COMP-SCI 5542 (SP17) - Big Data Analytics and Applications

Tutorial 4 Assignment (Due 02/16/17 by 11:59 PM)

***Dayu Wang* (45)**

1. Spark Programming

1.1. Description of Dataset

The dataset consists of **emoji** icons. The data is separated into two classes, the *smiling* icons and *crying* icons. I used the method mentioned in tutorial class and **decision tree algorithm** (instead of the random forest model in the TA's source code). Figure 1 lists some images in my dataset.

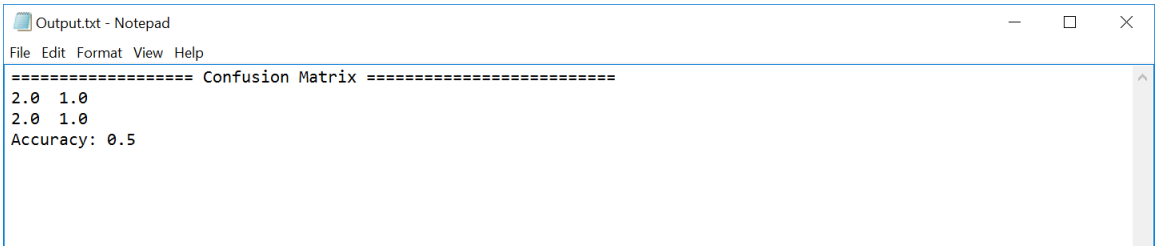
Figure 1. Examples of my dataset of **emoji** of smiling and crying.

There are **40** training images (20 smiles and 20 cries) and **6** test images (3 smiles and 3 cries).

1.2. Purpose behind Image Classification

Image classification applied on **emoji** impressions can significantly help researches on cross-platform emoji interpretation area. It opens a novel way to recognize and translate new impressions, with the rapid development of new platform/devices nowadays.

1.3. Accuracy and Confusion Matrix



2. Google Conversion API

Web server: Apache Tomcat 8.5.11

Please see the code in my GitHub account.