Deep Tomato

CS230 Deep Learning Winter 2018
Petra Grutzik, Kate Park, Leslie Tu
pgrutzik@stanford.edu, katepark@stanford.edu, leslietu@stanford.edu

Background / Problem

- Can we predict the greatness of a movie before we watch it? When we scroll through Netflix feeds and similar sites, we consider posters, titles, and genre in order to decide whether movies are worth watching.
- Neural networks have successfully predicted movie, book, and album cover *genre* from its poster and text, but can we predict a movie *rating*?
- Goal: Build a deep neural network to predict if a movie is thumbs up or down from visual and textual cues

Dataset & Features

- Scraped 40K movies from Rotten Tomatoes website with BeautifulSoup Package
- 23K with "fresh tomato" score (aggregate of % critics rating fresh / total number of reviews)
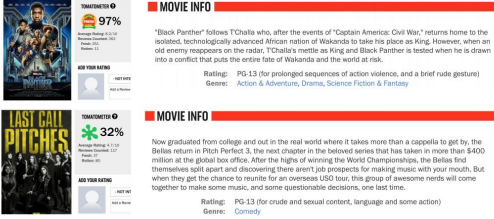
Features

- Colored movie posters sized: 224 x 224
  - Randomized saturation, brightness
- Movie Title, Movie Genre, Movie Description
- "Fresh Tomato" Score splits data 50/50 into 2 classes: score <=70, score > 70

Models

Text-based
- GloVe to represent words & LSTM with Dropout and Early Stopping, Adam optimizer

Image-based
- Conv 2D with Dropout, Data Augmentation, Early Stopping, and Adam optimizer

Cross-entropy Loss
$$-(ylog(p) + (1-y)log(1-p))$$

Models that did not work
- NN with Bag of Words & N-gram counts
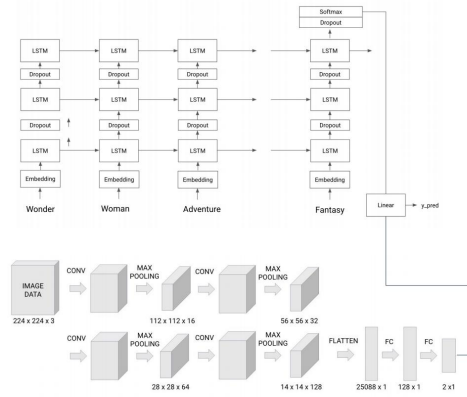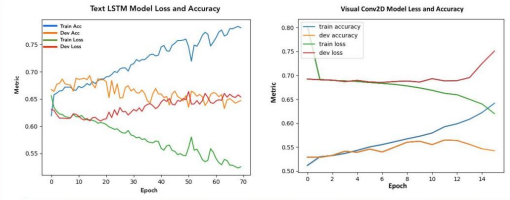- AlexNet

Figure 1. Combined Model Architecture

Figure 2. Model performances on train/dev sets (accuracy and loss)

Results

	Train Acc %	Dev Acc %	Test Acc %
Baseline (Random)	50	50	50
GloVe LSTM	69.37	67.59	66.51
Conv2D	61.22	56.59	56.85
AlexNet	53.9	52	--
Combined GloVe LSTM + Conv2D	69.80	67.99	**66.87**

Discussion

- Combined Model performed best suggesting images provide additional information to text features
- AlexNet did not work -- purpose to classify images, did not translate to judgement of movie
- Including description hurt performance
- Difficult problem because of subjective nature

Future

- Train larger model with description
- Reduce overfitting in CNN
- Further research in combining models

References

A. Krizhevsky, I. Sutskever, and GE Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," *NIPS*: 1106-1114. 2012.
C. Sun, "Predict Movie Rating," NYC Data Science. August 2017.
J. Pennington, R. Socher, and C. D. Manning. GloVe: Global Vectors for Word Representation. 2014.
J. Libeks and D. Turnbull, "You can Judge an Artist by an Album Cover: Using Images for Music Annotation," IEEE Multimedia 18(4):30-37. May 2011.