Module 11 Exercise Sheet

Your task is to try to improve the results of your machine learning application by changing elements in your configuration file. Start by filling in the tables below. Then, if time allows, you can use the suggestions that follow, along with your own inspiration and the online documentation, to further improve your result. There are some blank tables for further results; feel free to make your own new tables if you run out of space.

When you make a change to your configuration file, remember to reinitialise the PR! Right-click on the PR in the resources window and select "reinitialise".

Changing Parameters

Varying "thresholdProbabilityBoundary"

	Precision	Recall	F1
0.2			
0.4			
0.6			
0.8			

Varying "thresholdProbabilityEntity"

	Precision	Recall	F1
0.2			
0.4			
0.6			
0.8			

Changing Attributes

Varying the range of the "String" attribute

	Precision	Recall	F1
+/- 4			
+/- 3			
+/- 2			
+/- 1			

Varying the range of the "Orthography" attribute

	Precision	Recall	F1
+/- 4			
+/- 3			
+/- 2			
+/- 1			

Varying the range of the "Tokenkind" attribute

	Precision	Recall	F1
+/- 4			
+/- 3			
+/- 2			
+/- 1			

Varying the range of the "Gaz" attribute

	Precision	Recall	F1
+/- 4			
+/- 3			
+/- 2			
+/- 1			

Varying the range of the "Root" attribute

	Precision	Recall	F1
+/- 4			
+/- 3			
+/- 2			
+/- 1			

Adding further attributes

- Are there any other features you could add?

	Precision	Recall	F1

Changing the engine and the parameters of the engine

- Might it help to change the parameters to the PAUM engine? What parameters can you vary?
- In the documentation, you can find out about other machine learning engines that can be used with the learning PR. Might you get a better result with any of them?

	Precision	Recall	F1

Training time

- How does training time change as you vary elements of the configuration file?

- Do your changes increase training time? Application time? If so, do you think the improvement justifies the increase?

- To accurately determine training time, you might prefer to run the PR in training mode rather than evaluation mode. To accurately determine application time, you might prefer to run the PR in application mode.

	F1	Training Time	Application Time

Different approaches to Evaluation

You might also want to try out different approaches to evaluation:

- Experiment with k-fold cross-validation for different values of k

- Try holdout evaluation with different proportions of test to training data