SENTIMENT ANALYSIS USING TELUGU SENTIWORDNET

[1] Mrs. B. TRIVENI , [2] CH.SRI DURGA PRASANNA, [3] N.SAHITHI,

[4] SK.ABDUAL RASHEED

1.   *Associate Professor Department of Computer Science and Engineering, Teegala Krishna Reddy Engineering College, Rangareddy (TS).India.*

Email-: triveni.banavatu@gmail.com.

[2,3,4]B.Tech StudentstDepartment of Computer Science and Engineering, Teegala Krishna Reddy Engineering College, Rangareddy (TS).India.

Email-: [3] prasannachada443@gmail.com [2] sahithinalajala11@gmail.com,
[4] Rasheedshaik0508@gmail.com.

Abstract- In recent times, sentiment analysis in low resourced languages and regional languages has become emerging areas in natural language processing. Researchers have shown greater interest towards analyzing sentiment in Indian languages such as Hindi, Telugu, Tamil, Bengali, Malayalam, etc. In best of our knowledge, microscopic work has been reported till date towards Indian languages due to lack of annotated data set. In this project, we proposed a two-phase sentiment analysis for Telugu news sentences using Telugu SentiWordNet. Initially, it identifies subjectivity classification where sentences are classified as subjective or objective. Objective sentences are treated as neutral sentiment as they don't carry any sentiment value. Next, Sentiment Classification has been done where the subjective sentences are further classified into positive and negative sentences. With the existing Telugu SentiWordNet, our proposed system attains an accuracy of 74% and 81% for subjectivity and sentiment classification respectively.

KEYWORDS: SentiWordNet, Natural Language Processing, Sentiments, Emotions.

1. INTRODUCTION

In natural language processing (NLP), sentiment analysis is a technique that deals with analyzing the emotions, sentiments, opinions of an individual towards a product, movies, events, news or organizations, etc. The primary task of sentiment analysis is to identify the polarity of a text in a given document. The polarity may be either positive, negative or neutral. Sentiment analysis can be applied to text in three categories namely, sentence level, document level, and aspect level. Sentence level analysis focuses on identifying sentence-wise polarity value in a given document. Document level analysis determines the polarity value based on consideration of the whole document. In aspect level analysis, it identifies the polarity of every aspect (word-wise) in a given text. Telugu is the second most popular language in India after Hindi. According to Ethnologue list of most-spoken languages worldwide, Telugu ranks fifteenth in the list, and a total of 85 million Telugu native speakers exist across the world. In the Telugu language, several e-Newspapers are available which publish news on a daily basis such as Eenadu, Sakshi, Andhrajyothy , Vaartha , and Andhrabhoomi .

2. LITERATURE SURVEY
Liu and Bing

Opinions are central to almost all human activities and are key influencers of our behaviors. Our beliefs and perceptions of reality, and the choices we make, are, to a considerable degree, conditioned upon how others see and evaluate the world. For this reason, when we need to make a decision we often seek out the opinions of others. This is not only true for individuals but also true for organizations. Opinions and its related concepts such as sentiments, evaluations, attitudes, and emotions are the subjects of study of sentiment analysis and opinion mining. The inception and rapid growth of the field coincide with those of the social media on the Web, e.g., reviews, forum discussions, blogs, micro- blogs, Twitter, and social networks, because for the first time in human history, we have a huge

volume of opinionated data recorded in digital forms. Since early 2000, sentiment analysis has grown to be one of the most active research areas in natural language processing. It is also widely studied in data mining, Web mining, and text mining. In fact, it has spread from computer science to management sciences and social sciences due to its importance to business and society as a whole. In recent years, industrial activities surrounding sentiment analysis have also thrived. Numerous startups have emerged. Many large corporations have built their own in- house capabilities. Sentiment analysis systems have found their applications in almost every business and social domain.

3.   EXISTING SYSTEM:

In the recent past, researchers have shown their interest towards sentiment analysis in the context of Indian languages such as Hindi, Bengali, Telugu, Punjabi, Marathi, etc. . Das and Bandyopadhyay deployed a computational tech- nique on English sentiment lexicons and English-Bengali bilingual dictionary to developed a Bengali SentiWordNet. In their subsequent work , they have extended their work and added two more Indian languages such as Hindi and Telugu to the SentiWordNet through an

interactive gaming strategy called "Dr. Sentiment" to create and validate the SentiWord- Net(s) for three Indian languages with the help of Internet user

DISADVANTAGES   OF   EXISTING SYSTEM:

1. Less accuracy

2. Low Efficiency

4.   PROPOSED SYSTEM:

In this section, we proposed an automatic sentiment analyzer for Telugu e-Newspapers sentences. A model is It starts with data collection and annotation. Further, using Telugu SentiWordNet, it classifies the sentiment of each sentence in news corpus. Finally, it compares the classification result with the manually annotated result for error analysis.

SentiWordNet is a sentiment lexicon that associates the sentiment information to each and every word synset. We can represent SentiWordNet as Wordnet + sentiment information. In this paper, we have used Telugu SentiWordNet to perform the sentiment analysis. This SentiWordNet consists of four files which contain negative, positive, neutral and ambiguous words

861

respectively.

ADVANTAGES OF PROPOSED SYSTEM:

1. High accuracy

2. High efficiency

5. MODULES:

Tensorflow

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google.

TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache 2.0 open-source license on November 9, 2015.

Numpy

Numpy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter Notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

Scikit – learn

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use.

5. RESULTS:
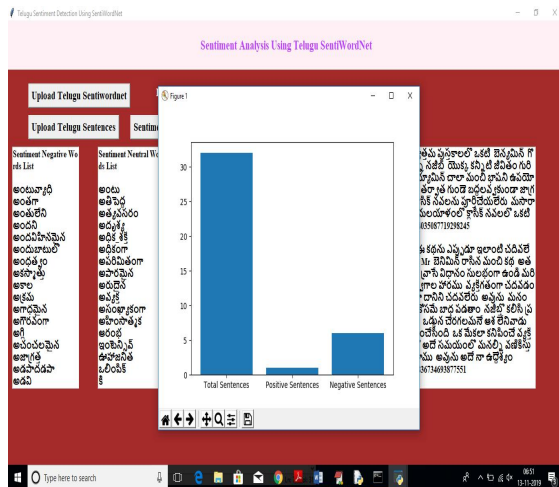
As a result its shows as in a graph x-axis represents sentence type as total, positive and negative and y-axis represents count.

7. CONCLUSION

In Telugu languages, it's hard to find annotated dataset to perform NLP tasks such as POS tagging, sentiment analysis, sarcasm analysis, text summarization, etc. There are few an- notated datasets available in this language. This paper exploits the available

Telugu SentiWordNet to perform sentiment anal- ysis for Telugu e-Newspapers sentences. The proposed system for sentiment analysis has attained an accuracy of 74% for subjectivity classification and 81% for sentiment classification in the domain of news data. In future, we need to improve the existing SentiWordNet to attains better accuracy and find an alternate way to make this SentiWordNet dynamic. It learns annotated data automatically and adds to the existing SentiWordNet. Acknowledgements: The authors would like to thank Bala Prakash, Manikanta, Vijay and Madhusudan for annotating the collected dataset. All the annotators are native to the states of Andhra Pradesh & Telangana and have a good knowledge of the Telugu language.

8. REFERENCES

[1] Liu and Bing, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, 2012, pp. 1–167.

[2] Ethnologue Languages of the world [online]. Available: https://www. ethnologue.com/statistics/size.

[3] Baccianella, Stefano, Andrea Esuli and

863

Fabrizio Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," LREC, 2010, vol. 10.

[4] Turney and Peter D., "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in Proceedings of the 40th annual meeting on association for computational linguistics, Association for Computational Linguistics, 2002.

[5] Pang Bo, Lillian Lee and Shivakumar Vaithyanathan, "Thumbs up?: sen- timent classification using machine learning techniques," in Proceedings of the ACL 2nd conference on Empirical methods in natural language processing Association for Computational Linguistics, 2002, vol. 10.