Library Progress International
Vol.44 No. 3, Jul-Dec 2024: P. 14991-14999

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

Optimizing Sentiment Classification: A hybrid approach combining NLP & advanced ML techniques

[1,]Yedugani Akanksha, [2,]Mekala kamala

[1]Student-CSE, CMR College of Engineering and Technology, Hyderabad, Telangana, India akankshayedugani@gmail.com
[2]Associate Professor-CSE, CMR College of Engineering and Technology, Hyderabad, Telangana, India kamala.mekala@gmail.com
.

How to cite this article: Yedugani Akanksha, Mekala kamala (2024) Optimizing Sentiment Classification: A hybrid approach combining NLP & advanced ML techniques. *Library Progress International,* 44(3), 14991-14999.

Abstract

A cornerstone of numerous message mining applications is mood analysis. The writing has introduced an extensive variety of mood characterization procedures, including profound learning-based and conventional methodologies. Indeed, even while mood examination has been effectively applied in various business applications and huge achievements have been achieved utilizing the strategies currently being used, its precision might in any case be expanded. We tackle this theme by using a few Machine Learning (ML) and Deep Learning (DL) techniques along with NLP approaches like count vectorizers and stopwords. Casting a ballot classifiers, Long Short-Term Memory (LSTM), Recurrent neural networks (RNN), Gated recurrent units, random forests, Ada boost, stochastic gradient descent, k nearest neighbor, decision trees, multinomial naïve bayes, support vector machines, gradient boosting, multilayer perceptrons, and voting classifiers are among the calculations that are utilized. Of these, the democratic classifier has achieved the most elevated precision. also, utilizing dictionary to break down feeling utilizing extremity scores.

Keywords *– Flipping, lexicon embedding, sentiment analysis, text classification*

INTRODUCTION

Text is utilized in numerous PC based knowledge applications. Message mining methods, for example, opinion investigation are fundamental for connection examination and general assessment perception. There are issues with articulation, segment, and record assessment, as shown by input messages. This examination analyzes sentences. Assessment investigation is oftentimes utilized into message request. Most of message request approaches are applied to assessment examination. Typical ways incorporate subject displaying, profound learning, sack of-words (BOW), and word reference (or rule) approaches. In spite of the fact that assessment examination is generally utilized in corporate applications, there is opportunity to get better concerning accuracy. To address this, we utilize normal language handling procedures, for example, count vectorizer, stopword launch, and ML and DL calculations.

Among calculations, for example, Ada help, Stochastic Point Fall, k Closest Neighbor, choice tree, multinomial gullible bayes, support vector machine, multi-facet perceptron, LSTM, RNN, Gated Intermittent Unit, and CNN, the democratic classifier has the most significant level of precision. Utilizing language to arrange feelings utilizing furthest point scores

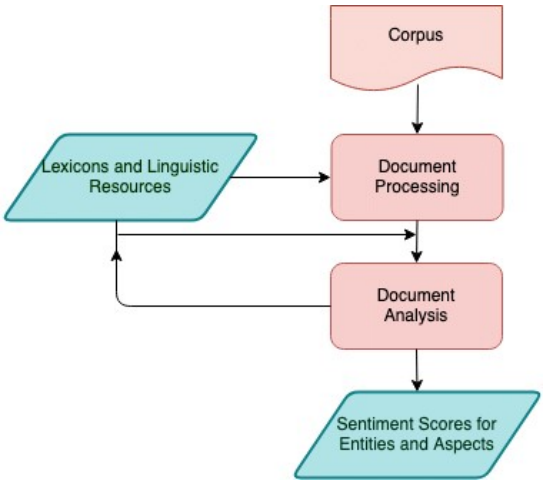

Fig.1: Example figure.

Machine Learning:

Artificial Intelligence is a subset of human insight, or the degree to which robots can mirror human way of behaving. Perplexed issues like people are tended to by modernized thinking structures. Man-made consciousness empowers a PC program to concentrate and concentrate significance from information in a semi-recursive way. Joining a few modalities is conceivable.

Deep Learning:

The deep learning field of computer based intelligence utilizes counterfeit brain organizations. Layers of neurons are utilized by a artificial neural network (ANN) to dissect and gain from input information.

In a completely connected deep mind engineering, something like one secret layer comes after a data layer. Every neuron gets data from the primary layer. One layer's result turns into a commitment to the neurons in the layer beneath it, etc, until the last layer creates the result of the association. Nonlinear changes in the layers of the mind network empower the association to procure confounded portrayals of approaching data.

Natural Language Processing:

NLP empowers PCs to grasp composed and communicated in human language. For PC based knowledge, it is pivotal.

NLP shows to computers typical language insight by various strategies. Text information is coordinated and cleaned for PC examination by data association. Preprocessing improves calculation cordial text elements and renders data available. A couple of techniques are accessible, like the going with

Tokenization: Tokenization substitutes a token for touchy information. Tokenization is a typical strategy utilized by portion exchangers to get Visa data.

Eliminate stop words: After normal catchphrases are taken out, the most considerate words that give the most data about the text are held.

Standard language handling for extraction, machine interpretation, text association, and different purposes.

LITERATURE REVIEW

Enhance gated graph neural network with syntactic for sentiment analysis

Examining the field of feeling examination is extremely useful for investigation into business, governmental issues, and different regions. The Graphic Neural Network (GNN) has drawn in a ton of interest of late in the field of normal language handling. Literary portrayals of chart structures are significant, particularly for GNN opinion investigation. In this field, two primary strategies have been created: the syntactic procedure, which considers syntactic data, and the word co-event strategy, which utilizes worldwide word co-event information. This review presents another strategy utilizing an updated diagram brain network model with syntactic edges that joins additional printed data from the two recently portrayed approaches. To incorporate worldwide word co-event with grammar, this model joins syntactic example data with the word co-event procedure to make the message diagram structure. It utilizes a gated diagram brain network plan to gain significant message bits of knowledge by removing message portrayals and using consideration processes. Tests led on different datasets show that the recommended model outflanks current calculations in opinion order assignments.

Sentiment Analysis of Financial Texts Based on Attention Mechanism of FinBERT and BiLSTM

A developing corpus of work is utilizing refined message based opinion calculations to work on our cognizance of social money among monetary market players. Monetary feeling investigation models are incorrect because of conventional opinion examination procedures that regularly depend on fixed emblematic encoders for message portrayal extraction. The Message Opinion Examination Model (BBiLSTM-Consideration) gives a response to these issues.

With an emphasis on finance, the model progressively separates relevant data from remarks utilizing FinBERT, an element extractor. By utilizing a few consideration processes and BiLSTM, separating opinion from monetary discourse is capable. Trial and error evaluations use datasets pertinent to remarks in the monetary business. Discoveries show improved execution and exactness, with a F1-score of 0.8068 and a precision of 79.33%.

Sentiment Infomation based Model For Chinese text Sentiment Analysis

Chinese opinion examination, as utilized in Regular Language Handling, includes evaluating the overall feeling extremity that is addressed in Chinese text. Despite the fact that profound brain network models have progressed, prior renditions were viewed as shaky since they couldn't precisely catch the nuances of feeling in undertakings including opinion examination. This paper presents an assessment mindful brain network model planned explicitly for Chinese feeling examination to close this hole. LSTM parts and a Transformer encoder make up the model engineering. Feeling acknowledgment in Chinese writing might be robotized by using an individual jargon written in Chinese. Feeling Surmising Brain Model, or SINM, utilizes a half breed task learning strategy to comprehend opinion designs and perceive human articulations simultaneously. With an emphasis on text based opinion perception, SINM sift through unessential material when profound information is available. In tests utilizing the ChnSentiCorp and ChnFoodReviews datasets, SINM fared better compared to most of different strategies regarding precision and execution.

ABSADM - Aspect-Based Sentiment Analysis using Distance Matrix:

Exhaustive assessment mining requires aspect-based sentiment analysis (ABSA) since messages often have numerous angles introduced in disconnection. Powerful ABSA expects that feeling articulations match their particular objectives. Despite the fact that cutting edge brain models as often as possible utilize consideration procedures to learn affiliations, they probably won't be adequate to deal with vague opinion phrases like "great" or "terrible." This paper recommends a two-CNN-based model for perspective level feeling classification to handle this issue. This model first tracks down viewpoint articulations in a specific setting, and afterward it searches for feeling articulations around those perspective articulations to gather the inclination related with that angle. To create a perspective guide, the primary CNN takes viewpoint articulations' area data. In spite of the fact that angles and sentiments don't straightforwardly correspond, the viewpoint guide can in any case have lacking management. This angle map is involved continuously CNN to sort feeling for the objective perspective. SemEval 2016 Errand 5 assessments show that the recommended model performs better compared to benchmark models, exhibiting its cutting edge execution on the dataset.

Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary

Since some feeling words are missing or catchphrases specific to a specific area are prohibited, customary opinion

investigation utilizing a standard jargon regularly fizzles. Lower exactness results from polysemic feeling words, which can communicate both good and pessimistic implications, making extremity identification more troublesome. To further develop feeling examination exactness, this study extends the opinion jargon to incorporate principal, field-explicit, and polysemic feeling ideas. Opinion values are processed by using an innocent Bayesian classifier to find out the setting of polysemic feeling terms. Opinion investigation is improved by using the lengthy jargon and feeling score rules. The methodology's exactness and feasibility are shown by trial discoveries, which stress the meaning of feeling acknowledgment in message examination.

METHODOLOGY

The ongoing engineering utilizes two-level LSTM. An enormous number of brief things with unadulterated assessment guidelines are at first named by annotators. Each model is set apart by one annotator. The following stage is to check a couple of message tests with blended assessment headings by various annotators. Second, we utilize two checked datasets to LSTM design to deal with incorporates and characterize feeling headings of message tests.

Disadvantages:

1. For unadulterated and composite inclination, a two-level Long Momentary Memory Model is utilized. The precision of this model is not exactly that of the recommended model.

2. Over 10% of Chinese datasets frequently contain mark mistakes. This blunder in judgment shows a significant feeling of disturbance.

We address the above depicted issues by using NLP procedures including stopwords, count vectorizers, and ML/DL calculations. Among calculations, for example, Ada help, Stochastic Incline Drop, k Nearest Neighbor, decision tree, multinomial blameless bayes, support vector machine, point helping, multi-layer perceptron, LSTM, RNN, Gated Irregular Unit, and CNN, the vote based classifier has the most significant level of precision. inspecting conclusions with an outrageous score while utilizing a word reference.

Advantages:

1. Incorporate useful lexical clues effectively.

2. Obtained the maximum accuracy on these three data corpora.

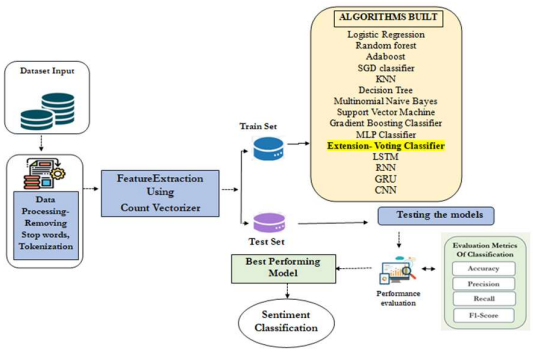

Fig.2: System architecture

MODULES:

To finish the task determined above, we fostered the modules displayed underneath.

- Data exploration: This module will be utilized to enter information into the framework.

- Processing: Utilizing this module, we will peruse information for handling.
- Splitting data into train & test: Information from this module will be parted into train and test classes.
- Model generation: Model building --> Logistic Regression --> Random Forest -->Adaboost --> SGD Classifier --> KNN --> Decision Tree --> Naive Bayes --> SVM --> MLP --> Gradient Boosting --> Voting Classifier --> RNN --> CNN -->LSTM --> GRU. Algorithms accuracy is calculated.
- User signup and login: At the point when you utilize this module, you should enlist and sign in.
- User input: Anticipated info will be created by utilizing this module.
- Prediction: final predicted shown

IMPLEMENTATION

ALGORITHMS:

Logistic Regression: Logistic regression decides the likelihood of an occasion happening, such as casting a ballot or not, in light of a bunch of free factors. With a likelihood as the result, the reliant variable's reach is 0 to 1. The connection between one ordinal, stretch, rate, or ward twofold element and no less than one evident free component might be portrayed and made sense of measurably utilizing calculated relapse.

Random forest: A classification strategy called random forests utilizes a few decision trees. It utilizes packing and component randomization to make every individual tree with an end goal to make an uncorrelated woods of trees whose board of trustees estimate is more precise than any one tree's.

AdaBoost Classifier: A meta-assessor known as an AdaBoost classifier first fits utilizing the first dataset as a classifier. At that point, it fits extra classifier copies on the comparing dataset, changing the heaps of cases that are erroneously ordered so the classifiers that follow focus more on testing circumstances.

SGD Classifier: For fitting direct classifiers and regressors to raised misfortune capabilities, similar to SGD is a clear yet staggeringly compelling method. The SGD Classifier is a SGD-advanced straight classifier (SVM, logistic regression, and so on). These are two separate ideas. SGD is an enhancement technique, though straight help vector machines, or strategic relapse, are AI models or calculations.

KNN: KNN is an essential calculation that protects each of the occurrences that as of now exist and uses a likeness measure to classify approaching information or cases. An information point is habitually classified utilizing the order of its neighbors. KNN chooses the quantity of models (K) that are nearest to the question and decisions in favor of the most well-known name (on account of arrangement) or midpoints the marks (on account of relapse) in view of the distances between each example in the information and the question.

DT: A decision tree is a diagram that involves a fanning strategy to show each conceivable outcome for a given information. Decision trees can be made manually or with the utilization of particular programming or a designs program. Decision trees can assist with centering conversations when a gathering needs to choose a decision.

MNB: In NLP, the MNB is a well known Bayesian learning strategy. The program decides the tag of a text, such an email or a paper article, by applying the Bayes hypothesis. For a given example, it computes the probability of each tag and returns the tag with the most noteworthy possibility.

SVM: A managed AI strategy for relapse and characterization is called Support Vector Machine (SVM). These are generally appropriate for characterization, regardless of whether we notice them as potential backslide issues. The objective of the SVM is to find a hyperplane in a N-layered space that accurately organizes the data centers. technique.

MLP: The multi-facet perceptron is another multi-facet counterfeit brain network approach (MLP). A solitary perceptron can take care of plainly straight issues, however it isn't very much adjusted for non-direct applications. MLP can be applied to these difficult circumstances. With regards to marked input order forecast undertakings, MLPs are a solid match. Relapse expectation issues that anticipate a genuine esteemed amount given a bunch of sources of info are likewise reasonable for them.

Gradient Boosting: Gradient boosting is a kind of AI supporting. It is predicated on the possibility that the ideal next model limits the general expectation blunder when joined with past models. To limit error, the key thought is to determine the planned outcomes for the following model.

Voting classifier: A voting classifier is a kind of AI assessor that makes forecasts by using the result of many base models or assessors that have been prepared. Casting a ballot choices for each assessor result can be utilized to shape collecting models.

RNN: A deep learning model intended to investigate and change over consecutive information inputs into explicit successive information yields is known as an recurrent neural network (RNN). Words, expressions, and time series information are instances of successive information. Successive parts of these kinds of information connect in view of complicated syntactic and semantic shows.

LSTM: A kind of recurrent neural network (RNN) that has the ability to learn and recollect data over the long haul is called long short-term memory (LSTM). This makes it a helpful strategy for time series guaging, discourse acknowledgment, and normal language handling.

CNN: A CNN is a kind of deep learning calculation network design that is generally utilized for pixel information handling and picture acknowledgment undertakings. While CNNs are the favored organization design for object distinguishing proof and acknowledgment, there are different assortments of brain networks utilized in profound learning. A deep gaining network engineering that gains from information straightforwardly is the convolutional brain organization, some of the time known as a ConvNet or CNN. With regards to distinguishing designs in pictures that might be used to separate between articles, classifications, and arrangements, CNNs are very valuable.

GRU: Kyunghyun Cho et al. presented gated recurrent units (GRUs) as a gating procedure for repetitive brain networks in 2014. Since it doesn't have a result entryway, the GRU works in much the same way to a LSTM with a neglect door however with less boundaries.

5. EXPERIMENTAL RESULTS

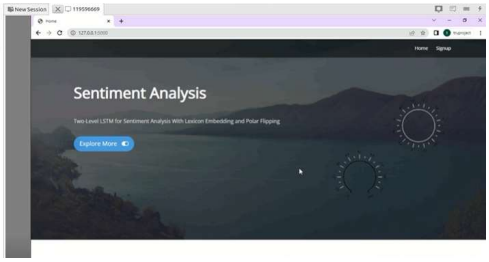

Fig.3: Home screen

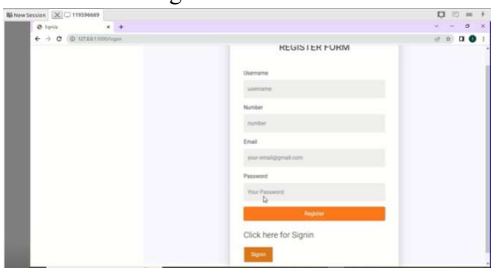

Fig.4: User signup

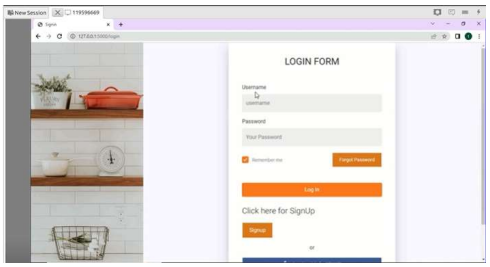

Fig.5: User signin

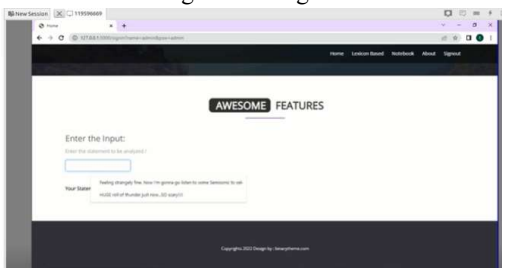

Fig.6: Main screen

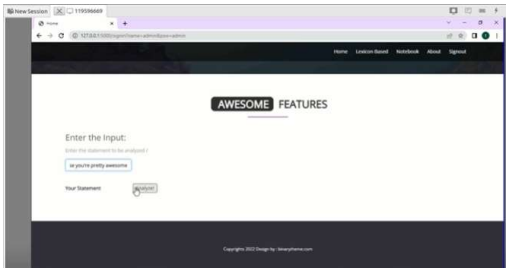

Fig.7: User input

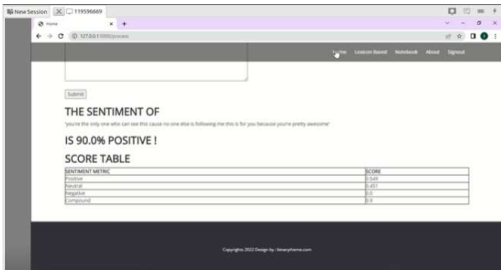

Fig.8: Prediction result

Algorithm	Accuracies
Logistic Regression	71.5
Random Forest	71.5
Ada Boost	69.3
Stochastic Gradient Descent	68.1
K-Nearest Neighbour	66.9
Decision tree	68.4
Multinomial Naïve Bayes	73.0
Support Vector Machine	69.6
Gradient Boosting	68.1
Multi Layer Perceptron	70.0
Voting Classifier	97
Long short term memory	69.4
Recurrent neural network	67.0
Gated recurrent unit	67.0
Convolutional Neural Network	67.0

Fig.9: Accuracy Scores of Various Models

6. CONCLUSION

This review closes by examining the challenges in feeling examination and focusing on making opinion order models with a serious level of accuracy.Consequently, the democratic classifier's exactness is expanded. Casting a ballot classifier has been chosen as the ideal strategy for feeling investigation location.

We mean to extend in future work by examining other datasets and building the model with state of the art apparatuses and techniques. By joining these parts and testing the outcomes on different datasets, it outflanks state of the art calculations, showing its handiness in functional text mining applications.

REFERENCES

[1] L. Niu, Q. Zheng and L. Zhang, "Enhance gated graph neural network with syntactic for sentiment analysis," 2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), Dalian, China, 2021.

[2] W. Cheng and S. Chen, "Sentiment Analysis of Financial Texts Based on Attention Mechanism of FinBERT and BiLSTM," 2021 International Conference on Computer Engineering and Application (ICCEA), Kunming, China, 2021.

[3] G. Li, Q. Zheng, L. Zhang, S. Guo and L. Niu, "Sentiment Infomation based Model For Chinese text Sentiment Analysis," 2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), Shenyang, China.

[4] S. R. Savanur and R. Sumathi, "ABSADM - Aspect-Based Sentiment Analysis using Distance Matrix," 2022 International Conference on Futuristic Technologies (INCOFT), Belgaum, India, 2022

[5] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng and X. Wu, "Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary," in IEEE Access.

[6] J. Wu, K. Lu, S. Su and S. Wang, "Chinese Micro-Blog Sentiment Analysis Based on Multiple Sentiment Dictionaries and Semantic Rule Sets," in IEEE Access, vol. 7, pp. 183924-183939, 2019

[7] W. Liu, G. Cao and J. Yin, "Bi-Level Attention Model for Sentiment Analysis of Short Texts," in IEEE Access, vol. 7, pp. 119813-119822, 2019,

[8] Z. Jianqiang, G. Xiaolin and Z. Xuejun, "Deep Convolution Neural Networks for Twitter Sentiment Analysis," in IEEE Access.

[9] M. Bouazizi and T. Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter," in IEEE Access, vol. 5.

[10] W. Meng, Y. Wei, P. Liu, Z. Zhu and H. Yin, "Aspect Based Sentiment Analysis With Feature Enhanced Attention CNN-BiLSTM," in IEEE Access, vol. 7, pp. 167240-167249, 2019,

[11] O. Wu, T. Yang, M. Li and M. Li, "Two-Level LSTM for Sentiment Analysis With Lexicon Embedding and Polar Flipping," in IEEE Transactions on Cybernetics, vol. 52, no. 5, pp. 3867- 3879, May 2022.

[12] B. Liu, Sentiment Analysis: Mining Opinions Sentiments and Emotions, Cambridge, U.K.:Cambridge Univ. Press, 2015.

[13] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining", Proc. Int. Conf. Lang. Resources Eval. (LREC), pp. 17-23.

[14] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, "Lexicon-based methods for sentiment analysis", Comput. Linguist., vol. 37, no. 2, pp. 267-307.

[15] Z. Teng, D.-T. Vo and Y. Zhang, "Context-sensitive lexicon features for neural sentiment analysis", Proc. EMNLP, pp. 1629-1638

[16] E. Cambria, S. Poria, A. F. Gelbukh and M. Thelwall, "Sentiment analysis is a big suitcase", IEEE Intell. Syst., vol. 32, no. 6, pp. 74-80, Nov./Dec. 2017.

[17] B. Wang and W. Lu, "Learning latent opinions for aspect-level sentiment classification", Proc. AAAI Conf., pp. 5537-5544, 2018.

[18] W. Zhao et al., "Weakly-supervised deep embedding for product review sentiment analysis", IEEE Trans. Knowl. Data Eng., vol. 30, no. 1, pp. 185-197, Sep. 2018.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory", Neural Comput., vol. 9, no. 8, pp. 1735-1780.

[20] K. Schouten, O. van der Weijde, F. Frasincar and R. Dekker, "Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data", IEEE Trans. Cybern., vol. 48, no. 4, pp. 1263-1275, Apr. 2018.

[21] M. Kamala , A.S. Srinath, Pranav Kumar, Arun Teja, "Facial Sentimental Analysis in Hospitals" in IJRASET, volume 12, March 2024.

[22] Nandireddy Srikanth Reddy, Mekala Kamala, "Interactive Chatbot with Songs Recommendation Based on Emotion" in IJIRSET, volume