CSA250: Deep Learning Project 3 Report

Rajat Nagpal
M.Tech SE
rajatnagpal@iisc.ac.in

Abstract

In this project, two types of features have been used, namely, infersent features by facebook research for deep learning task and TF-IDF features for logistic regression classification. The classification task is natural language inference task to get the similarity and dissimilarity between two sentences. Classifiers are trained using SNLI dataset.

1 Introduction

Natural language inference is basic building block for many important tasks in natural language processing. The task is divided into two parts, extraction of features and the choice of classifier.

In this project, we are going to look at different aspects of feature extraction for natural language inference task. We will be working with SNLI dataset with 549367 training examples, 9842 validation examples and 9824 test examples. CSV files have been created by preprocessing the data where first column refers to sentence 1 and second column refers to sentence 2 and the third refers to the label associated with it. This is a three class classification problem where classes are defined to be "Neutral", "Contradiction", "Entailment".

2 Analysis

2.1 TF-IDF

In this method, term frequency - inverse document frequency have been used as feature representation of the words present in the dataset. The term frequency of a word in a document is referred to the number of times it has appeared in the document and document frequency refers to the number of documents in which the word is present. The weighing formula is as follows :

$$w_{i,j} = tf_{i,j} \times log(\frac{N}{df_i})$$

$tf_{i,j}$ = Number of occurrences of i in j
df_i = Number of documents containing i
N = Total number of documents

After getting the TF-IDF features, we apply a logistic regression classifier on the three class classification problem. The accuracy for this model is 69.24 percent.

2.2 Deep model

The features used in this task are generated by the architecture given by facebook research, i.e., Infersent. This architectures takes word embeddings as input and uses sequential methods to generate sentence embeddings of dimensions 4096. We have used pretrained word embeddings known as fastText word embeddings to generate sentence embeddings. The produced sentence embeddings are then fed to a fully connected 3 layer neural network. Following three ways have been implemented to generate input vectors for this feed forward neural network:

- Element-wise subtraction.
- Element-wise multiplication.
- Concatenation

Following is the accuracy table for these three vector representations:

Methods	
Element-wise subtraction	81.34
Element-wise multiplication	78.44
Concatenation	81.32

Table 1: Accuracy score for different representations in deep learning model

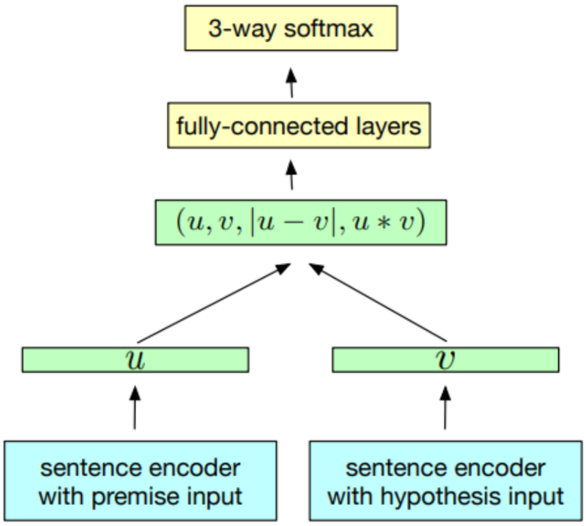

Figure 1: Architecture given in the paper [Con+17] by facebook research

Hyper-parameters and methods	
Learning rate	0.01
momentum	0.9
Batch-size	64
Epochs	30
Regularization factor	0
Hidden layers	3
Optimizer	Adam
Output	Softmax
Loss	Cross-entropy
Activation function	Relu
Neurons in HL 1	1024
Neurons in HL 2	256
Neurons in HL 3	64

Table 2: Experimental conditions used for the Fully-connected layer

We have used Bi-LSTM layer with mean/max pooling for the sentence encoder shown in the figure above. The hidden representation is the concatenation of the forward LSTM hidden layer and the backward LSTM hidden layer. We have used two ways to get a fixed size hidden representation, i.e., either by calculating the average of hidden vectors in each direction or by taking the maximum out of it.

The training loss and the validation set loss curve is given as under:

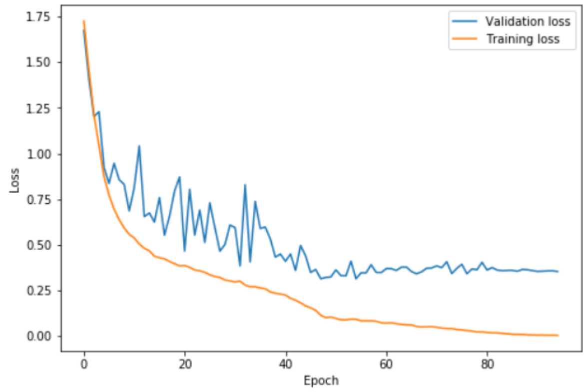

Figure 2: Training and validation loss for the deep learning model

The key thing to notice is how this curve is decreasing over time. Hence, our model is learning the representations and the decrease in validation loss shows that our model is generalized too. The model with best validation accuracy was taken to get the test set output file.

3 Conclusion

It is found that the semantic and syntactic representation of features in the deep model plays a vital role in determining the relation between two sentences. It is important to get a good representation of the sentences so that our network can efficiently learn the differences and the similarities. On the other hand, although TF-IDF features produce a satisfactory result but they are not good enough for this kind of classification task.

References

[Con+17] Alexis Conneau et al. "Supervised learning of universal sentence representations from natural language inference data". In: *arXiv preprint arXiv:1705.02364* (2017).