Prediction Stability of Text Classification Methods

Comparing Traditional and Deep Learning Methods

Orin Brown
University of Colorado Boulder
obbrown1@gmail.com

ABSTRACT

Text classification is a fundamental component of Natural Language Processing (NLP), which has a broad range of applications, including document search, sentiment analysis, chatbots and virtual assistants, machine translation, and many others. Text classification methods have evolved significantly over the past decades, going from rule-based approaches to the sophisticated machine learning models that are the current state of the art. This paper aims to investigate a key characteristic of these models: prediction stability.

INTRODUCTION

This paper aims to investigate the prediction stability of different text classification models. Prediction stability refers to the consistency and reliability of the classification results obtained from text classifiers when applied to different samples or variations in the input data. It aims to investigate the robustness of text classifiers and examine how consistent their predictions are across different instances of the same problem.

Prediction stability is crucial in text classification as it directly impacts the reliability and trustworthiness of the classification results. In real-world scenarios, it is essential to have text classifiers that produce consistent predictions that are robust in the face of small or insignificant variations in the input data or different samples

Existing text classification solutions often focus on maximizing accuracy without explicitly addressing prediction stability. While these solutions may perform well in terms of accuracy on specific datasets, they may lack consistency when applied to different samples or variations in the input data.

RELATED WORK

Text classification methods have evolved significantly over the years, driven by advancements in machine learning and natural language processing techniques. “From the 1960s until the 2010s, traditional text classification models dominated. Traditional methods mean statistics-based models, such as Naïve Bayes (NB), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM).”[1] Current research is largely focused on deep learning methods which, while they have drawbacks of their own, address many of the issues that plague traditional approaches.

PROPOSED WORK

Datasets: This project will focus on datasets containing categorized news articles.

Tools: The project will employ Python as the primary programming language, along with relevant libraries for text analysis and machine learning tasks. A Jupyter notebook will be used as the development environment to facilitate code experimentation and documentation.

Tasks: The primary task will be analyzing the prediction stability of traditional and deep learning classification methods. To this end, models of each category will be implemented and trained to a reasonable degree of accuracy. Once done, experiments will be run to determine how the gradual addition of new words affects model predictions.

Ancillary tasks will include: data preprocessing and warehousing, statistical analysis, and data visualization.

EVALUATION

The models will be evaluated across the dimensions of classification accuracy, training speed, and prediction stability.

Experimental setup: The project will set up experiments to observe the behavior of prediction confidence when additional words are added to text samples classified with high confidence.

The experimental setup will involve selecting a set of text samples that are classified with high confidence by the classifiers. Additional words will be added to these samples, and the classifiers' prediction confidence scores will be recorded. The behavior of the confidence scores, such as their variation or consistency, will be analyzed to understand how the classifiers handle modifications in the input and their level of confidence in the revised samples.

DISCUSSION

Timeline: There are roughly six weeks left in the term at time of writing, so I propose to divide the tasks as follows

- Data gathering – 1 week
- Research on models – 1 week
- EDA and data cleaning – 1 week
- Modeling and model tuning – 2 weeks
- Analysis and report writing – 1 week

Potential Challenges: Identifying an appropriate dataset could be a challenge. Identifying appropriate models that can produce classification estimates or probabilities may also take some time, but I am confident that there are at least two available. Sticking to the proposed timeline may also be a challenge.

Alternatives: If the proposed analysis proves to be impractical, I can shift my focus to some other metric in the text classification space. Perhaps investigating how well models can transfer learning from one corpus to another. Are the models sufficiently general to be able to do this well, or do they need training on each specific dataset?

CONCLUSION

TBD

[1] Li et al., A Survey on Text Classification: From Traditional to Deep Learning ACM Transactions on Intelligent Systems and TechnologyVolume 13Issue 2Article No.: 31pp 1–41https://doi.org/10.1145/3495162

REFERENCES

On the Sensitivity and Stability of Model Interpretations in NLP
https://aclanthology.org/2022.acl-long.188.pdf

A Survey on Text Classification Algorithms: From Text to Predictions
https://www.mdpi.com/2078-2489/13/2/83

Text Classification Algorithms: A Survey
https://medium.com/text-classification-algorithms/text-classification-algorithms-a-survey-a215b7ab7e2d

Li et al., A Survey on Text Classification: From Traditional to Deep Learning ACM Transactions on Intelligent Systems and TechnologyVolume 13Issue 2Article No.: 31pp 1–41https://doi.org/10.1145/3495162