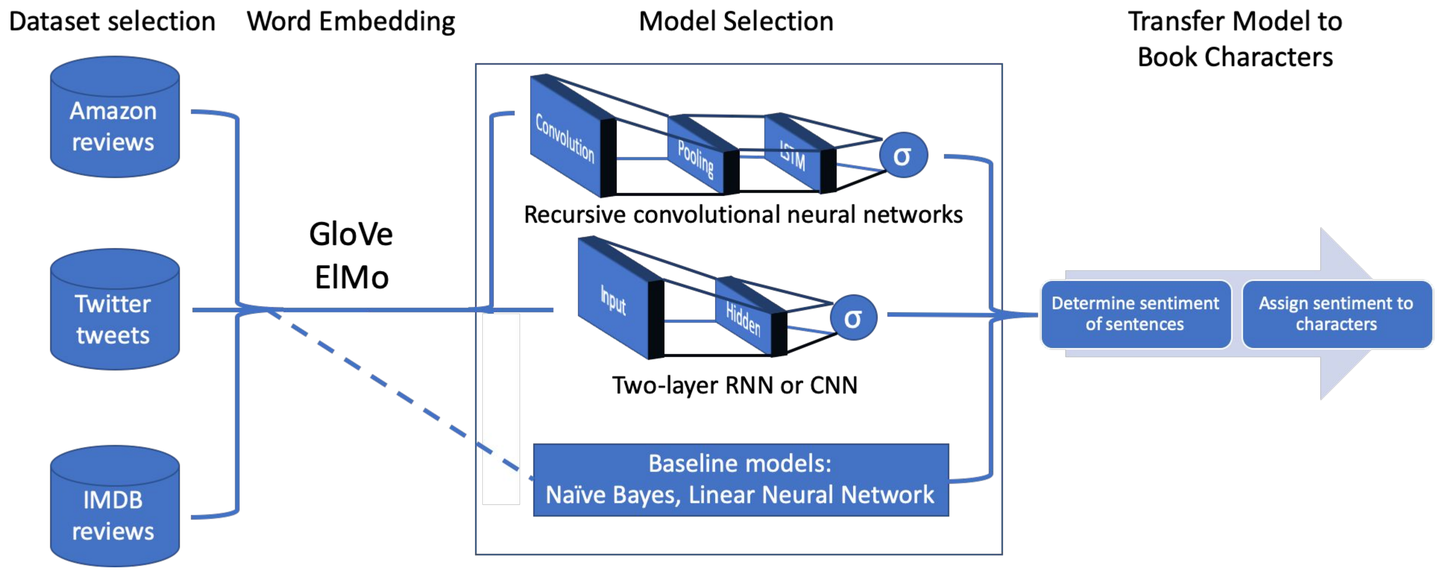
Application of Sentiment Analysis to Labeling Characters as Good or Evil

Ethan Curtis (curtie), Maria Karelina (mashka), Felipe Kettlun (fkettlun)
CS229 - Fall 2019

Introduction

Have you ever found yourself wondering whether Sauron or Pippin was the real villain of Lord of the Rings? As children, we were taught that everything could be divided into two categories: good and evil. We intend to use sentiment analysis to analyze characters from children's books on how good or evil they are.

We experimented with different machine learning models, training datasets, word embeddings and sentiment association algorithms to label characters as good or evil.



Results

From Ashputtel:
"*shake, shake, hazel-tree, gold and silver over me! then her friend the bird flew out of the tree, and brought a gold and silver dress for her, and slippers of spangled silk; and she put them on, and followed her sisters to the feast*"
sentiment = 0.86080
assigned to sisters

- Transferability to book characters is low independent of sentences/character assignment
- Sentences in complicated model seem to have average sentiment
- Character assignment in complicated models is also near 0.5 sentiment
- Multiple characters appear in same string
- Sentence sentiment doesn't reflect character values

Data and Features

Training datasets:
We used the following three datasets:
- **IMDb**: 50k movie reviews (50/50 positive/negative labels).
- **Amazon**: 970k book reviews.
- **Twitter**: 1.6 million tweets with emoticons removed.

Data set transferability test

Test/Train	IMDB	Amazon	Twitter
IMDB	0.776	0.673	0.499
Amazon	0.835	0.880	0.115
Twitter	0.563	0.117	0.955

Books dataset:
10 hand-labeled children's books.
Assigning sentences to characters:
- Split 1: Split book into sentences and assign to all characters which appear in that sentence
- Split 2: Make a sentence that surrounds the character name with 10 words on each side

Embeddings:
Pre-processing done with two different embeddings:
- **GloVe**: 200 dimensional embedding which produces vectors whose dot product corresponds to the log probability of co-occurence.
- **ElMo**: 512 dimensional context-dependent, character-based embedding where vectors are learned from a deep, bi-directional recursive model.

Embedding Testing

	CNN	CRNN
GloVe	0.675	0.682
ElMo	0.640	0.778

- Trained on subset of IMDb, test results averaged over IMDb and Amazon
- ElMo does not significantly outperform GloVe, except on Twitter tests (where all models failed to reach 50% accuracy)

Models

Naive Bayes: Class model trained using dictionary created from IMDd training set.
Linear NN: Neural network with five layers of 1k-2k neurons, ReLu activation and sigmoid function for the output.
CNN: Convolutional layer with 2 filters of size 64 and a max pooling layer of size 4.
CRNN: CNN+LSTM layer of size 70.

Sentiment Analysis Accuracy

Model	IMDB	Twitter	Amazon
Naive Bayes	0.842	0.586	0.801
Linear NN	0.893	0.580	0.759
CRNN + Glove	0.807	0.56	0.835

Transfer

Accuracy of sentiment labeling of character

Model	Books Split 1	Books Split 2
Naive Bayes	0.45	0.45
CRNN + Glove	0.40	0.50

Future Directions

- Investigate use of natural language processing for assigning sentiment to characters
  - Identify parts of speech
  - More complex models that can learn pronouns
- Identify labeled large databases that are more similar to novels
- Optimize parameters of neural networks
- Explore more feature embeddings

Conclusion

- It is easy to train models which perform well on data very similar to the training data
- Our efforts to transfer sentiment models to books were not successful
- We need to explore better methods of labeling characters and capturing context

References

Mika V. Mäntylä, Daniel Graziotin, Miikka Kuutila. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, Volume 27, 2018, Pages 16-32.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. GloVe: Global Vectors for Word Representation, 2014.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer. Deep contextualized word representations, *Proc. of NAAC*, 2018.

Alec Go, Richa Bhayani, Lei Huang. Twitter Sentiment Classification using Distant Learning, 2009.