AI Tools for Earnings Call Transcript Analysis

Stanford CS229 Project

Shawn Zhang
Department of Computer Science
Stanford University
shawnz@stanford.edu

Robert Michael Irelan
Department of Computer Science
Stanford University
rirelan@stanford.edu

Arman Akbarian Kaljahi
Department of Computer Science
Stanford University
akbarian@stanford.edu

Abstract

The abstract is optional, depending on your available space. It should consist of 1 paragraph consisting of the motivation for your paper and a high-level explanation of the methodology you used/results obtained.

1 Introduction

Explain the problem and why it is important. Discuss your motivation for pursuing this problem. Give some background if necessary. Clearly state what the input and output is. Be very explicit: "The input to our algorithm is an image, amplitude, patient age, rainfall measurements, grayscale video, etc.. We then use a SVM, neural network, linear regression, etc. to output a predicted age, stock price, cancer type, music genre, etc.." This is very important since different teams have different inputs/outputs spanning different application domains. Being explicit about this makes it easier for readers. If you are using your project for multiple classes, add a paragraph explaining which components of the project were used for each class.

Quarterly earnings call transcripts for publicly traded companies are often fairly large text documents that include lots of information about future growth of the company, its past performance and new initiatives and product launches. Once this information is released, if there is a strong and consistent positive outlook for the company in the content of the transcript, it results in a jump in stock price of the company in the next few days, while a negative sentiment and potential risks result in selloff of stock and price drop. In this project, we aim to explore a variety of machine learning tools within the domain of natural language processing to showcase how an ML model can extract relevant information and predict the "outcome" of the earning call (positive or negative stock price reaction) given the content of the text in the earning call. In particular, we will build a deep learning based sentiment analysis classifier predicting 3 classes of "positive", "negative" and "neutral" using publicly available sentiment annotated datasets. In addition we will explore other NLP approaches to segment and summarize the transcripts as they are often large and cannot be meaningfully represented as a vector for a classification task. Finally we will measure how different NLP components can generate signals to predict the outcome of the stock price reaction to earning calls.

2 Related Work

You should find existing papers, group them into categories based on their approaches, and discuss their strengths and weaknesses, as well as how they are similar to and differ from your work. In your opinion, which approaches were clever/good? What is the state- of-the-art? Do most people perform the task by hand? You should aim to have at least 5 references in the related work. Include

previous attempts by others at your problem, previous technical methods, or previous learning algorithms. Google Scholar is very useful for this: https://scholar.google.com/ (you can click "cite" and it generates MLA, APA, BibTeX, etc.)

3 Dataset and Features

Describe your dataset: how many training/validation/test examples do you have? Is there any preprocessing you did? What about normalization or data augmentation? What is the resolution of your images? How is your time-series data discretized? Include a citation on where you obtained your dataset from. Depending on available space, show some examples from your dataset. You should also talk about the features you used. If you extracted features using Fourier transforms, word2vec, histogram of oriented gradients (HOG), PCA, ICA, etc. make sure to talk about it. Try to include examples of your data in the report (e.g. include an image, show a waveform, etc.).

Shawn: how did you collect the Earning Call Transcripts?

To evaluate the reaction of stock price to earning calls, we need at minimum two types of raw data: earning call transcripts and historical stock price data. We initially collected a large number of earning call transcripts, most of which were Q4 2023 earning reports from January-February 2024, as PDF printouts from the financial website Seeking Alpha. Stock price data is collected from Yahoo Finance and is used to train the model to predict the stock price reaction to earning calls. Earning call transcripts are collected from the SEC website and are used to evaluate the performance of the model.

4 Methods

Describe your learning algorithms, proposed algorithm(s), or theoretical proof(s). Make sure to include relevant mathematical notation. For example, you can briefly include the SVM optimization objective/formula or say what the softmax function is. It is okay to use formulas from the lecture notes. For each algorithm, give a short description (1̃ para- graph) of how it works. Again, we are looking for your understanding of how these machine learning algorithms work. Although the teaching staff probably know the algorithms, future readers may not (reports will be posted on the class website). Additionally, if you are using a niche or cutting-edge algorithm (e.g. long short-term memory, SURF features, or anything else not covered in the class), you may want to explain your algorithm using 1/2 paragraphs. Note: Theory/algorithms projects may have an appendix showing extended proofs (see Appendix section below).

We will first explore the problem of text classification and machine learning methodologies used to represent text as a vector. In particular, there are a variety of publicly available sentiment analysis datasets (IMDB, SST-2 Binary classification, Yelp Binary classification) that we will benchmark our baseline and more advanced models on.

A key element of a text classifier is representation of words of the document as vectors and in turn constructing a vector representation of larger elements of text such as sentences and paragraphs. As our baseline model we have explored Glove vectors as our pre-trained semantic representation of words and if time permits will explore sub-word methodologies similar to the original BERT paper. We have established some baseline classification model performance metrics using simple averaging of word vectors as will be discussed in the next section. For more advanced methodologies, we will explore LSTM, convolutional neural networks for sentence classification as well as few pre-trained sentence embedding models such as universal sentence encoder and BERT (if time permits). We can fine-tune or use these pre-trained models without tuning and append fully connected neural networks with an N-class output layer at the end to train these models on sentiment datasets mentioned above.

5 Experiments / Results / Discussion

You should also give details about what (hyper)parameters you chose (e.g. why did you use X learning rate for gradient descent, what was your mini-batch size and why) and how you chose them. Did you do cross-validation, if so, how many folds? Before you list your results, make sure to list and explain what your primary metrics are: accuracy, precision, AUC, etc. Provide equations for the metrics if necessary. For results, you want to have a mixture of tables and plots. If you are solving

a classification problem, you should include a confusion matrix or AUC/AUPRC curves. Include performance metrics such as precision, recall, and accuracy. For regression problems, state the average error. You should have both quantitative and qualitative results. To reiterate, you must have both quantitative and qualitative results! This includes unsupervised learning (talk with your project TA on how to quantify unsupervised methods). Include visualizations of results, heatmaps, examples of where your algorithm failed and a discussion of why certain algorithms failed or succeeded. In addition, explain whether you think you have overfit to your training set and what, if anything, you did to mitigate that. Make sure to discuss the figures/tables in your main text throughout this section. Your plots should include legends, axis labels, and have font sizes that are legible when printed.

To be added

6 Conclusion / Future Work

Summarize your report and reiterate key points. Which algorithms were the highest-performing? Why do you think that some algorithms worked better than others? For future work, if you had more time, more team members, or more computational resources, what would you explore?

7 Appendices

This section is optional for theory/algorithmic projects. Include additional derivations of proofs which weren't core to the understanding of your proposed algorithm in the methods section. Note: All sections before this point must fit on five (5) pages. No exceptions. Supplemental material is not allowed. Anything else you want to add to your report (e.g. acknowledgements, author bios, funding sources) is included in the 5 page limit. The exception is the section describing the contributions of each team member. You will be penalized -10 points per page exceeding this limit. The max report score is 100.

8 Contributions

The contributions section is not included in the 5 page limit. This section should describe what each team member worked on and contributed to the project.

9 Note on citations

There are two citation commands:

Citation in parentheses The command `\citep{}` is for when you cite a paper at the end of a clause, and you could read the text out loud and not read the authors' names. For example "We also run our experiments on a multilingual language model (Rajpurkar et al., 2018)."

Citation in text The command `\citet{}` is for in-text citations, when someone reading the text out loud would have to read the names of the authors for the sentence to make sense. For example, "We also run our experiments on the multilingual model described in Rajpurkar et al. (2018)."

References

Pranav Rajpurkar, Robin Jia, and Percy Liang. 2018. Know what you don't know: Unanswerable questions for SQuAD. In *Association for Computational Linguistics (ACL)*.