e-ISSN: 2320-4230, p-ISSN: 2961-6085

Journal of Drug Discovery and Therapeutics

Available Online at www.jddt.in

CODEN: - JDDTBP (Source: - American Chemical Society)

Volume 11, Issue 06: 2023, 64-71

Evaluation of Natural Language Processing and Machine Learning

PV Ramana Murthy [1*] & Dr. Rakesh Kumar Giri [2]

[1] Research Scholar, Sunrise University, Alwar Rajasthan.

[2] Research Supervisor, Assistant Professor, Sunrise University, Alwar

Received: 03-08-2023 / Revised: 13-09-2023 / Accepted: 04-10-2023

Corresponding author: PV Ramana Murthy

Conflict of interest: No conflict of interest.

Abstract:

Natural language processing, or NLP, is a set of computer methods that are based on theories and are used to automatically understand and describe human conversation. **T**he main aim of the study is to Evaluation of Natural Language Processing and Machine Learning**.** Pre-processing of text features and pre-processing of numerical and category characteristics are the two primary components that make up the bulk of the data pre-processing procedure. In this study, a classification system for technical papers as well as an assessment strategy for document quality rankings are presented.

Keywords: Natural, Language, Pre-processing, Primary, Document, Numerical.

1. INTRODUCTION

Natural language processing, or NLP, is a set of computer methods that are based on theories and are used to automatically understand and describe human conversation. In the old days of punch cards and group processing, it could take up to seven minutes to analyze a sentence. Now, thanks to Google and other companies like it, millions of webpages can be analyzed in less than a second.

Computers can do many tasks related to natural language with the help of NLP. These tasks include analysis, part-of-speech (POS) tracking, machine translation, and conversation systems. Architectures and methods for deep learning have already made huge steps forward in areas like computer vision and pattern recognition. In line with this trend, new NLP study is focused more and more on using new deep learning techniques (see Figure 1). We've been using shallow models (like SVM and logistic regression) to solve NLP problems with machine learning for decades. These models were learned on very high-dimensional and sparse data. In the past few years, neural networks using dense vector representations have been doing better at many NLP jobs. This trend began because word embeddings and deep learning methods worked so well. Deep learning makes it possible for automatic feature representation learning at multiple levels. Traditional NLP systems, on the other hand, use a lot of hand-crafted features instead of machine learning. These kinds of hand-made features take a long time and are often not finished. Collobert et al. showed that a basic deep learning system does better than most

cutting-edge methods in a number of natural languages processing tasks, including POS tagging, named-entity recognition (NER), and semantic role labeling (SRL). Since then, many complicated methods based on deep learning have been suggested as ways to solve tough NLP problems.

1.2 CHALLENGES OF NATURAL LANGUAGE PROCESSING

Thanks to progress in artificial intelligence (AI) and machine learning, natural language processing (NLP) has come a long way in the last few years. Software that reads writing or recognizes speech has a hard time understanding real language because it has so many subtleties. This means that NLP still has a long way to go before it can really be said to understand human language. For instance, NLP systems have a hard time understanding humor, idioms, puns, and other types of non-literal language. Because they are taught on data sets that show these biases, they also tend to be biased against some groups of people, like women or people of color.There are many great things about Natural Language Processing (NLP), but there are also some issues and problems that need to be fixed.

2. LITERATURE REVIEW

Tsuruoka, Yoshimasa. (2019). Since deep learning methods came out a few years ago, the area of natural language processing (NLP) has made a lot of progress very quickly. Now, simple mixtures of general neural network models like recurrent neural networks and attention processes can do a wide range of NLP tasks, such as grammar parsing, machine translation, and summary. This paper gives a short outline of deep learning and the current NLP technology that is built on deep learning.

Vedantam, Vamsi. (2020). Deep Learning is one of the most important areas of AI to study over the next few decades. In the past few years, important areas of study and development have grown up in artificial intelligence, which has led to its evolution. The newest advances in Natural Language Processing have helped make machine translations, language models, speech recognition, and automatic text generation applications work well. This essay talks about the latest progress made in Natural Language Processing (NLP) using Deep Learning as well as some areas of NLP that people are excited to see grow in the coming years. The first part talks about the design of Deep Learning and Natural Language Processing methods. The second part shows how Deep Learning has changed NLP, and the last part concludes my piece by saying what you should remember from it.

Arkhangelskaya, E. & Nikolenko, S.. (2023). Deep learning has changed machine learning a lot in the last ten years. In many situations, neural network designs are now the best way to do things. In this study, we look at how deep learning can be used to solve problems in natural language processing (NLP). First, we take a quick look at the main ideas and models of deep learning, including some new developments that are especially important for natural language processing (NLP). Then, we look at how words are represented in different ways, showing how word embeddings can be used for sentences and paragraphs as well as how words can be broken down even more in character-level models. The last part of the survey is mostly about the different deep architectures that have been created or have become popular for natural language processing (NLP) tasks. These tasks include machine translation, sentiment analysis, answering questions, dialog and conversational models, and more.

Zhou, Di. (2021). In computer science and artificial intelligence, natural language processing is a big area of study. It looks into different ideas and methods that make it

possible for people and computers to talk to each other in normal language. In a broad sense, NLP can be broken down into the following areas: machine translation, conversation systems, text search, mood analysis, and semantic role labeling. Even though AI has made some progress in the conversation system recently, there is still room for improvement. This piece suggests a way to handle multiple conversations between a person and a machine. This way, the machine dialogue can take into account more of what was said before and after, and the behavior of the machine dialogue is more like real human dialogue.

Bharadiya, Jasmin. (2023). The goal of this study is to look into why transfer learning methods aren't used more in radio frequency machine learning and to suggest a custom classification for radio frequency uses. In this area, the goal is to make performance gains, better generalization, and cost-effective training data options possible. Methodology: A thorough review of all the available literature on transfer learning in radio frequency machine learning is used in this study's research plan. The researchers got relevant papers from trustworthy sources and read them to find patterns, trends, and new ideas. The main way the data was gathered was by reading and putting together current books. Finding the most important results and making a special classification for radio frequency uses were part of the data analysis. Findings: The study's results show that transfer learning methods aren't used very much in radio frequency machine learning.

3. METHODOLOGY

3.1 EVALUATION OF NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING TOOLS

3.1.1 Data Pre-processing

Pre-processing of text features and pre-processing of numerical and category characteristics are the two primary components that make up the bulk of the data pre-processing procedure. Different approaches are used, depending on the sort of method being utilized.

Text Pre-processing

In the dataset, most of the textual data is written in English. However, customers tend to submit their inquiries in the language spoken in the area, which results in some samples being written in languages other than English. Observations are also made on the remarks made by minor traders in the local languages. To make use of these data, a preliminary translation into English is carried out with the assistance of Amazon Translate, which is a text translation service that is based on the Neural Network and follows the encoder-decoder architecture. The attention mechanism that is explained here is used. A nation that is the origin of a disagreement is the one that identifies the original language.

Once it is complete, the text is cleaned up with the assistance of the tools that were discussed. The sequence of procedures for cleaning text is applied to all of the important text characteristics, which include free-text comments from an agent, a merchant, and a customer, predetermined replies from the merchant, a dispute cause, and a brand name. Regarding the algorithm that is being trained, the number of text cleaning steps is identified.

4. RESULTS

4.1 RESULTS AND DISCUSSION

All the findings from the experiments are shown in Table 4.1, which can be found here. Because the dataset is not evenly distributed, the balanced accuracy is used, and the precision, recall, and F1-Score are assessed for each class in a distinct manner.

Logistic Regression offers the best performance among the baseline classifiers that use BOW text representation. It achieves this performance in all three scenarios of the text, nontext, and combination features as input, with an accuracy of 71.5%, 62.3%, and 75.4%, respectively, during the first phase. The reason for this is because Logistic Regression and XGBoost, which is the second-best, are both retained for the trials that will take place during the second phase. This is done for the purpose of comparing the models. With an accuracy of 76.4% throughout the first phase, the DistilBERT model surpasses the baseline models that are based on the comments made by merchants and agents as text input. The expansion of the dataset during the second phase, on the other hand, does not result in an improvement in the performance of the models; rather, it diminishes for both the baseline and DistilBERT models, with the DistilBERT model achieving the highest score of 71.4%. Furthermore, the addition of additional comments from customers to the input does not result in a substantial improvement to the DistilBERT model, and in fact, it results in a deterioration in the performance of the baseline models.

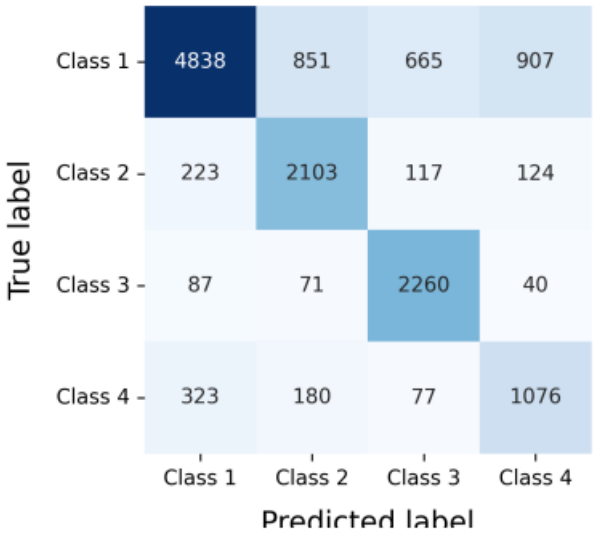
Table 4.1: balanced precision. Phase 2 is the expanded dataset, whereas Phase 1 is the original dataset. Text input is defined as follows: AM uses the comments from agents and merchants; C refers to customer comments; AMC refers to comments from all three parties; NT refers to non-text features alone; and combined refers to both text and non-text features.

Model	Phase 1	Phase 2	Phase 3
Logistic Regression$_{Softmax}$ (AM)	0.7147	0.7033	-
Logistic Regression$_{Softmax}$ (C)	-	0.5161	-
Logistic Regression$_{Softmax}$ (AMC)	-	0.7020	-
Logistic Regression$_{Softmax}$ (NT)	0.6226	0.6179	0.8361
Logistic Regression$_{Softmax}$ (Combined)	0.7543	0.7364	-
MultinomialNB (AM)	0.6493	-	-
MultinomialNB (NT)	0.5897	-	-
MultinomialNB (Combined)	0.7074	-	-
SVC$_{ovo}$ (AM)	0.6841	-	-
SVC$_{ovo}$ (NT)	0.6249	-	-
SVC$_{ovo}$ (Combined)	0.7362	-	-
XGBClassifier (AM)	0.7022	0.6731	-
XGBClassifier (C)	-	0.5074	-
XGBClassifier (AMC)	-	0.6722	-
XGBClassifier (NT)	0.7055	0.6847	0.8267
XGBClassifier (Combined)	0.7405	0.7216	-
DistilBERT (AM)	0.7637	0.7621	0.7844
DistilBERT (C)	-	0.5214	-
DistilBERT (AMC)	-	0.7620	-
DistilBERT (AM combined with raw NT)	0.7120	0.6988	-
DistilBERT (AM combined with processed NT)	0.7204	0.7104	0.8391
Separate NN (NT)	-	-	0.8449
DistilBERT (AM) + separate NN (NT)	-	-	0.8439
TOD-BERT (AM)	0.7636	-	-

The use of TOD-BERT does not contribute to an improvement in accuracy. 76.4% is the figure that was got from the remarks that were provided by the merchants and the agents. This is the reason why it will not be reviewed during the second part of the operation. Logistic Regression, XGBoost, and DistilBERT all have poor accuracy, with Logistic Regression having a 51.6% accuracy rate, XGBoost having a 50.7% accuracy rate, and DistilBERT having a 52.1% accuracy rate. The performance of the four baseline models using non-text characteristics as input ranges from 59.0%

for Naïve Bayes to 70.6% for XGBoost during the first phase. This variation is seen in the performance of the models. It has been noted that the accuracy of this collection of characteristics is diminishing during the second phase in comparison to the first phase. For Logistic Regression, it decreases by 0.5%, while for XGBoost, it decreases by 2.1%. Because of this, it does not affect the performance of Logistic Regression, but it does cause a slight decrease in the accuracy of XGBoost during the second phase, which is based on a mix of text and non-text data. Nevertheless, it is observable that the incorporation of non-text characteristics into the BOW text representations results in an improvement in the performance of all baseline models ranging from 1% to 6% for various models simultaneously throughout both stages. Adding non-text characteristics to the implementation of the DistilBERT model, on the other hand, does not result in any improvement in the accuracy of the model. However, throughout both periods, it drops to between 3 and 4 percent. When non-text features are further treated, DistilBERT performs better than it does when text characteristics are not handled.

(a) Confusion matrix for Logistic Regression



(b) Confusion matrix for the DistilBERT model

	precision	recall	f1-score	support
Class 4	0.47	0.63	0.54	1656
Class 1	0.91	0.63	0.74	7260
Class 3	0.69	0.91	0.79	2458
Class 2	0.62	0.85	0.72	2567
accuracy			0.72	13941
macro avg	0.67	0.75	0.70	13941
weighted avg	0.77	0.72	0.72	13941

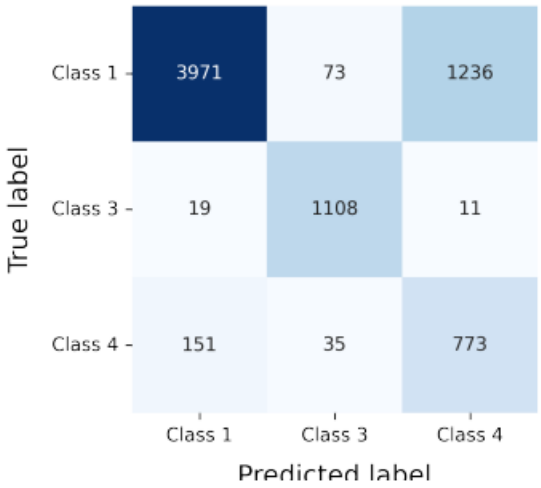
(c) Metrics' report for Logistic Regression

	precision	recall	f1-score	support
Class 1	0.88	0.67	0.76	7261
Class 2	0.66	0.82	0.73	2567
Class 3	0.72	0.92	0.81	2458
Class 4	0.50	0.65	0.57	1656
accuracy			0.74	13942
macro avg	0.69	0.76	0.72	13942
weighted avg	0.77	0.74	0.74	13942

(d) Metrics' report for the DistilBERT model

Figure 4.1 Metrics for the best baseline and BERT models: Logistic Regression based on combined features and DistilBERT based on text input. Both are based on agents' and merchants' comments.

The further in-depth examination of the data is carried out with the assistance of confusion matrices for the best baseline and BERT models, which are represented in Figures 4.2 (a) and (b). As far as "Class 2" and "Class 3" are concerned, both models provide outcomes that are satisfactory. However, a significant portion of the samples that were incorrectly categorized may be seen for the "Class 1" category that is the biggest. When it comes to "Class 4", the issue is much more dire since the greatest number of samples that have been incorrectly categorized are anticipated to be "Class 1." An further analysis into the metrics for each class, which are shown in Figures 4.2(c) and (d), reveals that the values of the accuracy, recall, and F1-Score are higher for two classes, namely "Class 2" and "Class 3." On the other hand, the metrics are much lower for one of the four classes, which is "Class 4."



	precision	recall	f1-score	support
Class 1	0.96	0.75	0.84	5280
Class 3	0.91	0.97	0.94	1138
Class 4	0.38	0.81	0.52	959
accuracy			0.79	7377
macro avg	0.75	0.84	0.77	7377
weighted avg	0.88	0.79	0.82	7377

Figure 4.2 Confusion matrix and metrics' report for the Neural Network combining representations from the DistilBERT model and the Neural Network based on non-text features

By modifying the configuration of the task and including metadata into non-text features during the third phase, it is possible to achieve an increase in the accuracy of the baseline models that include non-text features of up to 21% for Logistic Regression and 14% for XGBoost. With a level of accuracy of 84.5%, the Neural Network that was trained independently on non-text characteristics provides superior performance over the baseline models. Additionally, the performance of the DistilBERT model showed growth of up to 78.4%. The performance of the DistilBERT model is improved by 5.5% as a result of the addition of non-text characteristics because of the combined model that was introduced in the first phase and re-trained for the new assignment. The Neural Network, on the other hand, cannot be outperformed by it since it is based on nontext characteristics. When compared to previous models, the newly built model that combines the DistilBERT model with the Neural Network based on non-text data does not result in a gain in performance. Even though the metrics for "Class 1" and "Class 3" have been improved, the problem with the high rate of misclassified samples of "Class 1" and "Class 4" continues to exist, as shown by the study of the confusion matrix and the report of metrics in Figure 4.2.

5. CONCLUSION

In this study, a classification system for technical papers as well as an assessment strategy for document quality rankings are presented. The classification system makes use of machine learning techniques. An effort might be made to further enhance classification quality on a document level by aggregating sentences to document level based on estimated confidences of the resultant classification models. This would be an extension of the work that has been done on this subject. Because the algorithm's confidence for sentence-based classifications is taken into consideration, this strategy would result in a more fine-grained categorization of the documents. A phrase with the predicted label professional translation might have its prediction value translated to a confidence score of 0.873, but the value of a sentence with the predicted label automated translation could be transformed to 0.421. This is an example shown in the previous sentence. These confidences would be a representation of the algorithm's computed chance that the provided candidate translation was a sentence that had been translated by a professional. According to what was

discussed in chapter 5, several machine learning algorithms that were used were unable to be optimized to the extent that would have been ideal owing to restrictions in the resources available for computing. Therefore, about k-Nearest Neighbor classifiers and Artificial Neural Networks, it would be preferable to have a more in-depth optimization procedure in order to get closer to a global optimum in terms of classification outcomes. In addition, the preprocessing procedure that is involved in the generation of a sentence-based data collection is an additional topic of interest.

REFERENCES

1. Tsuruoka, Yoshimasa. (2019). Deep Learning and Natural Language Processing. Brain and nerve = Shinkei kenkyu no shinpo. 71. 45-55. 10.11477/mf.1416201215.
2. Vedantam, Vamsi. (2020). The Survey - Advances in Natural Language Processing using Deep Learning. 10.17762/turcomat.v12i4.611.
3. Arkhangelskaya, E. & Nikolenko, S.. (2023). Deep Learning for Natural Language Processing: A Survey. Journal of Mathematical Sciences. 273. 1-50. 10.1007/s10958-023-06519-6.
4. Zhou, Di. (2021). A New Training Idea of Machine Learning in NLP. Journal of Physics: Conference Series. 1861. 012083. 10.1088/1742-6596/1861/1/012083.
5. Bharadiya, Jasmin. (2023). Transfer Learning in Natural Language Processing (NLP). European Journal of Technology. 7. 10.47672/ejt.1490.
6. Sawicki, Jan & Ganzha, Maria & Paprzycki, Marcin. (2023). The State of the Art of Natural Language Processing—A Systematic Automated Review of NLP Literature Using NLP Techniques. Data Intelligence. 5. 1-47. 10.1162/dint_a_00213.
7. Gunasekaran, Karthick Prasad. (2023). Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review.
8. Bharadiya, Jasmin. (2023). A Comprehensive Survey of Deep Learning Techniques Natural Language Processing. European Journal of Technology. 7. 58-66. 10.47672/ejt.1473.
9. Khurana, Diksha & Koli, Aditya & Khatter, Kiran & Singh, Sukhdev. (2022). Natural Language Processing: State of The Art, Current Trends and Challenges. Multimedia Tools and Applications. 82. 10.1007/s11042-022-13428-4.
10. Ali, Muzaffar & Khan, Mudassar & Abbas, Asad. (2023). Deep Learning for Natural Language Processing: Current Trends and Future Directions.