MapReduce下朴素贝叶斯的实现

- 实验环境：本实验基于Hadoop2.6.0以及JDK1.8.0
- 测试数据：本实验测试数据须符合以下形式(以制表符分隔)：

  属性A　属性B　属性C ... 分类
- 使用方法：

启动Hadoop;

在HDFS分布式文件系统上进行必要准备：

```
/usr/local/hadoop/bin/hdfs dfs -mkdir /naivebayes
/usr/local/hadoop/bin/hdfs dfs -put <训练集> <测试集> /naivebayes/Data
```

编译Java文件并打包：

```
javac NaiveBayesClass.java
jar -cvf  NaiveBayesClass.jar NaiveBayesClass*.class
javac NaiveBayesAttribute.java
jar -cvf  NaiveBayesAttribute .jar  NaiveBayesAttribute*.class
javac NaiveBayesPro.java
jar -cvf  NaiveBayesPro.jar  NaiveBayesPro*.class
javac NaiveBayesTest.java
jar -cvf  NaiveBayesTest.jar  NaiveBayesTest*.class
javac NaiveBayesCount.java
jar -cvf  NaiveBayesCount.jar  NaiveBayesCount*.class
```

对分类进行统计并查看输出：

```
/usr/local/hadoop/bin/hadoop jar NaiveBayesClass.jar NaiveBayesClass /naivebayes/Data/<训练集> /naivebayes/Classify/class_output
/usr/local/hadoop/bin/hdfs dfs -cat /naivebayes/Classify/class_output/*
```

对属性进行统计并查看输出：

```
/usr/local/hadoop/bin/hadoop jar NaiveBayesAttribute.jar NaiveBayesAttribute /naivebayes/Data/<训练集> /naivebayes/Classify/attr_output
/usr/local/hadoop/bin/hdfs dfs -cat /naivebayes/Classify/attr_output/*
```

生成概率表并查看输出：

```
/usr/local/hadoop/bin/hadoop jar NaiveBayesPro.jar NaiveBayesPro /naivebayes/Classify/* /naivebayes/Pro_output
/usr/local/hadoop/bin/hdfs dfs -cat /naivebayes/Pro_output/*
```

对测试集进行预测、判断正确与否，并查看输出：

```
/usr/local/hadoop/bin/hadoop jar NaiveBayesTest.jar NaiveBayesTest /naivebayes/Data/<测试集> /naivebayes/Test_output
/usr/local/hadoop/bin/hdfs dfs -cat  /naivebayes/Test_output/*
```

统计正确和错误的个数，并查看输出：

```
/usr/local/hadoop/bin/hadoop jar NaiveBayesCount/NaiveBayesCount.jar NaiveBayesCount /naivebayes/Test_output/* /naivebayes/Count_output
/usr/local/hadoop/bin/hdfs dfs -cat /naivebayes/Count_output/*
```