

Using Attention on Movie Sentiment Classification

James Li, dawwctor@stanford.edu
Sam Kwong, samkwong@stanford.edu



Introduction

We implemented a platform to predict movie review sentiment as positive, negative, or neutral. We utilized various unidirectional RNN models, one with regular RNN cells, one with LSTMs, and one with LSTMs and single-context attention. We trained these models on cleaned review texts to output sentiment labels. Surprisingly, we found the simple LSTM model to perform the best.

Data

IMDB movie reviews: roughly 1.8 million reviews across 45,000 movies.

We used a 90/5/5 training/dev/test split by movies. The reviews were cleaned to remove all non-alphanumeric characters for parsing consistency.

Models

All models trained for 5 epochs with a word embedding size of 50.

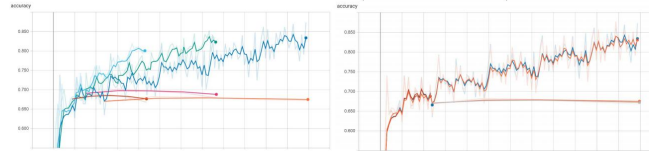
We used cross-entropy loss and measured accuracy using the maximum prediction probabilities.

1. Baseline: Unidirectional RNN cells with a final softmax layer.
2. LSTM: Unidirectional LSTMs with a final softmax layer.
3. Attention: Unidirectional LSTMs attended by one trainable context vector with a final softmax layer.

Searching hyperparameters, we consistently found a learning rate of 0.001, a dropout rate of 0.2, and a hidden size of 50 to perform best.

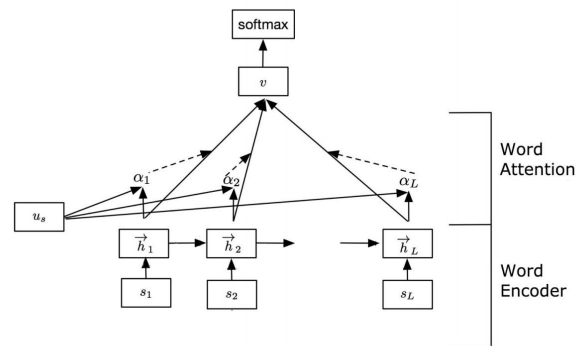
Results

Model	Test Accuracy	Loss
Baseline	0.6876	0.6963
LSTM	0.7016	0.6675
Attention (25 units)	0.6800	0.7056
Attention (50 units)	0.6824	0.7087
Attention (75 units)	0.6658	0.7228



Single-Context Attention

We were inspired by the “Hierarchical Attention Networks for Document Classification” paper [1], as it also dealt with using attention in a non-sequence-to-sequence context for sentiment analysis.



Discussion

Surprising Results:

- Our attention model performed worse than both our basic LSTM model and our baseline model.
- Even with hyperparameter testing, most of the models had roughly the same performance.

Model Analysis:

- Models converged quickly before overfitting severely, so regularization might have helped.
- Our attention model may not have performed well due to an overly complex model.

Data Analysis:

- Certain movie genres could have different types of reviews than others, making it a bad idea to neglect genre.
- Some reviews were inconsistent with given reviews, possibly confounding our models.

Future

In the future, we would like to train these models again with bidirectional RNNs, as sentiment could depend on all words throughout the sentence, not just the beginning.

In addition, we could reintroduce punctuation into the dataset to section reviews into sentences, so that we could more accurately implement a hierarchical attention model [1].

References

- [1] Yang, Zichao, et al. “Hierarchical Attention Networks for Document Classification.” Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, doi:10.18653/v1/n16-1174.