Understanding Workplace Sentiment

Aaron Marks — aamarks@stanford.edu
Dhruv Kedia — dkedia@stanford.edu

Project Overview

Motivation

Using natural language processing on consumer reviews, whether they are movie reviews, Amazon reviews, workplace reviews is a common occurrence in studying sentiment analysis. This project seeks to use natural language processing techniques, including sentiment analysis, to predict employee happiness and satisfaction in the workplace. Many job seekers use the website Glass-door to learn about employees' experiences at different companies before deciding to apply or accept job offers.

Predicting

The inputs for our models were the text from employee reviews on Glassdoor with corresponding ratings of the companies (1-5). Our task was to build classifiers to predict the associated class, based on rating, from the text alone of a review.

Data

- Uses the Google, Amazon and more Employee Reviews dataset from a Kaggle competition.
- Dataset included over 67k employee reviews of technology companies. All containing the employees rating of their experience at the company.
- The reviews are for Google, Amazon, Facebook, Microsoft, Netflix, and Apple
- There are 5 potentially ratings (1-5 stars)
- Here are some examples of reviews for different classes:
- 5 star: "Best Company to Work For!. Atmosphere that promotes the expression of fresh ideas!. It can be an overwhelming machine at times but there are no real drawbacks I encountered." it will feel annoying."
- 1 star: "Toxic and Retaliatory Environment. Great events, food, gym, perks etc. Toxic Working Environment Recruiting Management Bullying HR does not care about employees rights. No Work-Life balance"

Fig 1: Distribution of ratings for Firms

Approach

- **Embeddings:**
  - Created word-level embeddings for the different words observed in the comments by employees using 50D GloVe encodings.
  - Concatenated to form sentence level embeddings with a max length.
- **Baseline Models**
  - **Logistic Regression:** $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$
    - A sigmoid loss function.
    - Specifically the One vs. Rest for multiclass classification.
  - **Convolutional Neural Networks:**
    - Embedding layer → 2 Conv1D layers with ReLu activation → MaxPooling → Linear With ReLu → Softmax Layer

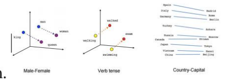
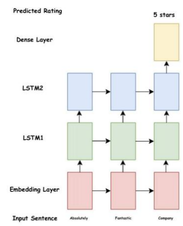


Fig 4: 3D representation of GloVe vectors

- **RNN Based Models**
  - RNNs are designed to work better on sequential data
    - Ideal for text data and a problem like sentiment classification.
    - Vanilla RNNs suffer from shrinking gradients so we make use of LSTMs to tackle longer reviews.
    - The model architecture we use is as follows:
      - LSTM layer with 128 neurons, 0.2 dropout
      - LSTM layer with 128 neurons, 0.2 dropout
      - Dense layer with softmax activation
      - Categorical cross entropy loss function and an Adam optimizer
  - **Only Summary:** Initially we believe summary comments will help provide the overall sentiment for the employees experience at the company
  - **Summary, pros, and cons**
    - **Concatenated:**
      - Concatenated all the 3 parts of the review to be 1 and repeated the same steps as above.
      - Concatenating all these 3 reviews would lead to 1 review with a very confusing sentiment. Hence, this would make it difficult for the model to learn and perform well.
    - **Individual Models:** Instead of combining the reviews into a large sentence embedded sentence, we feed the pieces of the review to separate RNNS, connecting their outputs later in an RNN LSTM layer -> dense layer -> softmax



RNN architecture

Experiments And Results

Baseline Results

Model	Train Accuracy	Val Accuracy
LR	0.36	0.34
CNN	0.46	0.46

RNN Results

Model	TT1A	VT1A	VT2A
Summary Only	0.62	0.58	0.77
All Reviews Concat	0.42	0.41	0.64
All Reviews Individual	0.78	0.62	0.88

Best Results: Individual RNN for all reviews

Fig 5: Loss vs. Epochs

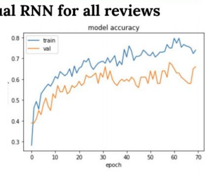


Fig 6: Accuracy vs. Epoch

Test Set Results: Top 1 Accuracy of 0.63 and Top 2 Accuracy of 0.86.

Qualitative Analysis

- Model successfully classifies following words in respective ratings:
  - 1 or 2 star - "poor", "stress", "burnout", "bad"
  - 4 or 5 star - "great", "best", "smart", and "friendly"
- Model incorrectly classifies following words in respective ratings:
  - 1 or 2 star - "long", "boring", and "free"
  - 4 or 5 star - "unique", "different", and "exciting"
- Also there were certain phrases that were ambiguous and led to poor classifications, for example: "work life balance", "technical understanding", and "company culture". This is due to humans not writing reviews in full sentences. (appearing in both pros and cons)
- It's clear that the signal from 5 star reviews is strong and from 1 star reviews is strong however, like many multi-class problems the signal becomes muddle in the classes between the extremes.
- We discovered that there is a high human error rate for this problem and dataset by predicting ratings ourselves. Many employees leave reviews with sentiment mismatching the ratings.

Conclusion

- Best model uses 3 different RNN based models with GloVe word embeddings.
- The model has a lot of scope to learn more and overfit the training data.
- Model performs significantly better when using top 2 accuracy as opposed to top 1 accuracy.
- Key limitations working on this project was training time. It prevented us from further hyper-tuning several of the parameters.

Future Work

- Better hypertuning of parameters.
- Integrating BERT and ELMO contextual word embeddings with our model as well as our own VSM.
- Experimenting with state of the art sentiment classification models.
- Experimenting with attention networks.
- More in depth error analysis.

References

1. Dos Santos, Cicero & Gatti de Bayser, Maira. (2014). Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts.
2. Google, Amazon and more Employee Reviews, Kaggle. https://www.kaggle.com/petersunga/google-amazon-facebook-employee-reviews
3. Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation.
4. Karpathy, Andrej. "Convolutional layer" CS231N Lecture Notes.
5. Ng, Andrew. "Recurrent Neural Networks" CS230 Lecture Notes.