Khoa Nguyen

Prof. Monogioudis

CS301

04/26/2020

Explanation on LIME (Task 3 and Task 4)

1. Summary of Task 3: Choosing classifier for the dataset.

I used the ***Random Forest Classifier*** with two classes, which is “0 for ‘not spam’, and one for ‘spam’.” for this Tubespam dataset. Here is the result that I get after doing these data with ***Random Forest Classifier***. My dataset is “Youtube01-Psy.csv”

Table

Description automatically generated

Figure 1: Result after training a classifier for Tubespam Dataset.

This classifier achieves a very high F-score, which is 0.96. It indicates that this classifier overfits this dataset by learning irrelevant stuff. Moreover, this classifier also has a high accuracy with this dataset. Therefore, ***Random Forest Classifier*** is a good classifier for this dataset.

1. Task 4: Lime Explanation

I do the explanation with at most 6 features for an arbitrary document in the test set:

A picture containing logo

Description automatically generated

Figure 2:Result of do the explanation with at most 6 features.

I choose the document id is seven. That might be because it’s showing us perfectly that my classification is working well. I also print out the content of that id which is below the probability(Spam). As we can see above, the classifier got this right (it predicted not spam)

Text

Description automatically generatedThen I print out all 6 features of this.

Figure : 6 features of the explanation.

Text

Description automatically generated with medium confidenceNow I tried to remove ‘views’ and ‘the’ from the document. Let’s see what the result will be in this case.

Figure : The difference between the original and the prediction removing.

As we can see above, the difference if small, which is 0.16. It is also a good evidence for my good ***Random Forest Classifiers.***

Chart, bar chart

Description automatically generatedTo be easier to understand, I will show you my visualize result of lime explanation.

Figure : The graph of the lime explanation.

(The result is right)