IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 1, January 2019

Sentiment Analysis of English Text and Emoticons

P Amuthabala[1], Sangeetha K R[2], Yogitha K R[3]

Assistant Professor, Dept. of Information Science & Engg., Atria Institute of Technology, Bangalore, India[1]

Student, Dept. of Information Science & Engg., Atria Institute of Technology, Bangalore, India[2]

Student, Dept. of Information Science & Engg., Atria Institute of Technology, Bangalore, India[3]

Abstract: Sentiment Analysis also called as opinion mining is a type of natural language processing for making out the mood of the people about a particular product or event etc., by building a system to extract and classify opinions on a product. For example human opinion can be positive and negative or both or neither. Opinions change every day. Natural language processing techniques are applied to extract emotions from unstructured data.

Keywords: Sentiment analysis, Natural Language Processing (NLP), Machine learning, Data Retrieval

INTRODUCTION

The Opinion Mining (or Sentiment Analysis) is the process that performs the analysis on the text that is written in human language. NLP (Natural Language Processing) and data mining are the two main components of opinion mining. NLP is used to provide interface between human languages and computer, the NLP is curious about establishing effective algorithms to process the text that is written in the human language and provide that information which is understandable to computer application.

Data mining can be applied into vivid areas one as image mining, multimedia mining, web-mining etc. One of the important fractions of data mining is Opinion mining, which transpires as a part of Web mining. The opinion Mining is used to examine and collect or categorize the people opinions, reviews, sentiments, emotions about the product, event, services etc. that are in the human language.
It classifies them into positive, negative or neutral depending on humans' sentiments, opinions, emotions that are expressed in it, for instance word that is positive in one situation could also be considered as negative in a different situation, take a word "lengthy". If customer says phone life is lengthy, that is positive opinion or if the customer says booting time of phone is lengthy, that would be a negative opinion. [1]

Computerized opinion mining ordinarily makes use of computing device learning, a form of AI (Artificial Intelligence), to mine textual content for sentiment. The opinion mining can be useful in various number of approaches, Blog, tweets or any social media sites where consumer can broadcast their opinions, and it helps marketers to examine which product or services are popular and liked. To find this type of information in a semantic way, provide the vendor a clear picture of opinions because the data is created by the people [2]. Day-by-day the growth of social media is increasing and it is difficult to process all the content of each website on web so that there are some approaches for mining sentiment from the online sites, most sentiment analysis systems use bag-of-words approach for mining from social media rather than complete sentence/paragraph for analysis. [3]And there is also some filtration uses to detect false information and rumors that are put about on social sites like twitter.

ARCHITECTURE DIAGRAM

Opinion Retrieval

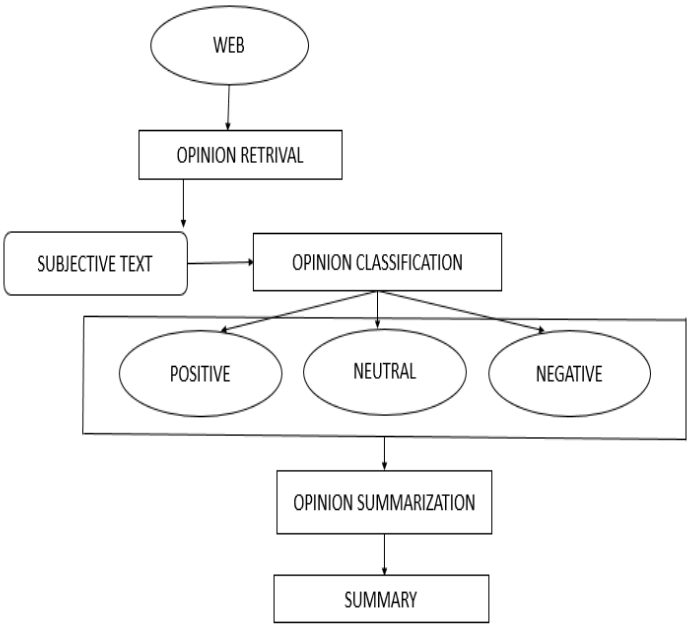
It is the procedure of gathering review textual content from review web sites. Different review web pages include studies for merchandise, movies, news and events. Information retrieval approach such as web crawler may be applied to collect the review textual content information from many sources and retailer them in database. This step contains retrieval of studies, microblogs, and comments of person.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 1, January 2019



Opinion Classification
Most important step is classification of review textual content into two types namely positive,neutral or negative . Given review document D = {d1.... di} and a predefined categories set C = {positive, negative,neutral}, sentiment classification classifies each and every di in D, with a label expressed in C.This process is achieved using appropriate algorithms.

Opinion Summarization
Summarization of opinion is a most important phase of opinion mining system. Abstract of reviews should be based on sub topics or features that are mentioned in reviews. Analysis is done on the reviews and the result is plotted.

The opinion summarization mainly contains the following two strategies-

Feature based summarization: It is a type of summarization which includes finding of frequent terms (features) which might be showing in lots of reviews. The summary is presented by deciding on sentences that contain detailed characteristic knowledge.

Term frequency: It relies on the time period of a word occuring in a document. If a term has better frequency it signifies that term is extra import for abstract presentation.

METHODOLGIES

Machine Learning Approach
Machine learning is a branch of computer science that gives electronic machines the capability to learn and understand by setting predictions on the given data without being explicitly programmed [4]. Some of the popular classifying algorithms are:

A. Naive Bayes classifier assumes that the value of a specific feature is independent of the value of any other feature [4]. This classifier is straightforward, uncomplicated and efficient for large datasets, without any complex iterative parameter estimation [5].

B. Maximum entropy is a classifier which is based on probability distributions of the data. The primary rule is that when no information is known then the distribution should have maximal entropy [6]. The labelled training data offers restriction on the distribution and find out where to have minimal non-uniformity.

C. Support Vector Machines (SVM) are supervised techniques together with learning algorithms that observe data used for classification provided with training examples, which are clearly labelled for belonging to one of the types [7]. An

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 1, January 2019

SVM training algorithm develops a system that assigns unique examples to each group, making it a non-probabilistic, binary linear classifier

Manual approach: It is a lengthy, labor intensive and an exhaustive technique to build a lexicon. Hence it is combined with one of the following two approaches.

A. Dictionary-based approach: In this approach, initially a basic group of seed words which have known polarity is assembled manually. Then, a program is run that collects synonyms and autonyms for these words and hence expand the dictionary. In each iteration, new words are added to the dictionary until no more new words can be found. Once the first cycle is completed, the list is manually examined for cleanup. Even though the approach seems simple, the key limitation of this method is that it generates general words which are independent of the background or context [4].

B. Corpus-based approach: In this approach, a domain specific lexicon is built. Such dictionaries can be built by two methods. A primary seed list of general-purpose sentiment words is generated and then the different opinion words along with their orientations are acquired [8].

The second method is to convert a general- purpose dictionary to corpus-based dictionary by a field corpus for opinion mining applications in the domain. Since a word in the same domain can be negative in one context and positive in another process is too complicated [9].

CONCLUSION

In this survey, importance of emoticons in sentiment analysis has been shown. Factors that affect sentiment analysis are discussed in brief. The survey also summarizes existing approaches for sentiment analysis. Text pre-processing, feature extraction and feature selection plays an important role for analyzing sentiments efficiently. Among existing techniques in sentiment analysis, Machine learning techniques are domain specific and work well for a specific domain (movie or product reviews) but not in general applications such as sentiment analysis on social networking data or twitter data set. Lexicon based approach are convenient for all domains as it emphasis on part of text present in the lexicon. Various machine learning techniques and lexicon based techniques can be combined to form a hybrid approach which may result into more accurate sentimental analysis

REFERENCES

[1] Chinsha T C and Shibily Joseph," A Syntactic Approach for Aspect Based Opinion Mining" IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015).
[2] Monika Arora and VineetKansal," A Framework for Informal Language: Opinion Mining", International Conference on Computing, Communication and Automation (ICCCA2015).
[3] MonishaKanakaraj and Ram Mohana Reddy Guddeti," NLP Based Sentiment Analysis on Twitter Data Using Ensemble Classifiers", 2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN).
[4] Mumtaz, Deebha, and BindiyaAhuja. "A Lexical Approach for Opinion Mining in Twitter",International Journal of Education and Management Engineering, 2016.
[5] Rennie, Jason D., et al. "Tackling the poor assumptions of Naïve Bayes text classifiers." ICML.Vol. 3. 2003
[6] 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, Dec 2014.
[7] Li, Kunlun, XuerongLuo, and Ming Jin. "Semi-supervised Learning for SVM-KNN." Journal of computers 5.5 (2010):671-678.
[8] Greene, Stephan Charles. Spin: Lexical semantics, transitivity, and the identification of implicit sentiment. ProQuest, 2007.
[9] Richard M. Tong. An operational system for detecting and tracking opinions in on-line discussion. In Proceedings of the Workshop on Operational Text Classification (OTC), 2001.

BIOGRAPHY

P. Amuthabala has obtained her B.E in Computer science Engineering at Avanishilingam University in Coimbatore, Tamilnadu, India in 2002 and her M.E degree in Software Engineering at Bangalore University in Bangalore, Karnataka, India in 2011.She is working as an Assistant Professor in Information Science Department at Atria Institute of Technology, Bangalore, Karnataka. Her research areas of Interest include Data Mining Data warehousing and Cloud Computing. She is currently doing her PhD in Karpagam University.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 1, January 2019

Sangeetha K R doing B.E in Information Science and Engineering at Atria Institute of Technology, Visvesvaraya Technology University, Bangalore.

Yogitha K R doing B.E in Information Science and Engineering at Atria Institute of Technology, Visvesvaraya Technology University, Bangalore.