JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A REVIEW ON SENTIMENT ANALYSIS AND NATURAL LANGUAGE PROCESSING

Nisha Yadav, Dr. Banita
Research Scholar, Associate Professor
Department of Computer Science and Engineering
Baba MastNath University, Haryana

Abstract:

It has been found in the research area that many of the words are having similar spellings and pronounced in a similar way but their meaning will remain different, this difference in the meaning of any word will be analysed using sentiment analysis. The duplicity of the meaning will be distinguished with the use of sentence in which the word is mentioned. Here in this study a novel analytical approach has been introduced by which one can easily detect the duplicity of the word with respect to positive or negative aspect of the sentence. The semantic analysis is used to divide the word into number of tokens such that the preference will be arranged in such manner that one can easily justify the exact meaning of the word with in a particular sentence.

Keywords: NLP, Sentiment Analysis, Natural Language Generator, Steps of NLP, Recent Development.

I. INTRODUCTION

NLP can be classified into two parts i.e., Natural Language Understanding and Natural Language Generation which evolves the task to understand and generate the text. Figure 1 presents the broad classification of NLP. The objective of this section is to discuss the Natural Language Understanding (Linguistic) (NLU) and the Natural Language Generation (NLG). In theory, it deals with the range of techniques that compute, analyze and represent naturally happening texts at multi-level analysis of languages for the purpose to make the machine process like human language for different disciplines and applications. NLP algorithms depend highly on machine learning with the majority being statistical. Older implementation of language- processing tasks normally required hard coding of big set of rules. By using machine learning, we can use normal learning algorithms usually in statistical inference, to learn rules by analysing large corpora of real-world examples. A corpus is a set of documents that were annotated by hand with the correct valuesto be learned. Types of Natural Language Processing are:

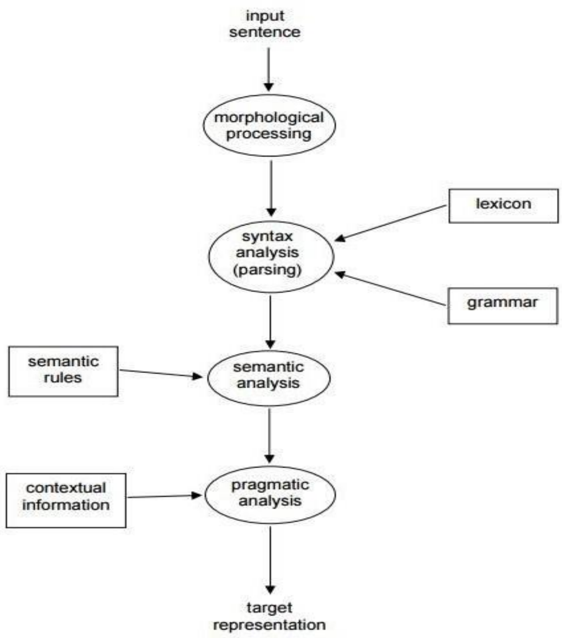

Figure 1.1: Steps of NLP

❖ **Syntax and semantic analysis:**

A processor that carries out different functions primarily based on syntax and semantic analysis. There are two uses of syntax analysis. One is to check if a sentence is well-formed and the other is to break it into a structure that gives syntactic relation between them. The same can be achieved by parser using a dictionary of word definitions and a set of syntax rules.

II. Natural Language Understanding

NLU enables machines to understand natural language and analyse it by extracting concepts, emotion, keywords etc. It is used in customer care applications to understand the problems reported by customers either verbally or in writing. Linguistics is the science which involves the meaning of language, language context and various forms of the language. So, it is important to understand various important terminologies of NLP and different levels of NLP. We next discuss some of the commonly used terminologies in different levels of NLP.

III. Sentimental Analysis:

Sentimental Analysis is a method of opinion mining to extract information about people's views, opinions, sentiments towards an everyday happening things. And each individual has a different opinion on same topic. The sentiment analysis task is technically more challenging but practically more useful. For ex- ample, Businessmen always want to know about the public opinion regarding that products and feedback from different customers. The customers also want to know the rating of that product which has given by other customers who had pur- chased earlier, and marketers also prefer sentiment analysis because they wanted to know the targeted customers. With the major development in social networking (i.e., Facebook, Twitter, LinkedIn, Stumble upon etc.,) on the Web, individuals and large associations are concentrating on publics opinion for their decision making.

Additionally, it is likewise realized that human analysis of content data is liable to significant preferences, e.g., peoples regularly give more priority to opinions that are reliable with their own preferences. There are other factors as well like human mental capacity and physical limitation that make humans inept to analyze large amount of data. Thus an automated opinion mining is required which will eventually help humans in sentiment analysis.

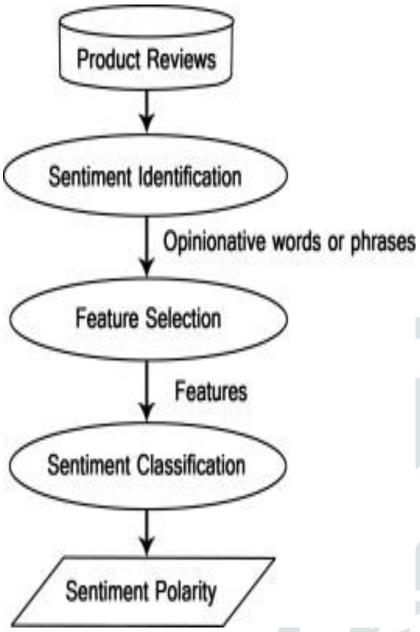

Figure 1.2: Classification of Sentiment Analysis

- **Semantic**

On a semantic level, the most important task is to determine the proper meaning of a sentence. To understand the meaning of a sentence, human beings rely on the knowledge about language and the concepts present in that sentence, but machines can't count on these techniques. Semantic processing determines the possible meanings of a sentence by processing its logical structure to recognize the most relevant words to understand the interactions among words or different concepts in the sentence. For example, it understands that a sentence is about "movies" even if it doesn't comprise actual words, but it contains related concepts such as "actor", "actress", "dialogue" or "script". This level of processing also incorporates the semantic disambiguation of words with multiple senses (Elizabeth D. Liddy, 2001). For example, the word "bark" as a noun can mean either as a sound that a dog makes or outer covering of the tree. The semantic level examines words for their dictionary interpretation or interpretation is derived from the context of the sentence. For example: the sentence "Krishna is good and noble." This sentence is either talking about Lord Krishna or about a person "Krishna". That is why, to get the proper meaning of the sentence, the appropriate interpretation is considered by looking at the rest of the sentence.

The goal of NLP is to accommodate one or more specialties of an algorithm or system. The metric of NLP assess on an algorithmic system allows for the integration of language understanding and language generation. The system incorporates a modular set of foremost multilingual NLP tools. The pipeline integrates modules for basic NLP processing as well as more advanced tasks such as cross-lingual named entity linking, semantic role labeling and time normalization. Thus, the cross-lingual framework allows for the interpretation of events, participants, locations, and time, as well as the relations between them. Output of these individual pipelines is intended to be used as input for a system that obtains event centric knowledge graphs. All modules take standard input, to do some annotation, and produce standard output which in turn becomes the input for the next module pipelines. Their pipelines are built as a data centric architecture so that modules can be adapted and replaced. Furthermore, modular architecture allows for different configurations and for dynamic distribution.

IV. Natural Language Generator:

Natural Language Generation (NLG) is the process of producing phrases, sentences and paragraphs that are meaningful from an internal representation. It is a part of Natural Language Processing and happens in four phases: identifying the goals, planning on how goals may be achieved by evaluating the situation and available communicative sources and realizing the plans as a text (Fig. 2). It is opposite to Understanding.

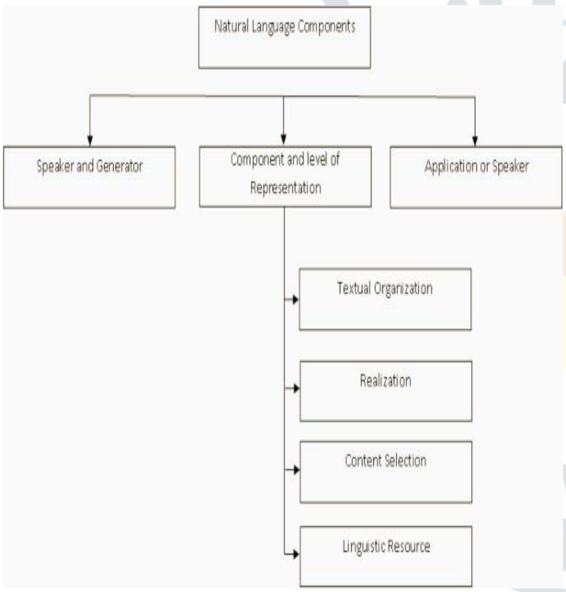

Fig1.3: Components of NLG

❖ Application or Speaker

This is only for maintaining the model of the situation. Here the speaker just initiates the process doesn't take part in the language generation. It stores the history, structures the content that is potentially relevant and deploys a representation of what it knows. All these forms the situation, while selecting subset of propositions that speaker has. The only requirement is the speaker must make sense of the situation.

❖ Applications of NLP

Natural Language Processing can be applied into various areas like Machine Translation, Email Spam detection, Information Extraction, Summarization, Question Answering etc. Next, we discuss some of the areas with the relevant work done in those directions.

Machine Translation

As most of the world is online, the task of making data accessible and available to all is a challenge. Major challenge in making data accessible is the language barrier. There are a multitude of languages with different sentence structure and grammar. Machine Translation is generally translating phrases from one language to another with the help of a statistical engine like Google Translate. The challenge with machine translation technologies is not directly translating words but keeping the meaning of sentences intact along with grammar and tenses. The statistical machine learning gathers as many data as they can find that seems to be parallel between two languages and they crunch their data to find the likelihood that something in Language A corresponds to something in Language B. As for Google, in September 2016, announced a new machine translation system based on artificial neural networks and Deep learning. In recent years, various methods have been proposed to automatically evaluate machine translation quality by comparing hypothesis translations with reference translations. It takes the information of which words are used in a document irrespective of number of words and order. In second model, a document is generated by choosing a set of word occurrences and arranging them in any order. This model is called multi-nominal model, in addition to the Multi-variate Bernoulli model, it also captures information on how many times a word is used in a document.

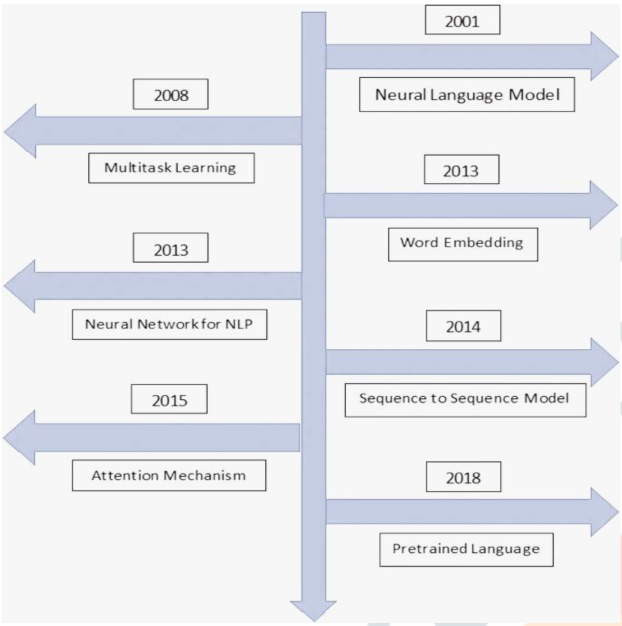

Fig 1.4. Recent Development of NLP

Discovery of knowledge is becoming important areas of research over the recent years Former one has higher accuracy but higher cost of implementation while latter has lower implementation cost and is usually insufficient for IR). Compound or Statistical Phrases (Compounds and statistical phrases index multi token units instead of single tokens.) Word Sense

The extracted information can be applied for a variety of purposes, for example to prepare a summary, to build databases, identify keywords, classifying text items according to some pre-defined categories etc. For example, CONSTRUE, it was developed for Reuters, that is used in classifying news stories (Hayes, 1992). It has been suggested that many IE systems can successfully extract terms from documents, acquiring relations between the terms is still a difficulty. PROMETHEE is a system that extracts lexico-syntactic patterns relative to a specific conceptual relation (Morin,1999). IE systems should work at many levels, from word recognition to discourse analysis at the level of the complete document. An application of the Blank Slate Language Processor (BSLP) (Bondale et al., 1999) approach for the analysis of a real-life natural language corpus that consists of responses to open-ended questionnaires in the field of advertising.

Sentiment analysis – otherwise known as opinion mining. In essence, it is the process of determining the emotional tone behind a series of words, used to gain an understanding of the attitudes, opinions and emotions expressed within an online mention. Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. The applications of sentiment analysis are broad and powerful. The ability to extract insights from social data is a practice that is being widely adopted by organizations across the world.

Consider the following sentence: ―My flight‘s been delayed. Brilliant, Most humans would be able to quickly interpret that the person was being sarcastic. We know that for most people having a delayed flight is not a good experience (unless there‘s a free bar as recompense involved). By applying this contextual understanding to the sentence, we can easily identify the sentiment as negative. Without contextual understanding, a machine looking at the sentence above might see the word ―brilliant‖ and categories it as positive.

That‘s not entirely dissimilar to how a linguist expert would teach a machine how to conduct basic sentiment analysis. As language evolves, the dictionary that machines use to comprehend sentiment will continue to expand. With the use of social media, language is evolving faster than ever before. 140-character limits, the need to be succinct and other prevailing memes have transformed the ways we talk to each other online. This of course brings with it many challenges.

We take all the words and phrases that imply positive or negative sentiment and apply rules that consider how context might affect the tone of the content. Carefully crafted rules help our software know the first sentence below is positive and the second is negative.

The above examples show how sentiment analysis has its limitations and is not to be used as a 100% accurate marker. As with any automated process, it is prone to error and often needs a human eye to watch over it. Sentiment Extraction (SE) deals with the retrieval of the opinion or mood conveyed in a block of unstructured text in relation to the domain of the document being analysed.

Conclusion:

It has been concluded in the study that the redundancy will be reduced upto an extent to make the sentence much more clear and having great vision of the understanding of meaning such that the actual meaning of the sentence shall not be affected.

Reference:

[1] Augustyniak, Łukasz, et al. "Comprehensive Study on Lexicon-based Ensemble Classification Sentiment Analysis." Entropy 18.1 (2015): 4.

[2] Bandgar, B. M., and Binod Kumar. "Real time extraction and processing of social tweets." International Journal of Computer Science and Engineering, E- ISSN 2347-2693 (2015): 1-6.

[3] Bhattacharyya, Pushpak, et al. "Facilitating multi-lingual sense annotation: Human mediated lemmatizer." Global WordNet Conference. 2014.

[4] Bird, Steven. "NLTK: the natural language toolkit." Proceedings of the COLING/ACL on Interactive presentation sessions. Association for Computational Linguistics, 2006.

[5] Bird, Steven, Ewan Klein, and Edward Loper. Natural language processing with Python. "O'Reilly Media, Inc.", 2009.

[6] Cambria, Erik. "Affective computing and sentiment analysis." IEEE Intelligent Systems 31.2 (2016): 102-107.

[7] Carrillo de Albornoz, Jorge, Laura Plaza, and Pablo Gervás. "A hybrid approach to emotional sentence polarity and intensity classification."Proceedings of the Fourteenth Conference on Computational Natural Language Learning. Association for Computational Linguistics, 2010.

[8] Carvalho, Jonnathan, Adriana Prado, and Alexandre Plastino. "A Statistical and Evolutionary Approach to Sentiment Analysis." Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02. IEEE Computer Society, 2014.

[9] Chowdhury, Gobinda G. "Natural language processing." Annual review of information science and technology 37.1 (2003): 51-89.

[10] "Company | About." Twitter. Twitter, 30 June 2016. Web. 04 Mar. 2017.

[11] Esuli, Andrea, and Fabrizio Sebastiani. "Sentiwordnet: A publicly available lexical resource for opinion mining." Proceedings of LREC. Vol. 6. 2006.

[12] Fernández-Gavilanes, Milagros, et al. "Unsupervised method for sentiment analysis in online texts." Expert Systems with Applications 58 (2016): 57-75.

[13] Firmino Alves, André Luiz, et al. "A Comparison of SVM versus naive-bayes techniques for sentiment analysis in tweets: a case study with the 2013 FIFA

confederations cup." Proceedings of the 20th Brazilian Symposium on Multimedia and the Web.

[14] ACM, 2014.

[15] Gastelum, Zoe N., and Kevin M. Whattam. "State-of-the-Art of Social Media Analytics Research." Pacific Northwest National Laboratory (2013).

[16] Gonçalo Oliveira, Hugo, António Paulo Santos, and Paulo Gomes. "Assigning Polarity Automatically to the Synsets of a Wordnet-like Resource." OASIcs-OpenAccess Series in Informatics. Vol. 38. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014.

[17] González, Cristóbal Barba, et al. "A Fine Grain Sentiment Analysis with Semantics in Tweets." International Journal of Interactive Multimedia and Artificial Inteligence 3.Special Issue on Big Data and AI (2016).

[18] Hemalatha, I., Dr GP Saradhi Varma, and A. Govardhan. "Case Study on Online Reviews Sentiment Analysis Using Machine Learning Algorithms." International Journal of Innovative Research in Computer and Communication Engineering 2.2 (2014):3182-3188.

[19] Hemalatha, I., Dr GP Saradhi Varma, and A. Govardhan. "Preprocessing the informal text for efficient sentiment analysis." International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) 1.2 (2012): 58-61.

[20] Hemalatha, I., Dr GP Saradhi Varma, and A. Govardhan. "Sentiment analysis tool using machine learning algorithms." International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) 2.2 (2013): 105-109.

[21] Hull, David A. "Stemming algorithms: A case study for detailed evaluation." JASIS 47.1 (1996): 70-84.

[22] Isah, Haruna, Paul Trundle, and Daniel Neagu. "Social media analysis for product safety using text mining and sentiment analysis." Computational Intelligence (UKCI), 2014 14th UK Workshop on. IEEE, 2014.