A Survey on Sentiment Analysis Approaches

Seema Chithore, D. A. Phalke,

P. G Student, Department of Computer Engineering, D.Y.Patil College of Engineering, Akurdi, Pune, India

Assistant Professor, Department of Computer Engineering, D.Y.Patil College of Engineering, Akurdi, Pune, India

ABSTRACT: : Different messages express opinions about products, events and services, political views or even their author's emotional state and mood. A basic task in sentiment analysis is classifying the polarity of a given text at the sentence, document or feature level whether the expressed opinion in a document, a sentence or an entity aspect is positive, negative or neutral. Sentiment analysis captures wide range of human moods, a large majority of studies focus on identifying the polarity of a given text that is to automatically identify if a message about a certain topic is positive or negative. In order to measure polarity, relative rate of positive and negative affects examined in the feeling categories. With this human labelled data, it can be measured the extent to which different sentiment analysis methods can accurately predict polarity of content. Polarity analysis used in numerous applications especially for real time systems that rely on analysing public opinions or mood fluctuations.

Data is collected from various social sites, from a document words are extracted those expresses emotions, their polarity and intensity are extracted which expresses the public wisdom and opinions. It is used in many applications like transportation system, Stock Market, Election criteria etc.

KEYWORDS: Sentiment analysis; Sentiment classification; Feature selection; Emotion detection.

I. INTRODUCTION

Sentiment analysis is the process of identifying the sentiment about the particular object. Textual information in the world can be broadly categorized into two main types: facts and opinions. Facts are objective expressions about entities, events and their properties. Opinions are usually subjective expressions that describe peoples sentiments, emotions etc [1]. Sentiment means feeling like Attitudes, Emotions, Opinions. It is a Subjective impression, not facts. Sentiment analysis is also called as opinion mining. Sentiment analysis is the process of automatically detecting if a text contains emotional or opinionated content and determining its polarity research that has received significant attention in recent years, both in academia and in industry [2]. The goal of sentiment analysis is typically to determine the polarity of a piece of natural language text [3]. Part-of-speech information is very often used effectively in sentiment analysis because POS tagging can be used for word sense disambiguation [4].

Microblog sentiment analysis has been a hot research area in recent years. Different important issues are studied like identifying whether a post is subjective or objective i.e. subjectivity classification, identifying whether a post is positive or negative i.e. polarity classification and recognizing the emotion in a particular post i.e. emotion classification [5]. Polarity analysis is the classification of a text document in positive, negative and neutral, ac-cording to a certain domain [6].

A large majority of studies focus on identifying the polarity of a given text that is to automatically identify if a message is positive or negative. Polarity analysis has many applications especially for real time systems that rely on analysing public opinions or mood fluctuations [7].

Sentiment analysis is one of the feedback mechanisms most used in Twitter data analysis, which provides a view about the sentiment expressed in the messages [5]. This sentiment is basically labelled according to the text polarity expressing message has a positive, negative or neutral connotation. For companies, this measure helps to observe the public and market opinion about itself. This task is named polarity analysis or sentiment polarity. There are three basic levels at which subjectivity can be calculated: document level, the sentence level and the aspect level Sentiment analysis. [8]. Sentiment analysis at document level consider whole document as a basic unit. Sentiment

analysis at sentence level consider sentence as a basic unit. In the aspect level sentiment analysis is done as per aspect of the entity. Different entities and their aspect is identified from the sentence.

Data used for sentiment analysis has a huge volume. Mostly it uses reviews of the product or a particular thing of which we want to calculate the sentiment. Various social networking sites are available which allows end to end communication between different users of different group of people. Microblogs such as Twitter, Facebook are the popular social media where people expresses their feelings, emotions [9]. Twitter sentiment analysis has become a important research topic in recent years. The goal of this task is to discover the attitude or opinion of the tweets [10]. Authors of those messages write about their life, share opinions on variety of topics and discuss current issues. Users write various Because of a free format of messages and an easy accessibility of microblogging platforms, Internet users tend to shift from traditional communication tools to microblogging services. As more and more users post about products and services they use, or express their political and religious views, microblogging websites become valuable sources of peoples opinions and sentiments. Such data can be efficiently used for marketing or social studies [11]. Sentiment analysis over microblogs like twitter faces several new challenges due to the typical short length and irregular structure of content. First direction is concerned with finding new methods to run such analysis, such as performing sentiment label propagation on Twitter follower graphs, and employing social relations for user-level sentiment analysis. The second direction is focused on identifying new sets of features to add to the trained model for sentiment identification, such as microblogging features including hashtags, emoticons etc. ,this can be done by finding semantic similarity [12].

People who exchange messages via e-mail and chat often use symbols for faces called as emoticons. Emoticons play an important role in identifying the emotion of sentence [13]. Use of emoticons in online conversation improve machine understanding of language used online, and contribute to the creation of more natural human-machine interfaces [14].

Figure 1 is the block diagram of system that gives the idea how the sentiment polarity is obtained; the input is read in the form of raw text which is converted into an array of sentences using Natural language Toolkit sentence tokenizer [15].
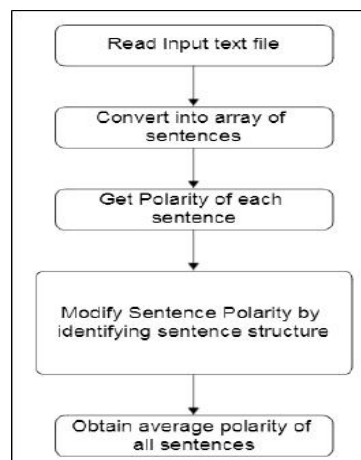

Figure 1: System design for Sentence Level Sentiment Calculation [15]

Sentiment analysis is used in various domains so to increase the pro t, strategy by knowing the people opinion about a particular product, event and social information. It is used in Business, Politics, Public Actions, Finance etc [16].

II. MAIN APPROACHES OF SENTIMENT ANALYSIS

In Broadly, there exist two types of methods for sentiment analysis: machine-learning-based and lexical-based. Machine learning methods often rely on supervised Classification approaches, where sentiment detection is framed as a binary (i.e., positive or negative). This approach requires labelled data to train classifiers. While one advantage of learning-based methods is their ability to adapt and create trained models for specific purposes and contexts. On the other hand, lexical-based methods make use of a predefined list of words, where each word is associated with a specific sentiment [17].

In Microblog sentiment analysis different important classification of post is done like subjectivity classification, polarity classification, emotion classification. It is assumed in general that the document being considered contains subjective information, such as in product reviews and feedback forms. Opinion orientation can be classified as belonging to opposing positive or negative polarities : positive or negative feedback about a product, favourable or unfavourable opinions on a topic or ranked according to a spectrum of possible opinions, for example on lm reviews with feedback ranging from one to five stars. Polarity analysis is a part of

sentiment analysis. Approaches to develop the polarity analysis are same as that of sentiment analysis. It is mainly classified as Machine learning and Lexicon based learning. The categorization is as shown in figure 2 Sentiment Analysis Approaches.

Machine Learning Approach:

Machine learning is one of the type of artificial intelligence. Due to it computer learned with-out being explicitly programmed. Machine learning approach uses the data to determine the patterns within the user's data and adjust the program actions accordingly [8].
Machine learning approach relies on the famous ML algorithms to solve the SA as a regular text classification problem that makes use of syntactic and/or linguistic features. Machine learning approach again categorized as follows:
1. Supervised algorithms
2. Unsupervised algorithms

Supervised is used to apply what has been learned in the past to new data. The supervised learning methods depend on the existence of labeled training documents [18]. In supervised learning, given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and the output.

Supervised learning problems are categorized into as follows:

a. Regression problems
b. Classification problems

In a regression problem, results are predicted within a continuous output, that mean input variables are mapped to some continuous function. In a Classification problem, Results are predicted discrete output. In other words, input variables are mapped into discrete categories.
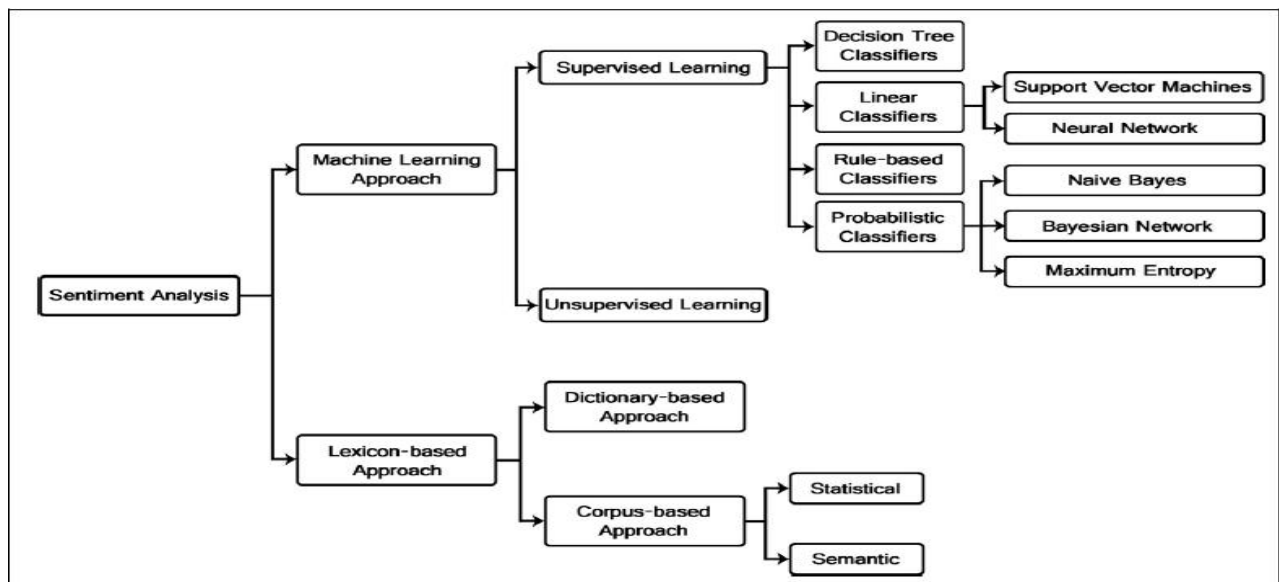

Figure 2: Sentiment Analysis Approaches [8]

Example:

Given data about the size of houses on the real estate market, try to predict their price. Price as a function of size is a continuous output, so this is a regression problem.

This example could turn into a Classification problem by instead making our output about whether the house "sells for more or less than the asking price." the houses based on price are classified into two discrete categories.

Unsupervised algorithms can draw conclusions from datasets. Unsupervised learning, on the other hand, allows us to approach problems with little or no idea what our results should look like. Structure is derived from data where the effect of the variables is not known. Structure is derived by clustering the data based on relationships among the variables in the data. With unsupervised learning there is no feedback based on the prediction results, i.e., there is no teacher to correct you. It's not just about clustering. For example, associative memory is unsupervised learning.

Example : Facebook's News Feed uses machine learning approach to personalize each member's feed.

Lexicon Based Approach :

Lexicon based approach extracts the sentiments from the text. The process of assigning a positive or negative label to a text that detects the text's opinion towards its main subject matter. This technique is based on assumption that polarity of the document or sentence is sum of polarity of individual words or phrases. Lexicon Based Approach is again categorized as Dictionary-based and Corpus based approach [8].

The lexicon-based approach calculates orientation for a document from the semantic orientation of words or phrases in the document [19].

The dictionary is the domain specific i.e. Polarity of the words are set according to a specific domain like political blogs, book reviews etc.

The corpus based approach begins with a basic list of opinion words and then find the other opinion words in a large corpus to help in finding the opinion words with context specific orientations. E.g. extract sentiments from Twitter data to know products review.

The hybrid Approach combines both approaches and is very common with sentiment lexicons playing a key role in the majority of methods. Hybrid approaches to affective computing and sentiment analysis, finally, exploit both knowledge-based techniques and statistical methods to perform tasks such as emotion recognition and polarity detection from text or multimodal data.

III. RELATED WORK

Sentiment analysis has been applied in various areas, knowing the public opinion about particular thing and work accordingly for our product is the approach followed by various owners. Intelligent transportation systems (ITSs) have stringent requirements of safety, efficiency, and information exchange. Author CAO et al proposed the traffic sentiment analysis (TSA) as a new tool for all these requirements, which provides a new prospective for modern ITSs [20]. Classification of information is important in the finance domain, where news about a company can affect the performance of its stocks [21].

In Sentiment analysis data is divided into multiple chunks, the preprocessing is done. Stop wards removals, converting the local words into standards, URLs removal all comes under preprocessing part. Then Classification is done using existing data on the new data.

IV. IDENTIFIED FUTURE SCOPE

Future works include the establishing Classification priorities by automatic assignment of weights to words and emoticons. Sentiment analysis in other than English language is a growing field due to lack of resources and researchers interest. Eliminating the manual work of assigning label to known data and more accurately assigning label is also a open research work.
.

V. CONCLUSION

Sentiment Analysis is a challenging task considering the volume of messages and speed of new message generation about the particular subject. Messages do not have polarity assigned with it and manually assigning the polarity to each message is a tedious task.

Text with emoticons in messages prompt the emotions more impact fully and help in faster sentiment calculation. So there could be automated task of polarity Classification of tweets. For that it is necessary to label known messages to train the classifier. This labelling task is very time consuming and thus might result in error. Various approaches work effectively to develop sentiment calculation tool as per requirement of the particular domain or business.

REFERENCES

1. B. Liu, "Sentiment analysis and subjectivity." Handbook of natural language processing, vol. 2, pp. 627-666, 2010.
2. A. H. Huang, D. C. Yen, and X. Zhang, "Exploring the potential effects of emoticons," Information & Management, vol. 45, no. 7, pp. 466-473, 2008.
3. F. Jiang, Y.-Q. Liu, H.-B. Luan, J.-S. Sun, X. Zhu, M. Zhang, and S.-P. Ma, "Microblog sentiment analysis with emoticon space model," Journal of Computer Science and Technology, vol. 30, no. 5, pp. 1120-1129, 2015.
4. A. C. E. Lima, L. N. de Castro, and J. M. Corchado, "A polarity analysis framework for twitter messages," Applied Mathematics and Computation, vol. 270, pp. 756-767, 2015.
5. J. Leskovec, "Social media analytics: tracking, modeling and predicting the flow of information through networks," in Proceedings of the 20th international conference companion on World wide web, pp. 277-278, ACM, 2011.
6. A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. de Jong, and U. Kaymak, "Exploiting emoticons in sentiment analysis," in Proceedings of the 28th Annual ACM Symposium on Applied Computing, pp. 703-710, ACM, 2013.
7. G. Paltoglou and M. Thelwall, "Twitter, myspace, digg: Unsupervised sentiment analysis in social media," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 3, no. 4, p. 66, 2012.
8. W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093-1113, 2014.
9. S.-H. Cho and H.-B. Kang, "Statistical text analysis and sentiment Classification in social media," in 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1112-1117, IEEE, 2012.
10. K.-L. Liu, W.-J. Li, and M. Guo, "Emoticon smoothed language models for twitter sentiment analysis.," in AAAI, 2012.
11. A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining.," in LREc, vol. 10, pp. 1320-1326, 2010.
12. H. Saif, Y. He, and H. Alani, "Semantic sentiment analysis of twitter," in International Semantic Web Conference, pp. 508-524, Springer, 2012.
13. M. Yuasa, K. Saito, and N. Mukawa, "Emoticons convey emotions without cognition of faces: an fmri study," in CHI'06 Extended Abstracts on Human Factors in Computing Systems, pp. 1565-1570, ACM, 2006.
14. M. Ptaszynski, J. Maciejewski, P. Dybala, R. Rzepka, and K. Araki, "Cao: A fully automatic emoticon analysis system based on theory of kinesics," IEEE Transactions on Affective Computing, vol. 1, no. 1, pp. 46-59, 2010.
15. N. Lalithamani, L. S. Thati, and R. Adhikesavan, "Sentence-level sentiment polarity calculation for customer reviews by considering complex sentential structures," IJRET: International Journal of Research in Engineering and Technology, vol. 3, 2014.
16. B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and trends in information retrieval, vol. 2, no. 1-2, pp. 1-135, 2008.
17. P. Goncalves, M. Araujo, F. Benevenuto, and M. Cha, "Comparing and combining sentiment analysis methods," in Proceedings of the rst ACM conference on Online social networks, pp. 27-38, ACM, 2013.
18. E. A. Hassan, N. El Gayar, and M. G. Moustafa, "Emotions analysis of speech for call Classification ," in 2010 10th International Conference on Intelligent Systems Design and Applications, pp. 242-247, IEEE, 2010.
19. M. Taboada, J. Brooke, M. To loski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," Computational linguistics, vol. 37, no. 2, pp. 267-307, 2011.
20. J. Cao, K. Zeng, H. Wang, J. Cheng, F. Qiao, D. Wen, and Y. Gao, "Web-based traffic sentiment analysis: Methods and applications," Intelligent Transportation Systems, IEEE Transactions on, vol. 15, no. 2, pp. 844-853, 2014.
21. J. Z. Ferreira, J. Rodrigues, M. Cristo, and D. F. de Oliveira, "Multi-entity polarity analysis in financial documents," in Proceedings of the 20th Brazilian Symposium on Multimedia and the Web, pp. 115-122, ACM, 2014.
22. E. Kouloumpis, T. Wilson, and J. D. Moore, "Twitter sentiment analysis: The good the bad and the omg!," Icwsm, vol. 11, pp. 538-541, 2011.
23. A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pp. 142-150, Association for Computational Linguistics, 2011.
24. B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in Mining text data, pp. 415-463, Springer, 2012.
25. E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," IEEE Intelligent Systems, vol. 28, no. 2, pp. 15-21, 2013.

26.   X. Hu, L. Tang, J. Tang, and H. Liu, "Exploiting social relations for sentiment analysis in microblogging," in Proceedings of the sixth ACM international conference on Web search and data mining, pp. 537-546, ACM, 2013.
27.   D. R. Recupero, V. Presutti, S. Consoli, A. Gangemi, and A. G. Nuzzolese, "Sentilo: frame-based sentiment analysis," Cognitive Computation, vol. 7, no. 2, pp. 211-225, 2015.

BIOGRAPHY

Seema Chithore Seema Chithore is currently pursuing Masters in Computer Engineering at D.Y. Patil College of Engineering, SPPU, Pune. She has received BE in Computer Science from Amravati University in 2009. Her area of interest is Information Retrieval.

Mrs. D. A. Phalke Mrs. D. A. Phalke is a Assistant professor in Computer Department at D. Y. Patil College of Engineering,Akurdi, Pune. She is currently pursuing phD and her area of specialization are Information retrieval and Data Mining.