

## SENTIMENT ANALYSIS AND ITS APPLICATIONS: AN COMPLETE OVERVIEW

Renu Dhaka<sup>1</sup>, Anil Prajapti<sup>2</sup>

<sup>1</sup>M.Tech Scholar, CSE, <sup>2</sup>Asst. Professor, CSE

Swami Keshvanand Institute of Technology, Management, and Gramothan , Jaipur

**Abstract:** *Sentiment Analysis is very important in the current trend as it help us to analyze the views of the people whether on the social networking sites , or related to some product. With tremendous growth of the websites related to such domain , the atomization of the process of the review analysis is very important. Our paper is dedicated to the study of the sentiment analysis and its applications.*

**Index Terms:** *Sentiment Analysis , Sentiwordnet, Wordnet.*

### I. INTRODUCTION

According to the Oxford dictionary, the definition for sentiment analysis 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. is positive, negative, or neutral. Simply put, for the purposes of social media, it is the process of determining the author's opinion conveyed in a post. In recent years, the World Wide Web has become an important and emerging source of information; this is because of its rapid growth due to the increasing phenomenon of social network contacts, online discussion forums, blogs, digital libraries and quick streaming news stories. The web content, in the form of unstructured text, is useful in decision making. Knowledge extraction from such online text is very important for planning market strategies and decision making process. KDT (knowledge discovery in texts) or text data mining or text mining are terms used for the mining of unstructured or semi-structured data. It is a slightly new sub-discipline of data mining that considers textual data. The fact is that, "text data mining" is an intermediate evolutionary lexical form [1]. The Majority of the online information about data mining is "misleading". Such ambiguous/misleading information implies to the mining metaphor that it is like "extracting precious nuggets of ore from otherwise worthless rock" as found in or finding "gold hidden in mountains of textual data" in (D orre et al. 1999). A more narrow definition of what text mining does is needed to distinguish it from the traditional IR, which is basically "information access" as argued by Hearst [1]. Retrieval of documents relevant to the information needs of a user, is the primary concern of the traditional IR (perhaps a more appropriate name would be data retrieval); however, the user is left on his/her own to find the desired information in the documents. In Hearst's opinion, data mining has not only directed dealings with the information, but it also attempts to uncover or glean previously unknown, information from the data (text). Different linguistic levels (words, sentences and documents) highlighting the key differences between supervised machine learning methods, that rely on annotated

corpora or corpus-based, and unsupervised/lexicon based methods in sentiment classification. Three main steps are always involved in the process of text mining and sentiment classification; they are (a) acquiring texts which are relevant to the area of concern usually called IR; (b) presenting contents collected from these texts in a format that can be processed, such as statistical modelling, natural language processing, etc.; and (c) actually using the information in the presented format [3] [4] communities [5] [6]. Sentiment analysis has now become the dominant approach used for extracting sentiment and appraisals from online sources. Separating non-opinionated, neutral and objective sentences and texts from subjective sentences carrying heavy sentiments is a very difficult job; however, it has been explored earnestly in a closely related yet separate field, (J. M. Wiebe, 1994). It concentrates on making a distinction between 'subjective' and 'objective' words and texts; on one hand, the subjective ones give evaluations and opinions and on the other, the objective ones are used to present information which is factual (J. Wiebe, Wilson, R. Bruce, Bell, & Martin, 2004) (J. Wiebe & Riloff, 2005). This is different than sentiment analysis in regards to the set of categories into which language units are classified by each of these two analyses. Subjectivity analysis focuses on dividing language units into two categories: objective and subjective, whereas sentiment analysis attempts to divide the language units into three categories; negative, positive and neutral.

Opinion mining, also called Sentiment analysis, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space. There are also many names and slightly different tasks, e.g. sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc. However, they are now all under the umbrella of sentiment analysis or opinion mining. While in industry, the term sentiment analysis is more commonly used, but in academia both sentiment analysis and opinion mining are frequently employed. They basically represent the same field of study.

Initial research in text mining focused on extracting factual information from documents. In recent times, focus is shifting towards opinion mining - also called sentiment analysis. One of the drivers for this shift is availability of opinionated text in the form of reviews, blog posts, social media comments and more recently, tweets. Such documents are also called User Generated Content (UGC).

## II. TYPES OF SENTIMENT ANALYSIS

Firstly we need to understand the methods that social media vendors use to determine sentiment. As I mention above, there are many types of sentiment analysis. However, for the purposes of this article, we will concentrate on three

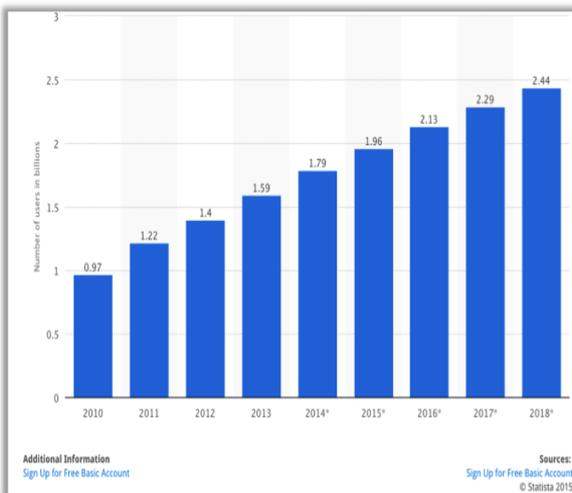
### A. Manual processing

Human interpretation of sentiment is definitely the most mature and accurate judge of sentiment. However, it still isn't 100% accurate. Very few vendors still use this process without the additional use of a tool. This is due to the prolific growth of social media. According to Seth Grimes, social is the fastest growing source of enterprise analytical data Source: Seth Grimes Social Media and the Enterprise Business Intelligence/Analytics Connection, A 2010 Study conducted by Seth Grimes.

According to Yellow Pages, 69% of Australians use social media. Source:

<https://www.sensis.com.au/content/dam/sas/PDFdirectory/Yellow-Social-Media-Report-2014.pdf>

Therefore, if you are going to use social media to determine sentiment, it is becoming less practical to use human processing and more likely you will need to automate the process.



### B. Keyword processing

Keyword processing algorithms assign a degree of positivity or negativity to an individual word, then it gives an overall percentage score to the post. For example, positive words, great, like, love or negative words: terrible, dislike

The advantages of this method are that it is very fast, predictable and cheap to implement and run.

However, there are numerous disadvantages including dealing with double negatives or positives or different meanings of words, for example, the use of a word such as 'sick' (to mean either 'ill' or to mean 'awesome'). Not to mention, different researchers may assign different percentages of positive or negative to a word. More often the issue is that it does not deal with multiple word/context issues or non-adjective words.

Most vendors represented in Australia use a keyword processing algorithm.

### C. Natural Language Processing

(NLP also called: text analytics, data mining, computational linguistics) NLP refers to computer systems that process human language in terms of its meaning. NLP understands that several words make a phrase, several phrases make a sentence and, ultimately, sentences convey ideas. NLP works by analyzing language for its meaning,

NLP systems are used for in a number of areas such as converting speech to text, language translation, and grammar checks Source: <http://mashable.com/2011/11/08/natural-language-processing-social-media/>

It can be likened to programming an algorithm to interpret the English language (or any language for the matter) with the rules that you were taught in English class.

Although NLP may seem to be far superior to keyword processing, it still has its limitations. Sarcasm a well-known Australian trait, is very difficult to detect using NLP as is hyperbole (exaggerated statements) and social media acronyms (e.g. OMG, BFF, BTW etc.) or social jargon such as:

- Youturn: To follow another person on social media with the intention of un-following them once they have you followed back, esp. on Twitter
- Wallflower: A person who regularly consumes the social media of others but never posts
- Face Crawling: Begging for Facebook likes, online or offline
- Hash-Browning: The excessive use of hashtags within a single post
- Metapals: Social media connections that have never personally met

People express opinions in complex ways for example: the difference between "I'm fine!!!" and "I'm fine." Also, changing topic mid post can be confusing.

## III. KEYWORD SEARCH

Information retrieval is the process of gathering information by using keywords from the relevant document and that document can be unstructured or structured data. It hides its complexity from user by providing abstract view. As user don't have any knowledge about schema and any other query processing language, he can search through abstract interface by putting keywords. By using Keyword Search user can submit keyword to search engines (Internet Search) or structured data and in turn it returns a list of documents to user according to ranking. Ranking of documents are provided based on the keywords match and occurrence of keyword match in particular document. Ranking is provided in descending order of occurrence of keyword match and the document with maximum occurrence get high priority.

Classification of Sentiment analysis/Opinion mining at different levels

Sentiment classification is a new field of Natural Language Processing that classifies subjectivity text into positive or negative.

The analysis levels can be done at three levels namely Document level, Sentence level and Feature level analysis.

#### A. Document Level Analysis

The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment. For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product. This task is commonly known as document-level sentiment classification. This level of analysis assumes that each document expresses opinions on a single entity. Thus, it is not applicable to documents which evaluate or compare multiple entities.

#### B. Sentence Level Analysis

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 level of analysis is closely related to subjectivity classification which distinguishes sentences (called objective sentences) that express factual information from sentences (called subjective sentences) that express subjective views and opinions. However, we should note that subjectivity is not equivalent to sentiment as many objective sentences can imply opinions.

#### C. Feature Level Analysis:

Both the document level and the sentence level analyses do not discover what exactly people liked and did not like. Aspect level performs finer-grained analysis. Aspect level was earlier called feature level (feature-based opinion mining and summarization). Instead of looking at language constructs (documents, paragraphs, sentences, clauses or phrases), aspect level directly looks at the opinion itself. It is based on the idea that an opinion consists of a sentiment (positive or negative) and a target (of opinion).

### IV. APPLICATION, CHALLENGES AND ISSUES

Opinion mining has various applications in different fields. It can be used in online advertising, hotspot detection in forums, search engines, recommendation systems, email filtering, questioning/answering systems, etc. Opinion mining application in daily life is most interesting as Opinion mining can be used to improve human-computer interactions, business intelligence, opinion poll that is voter can see opinion of other people before going to poll government intelligence, citation analysis etc. The following sample questions could be helpful in better understanding the applications of Opinion mining.

- What do individuals think about government strategies?
- Which feature of an item are loved or hated by overall population
- Who is a solid competitor for the general race body?
- Why has a sale of product declined?

Opinion mining suffers from several challenges, such as determining which segment of text is opinionated, identifying the opinion holder, determining the polarity strength of sentiment. Sentiment analysis is concerned with the human reviews, emotions and sentimental discussion. Everyone has their own perception and concern about a particular problem, issue, or topic. Opinionated text may be fake, irrelevant and

or ambiguous information. Opinions are far harder than facts to describe.

### V. CONCLUSION

Sentiment Analysis is a very interesting subject and there is a lot that can be done in this field. So we will also like to work in this field, for this we have taken the topic of analysis of documents related to women using the sentiment analysis for our thesis work.

### REFERENCES

- [1] A. Esuli and F. Sebastiani, 2006. "SentiWordNet: A publicly available lexical resource for opinion mining", In Proceedings of LREC-06, the 5th Conference on Language Resources and Evaluation, Geneva, Italy.
- [2] N. D. Valakunde, Dr. M. S. Patwardhan "Multi-Aspect and Multi-Class Based Document Sentiment Analysis of Educational Data Catering Accreditation Process" 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies.
- [3] Na Fan, Wandong Cai, Yu Zhao "Research on the Model of Multiple Levels for Determining Sentiment of Text" 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application.
- [4] Si Li, Hao Zhang, Weiran Xu, Guang Chen and Jun Guo, "Exploiting Combined Multi-level Model for Document Sentiment Analysis" 2010 International Conference on Pattern Recognition 1051-4651/10 \$26.00 © 2010 IEEE.
- [5] Raymond Y.K. Lau and Wenping Zhang, Peter D. Bruza "Learning Domain-specific Sentiment Lexicons for Predicting Product Sales" 2011 Eighth IEEE International Conference on e-Business Engineering.
- [6] Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment Classification using Machine Learning Techniques. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 79-86 (2002).
- [7] Si Li, Hao Zhang, Weiran Xu, Guang Chen and Jun Guo, "Exploiting Combined Multi-level Model for Document Sentiment Analysis" 2010 International Conference on Pattern Recognition 1051-4651/10 \$26.00 © 2010 IEEE.
- [8] Aurangzeb khan Baharum Baharudin, "Sentiment Classification Using Sentence-level Semantic Orientation of Opinion Terms from Blogs", 2011 IEEE.
- [9] Antonis Koukourikos<sup>1</sup>, Giannis Stoitsis<sup>2,3</sup>, Pythagoras Karampiperis<sup>1</sup> "Sentiment Analysis: A tool for Rating Attribution to Content in Recommender Systems" 2012.
- [10] V.K. Singh, R. Piryani, A. Uddin, "Sentiment Analysis of Textual Reviews" 2013 5th International Conference on Knowledge and Smart Technology (KST) 978-1-4673-4853-9/13/\$31.00

- ©2013 IEEE.
- [11] Dr. Muhammad Shahbaz<sup>1</sup>, Dr. Aziz Guergachi<sup>2</sup>, Rana Tanzeel ur Rehman<sup>3</sup>,” Sentiment Miner: A Prototype for Sentiment Analysis of Unstructured Data and Text” 978-1-4799-3010-9/14/\$31.00 ©2014 IEEE.
- [12] Chinsha T C, Shibily Joseph,” A Syntactic Approach for Aspect Based Opinion Mining;” Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015).