Natural Language Processing  
Homework and Programming Assignment 3

Parker Smith

10/24/2022

Paper 1: *Deep Extractive Text Summarization*

This paper covers extractive summarization using deep learning techniques to summarize text. The primary model is a mixed Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN). The model requires six parameters and ultimately has a testing accuracy of 75%. The primary model used by this paper is a mixed CNN and RNN. This was chosen because the CNN can quickly extract hidden features from the text by running it through convolution layers with kernels of different sizes. From there, the RNN can learn those features and select the sentences with the highest feature values. The dataset for this paper is called Para Multiling 2015 and was designed for Multilingual Single-document Summarization. Each document contains both the raw text and a human generated summary for testing evaluation. The performance metrics for this model are Training Accuracy, Validation Accuracy, Testing Accuracy, Testing Precision, Testing Recall, and Testing F-Score. The results have the model with a Training Accuracy of 96%, a Testing Accuracy of 75%, a Testing Precision of 25%, a Testing Recall of 42%, and a Testing F-Score of 31%. Some limitations to this paper include a lack of detailed information on the dataset and a weak idea of how the proposed summarization approach is implanted. No statistical information is given on the dataset such as the number of sentences, number of documents, or vocabulary size of the dataset. Additionally, the proposed summarization approach claims to divide the model into two phases, yet no implementation or discussion of those two phases are never shown. This paper’s methodology ultimately has nothing in comparison with the two methodologies I use in my paper. I am proposing methodologies that will execute with speed as they are not implementing and training an entire neural network. My methodologies only implement either a simple TF-IDF score, or a small naïve bayes classifier to determine important features.

Citation: R. Bhargava and Y. Sharma, “Deep Extractive Text Summarization,” Procedia Computer Science, vol. 167, pp. 138–146, 2020. [Online]. Available: https://doi.org/10.1016/j.procs.2020.03.191

Paper 2: *A Framework for Extractive Text Summarization Based on Deep Learning Modified Neural Network Classifier*

This paper covers extractive summarization using a basic Deep Learning Neural Network model. The model has fourteen explicitly predefined features and it uses those features to score sentences to determine their fitness for the summary. Ultimately, the model has a 91.21% accuracy. This paper uses several methods throughout their model: Preprocessing, Feature Extraction, Feature Selection (using the Improved Fruit Fly Optimization Algorithm (IFFOA) algorithm), Entropy Calculation, Classification using a Neural Network, and Sentence Extraction. Preprocessing cleans the input text, Feature Extraction and Selection allow the algorithm to select the most optimal features for summarizing the text, and Entropy Calculation, Classification, and Sentence Extraction are the final steps to outputting the final summarized text. Two algorithms were designed for this paper: Improved Fruit Fly Optimization Algorithm (IFFOA) and a Deep Learning Modified Neural Network Classifier (DLMNN). The IFFOA was designed to eliminate redundant features without affecting the predictive accuracy. The DLMNN Classifier was designed to classify sentences based on their score values: sentences with high scores classify to the positive class (sentences fit for summarization), sentences with low scores classify to the negative class (sentences not used for summarization). The dataset for this paper is from Document Understanding Conferences and contains 59 document sets. In total, the dataset contains 567 news articles with approximately 10 documents per category. Each document contains a human-generated summary of about 100 words per article. The performance metrics for this model included Sensitivity, Specificity, Accuracy, Precision, Recall, and F—measure. The average performance resulted in a sensitivity of 89.07%, a specificity of 89.35%, an accuracy of 93.93%, a precision of 92.63%, a recall of 90.29%, and an F-measure of 89.73%. Some limitations to this model are that all features have to be explicitly created before running the model. If certain features don’t exist, the model will not run. This paper’s methodology is similar to my Naïve Bayes model in that both models require features that have to be explicitly defined beforehand. The features themselves are automatically filled in by the model, but what exactly is being marked as a feature is predetermined.

Citation: B. Muthu, S. Cb, P. M. Kumar, S. N. Kadry, C. Hsu, O. Sanjuan, and R. G. Crespo, “A Framework for Extractive Text Summarization Based on Deep Learning Modified Neural Network Classifier,” Association for Computing Machinery, vol. 20, no. 3, pp. 1-20, 2021. [Online[. Available: https://doi.org/10.1145/3392048

Paper 3: *Extractive Hotel Review Summarization based on TF/IDF and Adjective-Noun Pairing by Considering Annual Sentiment Trends*

This paper summarizes hotel reviews in an attempt to simplify the process of sentiment analysis. Instead of finding the sentiment of long, exhaustive reviews, the short essence of the review is summarized and the sentiment of the entire review is based off the sentiment of the summarized review. For this project, extractive summarization is used with Term Frequency Inverse Document Frequency (TF-IDF) as the primary summarization method. No specific algorithm was designed for this model, however a generic TF-IDF algorithm was used to score sentences, and preprocessing was implemented through sentence segmentation, case folding, and tokenization. The dataset used for this paper is a list of hotel reviews scraped from TripAdviser.com. There are a total of 21,746 reviews, 1,370 of which contain text from a total of 167 different hotels. The performance metrics for this paper are the ROUGE and BLEU evaluation methods. The average ROUGE precision is 58.95%. The average BLEU precision is 61.32%. A major limitation of this paper is a complete lack of information on how the paper’s authors implement the algorithms. No information is given on how the TF-IDF score is generated and how the sentences are scored. This paper’s methodology is similar to my paper’s methodology in that we both implement TF-IDF as one of the primary ways of scoring sentences. However, there could be an important difference in how both TF-IDF algorithms are implemented. Unfortunately, it is impossible to figure that out as this paper does not showcase their algorithm.

Citation: G. Nathania H., R. Siautama, A. Claire I. A., D. Suhartono, “Extractive Hotel Review Summarization based on TF/IDF and Adjective-Noun Pairing by Considering Annual Sentiment Trends,” Procedia Computer Science, vol. 179, pp. 558-565, 2021. [Online[. Available: https://doi.org/10.1016/j.procs.2021.01.040