

Classifying Movie Reviews with Deep Learning Neural Networks

1st Alexander Fisher

Computer Science and Technology
Kean University
Union, NJ, USA
fisheral@kean.edu

2nd Matthew Fernandez

Computer Science and Technology
Kean University
Union, Nj, USA
fermatth@kean.edu

3rd Nicholas Moffa

Computer Science and Technology
Kean University
Union, NJ, USA
moffan@kean.edu

4th Kuan Huang

Computer Science and Technology
Kean University
Union, NJ, USA
khuang@kean.edu

I. INTRODUCTION

The data set used for this project is the IMDB data set of movie reviews from Kaggle.com [7]. It will be used to train a natural language processing model to classify the sentiment of reviews in a binary fashion as either positive or negative. The movie reviews data set enables us to explore the capabilities of natural language processing models. Movie reviews are often written with emotional and opinionated expressions which are occasionally articulated using sarcasm or euphemisms. It would be interesting to see how well different machine learning methods can classify reviews despite the aforementioned nuances of human language. The goal of this project is to train a model sufficient enough to be implemented into the review section of a web page to automatically and accurately place submitted reviews into their respective categories.

II. DATA SET

The IMDB movie reviews data set contains 50 thousand total records, each consisting of two columns, one for the text in the review, and one for the binary classification of that review's sentiment (positive or negative). The data set will be split into a training and testing set, with an 80-20% ratio.

III. LITERATURE REVIEW

A. *Is neuro-symbolic AI meeting its promises in natural language processing? A structured review [3]*

This article is about Neuro symbolic (NeSy) AI and what exactly it is. In short, NeSy AI is the combination of deep learning and symbolic reasoning, aka human reasoning [1]. The goal of NeSy is to address the weaknesses of both symbolic and sub-symbolic approaches while preserving their strengths. There are two most fundamental aspects of intelligent cognitive behavior: the ability to learn from experience, and the ability to reason on what has been learned. There has been a lot of progress made on the learning side, especially in the area of Natural Language Processing (NLP) in particular with deep learning. This AI is still very much in its infancy,

and researchers believe its weaknesses cannot be overcome through deep learning alone. Overall, the article provides an overview of the current state of Neuro Symbolic AI in NLP and stresses the lack of research that exists in order to make NeSy a viable intelligent agent to be used in nearly any industry.

B. *Artificial intelligence approaches using natural language processing to advance EHR-based clinical research [4]*

The article first highlights the issue of big data in the medical field; with so many hospitals converting their patient medical records to digital, leveraging the useful information of these records to assist clinical research is a top priority. This is where natural-language processing may have a part to play. An AI with sufficient training can examine millions of records in a matter of days if not hours and extract the relevant information. Such a model can drastically improve not only the accuracy but speed of new research for new medical breakthroughs, almost like an incredibly advanced ("Ctrl+F") function when reviewing a PDF. The article details some already existing NLP models utilized for both allergies and asthma, and a model for the asthma assessment shows how the EHR is broken down into structured data, such as patient information, and unstructured data, such as clinical notes by the physician. This is then put through the NLP model and is aggregated to give a summarized description of the asthma severity and possible treatments. As before, this article does mention that NLP of such a scale as to be used by the medical industry is still in its infancy, but strongly urges more and more industries to make use of automation via NLP to improve its viability.

C. *A Survey of the State of Explainable AI for Natural Language Processing [2]*

Explainable AI is any AI that can use natural language to explain its own processes and functions. This article serves to explain the capabilities and limitations of Explainable AI

(XAI) as of the time when it was written (2020). In brief, the article highlights three main obstacles in making NLP techniques for explainable AI more suitable for widespread use: Accuracy vs interpretability: If you want AI explanations to be more accurate, you run the risk of getting too confusing for the average viewer, and if you want explanations to be more interpretable, the user may not be given an accurate description. You have to train the NLP model to generalize and summarize, but not too much. Scalability: Large data sets become incredibly expensive to interpret (this can be said for any manner of data processing). And metrics: No standards for measuring an AI model's ability to balance accuracy/interpretability with natural language have been developed, or have at least been widely adopted.

D. Learning Word Vectors for Sentiment Analysis [5]

This Stanford-published research paper attacked the challenge of analyzing the sentiment of movie reviews by combining the use of two natural language processing techniques: unsupervised word vector clustering and supervised sentiment classification. The word vector technique makes use of an unsupervised probabilistic approach in clustering words in a high-dimensional space, where the vectors between two words are determined by measuring the semantic similarity between them. This approach could group together words of similar strength in terms of expression, however, was unable to distinguish the sentimental difference between positive and negative expressions. To compensate for this, the polarity annotations (star ratings) that labeled each movie review enabled a supervised sentiment model to be implemented in tandem with the word vector clustering model to build a word vector model capable of realizing the nuances in sentiment, thus imbuing those realizations into the vectors themselves. A linear SVM was used to then classify vectors in the clusters.

IV. METHODOLOGY

A. Implementation

The pipeline used for carrying out the research experiment consisted of initializing dependencies, data pre-processing, a pre-trained BERT model, fine-tuning the model using training data, and finally evaluating the model using test data.

1) *Initializing Dependencies:* The dependencies for our project consisted of two modules useful for machine learning and natural language processing: pandas & transformers (Trainer, Tokenizer, model loader). The one most used for our project was the “transformers” module, which gave access to Huggingface’s assortment of pre-trained transformers and NLP tokenizers.

2) *Data Preprocessing:* During the pre-processing stage, the IMDB data set downloaded from kaggle.com in CSV format was uploaded to the Python file into a pandas data frame for easier manipulation. Then, the data frame was separated into an 80-20 train-test split. From each split, a random sample of 10 percent of the split was taken to be used for the final training and evaluation data sets (40000 and 10000 records, respectively). For the BERT model to understand text

input, the input must be tokenized into encoding that the model can understand. With that in mind, we tokenized the train and test data frames using the ‘BertTokenizer’ from the transformers module.

3) *Fine-tuning the model:* Next, we loaded the pre-trained model called “bert-base-uncased” from the transformers module. The model was then fine-tuned by using a Python loop that calculated the model’s loss and optimized based on that loss for each batch, for each of the 5 epochs. Once the fine-tuning training was completed, the model was evaluated using the tokenized test data set.

A possible innovation that may be implemented in this work is distinguishing which sentences are relevant to the classification, (i.e., which sentences are relevant to the actual movie being reviewed).

4) *User Interface:* We wanted a simple-to-use web page where the semantic analysis feedback from our model could be displayed in numbers. This serves as a means of making the experience more user-friendly.

B. Evaluation

After fine-tuning the BERT-base-uncased NLP model to the 40000 sample records from the IMDB data set, it was evaluated using the 10000 record test data set to find the model’s accuracy, precision, and recall rate. The accuracy, a measurement of correctly classifying a review, was measured at 92.82 percent. The precision, a measurement of how much of the positively predicted sentiments actually were positive, was measured at 93.34 percent. Finally, the recall, a measurement of how many of the truly positive reviews were classified as positive, was measured at 92.23 percent.

REFERENCES

- [1] Bhatt, D. (2023, February 14). Neuro symbolic ai: Enhancing common sense in ai. Analytics Vidhya. Retrieved May 7, 2023, from <https://www.analyticsvidhya.com/blog/2023/02/neuro-symbolic-ai-enhancing-common-sense-in-ai/>
- [2] Danilevsky, M., Qian, K., Aharonov, R., Katsis, Y., Kawas, B., Sen, P. (2020, October 1). A survey of the state of explainable AI for natural language processing. arXiv.org. Retrieved March 31, 2023, from <https://arxiv.org/abs/2010.00711v1>
- [3] Hamilton, K., Nayak, A., Božić, B., Longo, L. (2022, January 1). Is neuro-symbolic AI meeting its promises in Natural Language Processing? A structured review. Semantic Web. Retrieved March 31, 2023, from <https://content.iospress.com/articles/semantic-web/sw223228>
- [4] Juhn, Y., Liu, H. (2019, December 26). Artificial intelligence approaches using natural language processing to advance EHR-based Clinical Research. Journal of Allergy and Clinical Immunology. Retrieved March 31, 2023, from <https://www.sciencedirect.com/science/article/abs/pii/S0091674919326041>
- [5] Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., Potts, C. (n.d.). Learning word vectors for sentiment analysis - Stanford University. Retrieved April 1, 2023, from https://ai.stanford.edu/amaas/papers/wvSent_acl2011.pdf
- [6] Nest, E. — T. R. (2020, April 12). Explainable AI - what is it and why do we need it? Medium. Retrieved May 7, 2023, from <https://medium.com/the-research-nest/explainable-ai-what-is-it-and-why-do-we-need-it-261509e48cc>
- [7] N, L. (2019, March 9). IMDB dataset of 50K movie reviews. Kaggle. Retrieved March 31, 2023, from <https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>