Advances in Computational Research
ISSN: 0975-3273 & E-ISSN: 0975-9085, Volume 7, Issue 1, 2015, pp.-259-261.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000187

REVIEW ARTICLE ON OPINION MINING USING NAÏVE BAYES CLASSIFIER

PISOTE A.* AND BHUYAR V.

MCA Department, Marathwada Institute of Technology, Aurangabad- 431 028, MS, India.
*Corresponding Author: Email- anitapisote@gmail.com

Received: December 18, 2014; Revised: January 05, 2015; Accepted: January 15, 2015

Abstract- Now a day's people are using Twitter, forum discussions, blogs, and Facebook to share their views, emotions and opinions. It may be positive, negative or neutral opinion. So the large amount of data is now available on web in text format and there is a need to analyze and interpret this data. Decision makers and researchers can make use of this interpreted data to make appropriate decision. This can be done with the help of sentiment analysis. When we summarize the document on the basis of their features we come to know whether it is positive, negative or neutral opinion. This document summarization is useful in feedback analysis Product review, business decision making. Different classifier are used to mine the sentiments, one of it is Naïve Bayes Classifier. This paper is aim to study the sentiment analysis and naïve bayes classifier to classify the document on the basis of their sentiments.

Keywords- Naïve Bayes Classifier, Opinion Mining

Citation: Pisote A. and Bhuyar V. (2015) Review Article on Opinion Mining Using Naïve Bayes Classifier. Advances in Computational Research, ISSN: 0975-3273 & E-ISSN: 0975-9085, Volume 7, Issue 1, pp.-269-261.

Introduction

Opinion mining is a process of tracking the opinion of the public about particular product and services. An opinion is simply positive or negative sentiment, view, attitude, emotion, or appraisal about an entity.[1] An entity is a product, person, event, organization, or topic. Automated opinion mining uses machine learning. Machine learning is a type of artificial intelligence (AI) to mine text for sentiment. Sentiment analysis is done at three levels [2,3].

Document Level: The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment. This level of analysis assumes that each document expresses opinions on a single entity (e.g., a single product). Thus, it is not applicable to documents which evaluate or compare multiple entities.

Sentence Level: The task at this level goes to the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion. This is closely related to subjectivity classification. Subjective expressions come in many forms, e.g. opinions, allegations, desires, beliefs, suspicions, speculations. A subjective sentence may contain a positive or negative opinion. Objective sentences can imply opinions too. Ex. "The machine stopped working in the second day" Sentiment classifications at both the document and sentence (or clause) levels are useful, but they do not find what people liked and disliked. Therefore we need to go for entity and aspect level.

Entity and aspect Level: Aspect level performs finer-grained analysis. Aspect level was earlier called feature level. Instead of looking at documents, paragraphs, sentences, clauses or phrases, aspect level directly looks at the opinion itself.

Steps Involved in Opinion Mining

In Opinion mining First step is Preparing Review database in this Reviews, comment, Remark, opinion of particular product or thing are stored in review database [Fig-1].

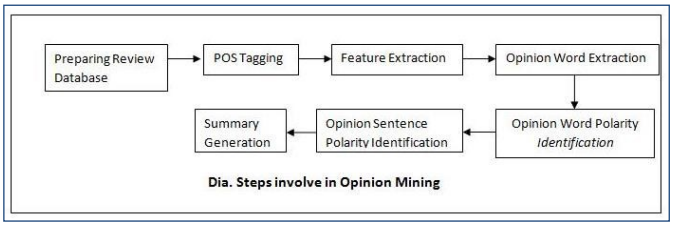

Fig. 1- Steps involved in opinion mining

Second step is part of speech tagging i.e. POS Tagging in which it find out product features and opinion words. In part-of-speech (POS tagging), each word in review is tagged with its part- of- speech (such as noun, adjective, adverb, verb etc).Third step is feature extraction in feature extraction, product features are extracted from each sentence. Fourth step is Opinion of word extraction in opinion word extraction, opinion words are identified i.e. one or more features or one or more opinion is extracted. Fifth step is Opinion Word Polarity Identification in this step semantic orientation of each opinion word is identified. Semantic orientation means identifying whether opinion word is expressing positive opinion, negative opinion or

neutral opinion. And sixth step is Opinion Sentence Polarity Identification in this step Opinion sentence polarity identification predicts the orientation of an opinion sentence. Consider following sentence- "Bincy is a very good student" Above sentence contains opinion word 'good' which expresses positive opinion. The last step is Summary Generation. Summary generation is generated after opinion sentence orientation identification. With the help of information discovered in previous steps summary can be generated.

Prior Work

Opinion Mining is recent research area where many people are working.

According to Tribhuvan, et al [2] For feature based opinion mining and summarization different tools like RapidMiner, WordNet, POSTagger, Crawlers and Parsers can be used. Raut, et al [3]discuss about machine learning approach works well for sentiment analysis of particular data such as movie, product, hotel etc., while lexicon based approach is suitable for short text in micro-blogs, tweets, and comments data on web. Osimo & Mureddu [4] focuses on Key challenges and gaps. Khairnar & Kinikar [5] discussed Machine learning techniques like Naïve Bayes, SVM is an excellent method for data classification. Pak & Paroubek [6] classify the sentiment on the basis of multinomial Na¨ıve Bayes classifier that uses N-gram and POS-tags as features. Angulakshmi & ManickaChezian [7] used a tool to track the opinion or polarity from the user generated contents are: Review Seer tool, Web Fountain, Red Opal, Opinion observer. Virmani, et al [8] done a case study where teachers give their remarks about the students and by applying the proposed sentiment analysis algorithm the opinion is extracted and represented. Ex. Remark 1.

This Student is very intelligent and hardworking. He is also participative in various co-curricular activities

Haseena [9] focuses on different application areas of opinion mining. OM is a few sequence of well known standard mechanisms as: Subjectivity Detection, polarity Detection, Degree of polarity identification [10].

Naïve Bayes Classifier

There is different classifier which is used for text classification. One of which is Naïve Bayes Classifier. Naive Bayes classifiers have worked well in many complex real-world situations. The naïve Bayes classification is a supervised learning technique as well as a statistical technique for classification. It constructs a model based on feature values which assign class labels to problem instances. There is different algorithm which work on common principle that value of a particular feature is independent of the value of any other feature, given the class variable [13].

Naïve Bayes classifier is based on Bayes Rule. Bayes rule applied to document d and class c.

$$p(c/d) = \frac{p(d/c)p(c)}{p(d)}$$

Where p(c/d) is a posterior probability

P(d/c) is a likelihood of evidence

P(c) is prior probability

P(d) is normal constant. It can be ignored. Classify the new instance of document d based 0n the tuple of attributes values into one of class cj Є C, d= {x1,x2…….xi}

C_{MAP}= argmax P(cj/x1,x2…..xi)

$$= argmax \frac{p(x1,x2....xi/cj)\,p(cj)}{p(x1,x2....xi)}$$

=argmax P(x1,x2……xi/cj) P(cj)

=argmax P(cj) ∏ P(xi/cj) MAP is Maximium a Posterior = Most likely class Bayes rule assumed that one feature is independence of the other. Therefore following equation occurs.

P(x1,x2…xi/c) = P(x1/c) * P(x2/c) …* P(xi/c)

If there is number of document and document consist of Bag of words (x1,x2……xi). In training phase, compute p(cj) for every class cj and P(xi/cj) for every term in vocabulary xi. Use the frequencies in data.

$$P(cj) = \frac{\text{document}(C = cj)}{N}$$

$$P\left(\frac{xi}{cj}\right) = \frac{\text{document}(X = xi,\ C = cj)}{N(C = cj)}$$

If any word occurs first time then its probability will be zero. So the numerator will be zero. To overcome this smoothing techniques are used to avoid over fitting. It is called as Laplace add 1 smoothing.

$$P\left(\frac{xi}{cj}\right) = \frac{\text{document}(X = xi,\ C = cj) + 1}{N(C = cj) + K}$$

where k is no. of values in xi.

Consider the example in which training dataset having 4 document and test set 1 as shown in table.

Type	Doc	Words	Class
Train	1	Chinese Beijing Chinese	C
	2	ChineseChineseShanghai	C
	3	Chinese Macao	C
	4	Tokyo Japan Chinese	J
Test	5	Chinese Chinese Chinese Tokyo Japan	?

Priors:

P(c) = 3/4

P(j) =1/4Conditional Probabilities:

P(Chinese/c) = (5+1)/(8+6)=6/14=3/7

P(Tokyo/c) =(0+1)/(8+6)=1/14

P(Japan/c)=(0+1)/(8+6)=1/14

P(Chinese/j)=(1+1)/(3+6)=2/9

P(Tokyo/j)=(1+1)/(3+6)=2/9

P(Japan/j)=(1+1)/(3+6)=2/9

Choosing a class:

C_{MAP} =argmax P(cj) ∏ P(xi/cj)

P(c/d5)=(3/4)*(3/7)^3*(1/14)*(1/14) =0.0003

P(j/d5)=(1/4)*(2/9)^3*(2/9)*(2/9) =0.0001

So the document d5 classify to class 'c' as it having maximum frequency using Naïve Bayes Classifier.

Conclusion

Opinion mining is emerging area where research going on rigorously. On web people share their opinion. Text classification is major

part of sentiment analysis. Naïve Bayes classifier is used to solve many complex problem. This paper gives study of Naïve Bayes classifier algorithm with its application in sentiment analysis.

Conflicts of Interest: None declared.

References

[1] Liu B. (2012) *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.

[2] Tribhuvan P.P., Bhirud S.G. & Tribhuvan A.P. (2014) *International Journal of Computer Science and Information Technologies*, 5(1), 247-250.

[3] Raut V.B. & Londhe D.D. (2014) *International Journal of Computer Science and Information Technologies*, 5(2), 1026-1030.

[4] Osimo D. & Mureddu F. (2012) *Research Challenge on Opinion Mining and Sentiment Analysis*, Universite de Paris-Sud, Laboratoire LIMSI-CNRS, Bˆatiment, 508.

[5] Khairnar J. & Kinikar M. (2013) *International Journal of Scientific and Research Publications*, 3, 153-161.

[6] Pak A. & Paroubek P. (2010) *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*, LREC. Universit´e de Paris-Sud, Laboratoire, France.

[7] Angulakshmi G. & ManickaChezian D.R. (2014) *International Journal of Advanced Research in Computer and Communication Engineering*, 3(7), 7483-7487.

[8] Virmani D., Malhotra V. & Tyagi R. (2014) *Sentiment Analysis Using Collaborated Opinion Mining*, arXiv preprint arXiv:1401.2618.

[9] Haseena R.P. (2014) *International Journal of Application or Innovation in Engineering & Management*, 3(5), 401-403

[10] Bhattacharyya D., Biswas S. & Kim T.H. (2010) *International Journal of Smart Home*, 4(2), 31-38.