

Natural Language Processing Assignment – 1

CRF Tagging in Movie Queries

Palak Dhanadia (210748141)

School of Electronic Engineering and Computer Science, Queen Mary
University of London, UK

ec21276@qmul.ac.uk

1 Split the training data into 80% training, 20% development set

In this part, we split the data into sets:

- Training Set – 80%
- Testing Set – 20%

The training set is used to train our CRF model and testing set is used for testing or checking the accuracy of our model. Also recorded the classification report output for our future reference .

2 Error Analysis 1: False Positive

This error shows that a condition exist when it doesn't. By performing this we can improve our application that's why we have first found out which classes have lowest precision and in our case it's B-Soundtrack, I-Soundtrack, I-Opinion, B-Opinion and B-Plot and based on that we have print out total number of sentences where there is a false positive for that class which is 198.

3 Error Analysis 2: False Negatives

This error shows when a condition doesn't exist when it does. In this we print out number of sentences by getting the false negative of that label as the label

is not predicted but it's there. 250 sentences with False Negative exist with their True Tag as well as Predicted Tag are printed.

4 Incorporating POS tags as features

In this, words in the training data are given POS tags using CRF part of speech (POS) tagger. As the CRF tagger only takes strings we concatenate the word and POS tag with @ and get their features by both POS tag as well as word.

5 Feature experimentation and other optimization for optimal macro average

In this, by adjusting the feature function by adding some function and removing some function to see optimal macro value.

When all the features are added then macro average value is 0.71 which is not optimal. When capitalization and number features are removed then then it gets reduced to a large amount. The optimal solution for different feature of macro average will be in between 0.60 – 0.68. As there can be cases of overfitting and underfitting as well.

By running the test on 20% data we get a result of 0.68 where as running a test on 100% data we get a result of 0.66.

accuracy			0.88	25271
macro avg	0.81	0.67	0.68	25271
weighted avg	0.88	0.87	0.88	25271
accuracy			0.85	39035
macro avg	0.69	0.65	0.67	39035
weighted avg	0.83	0.85	0.82	39035

By trying different tokens like sizes, tags etc we see different result. We again analyze the False Positive and False Negative error and try to find a optimized model which has a good accuracy as well as trying to get a optimized macro average value.

testing tagger...

done

	precision	recall	f1-score	support
B-Actor	0.92	0.95	0.94	727
B-Award	0.83	0.71	0.76	55
B-Character_Name	0.93	0.38	0.54	143
B-Director	0.90	0.82	0.86	260
B-Genre	0.85	0.73	0.79	490
B-Opinion	0.45	0.31	0.37	104
B-Origin	0.61	0.50	0.55	134
B-Plot	0.62	0.58	0.60	1004
B-Quote	0.72	0.68	0.70	19
B-Relationship	0.79	0.52	0.62	85
B-Soundtrack	1.00	0.11	0.20	9
B-Year	0.98	0.98	0.98	399
I-Actor	0.93	0.95	0.94	852
I-Award	0.73	0.94	0.82	100
I-Character_Name	0.88	0.48	0.62	96
I-Director	0.92	0.83	0.87	240
I-Genre	0.78	0.69	0.73	320
I-Opinion	0.64	0.32	0.42	107
I-Origin	0.71	0.82	0.76	539
I-Plot	0.89	0.97	0.93	9602
I-Quote	0.79	0.90	0.84	93
I-Relationship	0.63	0.59	0.61	188
I-Soundtrack	1.00	0.32	0.48	22
I-Year	0.78	0.78	0.78	36
O	0.93	0.87	0.90	9647
accuracy			0.88	25271
macro avg	0.81	0.67	0.71	25271
weighted avg	0.88	0.88	0.88	25271