

COMP4801 Final Year Project

HKUCS Graduate Admission Data Analysis:

A Multimedia Mining Approach

Interim Report

Supervisor: Dr. Reynold Cheng

Group Member: Song Yi Ting (3035124829)

Wang Michelle Yih-chyan (3035124441)

Submission Date: January 28th, 2018

Abstract

Recently, the Computer Science Department of the University of Hong Kong (HKUCS) brought in the concept for fully utilizing the admission data of graduate programs. A web application was developed for the purpose using data extraction, analysis and visualization. However, it only focused on analyzing text-based admission data submitted by the applicants without processing any multimedia data. Therefore, this project aims at finding new approaches to explore hidden information inside multimedia admission data collected during interviews and integrates them with the current system to facilitate a more comprehensive investigation of admission strategies. Multiple tools are leveraged for visual and audio data analysis respectively to extract related information, such as facial expressions and speech content of the applicants. Both visual and audio data analyzing system have been constructed and evaluated. In the future, the project will customize the original data mining system for both text-based and multimedia data and provide users with outcome prediction and intelligent matching functions.

Table of Contents

Abstract.....	2
List of Figures.....	4
List of Tables	4
Abbreviations	4
1 Introduction.....	5
2 Project Objectives	6
3 Methodology	7
3.1 Audio Data Analysis	7
3.2 Visual Data Analysis	9
4 Preliminary Findings	11
4.1 Audio Data Analyzing Tool	11
4.2 Visual Data Analyzing Tool.....	13
4.3 Difficulties Encountered	14
5 Future Plans.....	16
6 Conclusion.....	18
References.....	19

List of Figures

Figure 1 Affectiva API (Facial Expressions).....	10
Figure 2 Visual data analyzing system	14
Figure 3 Data selection for scatter plots visualization	17

List of Tables

Table 1 Selection criteria for candidate evaluation.....	6
--	---

Abbreviations

API	Application Programming Interface
HKUCS	The Computer Science Department of the University of Hong Kong
NLP	Natural Language Processing

1 Introduction

Every year, the Computer Science Department of the University of Hong Kong (HKUCS) receives hundreds of applications for graduate studies. In order to select the best students for the programs, HKUCS starts to utilize admission data collected in previous years and try to discover underlying admission strategies as well as to give advices to admission officers. At the moment, Wu and Xu have created a web-based analyzing system focusing on text-based datatype collected from applicant's CV [1]. Nevertheless, interview, an important component in admission process, is still left unutilized. As a result, this project aims at providing detailed insights into the relation between admission interview and the final admission result.

The report is structured as follows. It first outlines the scope and objectives of the project. Then, it will discuss the approaches adopted in different parts of the development and their preliminary results at the moment. Moreover, a future plan for the second semester is provided as the next step for the project.

2 Project Objectives

The main objective of this project is to equip HKUCS admission system with multimedia data mining functions and assist admission officers to make accurate and efficient decisions. To achieve the goal, the project has to set up general interview selection criteria for evaluating the performance of student candidates(See Table 1). In addition, tools for multimedia data processing/analyzing have to be developed to extract valuable information regarding the interviewee. After information extraction, the pre-processed data should be matched with selection criteria to produce corresponding indicators for interview performance of each student. These results are visualized by an interactive user interface with which admission officers can obtain informative factors for admission decisions.

Selection Criteria	Indicators
Communication Skills	English fluency, accent
Professional Impression	relevant background knowledge, emotion stability
Personal Characteristics	confidence, maturity

Table 1 Selection criteria for candidate evaluation

3 Methodology

Due to its complexity and nature, the project divides multimedia data into two major parts – audio data and visual data. They are processed separately using different information extraction tools and then stored in the database introduced in later parts for further analysis.

3.1 Audio Data Analysis

Existing libraries are used to extract the information and content of the audio files. There are different aspects extracted from audio files:

Speech Tone Interpretation

To extract the emotions and tone contained in the audio data, we utilize Python speech libraries and Sphinx, a tool consisted of the use of the deep neural network and extreme learning machine, classifies the emotions with the previously-trained model.

Speech Content Extraction

Tensorflow-speech-recognition, **Sphinx** and **Google Cloud Speech API** are utilized to retrieve the speech content of interviewers and candidates [2] [3]. It is trained with a large data set including synthetic Text to Speech snippets, Movies with transcripts, Gutenberg, YouTube with captions, with some extensions available. The packages first detect the number of speakers contained in the files and identify the switches of the speakers. And then, the packages output texts with speakers annotated.

Speech Fluency

The audio files are divided into a series of clips with timeframes. The audio files are divided according to the switches of speakers and re-combined as separated files. Speech contents of separated files are extracted and compared with the amounts of syllables. The speed of producing syllables are stored into the database as a numeric entry in the database.

Emotion stability

Portions of emotions are extracted with percentages within the whole audio file. Therefore, the audio file is divided with a static timeframe and the emotions contained are extracted accordingly. There are five kinds of emotions could be extracted: neutral, frustrated, happy, sad and excited. The switch rate of the audio file indicates the emotional stability of the particular candidate. The switch rate is then stored in the database as a numeric entry in the database.

3.2 Visual Data Analysis

Interview videos contain a variety of information that are valuable to admission officers. Among them, facial expression is able to convey the most accurate message regarding interviewees' emotion and to further imply to other desired criteria for suitable candidates. Therefore, the ability to accurately extract interviewee's facial expressions, or even emotions, is a key component of visual multimedia data analysis. Also, since the developing platform plays a crucial role in the system integration process discussed in the next section, system compatibility is of great importance when designing the visual data analyzing system.

Affectiva is chosen as the visual data analysis tool in the project due to the two reasons mentioned above [4]. It is a cross-platform API, supporting Android, iOS, Linux, JavaScript, etc. This provides more flexibility and extensibility for further development of the project. In terms of functionality, it is able to excerpt images from canvas element of HTML5 page and process them subsequently for facial expression detection. There are several types of information available through Affectiva API. Emotion and Expression are the two most important ones in regard to the project. Affectiva provides a value of 0 to 100 for each of the seven main emotions – joy, sadness, disgust, contempt, anger, fear and surprise. The higher the value is, the more likely the emotion is shown in the image. The same applies to the twenty one types of detailed facial expression (See Figure 1).

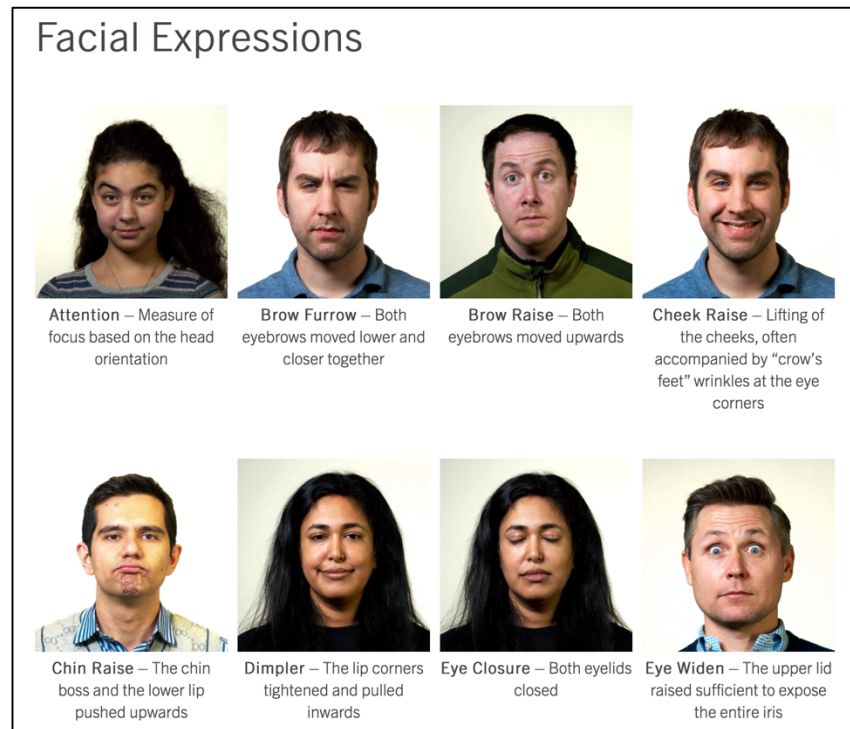


Figure 1 Affectiva API (Facial Expressions)

After collecting all of the information from each image in the interview video, the system will store them in the MySQL database for further processing. The processing procedure may include time-sequence analysis which indicates the changes of the interviewee’s emotions through different stage of the interview. Also, the corresponding facial expression response to particular type of questions raised by admission officers could be used to imply certain selection criteria.

4 Preliminary Findings

This chapter provides detailed information of the development progress as well as the difficulties faced by the project during the first semester. It first presents current project progress of audio and visual data analyzing tool respectively. Then, it describes several limitations the project had faced during implementation.

4.1 Audio Data Analyzing Tool

Tool Evaluation – Speech Content Extraction

Sphinx and Google Cloud Speech API are both used to convert speech content. Google Cloud Speech API has achieved high accuracy with efficiency. Converting a 1-minute wav file takes less than 5 seconds in average. However, Google Cloud Speech API has restrictions of the audio file length, which cannot exceed 1 minutes. Therefore, the audio files must be cut with 1-minute interval to get the entire result. Also, the Google Cloud Speech API must be used under stable connection. On the other hand, the model of Sphinx is able to be trained with custom input, but the efficiency and the accuracy are not as high as Google Cloud Speech API.

Tool Evaluation – Emotion stability

Sphinx is mainly used to retrieve the 5 emotion categories. The model is trained by custom input. The emotions in the audio files recognized are mainly neutral with small fluctuation of the stability.

Tool Evaluation – Speech Fluency

Because some files have a strong difference between the candidates' and interviewers' speech, the volume of the speech is averaged before the conversion. Also, the intervals of the speech are included in the speech. During the fluency analysis, the rate of syllables is calculated with the total time including intervals.

System Design

The audio file conversion consists of file reorganizations and transformation. The system first reads the files and detects the switches to reorganize the file with speakers accordingly. The switching time is divided evenly with the speakers. Next, the reorganized files are used to extract the speech content and the tones for analysis. The system utilizes the tools and packages to calculate the information for speech fluency and emotion stability. Then, after obtaining the required information, the system connects to the database and export the result to the database in numerical and text formats.

The system is constructed with python, since the tools are supported in Python and there are analyzing and extracting packages available in Python, such as NumPy, **Scikit-learn**, Sphinx, and MySQL-connector [5]. Moreover, it supports many functions to create a web service enabling other people to retrieve and interact with the analysis result. This helps the integration with the system built by previous students.

4.2 Visual Data Analyzing Tool

To obtain more representative information from interview videos, the project set up several criteria for tool evaluation before actually process the data. For example, as mentioned in the methodology section, system compatibility is likely to influence the development progress due to integration with current HKUCS admission analysis system in later parts of the project. Moreover, usability of the tool can also affect the difficulty and performance of multimedia data processing. It is critical to examine the ease of operation as well as information extraction functionality during the evaluation.

After a collection of potential candidates were selected based on the pre-requisites above, they were implemented and evaluated more thoroughly one by one. Finally, Affectiva were selected as a result of its comprehensive functionalities and support of multiple platforms, which significantly mitigate the system compatibility issue. On the other hand, **Emotime**, for example, was eliminated because of complex dependency issues on **CMake** and **OpenCV** which greatly complicates the development process [6] [7] [8].

Since the previous admission system is built as a web-based application, the project chose JavaScript Affectiva API for visual data analysis and constructed a front-end web application. The visual data analysis application takes videos as the source of data for analyzing. For each second of the video, the application breaks it into a certain number of frames and submits them through Affectiva API to generate a series of analyzed

information, e.g., emotions and facial expressions. The information generated are shown in the web interface in real-time for users to keep track of the changes in the interviewee's emotion and facial expression (See Figure 2).



Figure 2 Visual data analyzing system

4.3 Difficulties Encountered

The first difficulty of the project is the system compatibility issue between current admission system, local operating system and the tools leveraged. Due to the fact that this project is building on an existing system, it has to take the tools and libraries already in use into consideration. Also, given the limited time frame, it is difficult to create tools from scratch on our own. This leads to a more complicated tool evaluation criteria and limits the number of tools the project is able to leverage.

The second challenge encountered is conversion format restriction. Similar to the difficulty above, the existing tools are not sufficient in terms of multimedia processing. For example, the facial recognition tools give out numeric representation of emotions in each picture frame while speech recognition tools give out only the transformation in text format. These pre-processed data cannot be directly used in deriving deciding factors of admission strategies. As a result, transforming these primitive results to suitable parameterized data for the database storage and further data analysis is critical in future development.

Another difficulty regarding tool evaluation is confidentiality. Since the data source is recorded during the candidate interviews at HKUCS, it could neither be published nor uploaded to any cloud storage. However, many effective tools provide services through cloud computing. Therefore, tools that are eligible for the project is fairly limited and the time spent on this process largely exceeds the original estimation.

5 Future Plans

During the first semester, the project has developed audio and visual multimedia analyzing system and has verified their functionalities. For the second semester, a few more stages are scheduled to be implemented.

The first stage is to expand data analyzing to all of the available interview videos and add the pre-processed data into existing database. Since the MySQL database is a relational database, the project needs to decide the data storage format, i.e., relational model.

Secondly, the newly developed system will be integrated with the current admission system to construct a new HKUCS admission system supporting both text-based and multimedia datatype. The system integration procedure may involve evaluation of original data mining model and project code refactoring. By using multimedia dataset generated in the first stage, the project will be able to refine the data mining function inside the admission system and increase the accuracy of outcome prediction. On the other hand, code refactoring allows future developers to understand and modify the programming code more easily. After system integration, the web application interface also has to be updated according to the newly added features. For example, the attribute

selection for scatter plots visualization should be extended to multimedia categories as well (See Figure 3).

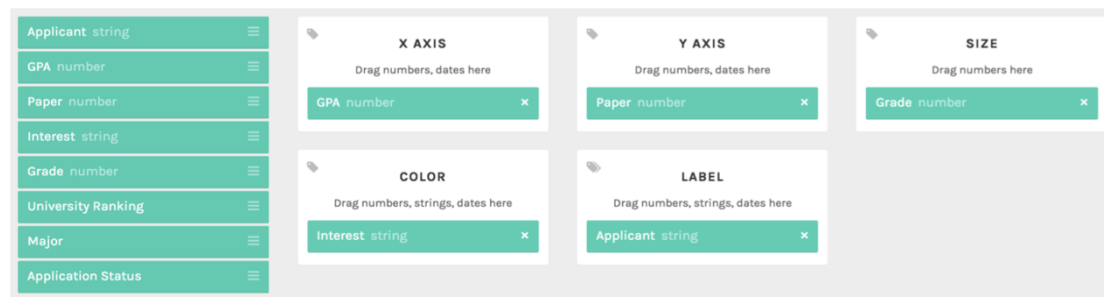


Figure 3 Data selection for scatter plots visualization

Another stage is further processing of the (pre-processed) multimedia data. As discussed in project objectives section, this project aims at providing insights into HKUCS admission strategies by inspecting multimedia admission data. Nevertheless, the pre-processed data from audio and visual data analyzing system could not provide us with informative conclusion for assessing the performance of the candidate at the moment. Therefore, it is necessary to extract high-level information which are directly related to the admission officers' selection criteria by re-processing the data collected in the first semester. For instance, the speech content obtained from speech content extraction can be further processed using Natural Language Processing (NLP) to examine the quality of the answer to a particular question.

6 Conclusion

In order to provide a deeper understanding and discover the underlying admission strategy for HKUCS postgraduate programs, the project applies multimedia data mining techniques to admission system and develops both audio and visual data analyzing tools. By exploring the interview videos, admission officers are able to efficiently assess candidates based on several attributes indicating their performances in the admission interview. In this stage, multimedia data analyzing tools have been developed and tested by sample interview videos. Integration with current HKUCS admission system and massive data processing are scheduled into several stages in the coming semester. The project hopes to re-examine the utilization of multimedia admission data and expand it to not only HKUCS Department but also other educational institution.

References

- [1] Wu, Y, Xu, F. Mining HKUCS Graduate Student Data: Extraction, Analysis, and Prediction. Hong Kong (PRC): The University of Hong Kong; 2017. Available at <http://www.cs.hku.hk/programme/projects/csfyp/csfyp.jsp>.
- [2] PocketSphinx [Internet]. cmusphinx/Pocketsphinx. [cited 2018 Jan 28]. Available from: <https://github.com/cmusphinx/pocketsphinx>.
- [3] Google Cloud Speech API [Internet]. Google Cloud Speech API. [cited 2018 Jan 28]. Available from: <https://cloud.google.com/speech/>.
- [4] Affectiva [Internet]. Affectiva. [cited 2018 Jan 28]. Available from: <https://www.affectiva.com/>.
- [5] Scikit-learn [Internet].: machine learning in Python — 0.18 documentation. [cited 2017 Dec 1]. Available from: <http://scikit-learn.org/stable/index.html>.
- [6] Emotime [Internet]. Emotime. [cited 2017 Oct 22]. Available from: <https://github.com/luca-m/emotime/>.
- [7] CMake [Internet]. CMake. [cited 2017 Oct 22]. Available from: <https://cmake.org/>.
- [8] OpenCV [Internet]. OpenCV library. [cited 2017 Oct 22]. Available from: <https://opencv.org/>.