorporated topic extraction. A practical system, in the form of a website portal, was design.

CONCLUSION

Opinion mining applications are the basic infrastructure of large scale collaborative policy making. They help making sense of thousands of interventions. They help to detect early warning of system of possible disruption in a timely manner, by detecting feedback from citizens. Traditionally adhoc survey are used to feedback in a structure manner. However this kind of data collection is expensive as it deserve an investment in design and data collection; it is difficult as people are not interested in answering survey and ultimately it is not

REFERENCE

- [1] Salloum, S. A., AlHamad, A. Q., Al-Emran, M., & Shaalan, K. (2018). A Survey of Arabic Text Mining. In Intelligent Natural Language Processing: Trends and Applications (pp.417-431). Springer, Cham.
- [2] Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twittersentiment analysis: The good the bad and the omg!. *Icwsn*, 11(538-541), 164.
- [3] Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: combined approach. *Journal of Informetrics*, 3 (2), 143-157.
- [4] Mudinas, A., Zhang, D., & Levene, M. (2012, August). Combining lexicon and learning based approaches for concept-level sentiment analysis. In Proceedings of the first international workshop on issues of sentiment discovery and opinion mining (p. 5). ACM.
- [5] Lin, C., & He, Y. (2009, November). Joint sentiment/topic model for sentiment analysis. In Proceedings of the 18th ACM conference on Information and knowledge management (pp.375-384). ACM.
- [6] Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision support systems*, 48 (2), 354-368.
- [7] McDonald, R., Hannan, K., Neylon, T., Wells, M., & Reynar, J. (2007). Structured models for fine-to-coarse

sentimentanalysis.In Proceedings of the 45th annual meeting of the association of computational linguistics (pp. 432-439).

[8] Saleh, M. R., Martín-Valdivia, M. T., Montejo-Ráez, A., &Ureña-López, L. A. (2011). Experiments with SVM to classifyopinions in different domains. *Expert Systems with Applications*, 38 (12), 14799-14804.vary valuable as it detect “known problems” through predefined questions but fails to define important problems.

