Naive Bayes and Logistic Regression for Text Classification ReadME

Lizhong Zhang

March 26, 2017

1 Compilation

Compile the project as: `javac Main.java`

2 Execution

Download the dataset and *stopword.txt* to specific folder. Executing the program

> `java Main` *<path of training dataset> <path of testing dataset> <path of stopword.txt> <number of iterations> <learning rate> <lambda>*

For example, with folder like this

```
|____src
| |____LogisticRegression.java
| |____LogisticRegression.class
| |____Main.java
| |____Main.class
| |____NaiveBayes.java
| |____NaiveBayes.class
| |____ParserOfLB.java
| |____ParserOfLB.class
| |____ParserOfNB.java
| |____ParserOfNB.class
| |____TestHamOfLB.java
| |____TestHamOfLB.class
```

```
| |____TestSetOfNB.java
| |____TestSetOfNB.class
| |____TestSpamOfLB.java
| |____TestSpamOfLB.class
|____stopword.txt
|____test
| |____ham
| | |____0003.1999-12-14.farmer.ham.txt
| | |____...
| |____spam
| | |____0073.2003-12-24.GP.spam.txt
| | |____...
|____train
| |____ham
| | |____0004.1999-12-14.farmer.ham.txt
| | |____...
| |____spam
| | |____0100.2003-12-28.GP.spam.txt
| | |____.....
```

We can execute like:

```
java Main ../train ../test ../stopword.txt 500 0.01
```

The accuracies using Naive Bayes and Logistic Regression with or without stop words will be calculated.