Deep Learning A3
Natural Language Processing

M&M:

Mehmet Ugurbil
Courant Institute
New York University
mu388@nyu.edu

Mark Liu
Courant Institute
New York University
ml4133@nyu.edu

Abstract

We try to use a convolutional neural network to predict the numerical star rating of a Yelp review when given only the corresponding text review.

1 Data set

We use the Yelp academic data set from the Yelp Dataset Challenge, which is a set of business reviews each labeled with a numerical star rating 1 through 5.

2 Pre-processing

We use the GloVe database to first turn every word in the review into a n-dimensional vector. Then we stack all these vectors in order to form a 2D $M \times n$ dimensional vector, where M is the number of words in the review. This matrix is then padded with vectors of 0s until it reaches a large fixed length K which is constant across reviews. In this way, we turn every review into a $K \times n$ dimensional vector.

3 Convolutional Neural Network

3.1 Architecture

Our neural network is 5 layers. The first three layers are convolutions followed by max pooling. Then there are two linear layers. The first convolutional layer has step size 5, the second has 4, third has 2. After each convolution there is max pooling with step sizes of 5, 4, 3 in order. Each convolutional layer doubles the feature size. Then we have the fully connected layer which turns takes in all the features and returns 16 new features. The last fully connected layer takes these 16 into 1 feature which is the classification. Both linear layers have dropout with probability 0.2. Then it is fed into a logsoftmax.

3.2 Learning Procedure

Stochastic Gradient Descent was used for the learning procedure of the neural net. We used a batch size of 128, momentum of 0.9, initial learning rate of 0.01 and learning rate decay of 10^{-7}.

The energy function here is set as the negative log likelihood function. The energy function is given as follows for input data $x = (x_1, ..., x_n)$, label $y = (y_1, ..., y_n)$, and the model function f, that is $f(x)$ is the output of the model.

$$E(x, y) = -\sum_k y_k \log f(x_k)$$

The loss function is given as the sum of the energy functions for the input data.

$$L(X) = \sum_i E(x_i, y_i)$$

4   Results

References

[1] LeCun, Y., Bottou, L., Genevieve B.O., & Muller K.R. (1998) *Efficient BackProp* New York: Springer.

[2] Zhang, X., LeCun, Y. (2015) *Text Understanding from Scratch*, ArXiv.

[3] Collobert, R. et al (2011) *Natural Language Processing (almost) from Scratch*, ArXiv.