UNIVERSITY OF TRENTO - Italy
Information Engineering
and Computer Science Department

Master Degree in Computer Science

Natural Language Processing
AA 2015-2016

Sentence compressor

Aliaksandr Siarohin

October 11, 2016

Abstract

This is a description of project on natural language processing.

1 Introduction to the designed system

In this project, I implement a neural network solution to deletion-based sentence compression where the task is to translate a sentence into a sequence of zeros and ones, corresponding to token deletion decisions. I use architecture similar to one proposed in [1]. My network is multiple LSTM layers stacked together. This network takes words embeddings from word2vec as an input and produces the probability of token deletion decision.

2 The analytic model

2.1 Pre-processing

The data were initially in json format. To parse json I use script provided by teaching assistant. After parsing the data become the list of lists with tokens. Each token is tuple with 4 field (word, tag, stem, token deletion decision). I populate this tuple with an additional field, word embedding. This embedding comes from google-news word2vec `https://code.google.com/archive/p/word2vec/`.

2.2 Formal description of neural network architecture

I use LSTM network for this problem. The input to the network is word embedding and one hot encoding of word tag. Then there is 2 bidirectional LSTM layers separated by dropout layer. Then there is a dense layer with a dropout. And finally dense layer with softmax non-linearity. The network structure is in fig. 1.

3 Software description

To implement this project I use python library's theano+lasagne. This framework can build a graph of execution, and then evaluate it on gpu.

3.1 Used Functions

I use different network layers from lasagne library (LSTMLayer, DropoutLayer, DenseLayer ...). Also, I use adam function from lasagne library for fitting the network.

4 Experiment Description

4.1 Data

Data consist of 10000 sentences. For each word in sentence there is stem form of this word and tag of this word. Each word has associated token deletion decisions (0 or 1). For the experiment I split the data in 3 sets: train, dev and test. The test set consists of 1000 sentences from the beginning, train and dev set is a random split of the rest 9000 sentences, 1000 for dev and 8000 for train.

4.2 Training process

I train the network using adam method. The batch in my case is 1 sentence. The sentences are feed to network in random order. The training process takes 4 epochs. After each 1000 sentences, I compute score for dev set and save network parameters. The network with the best score on validation set will be resulting.

4.3 Network parameters

The size of word embedding is 300 (Because in google-news dataset it is 300). The number of units in LSTM layer is 200, as well as in tag embedding. The size of last dense layer is also 200. The last layer gives us probability, so to get 0, 1 predictions we need some threshold. In my case this threshold maximize the f1 score on dev set, and it is equal to 0.505582.

4.4 Source code

Sorce code is single ipython notebook. I upload it on github. You can view it using the following link: `https://github.com/AliaksandrSiarohin/sentence-compressor`.

5 Result presentation

The result score on the test set:

- Per token accuracy: 0.827522
- Per token f1 score: 0.768665
- Fraction of right compression: 0.163000

The learning curves is in fig. 2. The Precision/Recall curve is in fig. 3.

The table with compression examples is in table 1 for good compression, and table 2 for bad compression.

6 Comparison with other solutions

This problem has already been addressed in [1] and [2].

The first one use approche similar to my own, but with 3 LSTM layers and additional features such as:

- The embedding vector for the parent word
- The label predicted for the last word
- A bit indicating whether the parent word has already been seen and kept in the compression
- A bit indicating whether the parent word has already been seen but discarded
- A bit indicating whether the parent word comes later in the input

In this paper, they also use a larger corpus for training about 2 million sentences. Their f1 score 0.81.

The second one also using 3 LSTM layers and same amount of data as my model, but they take eye-tracking recording as an additional feature. Their f1 score 0.8097.

7 Conclusion

In this project, I implement a neural network solution to deletion-based sentence compression. Resulting network shows good results on phrases with one subject and predicate. But not very good at phrases without subject, as well as complex phrases with 2 or more subjects.

References

[1] K. Filippova, E. Alfonseca, C. Colmenares, L. Kaiser, and O. Vinyals, "Sentence compression by deletion with lstms," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP'15)*, 2015.

[2] S. Klerke, Y. Goldberg, and A. Søgaard, “Improving sentence compression by learning to predict gaze,” *CoRR*, vol. abs/1604.03357, 2016. [Online]. Available: http://arxiv.org/abs/1604.03357

Table 1: Good compression

Starting sentence	Reference	Predicted
Actress NAOMI WATTS walked out of a radio interview as she was promoting her new movie DIANA	NAOMI WATTS walked out of a radio interview	NAOMI WATTS walked out of a radio interview
Nice Ride bikes will make their spring debut on city streets Saturday and riders can now keep them for twice as long	Nice Ride bikes will make their spring debut	Nice Ride bikes will make their spring debut
Not following its key ally Saudi Arabia 's on Syria Pakistan Thursday asserted that Syrian people should decide their future set up	Syrian people should decide their future set up	Syrian people should decide their future set up
The Egyptian embassy in Zambia has secured the release of 23 Egyptian nationals held in Zambian prisons for illegal migration and trade	The Egyptian embassy in Zambia has secured the release of 23 Egyptian nationals	The Egyptian embassy in Zambia has secured the release of 23 Egyptian nationals

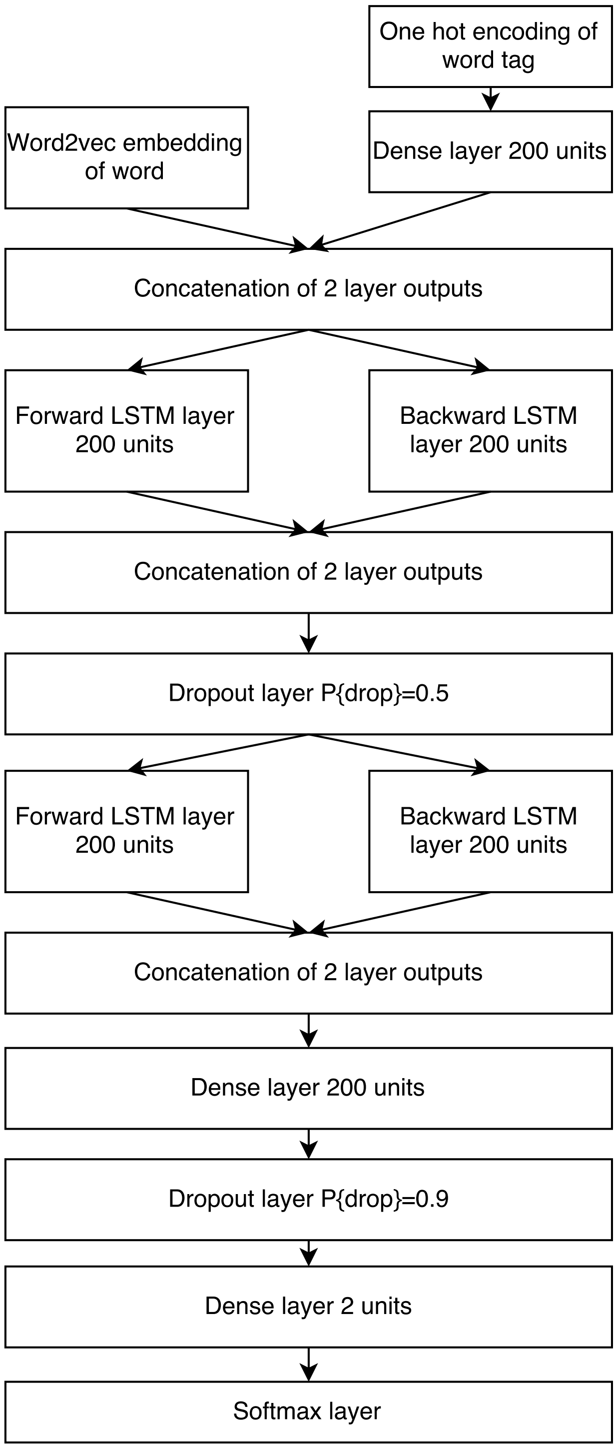

Figure 1: The network architecture

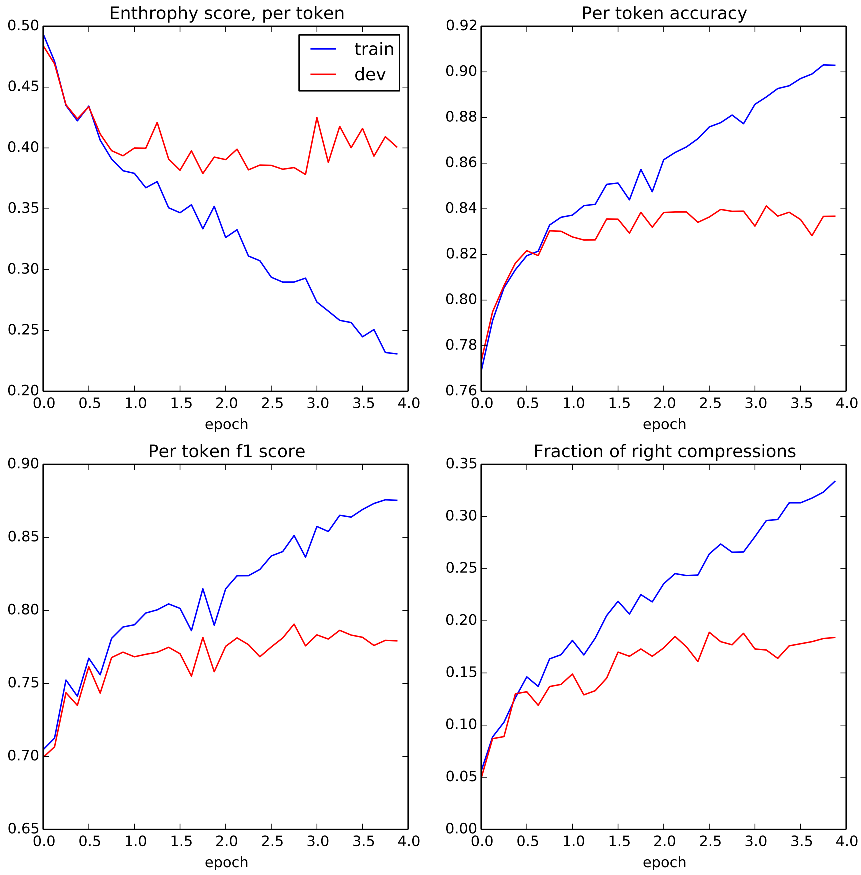

Figure 2: The learning curves on train and dev sets

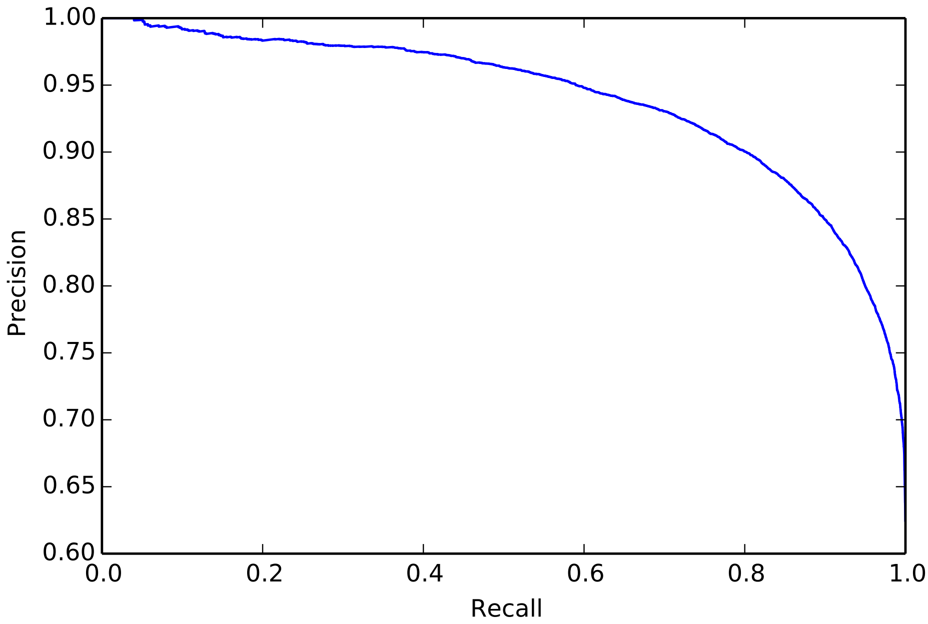

Figure 3: The Precision/Recall curve

Table 2: Bad compression

Starting sentence	Reference	Predicted
However that means high quality players are being left on the bench so which of this Newcastle quintet should feel most hard done by	that means so which of this Newcastle quintet should feel most hard done by	that means high quality players are being left on the bench
The second wild is the picture of all five of the Girls people roulette online chat Guns	people roulette online chat Guns	The second wild is the picture of all five of the roulette Guns
Stocks to watch on the Australian stock exchange at the close on Tuesday	Stocks to watch at the close on Tuesday	Stocks to watch on the Australian stock exchange
Stocks to watch on the Australian stock exchange at the close on Wednesday	Stocks to watch at the close on Wednesday	Stocks to watch on the Australian stock exchange
Boaters can become frustrated when a repair or upgrade takes a long time but delays are often a simple result of supply and demand	a repair takes but delays are often a simple result of supply and demand	Boaters can become frustrated