A

Major Project

On

Sentiment analysis using Telugu SentiWordNet

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

D S RITHIK (197R1A05K3)

ANKIT PATEL (197R1A05J5)

BANDARU SWAPNIL(197R1A05J7)

Under the Guidance of

J NARASIMHARAO

(Associate Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi)

RecognizedUnder Section 2(f) & 12(B) of the UGC.1956, Kandlakoya (V), Medchal Road, Hyderabad,

2019-202

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 Telugunews 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.

EXISTING SYSTEM:

For the sentiment analysis there is no such proper existing system. The existing system detects only that sentence is positive or negative .The existing system does not provide ratio of positive, negative and neural sentence. The existing system has many flaws which are covered in the proposed system. The Existing system was not accurate up to the mark and users were find difficult to get which one is positive and which one is negative sentence. The existing system failed to detect the sentiment from sentence which is positive and which is negative due to which many confusion arises. 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.

DISADVANTAGES:

1. The existing system detects only that sentence is positive or negative sentence.
2. The existing system does not provide ratio of positive, negative and neural sentence.

To avoid all these limitations and make the work more accurate the system needs to be implementedefficiently

PROPOSED SYSTEM:

Here we build an application for detecting positive or negative sentences from Telugu sentences, this detection consists of two parts in which using first part we can detect objective or subjective from sentences and if objective words appear in the neutral list of SentiWordNet then that sentence will be consider as Neutral, if words not appear in SentiWordNet Neutral list then sentence words will check inside positive and negative list of SentiWordNet, if sentence words found in positivelist then sentence will be consider as positive otherwise negative. 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.

ADVANTAGES:

The system is very simple in design and implementation. The system requires very low system resources and the system will work in almost all configurations. It has got the following features.

1. This help in analysis of positive/negative/neutral sentence.
2. If sentences contains words from both positive and negative list then we take ratio of both positive and negative words and if positive ratio higher then sentence will be consider as positive else negative.

HARDWARE REQUIREMENTS:

For the development of the application, the following are the Hardware

Requirements:System Architecture : 64-bit X86 with windows or Linux

CPU : Intel core 2 Quad CPU Q6600 @2.40GHz or greater

RAM : 4GB or greater

SOFTWARE REQUIREMENTS:

Operating system: Windows 7,8,10 ; MacOS

LinuxCoding Languages :Python 3.7 version

Tools & IDE’s :Jupiter, Chrome browser , Anaconda

CONCLUSION:

In this project we build an application for detecting positive or negative sentences from Telugu sentences, this detection consists of two parts in which using first part we can detect objective or subjective from sentences and if objective words appear in the neutral list of SentiWordNet then that sentence will be consider as Neutral, if words not appear in SentiWordNet Neutral list then sentence words will check inside positive and negative list of SentiWordNet, if sentence words found in positive list then sentence will be consider as positive otherwise negative

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

NOVELTY:

The novelty of the projects is detecting positive or negative sentences from Telugu sentences, this detection consists of two parts in which using first part we can detect objective or subjective from sentences and if objective words appear in the neutral list of SentiWordNet then that sentence will beconsider as Neutral, if words not appear in SentiWordNet Neutral list then sentence words will checkinside positive and negative list of SentiWordNet, if sentence words found in positive list then sentence will be consider as positive otherwise negative.

This project take help of available Telugu SentiWordNet to perform sentiment analysis 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.