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

I. Introduction

The dataset used for this project is the IMDB dataset of movie reviews from Kaggle.com [5]. 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 dataset 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 do in classifying reviews despite the nuances of human language such as those mentioned. The goal of this project is to train a model sufficient enough to be implemented into the review section of a webpage to automatically and accurately place submitted reviews into their respective categories.

II. Dataset

The IMDB movie reviews dataset contains 50k 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 dataset 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 [2]

This article is about Neuro symbolic (NeSy) AI and what it is exactly is the combination of deep learning and symbolic reasoning. The goal of NeSy is to address the weaknesses of each of symbolic and subsymbolic approaches while preserving their strengths. There are two most fundamental aspects of intelligent cognitive behavior that are the ability to learn from experience and the ability to reason what has been learned. There has been a lot of progress made in the learning side especially in the area of Natural Language Processing (NLP) in particular with deep learning. There are also some weaknesses but some researchers argue that it cannot be addressed by deep learning alone. Overall, the article provides an overview of the current state of Neuro symbolic AI in NLP and the main thing about it is that there still is a lot of research that needs to be done to overcome the challenges and limitations of this approach.

B. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research [3]

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 that can examine millions of records in a matter of days if not hours and extract the relevant information 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 the unstructured data which are the clinical notes enscribed by the physician. This is what is put through the NLP model, and is aggregated to give a summarized description of the asthma severity and possible treatments. The article does mention that this technology is still in infancy but strongly urges more and more industries to make use of automation via NLP.

C. A Survey of the State of Explainable AI for Natural Language Processing [1]

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 more suitable for widespread use: Accuracy vs interpretability: If you want to be more accurate, you run the risk of getting to confusing for the average viewer, and if you want to make it more interpretable, it may not give you the whole picture; You have to train the NLP model to generalize and summarize, but not too much. Scalability: Large datasets become incredibly expensive to interpret Metrics: No standards for measuring an AI mode's ability to balance accuracy/interpretability have been developed, or at least have been widely adopted.

D. Learning Word Vectors for Sentiment Analysis [4]

This Stanford-published research paper attacked the challenge of analyzing 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

The pipeline we are planning to use for carrying out the research experiment is as follows. The IMDB dataset will be uploaded to a Google Collab notebook into several python lists and numpy arrays distinguishing the training and testing data. The Bidirectional Encoder Representations from Transformers (BERT) algorithm by Google will then be used to train the deep learning model to classify reviews as having positive or negative sentiment. 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).

V. Experimentation

A. Simulation plans

We intend to train the deep learning model with the given dataset of 50K movie reviews, and then record the performance of it in classifying reviews in that same dataset. Afterwards, we plan to demo its performance in classifying new, custom-made movie reviews on the fly using a simple input interface.

B. Evaluation Methods

We are going to be using the accuracy, error rate, precision and recall for the evaluation methods. The accuracy is the number of correct classifications and error rate is the number of incorrect classification. Precision is classifying positive samples correctly and recall is the ability of finding out all positive samples.

References

[1] 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

[2] 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

[3] 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

[4] 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

[5] 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