Journal of Advanced Research in Embedded System
Volume 7, Issue 3&4 - 2020, Pg. No. 29-31
Peer Reviewed Journal

Article

Development of a Natural Language Processing System using the concepts of Machine Learning

Tian Jipeng[1], Mahesh B Neelagar[2], Achyutha Prasad N[3], Rekha VS[4], TC Manjunath[5]

[1]Professor, Department of Computer Science and Engineering, No. 41, Zhongyuan Road (M), Zhongyuan University of Technology, Henan, China.
[2]Assistant Professor, Department of Electronics and Communication Engineering, PG Centre-VLSI Design & Embedded Systems, Centre for Post-Graduate Studies, VTU Visvesvaraya Technological University, Macche, Jnana Sangama, Belagavi, Karnataka, Andhra Pradesh, India.
[3]Research Scholar, Assistant Professor, Department of Computer Science Engineering, Sri Siddhartha Institute of Technology, Siddhartha Univ, Maralur, Tumkur, Karnataka, Andhra Pradesh, India.
[4]Assistant Professor and Research Scholar (VTU), Department of Computer Science Engineering, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, Andhra Pradesh, India.
[5]Ph.D. (IIT Bombay), Sr. Member IEEE, Fellow IETE, Fellow IE, Chartered Engineer Professor and Head, Electronics Communication Engineering, Department Dayananda Sagar College of Engineering, Bangalore, Karnataka, Andhra Pradesh, India.

I N F O

Corresponding Author:
Achyutha Prasad N, Department of Computer Science Engineering, Sri Siddhartha Institute of Technology, Siddhartha Univ, Maralur, Tumkur, Karnataka, Andhra Pradesh, India.
E-mail Id:
achyuthaprasadn@ssit.edu.in
Orcid Id:
https://orcid.org/0000-0002-2356-4146
How to cite this article:
Jipeng T, Neelagar MB, Prasad AN et al. Development of a Natural Language Processing System using the concepts of Machine Learning. *J Adv Res Embed Sys* 2020; 7(3&4): 29-31.

Date of Submission: 2020-08-28
Date of Acceptance: 2020-09-03

A B S T R A C T

Sentiment Analysis is a branch of Natural Language Processing, which can be described as the process of fortitude of the emotional tone signified by a series of words, which are used to discern the opinion of the writer. In this paper, we introduce the development of Methodology for the Sentiment Analysis of Kannada Movie Reviews using Machine Learning utilizing the concepts of Natural Language Processing (NLP). The model results show the effectualness of the method created using programing skills.

Keywords: Sentiment Analysis, Kannada, Reviews, NLP, Machine Learning

Organization of the Paper

The paper is planned into the five sections: Section I contains the Introduction, Section II is about the Related Research Work done in the past, Section III provokes the research areas we have focused on, Section IV describes our Methodology, Section V contains the Employment and Results of our observed investigate and finally, Section VI details the final remarks and future scope of the paper.

Introduction

Natural Language Processing (NLP) is an area of artificial intelligence that investigates the use of computers to identify and influence natural language text or speech to

do useful things. Sentiment analysis is a well- known area in the field of Natural Language Processing. Given a set of texts, the purpose is to determine the polarity of that text. Research has been performed over the years on numerous methods, benchmarks, and resources of sentiment analysis and opinion mining.

Related Work

In the modern-day technological world, automation plays an especially important role in the human life varying from domestic applications to the industrial uses. This automation makes use of various technological devices such as machines, computers & its accessories, etc. which could be used by the humans for various applications. In this context, a review of the explanation based Natural Language Processing System using semi-supervised bootstrapping, ML approaches of Support Vector Machine (SVM) and Random Forest (RF) methods is being produced. A brief insight into the design and development, i.e., the work done by various authors till date are presented here in the form of an extensive literature survey.[1-10]

Proposed Methodology

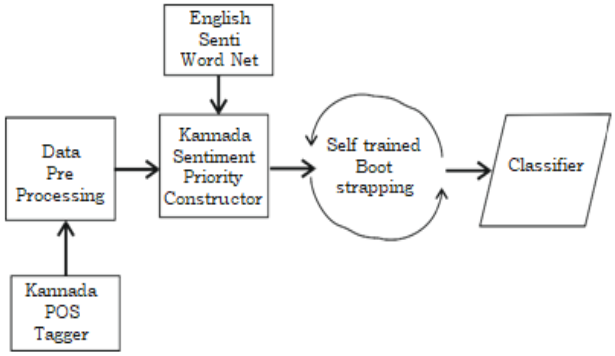

Figure 1.Block Diagram

The process, as shown in the flowchart, is conceived by implementation of the following steps: (Figure 1).

- Data Collection – Collation of Kannada movie assessments from different sources.
- Data Pre-processing – Normalization of collected data to formulate it for classification.
- Lexicon Construction – Use of SentiWordNet and POS Tagging for segmentation into positive, neutral, and negative.
- Construction of Training Dataset – Labelling using self-trained bootstrapping.
- Testing – Presentation assessment and contrast using the machine learning classifiers SVM and Random Forest.

Implementation and Result

We have utilized the techniques of classification using SVM and Random Forest in both positive and negative assessment sets and have achieved results as shown in Table 1. To calculate the performance of the two classifiers with the created dataset, we use execution metrics. The parameters used to evaluate and compare the two methods are - Precision, Recall, F-measure, Accuracy. These parameters are directly proportional to the execution i.e., better the result, better the algorithm. These parameters are calculated using the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values. The accuracy of the classifiers of the Kannada reviews is assessed based on the subsequent metrics:

Precision: The ratio of correctly predicted positive observations to the total predicted positive observations;

$$P = \frac{TP}{TP + FP}$$

Recall (Sensitivity): The ratio of correctly predicted positive observations to all observations in actual class given by;

$$R = \frac{TP}{TP + FN}$$

Accuracy: The ratio of correctly predicted observation to the total observations, measures the closeness of the measured value to the known value & is given by;

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

F-measure: the weighted average of Precision and Recall & is given by;

$$F - measure = 2 \times \left(\frac{P * R}{P + R}\right)$$

Table 1.Empirical results of the experiment

Classifier	Metrics	Value
SVM	Precision	86.77
	Recall	87.99
	F-measure	88.64
	Accuracy	89.08
Random Forest	Precision	86.77
	Recall	87.97
	F-measure	88.64
	Accuracy	86.73

Conclusion

In this paper, we submitted two machine learning methods to discern user sentiments. As seen in Table 1, containing the empirical results of our experiment, Random Forest and SVM yield almost similar results, other than in the factor of accuracy, where Random Forest classification outperforms SVM approach. Sentiment Analysis in this domain aids the business websites and helps them classify Kannada reviews necessarily without any human interaction. This paper delivers implementation of methods that are intended to

assist the section of Kannada speaking readers who prefer Kannada over English or are unable to read English.

References

1. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. *Journal of Computational Science* 2011; 2(1): 1-8.
2. Tumasjan A, Sprenger TO, Sandner PG et al. Predicting elections with twitter: What 140 characters reveal about political sentiment. ICWSM, Washington, DC 2010; 10: 178–185.
3. Nagy A, Stamberger J. Crowd sentiment detection during disasters and crises. In: Proceedings of the 9th International ISCRAM Conference, Vancouver, Canada. 2012; 1–9.
4. Nair DS, Jayan JP, Rajeev RR et al. Senti Ma-Sentiment Extraction for Malayalam. 2014.
5. Pang B, Lee L. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval* 2008; 2(1&2): 1–135.
6. Das A, Bandopadaya S, Senti Word Net for Bangla, Knowledge Sharing Event -4: Task, Volume 2, 2010.
7. Yakshi Sharma, VeenuMangat, MandeepKaur, A practical Approach to Semantic Analysis of Hindi tweets", 1st International Conference on Next Generation Computing Technologies (NGCT-2015), Dehradun, India,Page No(677-680). 2015.
8. Vijayalaxmi F. Patil, Designing POS Tagset for Kannada, LDC-IL, CIIL Mysore.
9. Cross Language POS Taggersand other Tools for Indian Languages: An Experiment with Kannada using Telugu Resources by Siva Reddy.