EMOLNT

ABSTRACT

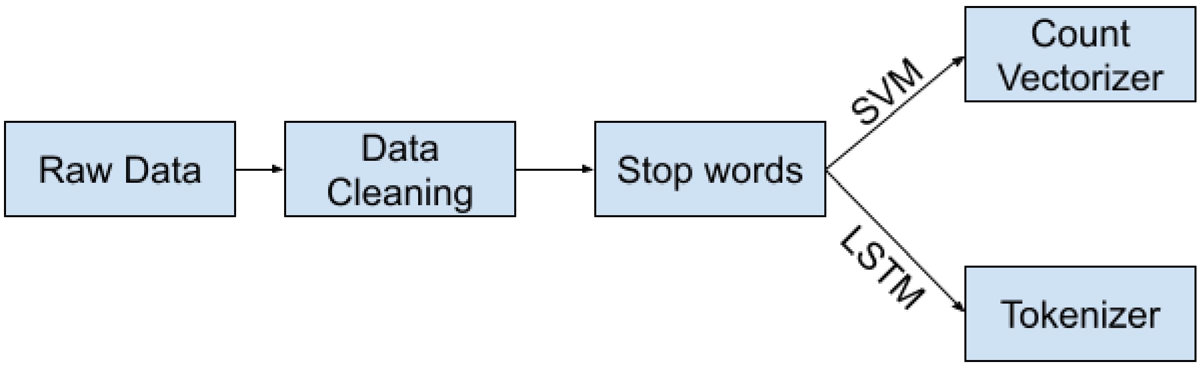
Emotion intensity analysis has become a crucial aspect of natural language processing, finding applications in various domains such as sentiment analysis, customer feedback, and human-computer interaction. This paper presents a comparative study between two emotion analysis models developed for the WASSA-2017 Shared Task on Emotion Intensity (EmoInt). The first model employs a traditional machine learning approach using Support Vector Machines (SVM) with count vectorization, while the second model leverages Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM) cells using TensorFlow tokenizer. The study explores the impact of different preprocessing techniques on the performance of the models, achieving accuracy scores of approximately 60% and 70%, respectively.

INTRODUCTION

The growing importance of emotion analysis in understanding textual data has led to the development of various models for tasks like sentiment analysis and emotion intensity prediction. The WASSA-2017 Shared Task on Emotion Intensity (EmoInt) provided a platform to explore and compare different approaches to address this challenge. In this paper, we present two models, one based on traditional machine learning techniques and the other utilizing deep learning, to analyze and predict emotion intensity in textual data.

DATA PRE-PROCESSING

The success of emotion analysis models heavily relies on effective data preprocessing. In our approach, we performed a comprehensive preprocessing step to enhance the quality of the input data. This involved the removal of unnecessary information, hashtags, emojis, punctuations and the application of stop words to refine the dataset. For the SVM model, we utilized count vectorization to represent textual data numerically, while for the LSTM model, TensorFlow tokenizer was employed. Along with these, few modulations like pad_sequence is used in order to improve the contribution of the data to the model. Pad sequences are a crucial aspect of natural language processing and sequence-based tasks, especially when dealing with variable-length sequences such as sentences or paragraphs. These preprocessing steps played a crucial role in shaping the input features and capturing relevant information for both models. I found a better result and performance after the data being passed onto several processing stages.



MODEL CREATION

Our first model employed a traditional machine learning approach, specifically Support Vector Machines (SVM), which has proven to be effective in various text classification tasks. The input data was transformed into numerical features using count vectorization, capturing the frequency of each term. The SVM model was trained on this representation to predict emotion intensity. On using linear kernel , i noticed a drastic change in the model score. I haven't made many changes to hyperparameter, keeping the default parameters contributed more to the accuracy score.

The second model utilized Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM) cells, a deep learning architecture known for its ability to capture sequential dependencies in data. I employed the TensorFlow tokenizer to preprocess the data and convert it into a format suitable for training the LSTM model. This allowed the model to learn complex patterns and dependencies within the textual data, enhancing its performance in emotion intensity prediction.Dense,Sequential,Embedding layers builds up the entire architecture of the LSTM model.The number count of embedding layer is determined by the brute-force method.

CONCLUSIONS

In this study, we presented two distinct models for emotion intensity prediction, one based on traditional machine learning (SVM) and the other on deep learning (RNN LSTM). Through thorough data preprocessing, including the removal of unnecessary information and the application of stop words, we aimed to enhance the accuracy of both models. Our findings indicate that the LSTM model, leveraging the power of deep learning, outperformed the SVM model, achieving a higher accuracy score of approximately 70%. This highlights the potential of deep learning approaches in capturing intricate patterns in textual data for emotion analysis tasks. Future work may involve further exploration of advanced deep learning architectures and optimization techniques to improve model performance.

REFERENCES

https://medium.com/@canerkilinc/padding-for-nlp-7dd8598c916a

https://huggingface.co/