Khanh Luu

CSCD429

Dan Li

Feb 15, 2020

Homework 2: Classifier implementation(Naïve Bayesian)

My program is hard coded at this point. You must modify the file name to run other dataset, other wise when you run the main it will take in the default name that I have chosen which:

File file = new File("Genes\_relation.data");

For training data

file = new File("Genes\_relation.test");

for testing data

File writeFileName = new File("output.txt");

For writing out the output into a text file.

I think I can use the missing as their own element for my prediction as well because Naïve Bayesian would handle it by its calculation. For example: in Phenotype there are 13 indexes included the “?” and it is appeared the most(1065/4636), why don’t you use it to predict the localization? Like calculate the P(“?”|nucleus) and so on. The logic if this idea sounds wrong but I want to try and see its accuracy…

It turns out I got very close to the one in RapidMiner noticed here I use the raw input data without pre-process anything. I would think about that, but I will save the ideas for our project.