IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

Overview of Sentimental Trend Analysis and Text Pattern Fetching

S. Maguda Gowreeswari[1], Dr. R. Manickachezian[2]

Research Scholar, Dr. Mahalingam Centre for Research and Development, NGM College, Pollachi, India[1]

Department of Computer Science, N G M College (Autonomous), Pollachi, Coimbatore, India[2]

Abstract: The advent of internet and World Wide Web the field of Sentiment Analysis is growing rapidly. There are numerous websites available on internet which provides analysis to users to give reviews about specific product. However the reviews expressed are mostly disorganized. An accurate method for sentiments could help us, to extract suggestions from the internet and predict customer's preferences which could prove valuable for Social Networks, Bulk suggestions and marketing research. There are various algorithms available for Sentimental analysis. Generally Sentimental analysis has three levels of granularities: Document level, Sentence level and Aspect level. In this paper, we study and analyze different issues, data sources, classification methods and evaluation metrics for Sentiment Analysis and text pattern fetching methods.

Keywords: Sentiment Analysis, Text classification, Machine Learning, Dataset, Lexicon Approach.

I. INTRODUCTION

Sentimental analyses are important to all humans as they influence ones behaviour. In today's competitive world, businesses and organizations always want to find consumer or public sentiments about their products and services. Consumers also want to know the suggestions of existing users of a product before purchasing it. In the past, when an individual needed suggestion, he/she asked friends and family. When an organization or a business needed public or consumer suggestion, it conducted surveys, opinion polls, and focus groups [7] With the explosive growth of social media (e.g., reviews sites, forum discussions, blogs, micro-blogs, Twitter, comments, and postings in social network sites) on the Web, individuals and organizations are increasingly using the content in these media for decision making. How to analyze and summarize the sentimental expressed in these huge opinionated text data is a very interesting domain for researchers. This new research domain is usually called Sentiment Analysis [9]. This paper is organized as follows: Section 2 presents the Introduction to Sentiment Analysis, Section 3 includes sources used for opinion mining, Section 3 introduces classification for sentiment Analysis. Then we present some comparison of different research on Sentiment Analysis and Last section concludes our study [12]. Main fields of research in sentiment analysis are Subjectivity Detection, Sentiment Prediction. Aspect based Sentiment Summarization, Text summarization for Opinions, Contractive viewpoint Summarization, Product Feature Extraction, Detecting opinion spam. Subjectivity Detection is a task of determining whether text is opinionated or not. Sentiment prediction is about predicting the polarity of text whether it is positive or negative. [3] Aspect based Sentiment summarization provides sentiment summary in the form of star ratings or scores of features of product. Text summarization generates a few sentences that summarize the reviews of a product. Contrastive viewpoint summarization puts an emphasis on contradicting suggestions. Product feature Extraction is a task that extracts the product features from its review. [6] Detecting opinion spam is concern with identifying fake or bogus opinion from reviews.

II. BACKGROUND STUDY

Sentiment analysis, also called suggestion finder, is the field of study that analyzes people's sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. The term sentiment analysis perhaps first appeared in Nasukawa and Yi, 2003, and the term opinion mining first appeared in Dave, Lawrence and Pennock,2003 [5]. Basically information is a single document of opinionated text. A single review about a single topic is considered in this document level classification. But comparative sentences may appear in the case of forums or blogs. In forums and blogs sometimes document level analysis is not desirable when customer may compare one product with another that has similar characteristics. [1] The challenge in the document level classification is that the entire sentence in an entire document may not be relevant in expressing the opinion about an entity. So subjectivity/objectivity classification is very much important in this type of classification. The text classification methods using Machine learning approach can be divided into supervised and unsupervised learning methods. The supervised methods use a large number of labelled training documents. The

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

unsupervised methods are used when it is difficult to find these labelled training documents. The lexicon-based approach depends on finding the opinion lexicon which is used to analyze the text. The dictionarybased approach which depends on finding opinion seed words, and then searches the dictionary of their synonyms and antonyms. The corpus-based approach starts with a seed list of opinion words, and then finds other opinion words in a large corpus to help in finding opinion words with context specific orientations. This could be done by using statistical or semantic methods.

a). Definition
Definition (Sentiment): A sentiment is a quintuple,
(ei , a ij , s ijkl , h k , t l),
where e i is the name of an entity, a ij is an aspect of e i , s ijkl is the sentiment on aspect a ij of entity e i , h k is the opinion holder, and t l is the time when the opinion is expressed by h k .

b). Issues in Sentiment Analysis
Before exploring Sentiment Analysis in depth we need to understand following issues in Sentiment analysis:
1) A sentence containing sentiment words may not express any sentiment., e.g., "Can you tell me which Canon camera is good?" such sentence contain the sentiment word "good", but neither expresses a positive or negative opinion on any specific camera.
2) A positive or negative sentiment word may have opposite orientations in different application domains. For example, "Not working" usually indicates negative sentiment, e.g., "This camera Not working," but it can also imply positive sentiment, e.g., "This vacuum cleaner really sucks."
3) Sarcastic sentences with or without sentiment words are hard to deal with, e.g., "What a great car! It stopped working in two days."
4) Many sentences without sentiment words can also imply suggestions. For example The sentence "This washer uses a lot of water" implies a negative sentiment about the washer since it uses a lot of resource (water).

III. LEXICON APPROACH

Input words are divided in many categories. Positive opinion words are used to express some necessary things, and negative opinion words are used to describe unnecessary things. Opinion phrases and idioms are also there which together are called opinion lexicon. To collect the opinion word list there are three main methods. One of them is Manual method not used alone and which is very time consuming. The basic steps of the lexicon based techniques are outlined below.

1. Pre-process each text (i.e. remove noisy characters and HTML tags)
2. Initialize the total text sentiment score: s<-0.
3. Tokenize text. For each token, check if it is present in a sentiment dictionary. (a). If token is present in dictionary, I. If token is positive, then s<-s+w. II. If token is negative, then s<-s-w.
4. Look at total text sentiment score s,
(i). If s> threshold, then classify the text as positive.
(ii). If s< threshold, then classify the text as negative.

IV. DATA SET AND SOURCE

Blogs, review sites, data and micro blogs provide a good understanding of the reception level of the products and services.
a) Blogs:
With an increasing usage of the internet, blogging and blog pages are growing rapidly. Blog pages have become the most popular means to express one's personal suggestions.

b) Review sites:
The reviews for products or services are usually based on sentimental expression in much unstructured format websites like www.amazon.com(productreviews),www.yelp.com(restaurant reviews), www.CNETdownload.com (product reviews) [8].

c) DataSet:
Movie review data are available as datasethttp://www.cs.cornell.edu/People/pabo/movie-reviewdata). Other dataset which is available online is multi domain sentiment (MDS) dataset. (http://www.cs.jhu.edu/mdredze/datasets/

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

sentiment)The MDS dataset contains four different types of product reviews extracted from Amazon.com including Books, DVDs, Electronics and Kitchen appliances, with 1000 positive and 1000 negative reviews for each domain. [8].

d) Micro-blogging:
Facebook is a popular micro blogging service where users create status messages called "posts". These posts sometimes express suggestions about different topics. Facebook messages are also used as data source for classifying sentiment [8].

V. METHODOLOGIES

There are different types of algorithms to analyze sentiments. Sentiment Classification techniques can be roughly divide into machine learning approach and lexicon based approach. The Machine Learning Approach (ML) applies the famous ML algorithms and uses linguistic features. The Lexicon-based Approach relies on a sentiment lexicon, a collection of known and precompiled sentiment terms. We discuss them below in brief.
a)Naive Bayes:
It is a simple but effective Learning & Classification algorithm. It is mostly used in Text Classification. The Classification method is based on theory of probability. It plays a vital role in probabilistic classification. It is also used in statistical method for classification and Supervised Learning method.
b) K-Nearest Neighbour (KNN):
It is also referred as Lazy Learning, Case-based Reasoning or Memorybased Reasoning. Given a test document d, the system finds the k nearest neighbours among training documents. The similarity score of each nearest neighbour document to the test document is used as the weight of the classes of the neighbour document.
c) Support Vector Machine (SVM):
It is a supervised learning model. This model is associated with a learning algorithm that analyzes the data and identifies the pattern for classification [11]. For example, consider an instance which belongs to either class Circle or Diamond. There is a separating line which defines a boundary. At the right side of boundary all instances are Circle and at the left side all instances are Diamond.
d)Dictionary Based Approach:
In this approach first of all a small set of sentiment words which are known as seed words are collected manually with their known positive or negative orientations. Then this set is grown by searching their synonyms and antonyms in WordNet or another online dictionary. The new words are added to the existing seed list. Then next iteration is started. The iteration should be stopped when no new words are found. [10]

VI. RESULT AND ANALYSIS

The performance of different methods is evaluated by calculating various metrics like precision, recall and F-measure. Accuracy Measurement shown in Table1.1. Figure1.1 represents the performance of machine learning algorithms.

Table 1.1 Accuracy Measurement

Techniques used	Year	Data source	Accuracy
Naive bayes	2015	Social Media reviews	92 %
SVM	2013	Amazon Reviews	89 %
Dictionary Based Approach	2010	Blog Post	85 %
KNN – K nearest neighbour	2007	Car Reviews	62 %

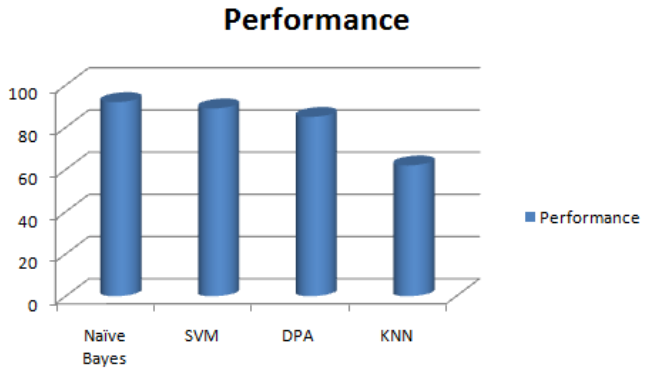

Figure 1.1 Machine learning algorithms

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

1) Precision is the fraction of retrieved instances that are relevant.
2) Recall is the fraction of relevant instances that are retrieved.
3) The two measures are sometimes used together in the F1 score (also F-score or F-measure) is a measure of a test's accuracy.

VII.CONCLUSION

Sentiment analysis has become very popular field of research. A lot has been researched in this field but still there are many issues as sentiment analysis processes text based unstructured data. Dictionary based approach takes less processing time than supervised learning approach but accuracy is not up to the mark. Supervised learning approach provides better accuracy. It is found that sentiment classifiers are severely dependent on domains or topics. From the above work it is evident that neither classification model consistently outperforms the other, different types of features have distinct distributions. It is also found that different types of features and classification algorithms are combined in an efficient way in order to overcome their individual drawbacks and benefit from each other's merits, and finally enhance the sentiment classification performance.

REFERENCES

[1] Kennedy and D. Inkpen,(2006) "Sentiment classification of movie reviews using contextual valence shifters," Computational Intelligence, vol. 22, pp. 110–125.
[2] NamrataGodbole,ManjunathSrinivasaiah, Steven Skiena,(2007) "LargeScale Sentiment Analysis for News and Blogs", ICWSM Boulder, Colorado, USA.
[3] Qingliang Miao, Qiudan Li, RuweiDai,(2009) "AMAZING: A sentiment mining and retrieval system", Expert Systems with Applications 36 7192–7198.
[4] GamgarnSomprasertsriPattarachaiLalitrojwong,(2010), Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization, Journal of Universal Computer Science, vol. 16, no. 6, 938-955.
[5] Gang Li ,Fei Liu ,(2010) "A Clustering-based Approach on Sentiment Analysis",978-1-4244-6793-8/10.
[6] KaiquanXu , Stephen Shaoyi Liao , Jiexun Li, Yuxia Song,(2011) "Mining comparative opinions from customer reviews for Competitive Intelligence", Decision Support Systems 50 743–754.
[7] Bing Liu,(May 2012)Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers,
[8] G.Vinodhini, R.M.Chandrasekaran,(June 2012) "Sentiment Analysis And Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering.
[9] Mr.SaifeeVohra, Prof. Jay Teraiya,(2013), "Applications and Challenges for Sentiment Analysis : A Survey" International Journal of Engineering Research & Technology (IJERT).
[10] WalaaMedhat, Ahmed Hassan, HodaKorashy, (2013)Sentiment analysis algorithms and applications: A survey.
[11] Pravesh Kumar Singh, MohdShahid Husain,(February 2014)"Methodological study of opinion mining and sentiment analysis techniques" International Journal on Soft Computing (IJSC) Vol. 5, No. 1.
[12] AsmitaDhokrat, Sunil Khillare, C.NamrataMahender(2015)"Review on Techniques and tools used for opinion mining" International Journal of Computer Applications Technology and Research.
[13] S.Kasthuri, Dr.L.Jayasimman, Dr. A. NishaJebaseeli,(2016) "An Opinion Mining and Sentiment Analysis Techniques: A Survey" International Research Journal of Engineering and Technology (IRJET).
[14] International Conference On Emanations in Modern Technology and Engineering (ICEMTE-2017) ISSN: 2321-8169 Volume: 5 Issue: 3, 69 – 73.

BIOGRAPHIES

S. Maguda Gowreeswari received her BCA (Computer Application)from NGM College, Pollachi, India. She completed her Master of Computer Application (MCA) from NGM College, Pollachi, India. Currently, she is a Research Scholar at Department of Computer Science, NGM College, Pollachi, India. She participated in a International Conference. Her area of interest includes Data mining, Sentimental analysis.

Dr. R. ManickaChezian received his M.Sc., degree in Applied Science from P.S.G College of Technology, Coimbatore, India in 1987. He completed his M.S. degree in Software Systems from Birla Institute of Technology and Science, Pilani, Rajasthan, India and Ph.D degree in Computer Science from School of Computer Science and Engineering, Bharathiar University, Coimbatore, India. He served as a Faculty of Maths and Computer Applications at P.S.G College of Technology, Coimbatore from 1987 to 1989. Presently, he has been working as an Associate Professor of Computer Science in N G M College (Autonomous), Pollachi under Bharathiar University, Coimbatore, India since 1989. He has published one-fifty papers in international/national journal and conferences: He is a recipient of many awards like DeshaMithra Award and Best Paper Award. Recently he received the award "Best Computer Science Faculty of the Year 2015" from Association of Scientists, Developers and Faculties. His research focuses on Network Databases, Data Mining, Distributed Computing, Data Compression, Mobile Computing, Real Time Systems and Bio-Informatics.