

[shagunsodhani](#) / **Learning to Generate Reviews and Discovering Sentiment.md**

Last active 2 years ago • Report abuse

 Star

<> Code

 Revisions 3 Stars 8 Forks 2

Notes for "Learning to Generate Reviews and Discovering Sentiment" paper

 [Learning to Generate Reviews and Discovering Sentiment.md](#)

Learning to Generate Reviews and Discovering Sentiment

Summary

The authors train a character-RNN (using mLSTM units) over Amazon Product Reviews (82 million reviews) and use the char-RNN as the feature extractor for sentiment analysis. These unsupervised features beat state of the art results for the dataset while are outperformed by supervised approaches on other datasets. Most important observation is that the authors find a single neuron (called as the sentiment neuron) which alone achieves a test accuracy of 92.3% thus giving the impression that the sentiment concept has been captured in that single neuron. Switching this neuron on (or off) during the generative process produces positive (or negative) reviews.

Notes

- The paper aims to evaluate if the low level features captured by char-RNN can support learning of high-level representations.
- [Link to the paper](#)
- [Link to the blog by OpenAI](#)
- The paper mentions two possible reasons for weak performance of purely unsupervised networks:
 - Distributional issues - Sentence vectors trained on books may not generalise to product reviews.
 - Limited Capacity of models - Resulting in representational underfitting.

- Single layer with 4096 units.
- Multiplicative LSTM units are used instead of standard LSTM units as they are observed to converge faster.
- Compact model with a high ratio of compute to total params (1.12 butts per byte)
- L1 penalty is used instead of L2 as it reduces sample complexity when there are many irrelevant features.
- Found a single neuron (sentiment neuron) which alone captures most of the sentiment concept.
- **Capacity Ceiling**
 - Even increasing the dataset by 4 orders of magnitude leads to a very small improvement in accuracy (~1%).
 - One possible reason could be the change in data distribution - trained on Amazon Reviews and tested on Yelp Reviews.
 - Similarly, the linear model (trained on top of feature vectors) has its own limitations in terms of capacity.
- The model does not work well on out of domain tasks like semantic relatedness over image descriptions.
- The paper shows that positive (or negative) reviews can be generated by switching the sentiment neuron on (or off) during the generative process.
- A [tweet](#) by [@AlecRad](#) says that zeroing the sentiment neuron drops the performance only by 2% on SST and 10% on IMDB indicating that the network has still learnt a distributed representation.

Open Questions

- Is this phenomenon of disentangling of high level concepts specific to sentiment analysis?
- How do we explain the compression of almost all the sentiment in a single unit?
- Use of hierarchical models for increasing the capacity of char-RNN.