



SENTIMENT ANALYSIS USING TELUGU SENTIWORDNET

1. Soumya Sunkapaka, 2. Dr. Nageshwar V

Department of Electronics and Instrumentation, VNR Vignan jyothi institute of technology,
soumyasunkapaka50@gmail.com

Associate Professor, Department of Electronics and Instrumentation, VNR Vignan jyothi institute of technology,
nageshwar_v@vnrvjiet.in

ABSTRACT: Sentiment analysis in dialects with few assets and territorial dialects has turned into another area in regular language handling lately. Scientists are more keen on sorting out how individuals feel in Indian dialects like Hindi, Telugu, Tamil, Bengali, Malayalam, etc. Supposedly, no nitty gritty work has been finished on Indian dialects so far in light of the fact that there aren't an adequate number of marked informational indexes. In this work, we recommended that Telugu SentiWordNet be utilized to break down the mind-set of Telugu news words in two stages. In the first place, it recognizes the meaning of subjectivity, which is the way words are marked as abstract or goal. Since they have no profound worth, objective words are treated as though they have no close to home estimation. Then, Feeling Order was finished, where the abstract sentences were additionally arranged into good and pessimistic expressions. With the ongoing Telugu SentiWordNet, our recommended framework gets a predisposition grouping accuracy of 74% and a temperament characterization exactness of 83%.

Keywords – *Natural Language Processing, Sentiment Analysis, Telugu, SentiWordNet, News sentences.*

DOI: 10.31838/ecb/2023.12.9.171

INTRODUCTION

In natural language processing (NLP), sentiment analysis is a strategy for sorting out how an individual feels about an item, film, occasion, piece of information, an association, and so on by taking a gander at how they discuss it. [1]. The primary objective of mind-set examination is to sort out which side of a contention a piece of text is on in a given report. Extremity can be either great or negative, or it tends to be nonpartisan. Message can be examined for its tone in three ways: at the sentence level, the archive level, and the perspective level. line-level examination takes a gander at the extremity worth of each line in a text. text-level investigation takes a gander at the entire text to sort out the extremity number. In outlook level test, it shows all-inclusive-saying course of each frame of reference in a quotation. After Hindi, Telugu is the second uttered in dialect in India. Telugu is the fifteenth uttered in accent on the earth, as per Ethnologue. There are 85 heap local speakers of Telugu everywhere the sphere [2]. There are many e-papers that print information usually in the Telugu language, like Eenadu, Sakshi, Andhrajyothy, Vaartha, and Andhrabhoomi, thus. SentiWordNet is a discussion remark that was fashioned definitely to help requests for spirits test and appraisal excavating [3]. "SentiWordNet is the aftereffect of WordNet's synsets

being naturally clarified for energy, pessimism, and impartiality," say Esuli and Sebastiani [3]. Every synset has three numbers: pos(s), neg(s), and obj(s), which mean "positive," "negative," and "objective," or "neutral," individually. There are various mind-set investigators for the English language [4-8], however not much has been finished for Indian dialects [9-20]. The fundamental explanation is on the grounds that there aren't an adequate number of apparatuses in Indian dialects.

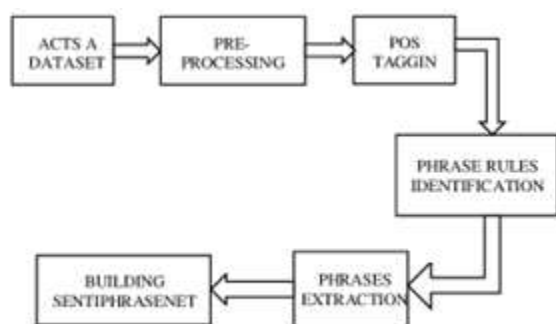


Fig. 1: A rule-based approach

Fig.1: Example figure

Message can be dissected for its tone in three ways: at the sentence level, the report level, and the viewpoint level. line-level examination takes a gander at the extremity worth of each line in a text. text-level examination takes a gander at the entire text to sort out the extremity number. In perspective level examination, it shows the word-by-word direction of each and every viewpoint in a text. Sentiment analysis in dialects with few assets and territorial dialects has turned into another area in normal language handling lately. Specialists are more keen on sorting out how individuals feel in Indian dialects like Hindi, Telugu, Tamil, Bengali, Malayalam, etc. Apparently, no definite work has been finished on Indian dialects so far on the grounds that there aren't an adequate number of marked informational collections. We proposed

utilizing Telugu SentiWordNet to break down the temperament of Telugu news words in two stages. In the first place, it recognizes the meaning of subjectivity, which is the way words are named as emotional or objective.

1. LITERATURE REVIEW

Sentiment Analysis and Opinion Mining:

Suppositions are at the focal point of nearly all that we do and hugely affect how we act. How others see and judge the world hugely affects what we accept and how we see reality, as well as the choices we make. Along these lines, when we want to pursue a decision, we frequently ask others their thought process. This is valid for individuals, yet additionally for organizations and different gatherings. Feeling investigation and assessment mining both gander at conclusions and connected thoughts like sentiments, evaluations, mentalities, and feelings. The field began and is developing rapidly simultaneously as virtual entertainment Online, similar to audits, discussion talks, websites, microblogs, Twitter, and informal communities. This is on the grounds that, without precedent for mankind's set of experiences, we have an immense measure of one-sided information put away in computerized structures. Starting from the start of the 21st hundred years, mind-set examination has become one of the most famous areas of concentrate in normal language handling. In data mining, Web mining, and text mining, it is likewise concentrated on a ton. Since it is so critical to business and society all in all, it has moved from software engineering to the board science and sociology. In the beyond couple of years, organizations that arrangement with mind-set research have additionally developed. There are presently a great deal of new organizations. A great deal of large organizations have constructed their own

abilities in-house. Frameworks that gander at individuals' sentiments have been utilized in pretty much every business and social region.

SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining :

In this paper, we show SENTIWORDNET 3.0, a word reference that was made explicitly to help applications for mind-set order and assessment mining. SENTIWORDNET 3.0 is a refreshed type of SENTIWORDNET 1.0, a word reference that can be utilized for research and is currently authorized to in excess of 300 examination gatherings and utilized in an extensive variety of exploration projects all over the planet. Both SENTIWORDNET 1.0 and 3.0 are made by marking all WORDNET synsets in a flash in light of how positive, negative, or nonpartisan they are. SentiWordNet 1.0 and 3.0 are different on the grounds that (a) they explain various variants of WORDNET (WORDNET 2.0 and 3.0, separately) and (b) the calculation used to naturally clarify WORDNET presently incorporates an irregular walk step for refining the scores, notwithstanding the semi-directed learning step that was utilized previously. In this article, we discuss SENTIWORDNET 3.0, with an extraordinary spotlight on the way things are superior to form 1.0 as far as (b).

Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews:

This paper shows a basic unaided learning strategy for arranging surveys into those that are proposed (approval) and those that aren't (disapproval). The typical importance course of the sentences in a survey that contain descriptive words or modifiers lets us

know how to characterize it. At the point when a word has great affiliations (like "inconspicuous subtleties"), it has a positive semantic direction. At the point when it has terrible affiliations (like "extremely dismissive"), it has a negative semantic direction. In this review, the pertaining to syntax route of an verbalization is establish by deducting the joint dossier between the verbalization and "first-rate" from the accepted data between the verbalization and "weak." When a abundant portion of mandate in a survey are excellent, it is pronounced that the audit is "submitted." When tried on 410 audits from Epinions from four distinct classifications (vehicles, banks, motion pictures, and excursion puts), the recipe is 74% exact overall. The surveys of vehicles are 84% precise, while audits of movies are just 66% exact.

Thumbs up? Sentiment Classification using Machine Learning Techniques :

We contemplate the issue of arranging papers not by what they are about, but rather by how they affect us, such as sorting out whether or not a survey is fortunate or unfortunate. Utilizing film surveys as information, we find that ordinary ML techniques are plainly better compared to baselines made by people. In any case, the three ML strategies we utilized (Naive Bayes, maximum entropy classification, and support vector machines) don't get along nicely at ordering opinions as they do at placing things into points. Eventually, we take a gander at the things that make it harder to group individuals' sentiments.

A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts:

Feeling examination attempts to sort out what perspective is behind a piece of text. One way this is done is by giving a film survey "approval" or "disapproval" rating. We propose another ML strategy that utilizes text-order methods on the abstract pieces of the record to sort out the course of the inclination. You can involve effective strategies for finding least slices in charts to take out these parts. This makes it a lot more straightforward to add cross-sentence relevant cutoff points.

3. METHODOLOGY

In this paper, the creator utilizes SentiWordNet to sort out whether Telugu sentences are positive or negative. This discovery has two sections. In the initial segment, we can sort out whether a sentence is evenhanded or emotional. On the off chance that goal words show up in the nonpartisan rundown of SentiWordNet, the sentence is viewed as unbiased. In the event that goal words don't show up in the impartial rundown, then the words in the sentence are checked against the positive and negative records.

Benefits:

1. In the event that a sentence has words from both the positive and negative records, the proportion of positive to negative words is utilized to conclude whether the sentence is positive or negative. If the proportion of positive to negative words is higher, the sentence is positive.

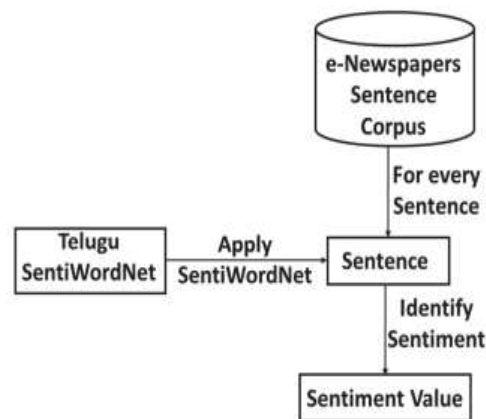


Fig.2: System architecture

4. IMPLEMENTATION

Data Collection & Annotation:

In this paper, data was captured from the Telugu e-Papers Eenadu, Sakshi, Andhrajyothy, Vaartha, and Andhrabhoomi. These are notable documents in Telugu-speaking states like Andhra Pradesh and Telangana. Our revelation variety has 1400 Telugu conversation from all of the e-Papers, as we pronounced earlier, from December 1 to December 31, 2016.

SentiWordNet for Sentiment Analysis:

SentiWordNet is a word reference of sentiments that connects the sensations of each word to its synset. SentiWordNet resembles Wordnet in addition to data about how individuals feel about words. In this review, we did the mind-set examination utilizing Telugu Senti-WordNet [12-14]. This SentiWordNet has four records, one for each kind of word: negative, positive, impartial, and muddled. Each document's words are set apart with labels for five grammatical features:

thing, action word, descriptor, modifier, and phrasal action word.

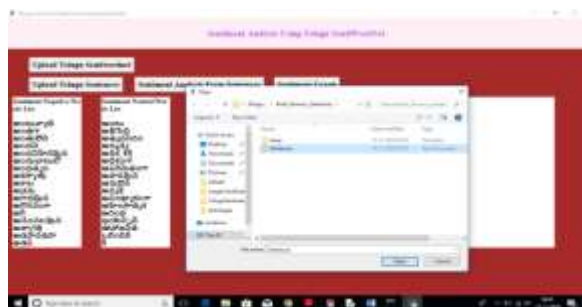
5. EXPERIMENTAL RESULTS



On the above screen, click the "Upload Telugu SentiWordNet" button to stack words from the SentiWordNet information base, for example, "neutral," "positive," and "negative."



On the screen over, every one of the three records are perused from SentiWordNet and displayed in isolated message regions. Look down the text region to see the entire rundown. Presently, click the "Upload Telugu Sentences" button to transfer the sentences record.



On the screen above, you can share a document of words. At the point when you do, you'll see the screen beneath.



Presently, click the "Sentiment Analysis From Sentences" button to figure out how each line causes you to feel.



On the screen above, we can see each sentence in the fourth message region. Underneath each sentence, we can check whether it is great or negative, and we can likewise see our number. Look down the text region to see the lines in general and sentiment.



On the screen above, we likewise saw the exactness score. Presently, click the "Sentiment Graph" button to see a chart of the quantity of positive and negative words.



On the above diagram, the x-hub shows the sort of sentence as "all out," "positive," or "negative," and the y-pivot shows the number.

6. CONCLUSION

For NLP tasks like POS tagging, desire study, snark study, textbook summary, etc., it's hard to find described datasets in Telugu. There aren't many samples accompanying annotations in this place style. This paper uses the Telugu SentiWordNet database to analyse the attitude of dispute from Telugu e-Newspapers. The submitted arrangement for analysing belief has an veracity of 74% for classifying belief and 83% for classifying belief in the area of news data.

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 Fabrizio Sebastiani, "Senti- WordNet 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?: sentiment 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

[6] Pang Bo and Lillian Lee. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in Proceedings of the 42nd annual meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2004.

[7] Hatzivassiloglou, Vasileios and Kathleen R. McKeown, "Predicting the semantic orientation of adjectives," in Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 1997.

[8] Taboada and Maite, "Lexicon-based methods for sentiment analysis," Computational linguistics, 2011, pp. 267-307.

[9] Das, Amitava and Sivaji Bandyopadhyay, "Sentiwordnet for bangla," Knowledge Sharing Event-4: Task 2, 2010.

- [10] Das, Amitava and S. Bandyopadhyay, "Dr sentiment creates Senti- WordNet (s) for Indian languages involving internet population," in Proceedings of Indo-wordnet workshop, 2010.
- [11] Das, Amitava and Sivaji Bandyopadhyay, "SentiWordNet for Indian languages," in Asian Federation for Natural Language Processing, China, 2010, pp. 56-63.
- [12] Das Amitava and Sivaji Bandyopadhyay, "Dr Sentiment knows everything!" in Proceedings of the 49th annual meeting of the association for computational linguistics, human language technologies, systems demonstrations, Association for Computational Linguistics, 2011.
- [13] Das Amitava and Bjrn Gambek, "Sentimantics: conceptual spaces for lexical sentiment polarity representation with contextuality," in Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, Association for Computational Linguistics, 2012.
- [14] D Das, S Poria, CM Dasari and S Bandyopadhyay, "Building resources for multilingual affect analysis A case study on Hindi, Bengali and Telugu," Workshop Programme, 2012.
- [15] BG Patra, D Das, A Das and R Prasath "Shared task on sentiment analysis in Indian languages (SAIL) tweets-an overview," in International Conference on Mining Intelligence and Knowledge Exploration, Springer International Publishing, 2015, vol. 9468.
- [16] Kumar S.S., Premjith B., Kumar M.A. and Soman K.P, "AMRITA CEN-NLP@ SAIL2015 Sentiment analysis in Indian Language using regularized least square approach with randomized feature learning," in International Conference on Mining Intelligence and Knowledge Exploration, Springer International Publishing, 2015, vol. 9468.
- [17] SS Prasad, J Kumar, DK Prabhakar and S Pal, "Sentiment Classification: An Approach for Indian Language Tweets Using Decision Tree," in International Conference on Mining Intelligence and Knowledge Exploration, Springer International Publishing, 2015, vol. 9468.
- [18] Sarkar, Kamal and Saikat Chakraborty, "A sentiment analysis system for Indian language tweets," in International Conference on Mining Intelligence and Knowledge Exploration, Springer international Publishing, 2015, vol. 9468.
- [19] Venugopalan Manju and Deepa Gupta, "Sentiment Classification for Hindi Tweets in a Constrained Environment Augmented Using Tweet Specific Features," in International Conference on Mining Intelligence and Knowledge Exploration, Springer International Publishing, 2015, vol. 9468.
- [20] SS Mukku, N Choudhary and R Mamidi, "Enhanced Sentiment Classification of Telugu Text using ML Techniques," in 25th International Joint Conference on Artificial Intelligence, 2016.