Hashtag Prediction using Convolutional Neural Networks

March 9, 2017

1 Introduction

The task of predicting hashtags given microblogs have received a considerable amount of attention over the past few years. However, the task of recommending hashtags typically rely on hand-crafted features. With the advancement of deep learning, several models have been proposed specifically for this task. In recent years, Convolutional Neural Networks (CNNs) have been used in several NLP tasks. In this context, CNNs would help to identify the feature maps required to encode a microblog so as to predict hashtags with a considerable amount of accuracy. Our work is along the lines of [?]. The proposed method uses an attention based mechanism alongside CNNs to incorporate the importance of trigger words for a microblog directly into the model. In the next section, we present a summary of the methodology required for this project.

2 Objective

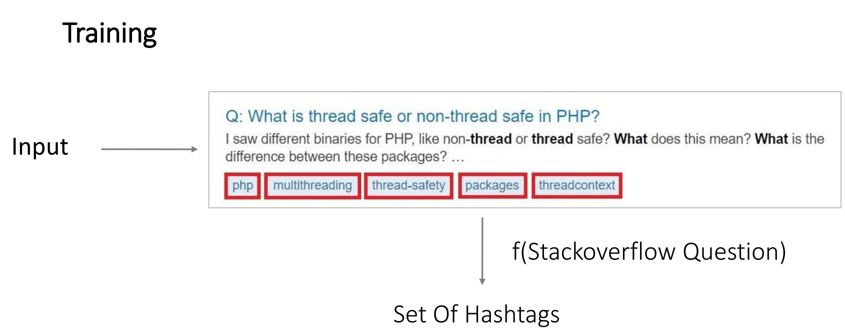

Figure 1: Training Input and Output

Given a microblog m, our objective is to learn a function $f : M \rightarrow h$ where M represents the domain of all possible microblogs and h represents a ranked set of hashtags. Our input is a set of stack overflow questions and corresponding sets of hashtags for each question. We shall adopt a supervised learning procedure which is depicted in Fig. 1.

3 Dataset

We are using the dataset provided by our mentor for the aforementioned work. The given dataset contains a total of 109886 samples. The average length of the questions per sample is 170 and the average number of tags per question for the dataset is 4. We shall first learn the word embeddings following the method proposed in [**?**]. This will be followed by implementing the work reported in [**?**]

4 Methodology

Given the set of questions and tags, the natural question that follows is how do we encode text data so that it can be used while training CNNs. This can be achieved by learning word embeddings which is discussed as follows.

4.1 Learning Word Embeddings

Learning word embeddings require training a simple neural network with a single hidden layer. The objective is to learn the weights of the hidden layer W which will actually serve as the word vectors. This is depicted in Fig. 2. Once we

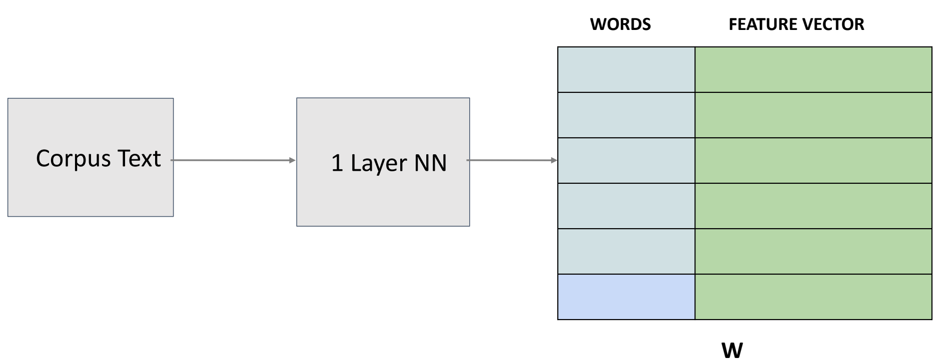

Figure 2: Learning Word Embeddings

have obtained the word vectors for each of the distinct words in the corpus, we shall collectively train using two different methodologies -

- **Attention Based Local CNN** - We shall use an attention-based CNN architecture which will be used to highlight specific parts of one input microblog.
- **Global CNN** - This entails using a CNN architecture to model the entire microblog.

Details regarding the above two methodologies are discussed as follows.

4.2 Attention Based Local CNN

Given an input microblog m, the word vectors $w_i \in \mathcal{R}^d$ are obtained using the word embedding approach. Once we have the word embeddings we slide an attention window M^l of size $h \times d$ across the set of word vectors and compute scores for each word. This is depicted in Figure. 3.

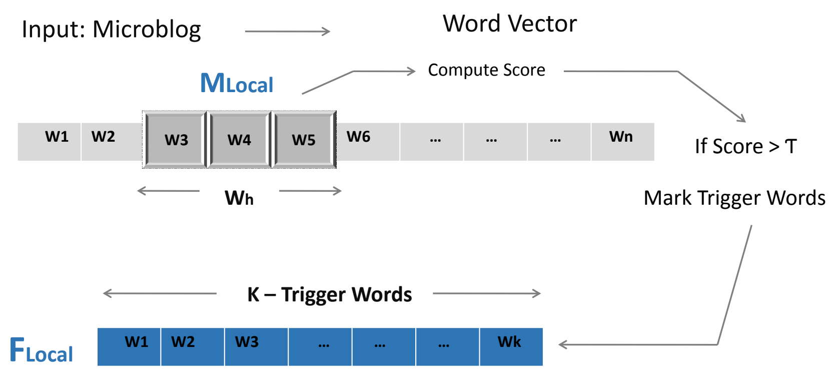

Figure 3: Attention Based Mechanism

If the score is greater than a particular threshold η, then that word is included in the trigger word set. Once all the trigger words have been obtained, a folding operation is applied to abstract out a fixed length feature vector F_{local} which encodes the trigger word set for microblog m

4.3 Global CNN

The global CNN is used to model the entire microblog. This is depicted in Figure. 4.

A window of size $l \times d$ is slid across the entire microblog to obtain a feature map. This is followed by a max pooling operation to yield one feature. This process is repeated with different filter windows to obtain the feature map F_{global} for the microblog.

4.4 Training

Once the feature maps F_{local} and F_{global} are obtained, they are concatenated and passed through a convolutional layer followed by a softmax layer to obtain the final set of ranked hashtags pertaining to the given microblog. This is depicted in Figure 5.

The training procedure will be done to optimize the values of the parameter matrices for both the local(M^l) and global mechanisms(M^g) as well as the ma-

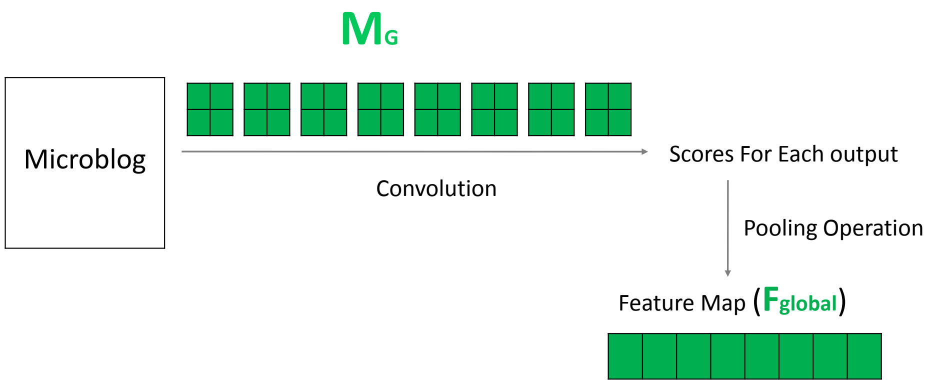

Figure 4: Global CNN

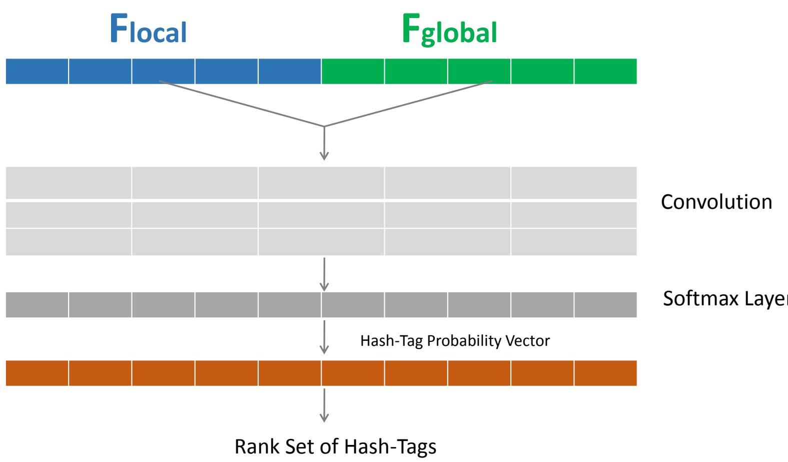

Figure 5: Final Training

trix for learning the word embeddings(W). The learning method to be employed in our project is that of AdaDelta.

5 Experimental Evaluation

We have currently processed our training dataset and have ascertained that the word vocabulary consists of 1668114 words and the hashtag vocabulary consists of 38197 words. Our objective would to implement the workflow depicted in the previous section and test with various configurations of the local and global convolution mechanisms. The hyper-parameters to be tuned in this context are the size of the windows M^l and M^g. We shall conduct experiments by varying the window sizes of both the filters used in the local and global attention mechanisms. Our evaluation policy will be based on the standard metrics of

Precision, Recall and F-Score for all the methods used.