

International conference on “Recent Advances in Interdisciplinary Trends in Engineering & Applications

Sentiment Analysis based on Hybrid Approach: A Survey

Shahida Khan^a, Kamlesh Chopra^b, Vijay Malviya^c

^aMalwa Institute Of Technology, Indore, s.shahida.k@gmail.com ^bMalwa
Institute Of Technology, Indore, kamleshchopra@mitindore.co.in ^cMalwa
Institute Of Technology, Indore, vijaymalviya@gmail.com

Abstract

Sentiment Analysis is the most widely used text classification technique which is used to analyze any text and result in terms of whether the sentiment of particular text is positive, negative or neutral. Micro blogging is used widely to express opinions like on twitter or Face book with limited character space of 140, one liner is best way to express your opinion. These messages are usually short texts like single sentences and opinions with limited contextual information. Retrieving the sentiment of such short text is a challenging task. It requires a strategy and prior knowledge to extract the correct emotion of post. Like any identification task this will also need a good feature extractor and classification approach. In this paper we are focusing on comparing the feature extractors and also utility of hybrid approach in sentiment analysis.

Keywords: Sentiment Analysis; Machine Learning; Natural Language Processing

1. Introduction

Sentiment analysis is a powerful tool used in social media monitoring as it provides us opinion of people on a certain topic. How all this evolved, why these reviews are of so much importance? We analyzed that there is a shift in way of buying or selecting any particular thing. We search Google or any search engine to know the reviews of people if we are buying any product or visiting a new place. We go by reviews even if we are dining out. To make it an automatic process we create tools like sentiment analysis.

Sentiment analysis is being used a lot because its result is efficient in most cases. A large number of text documents are being processed for extracting the sentiment in small timeframe, which is comparatively very fast if we consider humans processing as opponent. The efficiency is approximately 80%. The applications of sentiment analysis are spread over a broad spectrum. The ability to extract insights from social data is a practice that is being widely adopted by organizations across the world. [1] Shifts in sentiment on social media have been shown to correlate with shifts in the stock market. [2]

Opinion of a person is not always in systematic language, People use natural way of expression which makes social media data, analysis a tedious and complex task. It requires deep understanding to know what they are actually exhibiting in their own natural language: Natural Language Processing (NLP) is the

best way to understand the language used and identify the sentiment behind it. [3] Sentiment is usually considered in terms of polarity i.e. Positive, Negative or Neutral. In real time these sentiments are choice of customer, likes, dislikes, feedback and even their decisions. Natural language processing for sentiment analysis focused on intent of person and is genuinely compatible for text mining.

1.1. Applications of Sentiment Analysis

Sentiment Analysis in business can be used to know the tone of information conveyed. This text can be in any form ranging from tweets, feedback, Facebook post or comments with associated sentiments. Every business needs to keep an eye on customer's sentiments. We believe that only positive, negative and neutral does not help business to grow. The accuracy of such systems is not high because the machines can't understand sarcasm. Even if the accuracy is very low the amount of data needs to be analyzed is too huge that it's impossible for human to perform, manually analyzing is not an option. Hence, sentiment analysis in business is much more than just a trend. Few important and widely accepted applications are:

Social Media Monitoring

Social Media Monitoring provides solution to businesses and text mining extract the opinion from millions of tweets or messages floating over social media. These solutions give an overview of how their brand is viewed and also the value of their brand. This kind of analysis can be completed in minutes with speed and power of machines which is never an easy task if performed manually [4].

Product Management

Products are being managed either by creators thinking or it can be better managed by consumers opinion. The product managers can be at ease in merging data from varied sources and process the data commonly in a very less amount of time. Customer's insights make it easier for the decisions makers to analyze reviews and develop new strategies to improve the product experience and enhance product performance by applying necessary amendments [5].

Government Policy Review

Government releases new policies in certain time duration for betterment and benefit of people, these policies are actually assessed by opinion of public. The shortcomings and positive feedback both are useful to know the overall impact of scheme. Many applications are being used now days which are based on sentiment analysis such as brand value monitoring, enterprise search, legal assessments etc [6].

Analysis of customer sentiment is not only useful for business but it has widen the scope in health care, sports, politics, recommendation system and various domains are using such techniques in direct or indirect mode.

2. Literature Survey

Sentiment analysis has been considered an extended feature using Natural Language Processing task at many levels of granularity. Starting from being a document level classification (year 2002-2004) it has been handled at the sentence level (year 2003-2005) and more recently at the phrase level(year 2005 Onward) . Micro blog data like Twitter, on which users post real time reactions to and opinions about "topic", poses newer and different challenges. Methods used are

- ☐ Key-word based Approach
- ☐ Lexicon based Approach
- ☐ Machine Learning based Approach
- ☐ Hybrid Approach

A survey on recent micro blog data based sentiment analysis articles were used to analyses the recent development and scope of further research presented in table below.

Table 1. Summary of Research Papers

S.No	Title	Authors	Method Results	Year of Publication	Key-Points
1.	Improved lexicon-based sentiment analysis for social media analytics[7]	Anna Jurek et. al.	Lexicon based 69.1%With normalization 77.3%	2015	Sentiment normalization provides better accuracy as it uses intensity of sentiment rather than sum function.
2.	Identifying the Overlap between Election Result and Candidates' Ranking Based on Hashtag-Enhanced, Lexicon-Based Sentiment Analysis[8]	<u>Rezvaneh</u> <u>Rezapour</u>	Classification accuracy increased by 7%	2017	Manually added corpus-based hashtags to a sentiment lexicon, finding that this step in combination with negation detection
3.	Sentiment analysis system adaptation for multilingual processing[9]	Alexandra Balahuret. al.	use of hybrid features sentiment classification	2015	Lexicon based method used with semi supervised Learning Arabic Sentiment Lexicon
4.	A Review of Relational Machine Learning for Knowledge Graphs[10]	Maximilian Nickel et. al.	Utility of knowledge graphs in sentiment analysis	2016	statistical relational learning (SRL) methods applied to large knowledge graphs
5.	Twitter sentiment analysis using hybrid cuckoo search method[11]	Avinash Chandra Patel et. al.	CSK random initialization population K-means	2017	CSK and K-means to resolve the problem of random initialization.
6.	A hybrid method for bilingual text sentiment classification based on deep learning[12]	Guolong Liu ; Xiaofei Xu ; Bailong	RNN and RB-SVM	2016	deep learning features and shallow learning features

3. Datasets

A more in-depth analysis was done regarding the sources of datasets from 50 articles which are shown in Figure 1. The main source of data is Twitter with twenty three (23) articles, followed by movie reviews with eight (9) articles, Facebook data (6) articles, social media data (7) and 5 from other miscellaneous sources.

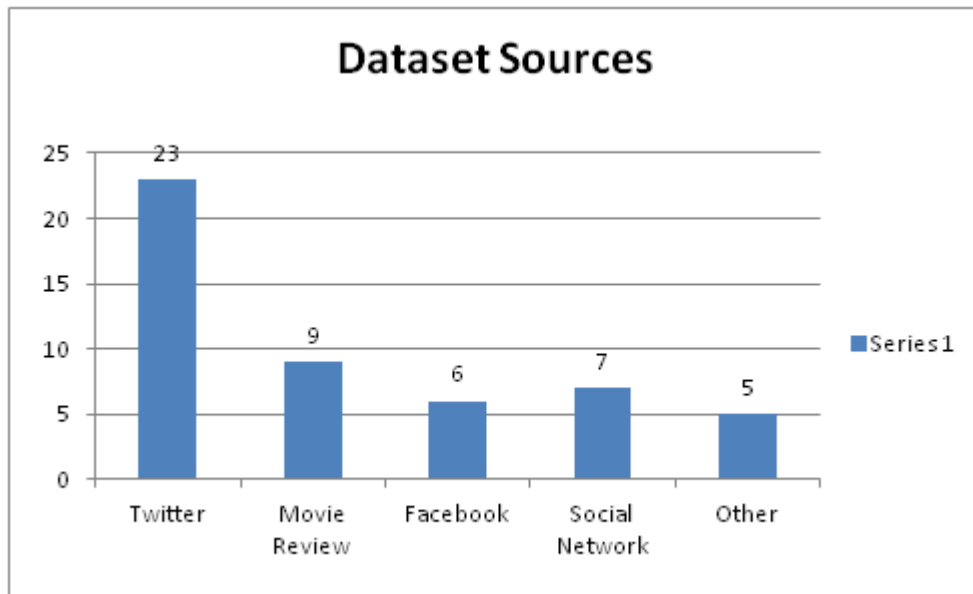


Fig 1. Sources of Datasets from 50 Articles

4. Conclusion

A few conclusions can be made based on the findings of the systematic literature which was conducted on articles published past few years using all approaches of sentiment analysis. In addition, more advancement on social media analytics left footprints since 2015. The emergence of the significance of sentiment analysis matches the development of social media usage, including reviews, forum blogs, micro-blogs, Face book, Twitter and other social networks. The social network analysis is basically done on opinionated data available in humungous amount and which can further be used for different methods of analysis.

In terms of applications, more research has been done on the assessment or evaluation of the various methods of opinion based on sentiment analysis. Although these refer to the evaluation of the techniques used, the data sets extracted from users'. Marketing-related activities still dominate the applications followed by the financial, healthcare, and hospitality and tourism industries. Amongst various approaches all has its benefits and drawback we are considering hybrid method to be the most useful method as we can customize the operations based on data and incorporate benefits of more than one approach.

References

- [1] <https://www.brandwatch.com/blog/understanding-sentiment-analysis>
- [2] Piñeiro-Chousa, J., Vizcaíno-González, M., & Pérez-Pico, A. M. (2016). "Influence of Social Media over the Stock Market". *Psychology & Marketing*, 34(1), 101–108. doi:10.1002/mar.20976
- [3] <https://www.expertsystem.com/natural-language-processing-sentiment-analysis>
- [4] F. Neri, C. Aliprandi, F. Capecci, M. Cuadros and T. By, "Sentiment Analysis on Social Media," 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, 2012, pp. 919-926. doi: 10.1109/ASONAM.2012.164
- [5] C. Chauhan and S. Sehgal, "Sentiment analysis on product reviews," 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, 2017, pp. 26-31. doi: 10.1109/CCAA.2017.8229825
- [6] Bharat R. Naiknaware et al. "Sentiment Analysis of Indian Government Schemes Using Twitter Datasets". IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, PP 70-78
- [7] Anna Jurek , Maurice D. Mulvenna and Yaxin Bi. "Improved lexicon-based sentiment analysis for social media analytics". *Secur Inform* (2015) 4:9 DOI 10.1186/s13388-015-0024-x
- [8] R. Rezapour, L. Wang, O. Abdar and J. Diesner, "Identifying the Overlap between Election Result and Candidates' Ranking Based on Hashtag-Enhanced, Lexicon-Based Sentiment Analysis," 2017 IEEE 11th International Conference on Semantic Computing (ICSC), San Diego, CA, 2017, pp. 93-96. doi: 10.1109/ICSC.2017.92
- [9] Alexandra Balahur et. al.. " Sentiment analysis system adaptation for multilingual processing: The case of tweets". <https://doi.org/10.1016/j.ipm.2014.10.004>
- [10] Maximilian Nickel et. al. "A Review of Relational Machine Learning for Knowledge Graphs". arXiv:1503.00759 [stat.ML]
- [11] Avinash Chandra Patel et. al. "Twitter sentiment analysis using hybrid cuckoo search method". <https://doi.org/10.1016/j.ipm.2017.02.004>
- [12] G. Liu, X. Xu, B. Deng, S. Chen and L. Li, "A hybrid method for bilingual text sentiment classification based on deep learning," 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Shanghai, 2016, pp.93-98. doi: 10.1109/SNPD.2016.7515884