Brain Storming Answers

Adil Kaan Akan

2171155

1 What if we were very lucky to get these accuracy results depending on training and test set distribution? What is the correct way to measure the accuracy?

There is a possibility to that we were very lucky to get these accuracy results. The correct way to measure accuracy applying cross validation. If we choose one training and test, we could get lucky in that training and test set. However, the model cannot generalize well. Instead of that, we should use cross validation. We split the dataset in to groups, and choose traning and test set, after training and testing, we compare the results in groups. In that way, measured accuracy is more accurate than choosing just one training and test set.

2 Why did we use top 2 and top 3 accuracy and they are significantly higher than categor- ical accuracy?

Because in top 2 and top 3 accuracy, we compare the true label and one of the top 2 or top 3 higher predictions. For example, our model can misclassify 'Crime' class as 'Action' by very little chance, in that situation, if we use the categorical accuracy, we will say "misclassification", but when we use top 2 or top 3 accuracy, we can say "true prediction". That is why accuracy is significantly higher when we use top 2 or top 3 accuracy instead of categorical accuracy.

Looking at the confusion matrix of best model with 500 epochs, the model had a tendency to misclassify similar classes, such as "Action" and "Crime", "Romance" and "Comedy". It is expected result since we use bag-of-word approach. It may be better classified if we use better approach to encode words, such as doc2vec method.

3 What are the disadvantages of using Bag-of-Words approach?

In bag-of-words approach, we just count the words and based on that count we encode the word. However, bag-of-words approach do not care the ordering of the words. That is really big problem. For example, consider two sentences, "Course is good" and "Is course good", in bag-of-words approach, both sentences we the have encoded vector, however, the semantic is totally different. Also, when the vocabulary goes very big, our feature vector will be very big. With very big feature vector, models can overfit easily.

4 What makes the difference between tf-idf and count? What is the importance of idf?

The difference is that tf-idf looks all documents for each term, whereas count looks just for that document. This difference comes from idf. IDF means inverse document frequency. It gives us a number according to how many times the term appear in all documents. When we use tf-idf instead of count, we will assign accurate values for terms with higher occurences and terms with lower occurences.

5 What was the effect of changing min df?

Min df variable determines which words we should add to our vocabulary. When we change it, our vocabulary changes, because we omit the words with smaller df than our min df variable. When we make min df variable small, our vocabulary will contain much more words, because we will add more words to our vocabulary. When our vocabulary is dense, we can encode sentences more accurate. Therefore, we can get better feature vectors. The training experiments support the claim. When we lower the min df variable, accuracies get higher.