Report

Anonymous ACL submission

In this project we implemented Naive base classifier for relation extraction task. The developed program classify a sentence as one of the following classes: publisher, director, performer, and characters.

1 Performance Report

1.1 Accuracy

The 3 fold cross validation is implemented on the training data and the average precision is equal to 0.890. Furthermore, we trained the classifier on the whole training data and implemented the classifier on the test data which resulted in the accuracy equal to 0.885.

Table 1: Accuracy

Data set	Accuracy
train	0.890
test	0.885

1.2 Confusion Matrix

Table 2: Confusion matrix

Sys/Gold	characters	director	performer	publisher
characters	87	6	7	3
director	6	83	3	0
performer	4	2	90	3
publisher	6	3	3	94

1.3 Pooled Precision

Pooled precision is equal to the micro average precision which is equal to 0.885 which is provided in the table (3) as micro average precision.

2 Justification of Design

We implemented NB classifier and represented text document as bag of words and used frequency of words as a feature for classifying sentences. In bag of words we assume that position of words does not matter, and probability of each word in a sentence can be calculated only based on word frequency. The accuracy on the test set is very close to the average accuracy calculated based on 3 fold cross validation on the training data which shows the model is not under fitting. Therefore, adding more features such as head and tail position may not be necessary and even it may lead to over-fitting and reduce the accuracy. In addition, as expected and due to the reasonable test accuracy compared to training accuracy the model is not over fitting to the training data and generalize well. Unknown words in the test data are ignored; however, we used Laplace smoothing for preventing zero probability for words that are not in a specific class. We also chose to leave out non-alphanumeric tokens. This was done to help with implementation and also to ensure that things like possessive nouns were not differentiated. We made the assumption that it would have little impact on accuracy in the negative and our accuracy was quite high.

Table 3: Classification Report

	Precision	Recall	F1-score	Support
characters	0.845	0.845	0.845	103
director	0.902	0.883	0.892	94
performer	0.909	0.874	0.891	103
publisher	0.887	0.940	0.913	100
accuracy			0.885	400
macroAvg	0.886	0.885	0.885	400
microAvg	0.885	0.885	0.885	400

3 Error Analysis

Based on table (2) it can be seen that class ”director” has more tendency to be mis-classified as

”characters”. The reason can be the similar words such as ’play’ or ’film’ that are used in some sentences of both classes interchangeably. Since in this classifier the position of words are not considered, any presence of words that common in another class can fool the classifier.

Characters is most often incorrect, likely due to it sharing the most words with other classes, which can be expected as directors and performers sentences often differ primarily in the name of the individual and can have similar structure and themes in both test and eval dataset. Maybe restricting the fragment based on head and tail position would have increased the accuracy but the accuracy was already quite high so we left it out.

In the opposite sense publisher has the least amount of mistakes, likely due to the nature of publisher sentences.They tend to have more dates and are less likely to be based on a singular entity and more likely to be based on a plural entity. This is distinct from the other sentences in both eval and test dataset.

In general adding more training data and more features should help the classifier. Adding more featured without having enough data may lead to over-fitting. However, in our case only using bag-of-words, the provided data seems sufficient.