

Latest trends in sentiment analysis - A survey

Anju Rose G Punnelparambil

PG Scholar

Department of Computer Science & Engineering

Govt. Engineering College, Thrissur, India

anjurose.ar@gmail.com

Abstract : Nowadays, people express their reactions to various public issues, events or products in social media applications. An organization can analyze such reactions of people to take an action on the event. Sentiment analysis helps to do that. It is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. There are several methods evolving to do sentiment analysis in the era of Big data. This survey deals with latest trends in sentiment analysis. Sentiment computing of news events and shop reviews are some applications.

Keywords : Social media applications, Sentiment analysis, Big data.

1. Introduction

Nowadays, people express their opinions and emotions through tweeting, sharing images, commenting on social sites. The huge amount of this content gives opportunities for understanding social behavior and building socially intelligent systems to investigate and extract information with text analysis methods from social media data. Sentiment analysis aims to find emotion of people from their text. This survey deals with recent sentiment analysis techniques which are applicable for analysis of shop reviews[1] , sentiment computing of news events[2]. Sentiment analysis becomes more challenging when language of data is other than English. One such challenging method for sentiment analysis described in [1], Sentiment Analysis System for Indonesia Online Retail Shop Review. The sentiment computing of news event is a significant component of the social media big data. News events, which is a significant component of the social media big data on the web, are news stories which have occurred in the society or on the web and are reported or discussed by a number of web pages on the web [3]. After the occurrence of this web event, many people discuss it on the web through the social media. Among lots of news event analysis, one of the most challenge tasks is the sentiment computing of the news events, which aims to discover the emotions of the texts from the users. This is another challenging approach the survey considered [2]. The last participating item in survey is an extended study of sentiment analysis using n-grams [4]. Tackling the challenges posed by Social Networking content and addressing its casual nature, n-gram graphs technique provides a language-independent supervised approach for text mining.

The persisting paper is devised as follows: Section 2 describes the Sentiment Analysis System for Indonesia Online Retail Shop Review, Section 3 gives the overview of Sentiment Computing for the News Event Based on the Social Media Big Data, N-gram graphs for sentiment analysis will be discussed in Section 4, Section 5 includes the comparison of the methods and Section 6 reaches the conclusion of this paper.

2.Sentiment analysis for shop reviews[1]

The rapid growth of internet user and the popularity of social media network has led to big data of online opinion. Analysis on these opinions is very important because it can extract knowledge that can be the basis in making business decisions for the organizations. The problem is Indonesian citizen communicate in Bahasa and local languages, not to mention slang languages. So the model built for Indonesian online distro classification opinions in three major groups, which are:

- *Target object* : This group would give knowledge regarding which aspect of online distro is most popular and what not.
- *Polarity of the sentiment* : This group would give knowledge regarding polarity of the opinion. Usually in this group using Indonesian adjective words and classifying them as negative, positive and neutral sentiment.
- *Polarity of the target object* : This group would give knowledge on sentiment of each aspect on online distro.

The system is divided into two subsystems as in the description on Fig. 1. The first subsystem is learning process, so the aim of the first subsystem is to build models for the system. The second subsystem goal is to identify and determine the polarity of the input dataset. This subsystem uses to test the model that has been built on the first subsystem, by doing the classification process on input dataset. In Pre-processing first, changing up all uppercase to lowercase in the input train dataset. Then splitting the sentence into words or token, this process called tokenization. Then building dictionary to replace token that related to online distro aspects and sentiments. Since online review prone to misspell and grammatical error, this process becomes crucial. Here used Naive Bayes Classifier (NBC) in order to get sentiment and aspects classification of online retail business. With this approach using prior online reviews from OSNs of Online retail business as a knowledge to classify the aspects and sentiment of the online retail business. This utilizes NBC technique to gain knowledge from online opinions.

Feature extraction and selection to select words from learning dataset of online review and then classifying them to the respective class of target objects and sentiment. So dimensionality reduction is done by selecting features that are capable of discriminating words (token) that belong to different classes. In the first step the model need to identify an overall Distro character which represents one of six store dimensions: product assortment and variety, value of the merchandise given its price, service, location, facilities and store atmosphere. The second step is to build an annotated corpus based on their respective class. In this stage, first analyze

all important aspect related to online Distro in order to get targeted object's class. Afterward, extract all of the aspect terms related to each class of online Distro classification from all token that got from the preprocessing stage. This token usually noun and other predefined words. Then calculated the number of tokens per opinion (Bag of Word) and count the number of keywords into a particular label. The same process goes to extract the polarity of sentiment (positive, negative and neutral) and the polarity of the target object. For polarity of target object, developers combine token into two subsets which are adjective, noun and predefined words.

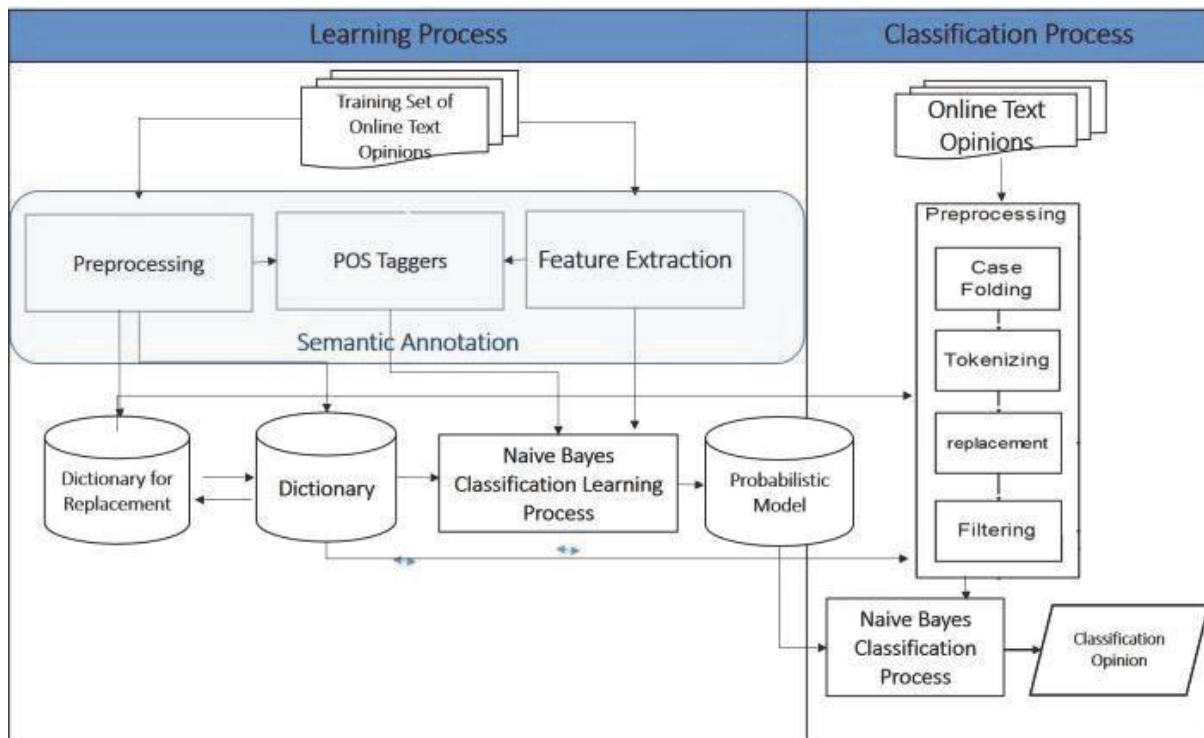


Fig. 1 overview diagram

Naive Bayes Classification learning process goal is to get probabilistic value for each word on each classification domain group as mention above. In the system, adopting naive Bayes to classify existing opinion. There are several steps this calculation process :

- Calculate probability for each class of Indonesia online retail shop aspects
- Calculate likelihood probability
- Calculate the highest probability of Distro's aspect and sentiment

2. Sentiment computing for news events[2]

The sentiment computing of news event is a significant component of the social media big data. It has also attracted lots of researches which could support many real-world applications, such as public opinion monitoring for governments and news recommendation for websites. However, existing sentiment computing methods are mainly based on the standard emotion thesaurus or supervised methods, which are not scalable to the social media big data. However,

the methods proposed almost classify texts into two categories: positive and negative, which doesn't conform the characteristic that the public sentiment is complex. Nowadays, some researchers tend to compute the text multidimensional emotions. According to commonly used emotions, the emotion is a six dimensional vector on: joy, love, surprise, fear, sad and

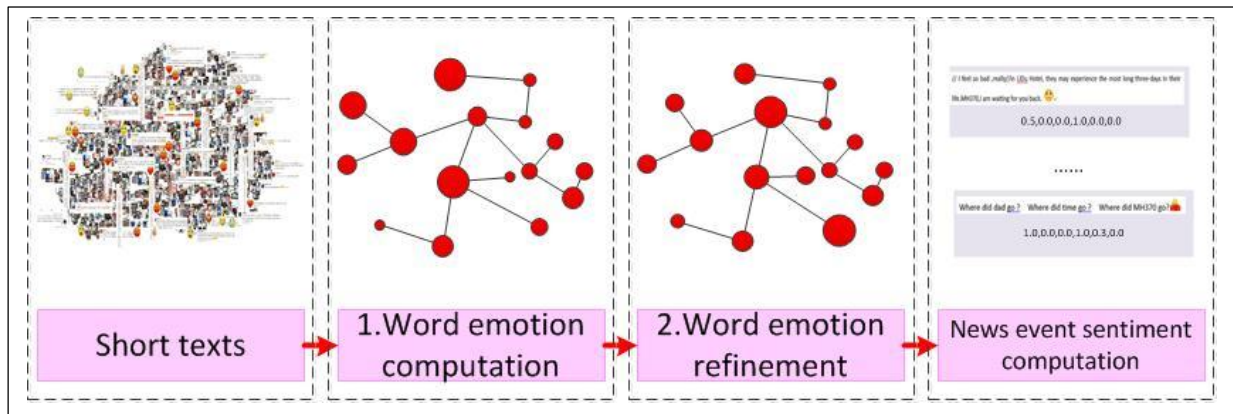


Fig.2 framework of text sentiment computation

As in Fig. 3, the main part of the news event sentiment computing task is the word emotion computation, which can be splitted into two procedures:

Word emotion computation through word emotion association network and word emotion refinement through standard sentiment thesaurus. For the first part of word emotion computation, a Word Emotion Association Network (WEAN) is built to jointly capture its semantics and emoticons, which is the basis for both word and text emotion computations. Assumption of the paper[2] is that the words semantically associated will be more possible with the similar emoticons - symbols. An iterative process with its convergence proof is designed to optimize the emotional weights assignment for the links in WEAN. After this process, initial word emotions obtained. but they may not be consistent with existing common knowledge. For example, the word 'happy' should have a large weight on joy, but it may obtain wrong emotion after the iterative process. So, in the second part of word emotion computation, have designed a mechanism to refine the initial word emotions by incorporating the common prior knowledge: standard emotion thesaurus.

3. Sentiment analysis using N- gram graphs [4]

Tackling the challenges posed by Social Networking content and addressing its casual nature, n-gram graphs technique provides a language-independent supervised approach for text mining. Adopting this data analysis model, the paper[4] provides an extended study of sentiment analysis, using a multi-lingual and multi-topic environment, employing and

combining different classification algorithms, and attempting various configuration approaches on classification parameters to increase the efficiency.

The method uses a supervised machine learning model to examine extensive experiment results with a multilingual corpus of manually annotated posts from Twitter. The n-gram graphs technique employed successfully deals with SNSs specificities presented above and yields promising results concerning accuracy, with respect to big data workload processing time limitations. More specifically, the contribution provided by the work corresponds to the following points:

- An innovative language-agnostic and noise tolerant technique for Sentiment Analysis is presented, improving the classification accuracy compared to the current State of the Art.
- Experiments were performed on an extended manually annotated dataset, multi-lingual (Spanish, English, Portuguese, Dutch, German and French included) and multi-topic. This aims at the generality, pragmatism and validity of the method results.
- An extra analysis for the best Split Ratio of training and testing sets is provided, given that the same analogy is used for each contributing dataset, to ensure the validation process is not biased.

4. Comparison

Three participating papers/systems in the survey similar by task done, ie sentiment analysis . But they differ by 6 attributes. The parameters are language of text to be processed, features of method, categories of sentiment, data used for system, storage of data and accuracy of system mentioned in paper. Comparison between the papers is shown in Table 1.

Citation of paper	Language Of processing	Feature	Categories	Data	Storage	Accuracy
[1]	English Bahasa	Naive Bayes Classification	3 : positive Negative Neutral	Facebook review	Database	89%
[2]	English	WEAN	6 : joy,love, surprise, fear, sad, anger	Malaysia Airlines MH370 on March 8, 2015	Database	78%
[4]	Spanish English Portuguese Dutch German French	N Gram Graphs	3 : positive Negative Neutral	Twitter dataset	Twitter API	73%

Table 1 shows that some systems with great accuracy do not reveal about implementation details and these systems are independent of each other. So their accuracy cannot be taken into account for this survey. But this analysis shows different methods for sentiment analysis and it enables to do more experiments with the same.

5. Conclusion

Different methods for sentiment computing handled in this article and analysis of these methods is tabulated. This survey helps to adapt the best method for an application. For example, if there is a need to build an application that analyze shop reviews in our area, then sentiment analysis using Naive bayes method[1] can be adapted. Analysis of reactions of people on an event can be done by WEAN[2].

REFERENCES

- [1] Cut Fiarni et al., Sentiment Analysis System for Indonesia Online Retail Shop Review Using Hierarchy Naive Bayes Technique, *2016 Fourth International Conference on Information and Communication Technologies (ICoICT)*, ISBN: 978-1-4673-9879-4.
- [2] Dandan Jiang, Xiangfeng Luo , Junyu Xuan, Zheng Xu, Sentiment Computing for the News Event Based on the Social Media Big Data , *IEEE Transactions*,2169-3536 2016.
- [3] J. Xuan, X. Luo, G. Zhang, J. Lu, and Z. Xu, Uncertainty analysis for the keyword system of web events, *Systems, Man, and Cybernetics: Systems*, *IEEE Transactions on*, vol. PP, no. 99, pp.1-1, 2015.
- [4] Fotis Aisopos et al., Using n-gram graphs for sentiment analysis: an extended study on Twitter, *2016 IEEE Second International Conference on Big Data Computing Service and Applications*, 978-1-5090-2251-9/16.