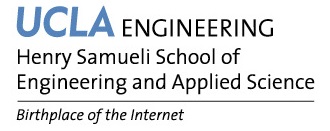
**Joint Research Center in Ubiquitous Computing**

**普適計算聯合研究中心**



Social Network Users Clustering using Tensor-Flow

Aug-2017

Advisor: Giovanni Pau, PhD Group Member: Meng Lai

(gpau@cs.ucla.edu) Tianyi Zheng

Rita Tse, PhD Chris Zhou

(ritatse@ipm.edu.mo)

Steven Liu

(steven.lyl147@gmail.com)

**Table of content**

[**1.** **Introduction** 3](#_Toc490419368)

[1.1 Overview 3](#_Toc490419369)

[1.2 Tools 3](#_Toc490419370)

[**2.** **Implementation** 4](#_Toc490419371)

[2.1 Build up classification function 4](#_Toc490419372)

[2.2 Image Recognition Application 5](#_Toc490419373)

[2.3 Data Mining 6](#_Toc490419374)

[**3.** **Result** 7](#_Toc490419375)

[**4.** **Experience** 8](#_Toc490419376)

1. **Introduction**

## 1.1 Overview

Nowadays machine learning across a range of tasks is very powerful and can serve our life. 2 years ago, our information-gathering behavior had been to find out what other people think through “Opinion Mining and Sentiment Analysis”. So we have a hypothesis about whether we can figure out the the meaning of an image. Tensor-Flow is such an open-source software library for us to detect and decipher patterns and correlations. Thus, the module which was trained based on Big Data should be accurate and the data must be rich. In our life, we always type a keyword in Google and Google will show the photos about it. However, if we have a photo and we don’t understand what is that, it’s hard for us to figure it out. Therefore, one of the example by using Tensor-Flow is to classify the image. Also, you can use Tensor-Flow to classify voices and signals. Tensor-Flow can be written in several programming languages such as Python, C++, Java, GO and so on. By using Tensor-Flow, our life will become more and more convenience. Based on that, we can take use of Tensor-Flow to find out what an image’s meaning is. Thus, our project aims to analyze and recognize an amount of user’ profile photo, which is provided by Weibo API from ***MPI UB Lab***. Based on these result, we may know some preference of the users[1], such as what clothes the user dressed, what jewelry the user wear. After that, we can offer the result to Weibo so that Weibo are able to provide the related advertisement according to user’s preference.

## 1.2 Tools

Programming language: Python, SQL

Database: MySQL

Server: Apache HTTP Server

Library: Tensor-Flow, PyQt5

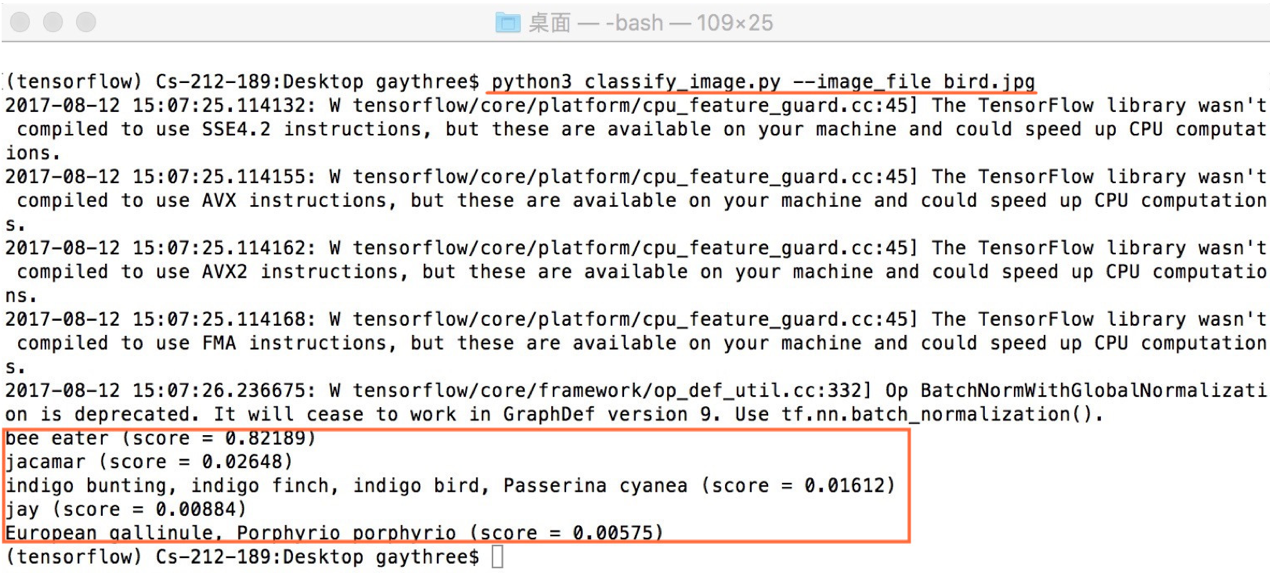
1. **Implementation**

## 2.1 Build up classification function

In order to build up a classification function, selection of module is the most important. The module was trained from raw data so that the quality and the quantity of raw data is the key of training a module. Generally speaking, to train a module is based on Big Data. With Tensor-Flow, Image classification is simply a comparison between your input and module. In short, the accuracy of the image classification depends on how powerful the module we use. For this research, we just use the module from ‘www.image-net.org’ because we think that is the most powerful one. Furthermore, we find a API called “classify\_image” in Github so that we can easily use it. In ***Figure 2***, we just simply input a photo of bird.jpg (***Figure 1***), and after the processing, we can get the top 5 possibilities with accurate rate about what is that photo (as shown in the orange box in ***Figure2***). As a result, with around 0.822 accurate rate the input photo is a “bee eater”, and with around 0.026 accurate rate the input photo is a “jacamar”. Base on that, we can almost confirm that the input is a “bee eater”. The bee-eaters are a group of near-passerine birdsin the family Meropidae containing three genera and 27 species. The reason why our result can be a specialization of bird instead of bird, is the module we used is so much powerful.



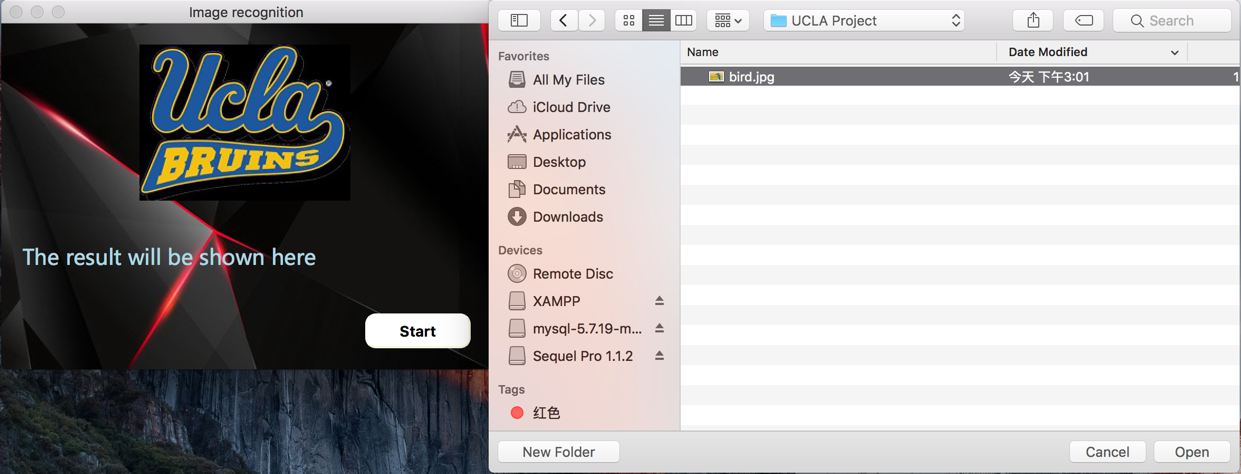
***Figure 1***

******

***Figure 2***

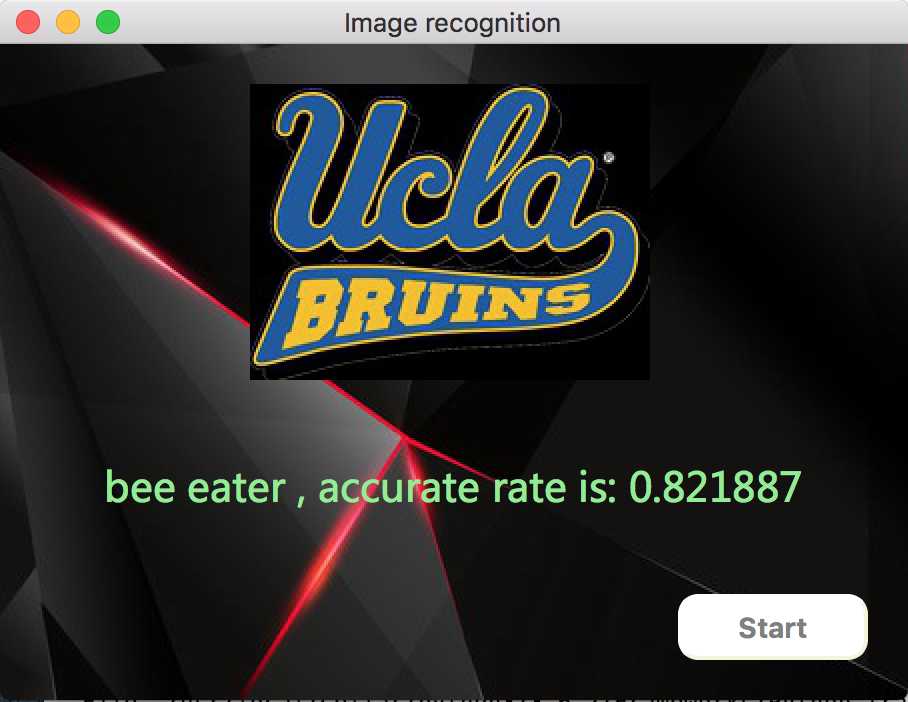
## 2.2 Image Recognition Application

Since we already know the function to classify an image, we decide to create a simply application for image recognition (as shown in ***Figure 3***).



***Figure 3***

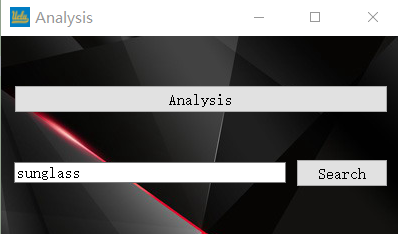
In order to visualize the function, we make a concise and friendly user interface, which can be easily used by everyone even a child. In python, PyQt5 is a good choice for design GUI. There is a start button for you to choose an image to input to the application. After clicking it, a file open dialog will show up for choosing an image. Then the chosen image (here we just simply use the ***Figure 1*** as an example) will be processed immediately, analysed by tensorflow and the result will be sent back (as shown in ***Figure 4***), telling us what the thing is in the image and the accuracy rate of recognition.



***Figure 4***

## 2.3 Data Mining

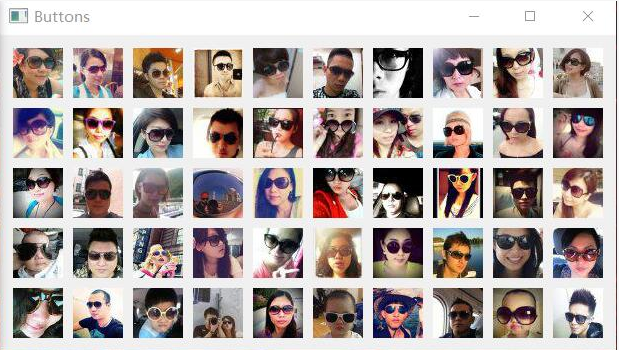
By the use of the classify-image function demonstrated above. We have 430-thousand information of the Weibo users. We processed and classified approximately the users’ profile photo according to their physiological features, and input these data into database. By the use of PyQt5 and QSS, we simply built a system (as shown in ***Figure 5***) for clustering the data with interactive user interface.



***Figure 5***

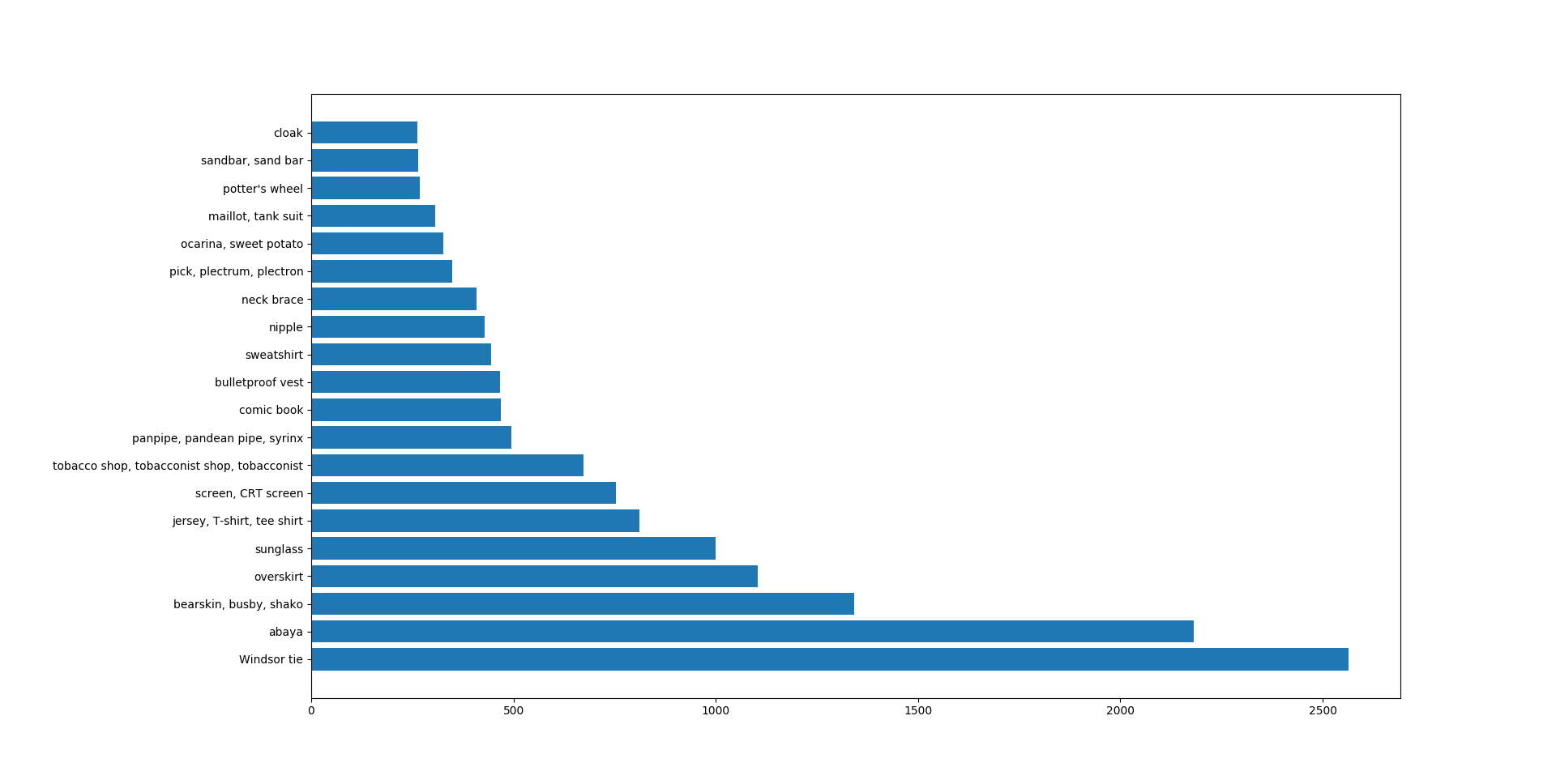
If you click the analysis button, you will get an overview of all these users. In order to make the result more reliable and precise, we only selected the data which the accurate rate is higher than 0.50. Then we made an application to count the number of images in different types, and show it in a histogram figure (will be shown in the next slide – 3. Result), which can make the result easier to be understood.

In this application, we also provide a function called “analysis” that can search the name of a certain type and then show the images of this type in our database.



***Figure 6***

1. **Result**

During two weeks hard working, we discussed together, worked together. Finally, we have processed those Weibo data and got a result (as shown in ***Figure 7***) by our data-mining application. The result can be used for various purposes, like advertising, marketing, future studying, statistics and improving the user experience, etc. For example, people in Macao are likely to upload the selfie with their windsor tie as their profile photos. And according to this information, Gucci can offer their advertisements to those people wearing windsor tie.

***Figure 7***

1. **Experience**

It’s our honor to be chosen to this summer exchange program. All of us hadn’t been to the USA before, so it’s a good chance for us to broaden our horizon and learn something new.

Honestly, the UCLA campus is wonderful and unforgettable, and here was where we worked. It’s our first time to deal with “big data”. It really cost us a lot of time to process those images by using our PC. At the same time, we learnt how the Tensor-Flow works, how the machine recognized images, and how valuable the big data is. We’ve encountered some problem and finally we solved these problem with our teamwork. At the end, the result is pretty good and useful, we realize that the “big data” field has huge potential, and we will always stay hungry and stay foolish.