

NLP lab final presentation: predicting IMDB movie score of an individual based on their review

Seyithan Dag, Universität Bonn

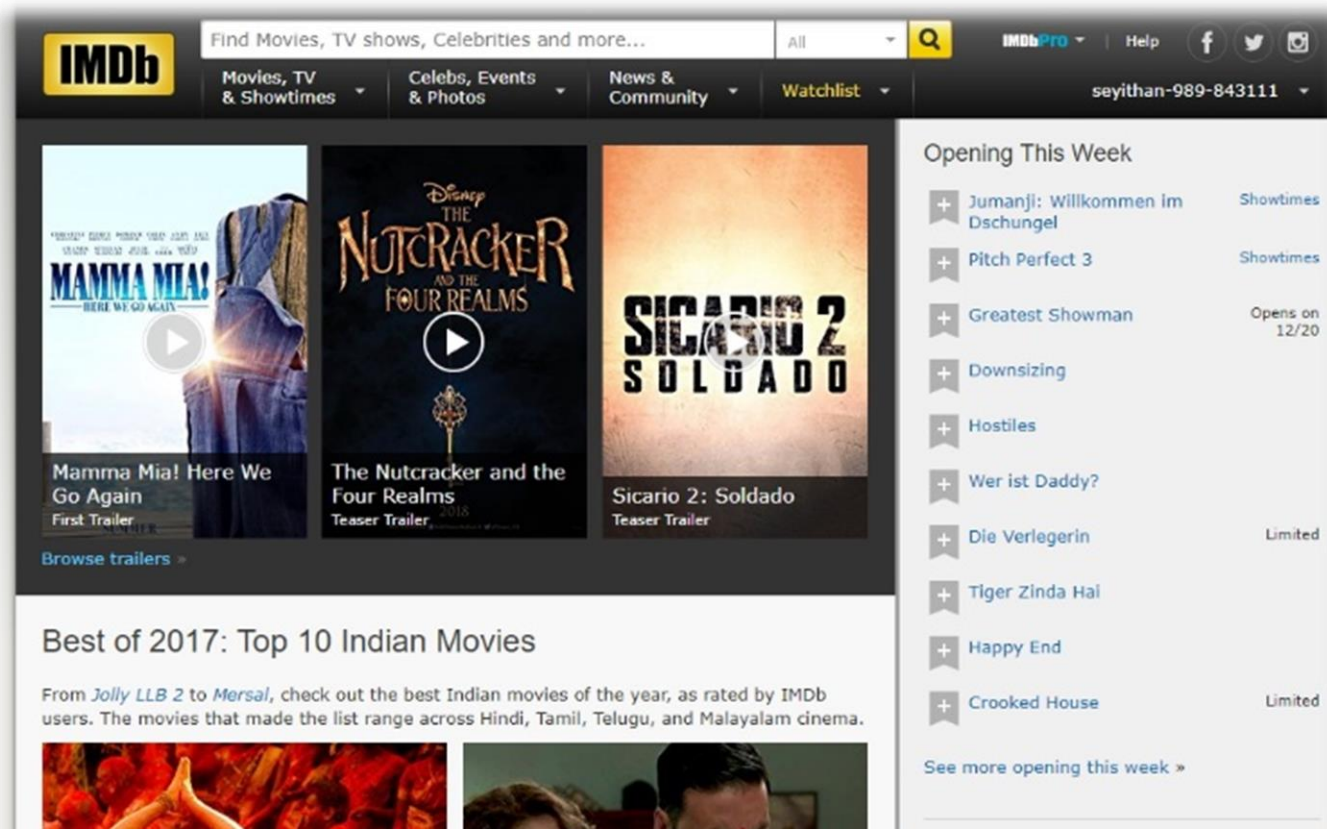
15.02.2018

today's talk

- topic / problem statement
 - IMDB
- drawbacks and problem refinement
- long-short-term memory
- convolutional neural networks
- sentiwordnet
- proposed model
- thoughts on the proposed model
- conclusion

topic / problem statement

predicting IMDB movie score of an individual based on their review



topic / problem statement

predicting IMDB movie score of an individual based on their review



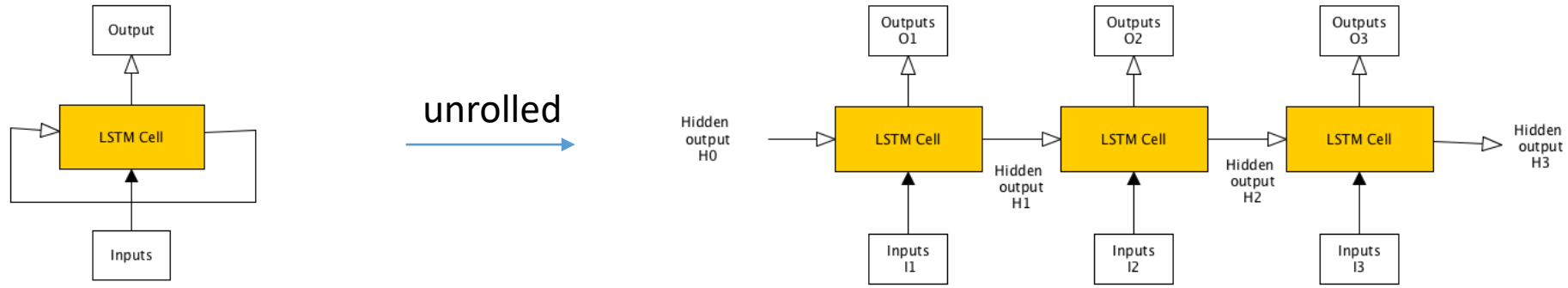
drawbacks and problem refinement

- no datasets available with review + scores
- datasets are all concentrated on reviews being positive or negative
- the problem was reduced to finding out whether a review being positive or negative, i.e., having the score of [0,5] or [6,10]

long-short-term memory (LSTM)

- specialized recurrent neural networks (RNNs)
- capable of “remembering” for long periods of time/timesteps
- common architecture composes of memory cell, input/output gates and a forget gate
- unlike conventional RNNs, the architecture solves the problem of vanishing gradient during training
- very capable of sequence classification

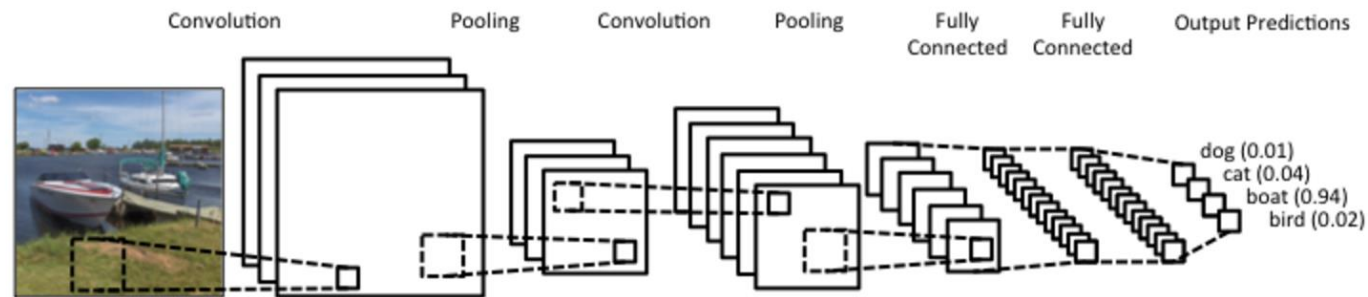
long-short-term memory (LSTM)



convolutional neural networks (CNN)

- deep, feed-forward neural networks
- consists of an input and an output layer, as well as multiple hidden layers. hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers and normalization layers
- very capable of image recognition/classification

convolutional neural networks (CNN)



sentiwordnet [1]

- sentiwordnet is a lexical resource for opinion mining.
sentiwordnet assigns to each synset of wordnet three sentiment scores: positivity, negativity, objectivity
- in short, capable of determining the sentimental “value” of a word
- works for nouns, verbs, adjectives and adverbs
- included in NLTK

technical details

- all implementation done in python
- tflearn (a tensorflow API) was used for all of the learning duties, due to extremely friendly structure
- nltk was used for sentiwordnet databank

today's talk

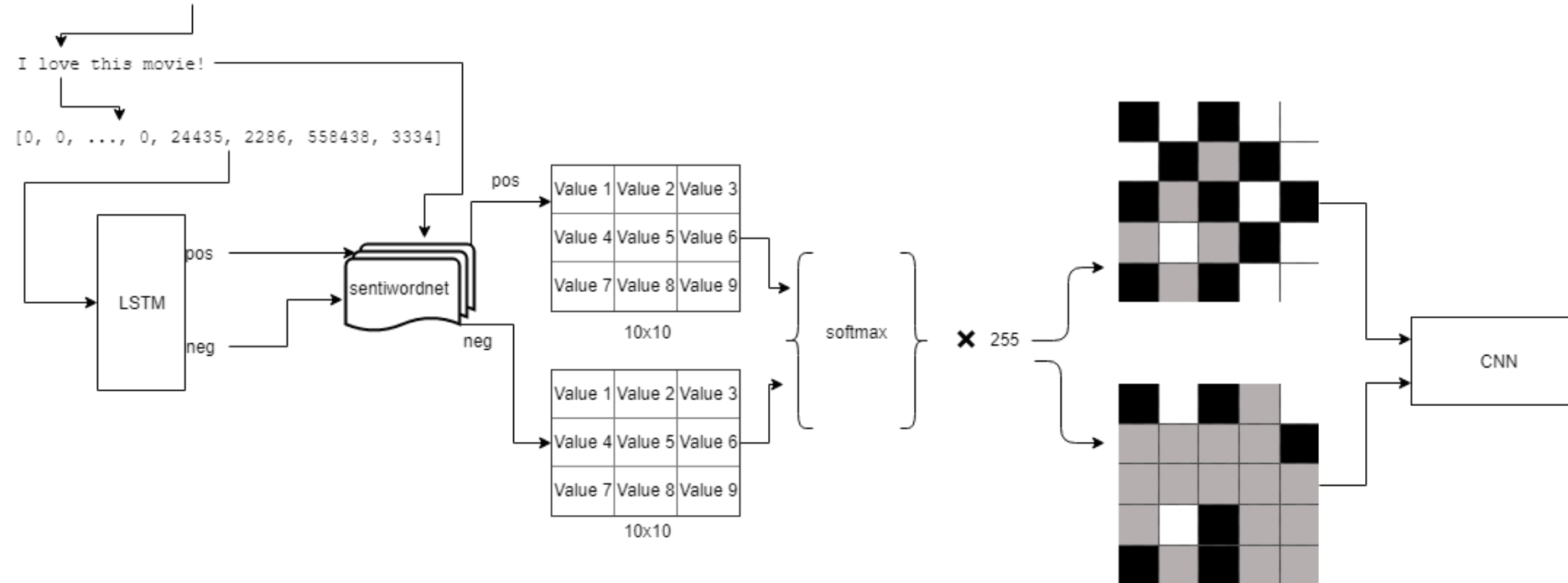
- topic / problem statement
 - IMDB
- drawbacks and problem refinement
- long-short-term memory
- convolutional neural networks
- sentiwordnet
- proposed model
- thoughts on the proposed model
- conclusion

proposed model

- combining sentiwordnet, LSTMs and CNN
- basic idea is turning text into sentiment maps/images through the help of sentiwordnet and LSTMs, then feeding to a CNN for classification
- three steps:
 1. LSTM vote on polarity
 2. construction of 10 x 10 sentiment map
 3. CNN feed

proposed model: overlook

It was amazing. The fast-paced action, the screenplay... I love this movie! It is, in fact, surprising that carpenter managed to fit so many pieces of his previous work into this film in such an admittedly novel way...



proposed model: LSTM training

model architecture:

- sequence length fixated at 100
- therefore, 100 cells (when unrolled)
- 128 hidden units in a single cell
- dropout layer after each cell, with 0.8 dropout probability
- fully connected layer at the end
- max 50 length sentences, each word encoded as a number by indexing of ~470,000 words in the english language
- each 50 word long encoded sequence turned into 100 long embedding vector

proposed model: LSTM training

training, first trial:

- twitter sentences
 - positive_emotions = [*'Neutral', 'Optimism', 'Anticipation', 'Love', 'Surprise', 'Joy', 'Submission', 'Ambiguous'*]
 - negative_emotions = [*'Anger', 'Disgust', 'Sadness', 'Aggression', 'Contempt', 'Disapproval', 'Remorse', 'Fear'*]
- Results: LSTM not capable of even learning.

proposed model: LSTM training

training, second trial:

- stanford sentence polarity dataset
- each sentence labelled as being positive or negative
- results:
 - very high accuracy on test data
 - however, failing in even simple sentences

proposed model: LSTM training

training, second trial:

- results:
 - very high accuracy on test data
 - however, failing in even simple sentences

```
This is the worst movie ever 1  
This movie is bad 1
```

```
I would rather sleep on my couch than go see this movie 0  
If theres a heaven for bad movies, Deuces Wild is on its way. 0  
This is the best movie ever made, action never stops! 1  
The acting is good but the directing sucks 0
```

proposed model: LSTM training

training, third trial:

- enriching the dataset with more reviews from yelp and amazon
- same structure, but for different topics
- results:
 - very high accuracy on test data
 - very high accuracy on simple sentences, or any arbitrary sentence!

proposed model: LSTM training

training, third trial:

- results:
 - very high accuracy on test data
 - very high accuracy on simple sentences, or any arbitrary sentence!

```
- Thats the last time I will rush out to watch a Star Wars movie 0
- one word describes this movie - POINTLESS. 0
- I would rather sleep on my couch than go see this movie 0
- There is so much wrong with it that you really just have
to see it for yourself, although I would recommend not paying money for it. 0
- By the end of the night, hours after we left the theater, I felt totally gutted and crestfallen
  to realize what the Star Wars sequels had become, because Rian Johnson had to "Age of Ultron" the series 0
- I loved it 1
- Not only was his behavior completely uncharacteristic,
but his fight scene against Emo Ren was a sham and his death was utterly meaningless 0
```

proposed model: sentiment maps

- there are multifarious ways of construction
- the method used is to get the average sentiment score of a given sentence, turn in into a probability and multiply it by 255 to get the pixel value in the map
- map is capable of handling 100 sentences, which is fairly good amount for any given review as reviews tend to have 20-30 sentences on average
- usage of padding (0 value, i.e. white) for the leftover pixels
- generality in question

proposed model: CNN training

model architecture:

- in total: 3 convolution layers, 2 max pooling layers, 2 fully connected layers, 1 dropout and 1 regression layer (at the end)
- rectified linear units (ReLU) used in all layers but the regression layer.

proposed model: CNN training

training:

- imdb movie review dataset, 1000 positive and 1000 negative reviews
- each review turned into aforementioned sentiment maps

results:

- to come! (first results are not that good, needs tweaking)

thoughts on the proposed model: sentiment maps

- is the proposed sentiment maps capable of capturing complex sentimental relationships?
- what would be the results if instead of average sentence scores, the sentiment score of each word was turned into a map?
- since the concentration is only on the movies, how would it be to create the sentiment map in terms of movie features, i.e., a row for director, screenplay, cinematography, actors, score?

any

questions?

references

- [1] esuli, a., sebastiani, f., (2010). sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. *proceedings of the 7th conference on language resources and evaluation (LREC 2010), valletta, my, 2010, pp. 2200-2204*