Product Aspect Extraction for Sentiment Analysis without using Parsers

Samik Some & Narendra Roy

Under: Prof. Amitabha Mukherjee

Motivation

- Sentiment analysis and summarization of customer reviews is important for automatically gauging the popularity/demand and customer satisfaction of a product.
- Current algorithms for aspect extraction from reviews depend on some form of parsing, such as POS tagging or dependency parsing.
- This inherently introduces language related dependencies in the algorithms making them hard to use for languages which do not have proper parsers.
- We propose a method to extract aspects without parsing the reviews in any manner.

Previous Works

Some previous works in this area are:

- Mining and summarizing customer reviews [2]
- Rule-Based Approach to Aspect Extraction from Product Reviews [3]
- Aspect Term Extraction for Sentiment Analysis: New Datasets, New Evaluation Measures and an Improved Unsupervised Method [4]

All of these use some form of parsing of reviews before aspect extraction.

Proposed Algorithm

Our algorithm is based on Zhou and Xu's implementation of end-to-end SRL without parsers. [1]. We propose a multilayer LSTM network for binary classification of aspects and non-aspects.

- The input to the first LSTM layer is a 300-dimensional word vector representation of a word in the sentence/phrase in which we wish to tag aspects and non-aspects.
- More LSTM layers stacked on top of this to form a deep network. Each LSTM layer has 300 dimensional input and output.
- The output from the last LSTM layer is passed through a Dense layer to reduce it to a single dimension, followed by a sigmoid activation for binary classification.

We tested the network with two LSTM layers with sigmoid and hard sigmoid activations. Word2Vec was used to obtain the word vector representations and the LSTM model was implemented in python using Keras framework.

Following this sentiment analysis can be done based on the aspects obtained and finally a summarized review may be generated based on the results of aspect based sentiment analysis.

Data Preprocessing

We preprocess the review data for training by.

- Removing all punctuations.
- Splitting reviews at fullstops, commas, or whenever we get 10 non-aspects after an aspect.
- Making sure only phrases with at least one aspect are included for training

Datasets

We used the following data set for training and testing our algorithm.

- Aspect tagged Amazon product reviews dataset [5]

For Word2Vec we used the pre-trained GoogleNews model available on their website.

Results

We trained our network on aspect-tagged review data of 3 products, Nokia 6610, Nikon coolpix 4300 and Canon G3.
For testing we used reviews for two other products, Creative Labs Nomad Jukebox Zen Xtra 40GB and Apex AD2600 Progressive-scan DVD player.
The results are as shown:

Product	Precision	Recall	F1-Score	Accuracy
Zen 40GB	74.58	63.00	68.30	97.24
Apex DVD	74.33	51.85	61.08	95.54

Table 1: Results

The precision, recall and F1-score are defined with respect to aspects.

Example

An example of what aspect extraction does is as follows:

Input: "The size of the mobile is very handy however the screen is too small."
Output: Aspects - size, screen

The actual output is a set of labels which identify every word as either an aspect or non-aspect

Structure

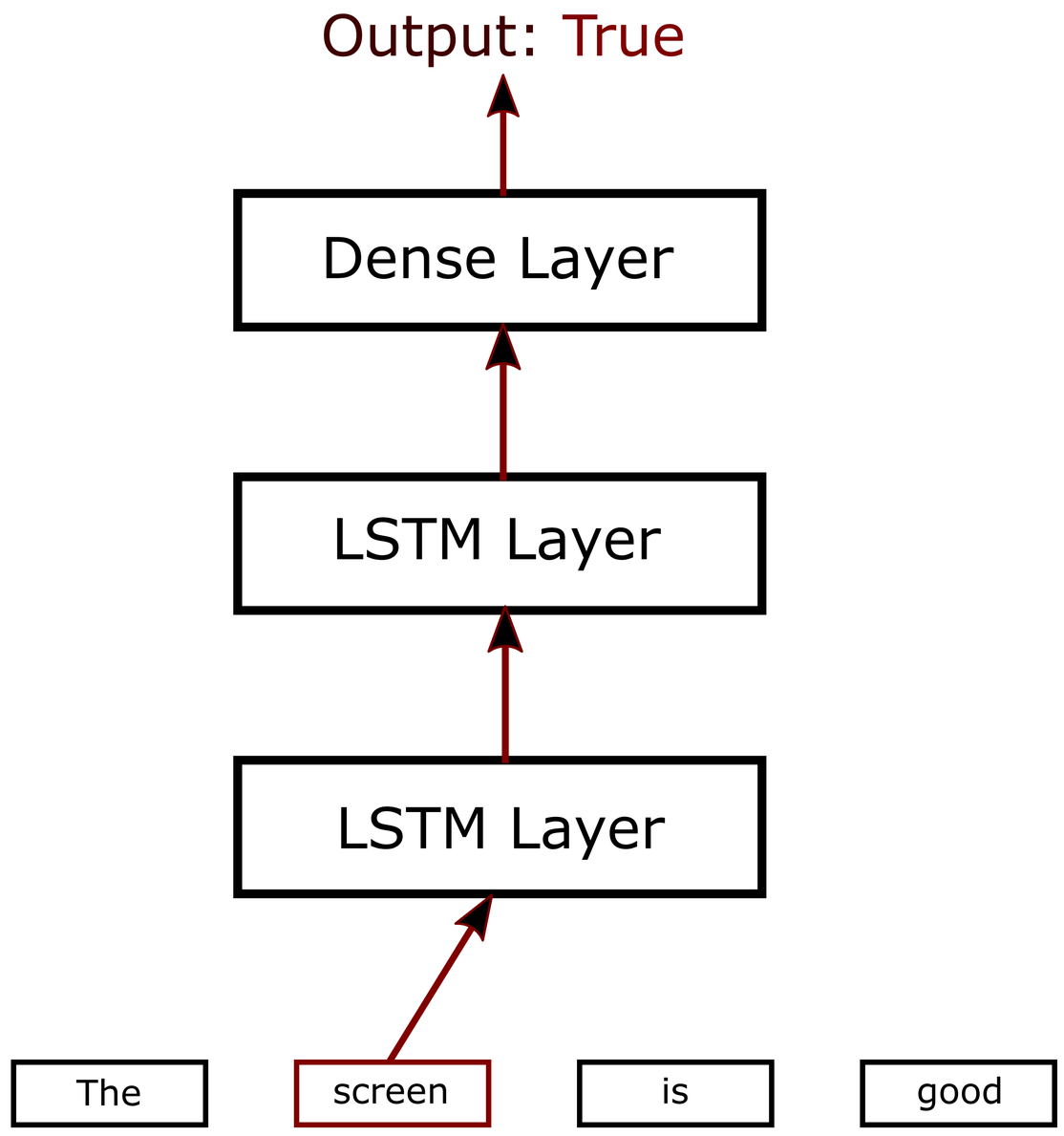

Fig. 1: The structure of the proposed network

This figure demonstrates structure of the network we use in our algorithm. The LSTM layers can be stacked more than twice if needed, and better results may be obtained with BD-LSTMs instead of simple LSTM layers.

Conclusion

We conclude by observing that using stacked LSTMs for end-to-end aspect extraction without parsers can work. Our dataset however was very small and the results can likely be improved by using a larger training dataset and also by using BD-LSTMs instead of simple LSTMs.

Since this method is completely parsing free and uses only LSTM layers, this method can be used for reviews in any language, by simply training the model on the respective language's dataset.

References

[1] Zhou, Jie and Xu, Wei. *"End-to-end Learning of Semantic Role Labeling Using Recurrent Neural Networks"* (2015)

[2] Minqing Hu and Bing Liu. *"Mining and summarizing customer reviews"* (2004)

[3] Poria, Cambria, Wei Ku et al. *"Rule-Based Approach to Aspect Extraction from Product Reviews"* (2014)

[4] Pavlopoulos and Androutsopoulos. *"Aspect Term Extraction for Sentiment Analysis: New Datasets, New Evaluation Measures and an Improved Unsupervised Method"* (2014)

[5] Amazon product reviews dataset , `http://curtis.ml.cmu.edu/w/courses/index.php/Amazon_product_reviews_dataset`