
Machine Learning Based Graduate Admission Prediction

Tianbao Li

Department of Computer Science
University of Toronto
Toronto, Ontario M5S 2E4
tianbao@cs.toronto.edu

Abstract

Graduation application to the US, Canada and the UK is a hot topic for students around the world, especially in China during the past years. Admission committee always gives out a final decision according to the student's background in many perspectives. Such decision can be considered as a classification problem and there is no much work on it by now. In this project, we applied several machine learning algorithms on our self-built dataset and trained models with relatively high accuracy.

1 Introduction

Nowadays, more and more Chinese students start to take graduate study overseas, usually in the US, Canada, the UK and somewhere else. Then, the problem just comes with it. How could they know whether a graduate school is going to offer admission or which one is the best fit?

Students usually finish their graduate application in two ways. The first one is to find an agency for help. Such agencies collect application data for year and give advice based on history cases and percentage on multiple indicators. However, misjudging always happens just because graduate admission consists of complex evaluation in many areas. The second type of application is called DIY-application, finished by students themselves. Due to lack of information, many students lose better offers.

During the application evaluation, information in many fields of one student is considered, including TOEFL score, GRE score, GPA and other supplementary materials as undergraduate school, work experience, research experience. Looking in the machine learning way, these indicators can be features of a certain model. We can use massive admission and rejection cases as training data to fit the admission model of a certain graduate program. Till now, decision is mainly made by human experience in this area. Comparing to human judgement and finding similar cases, many machine learning models seem to have a better prediction such as neural networks and decision tree. red A few related research

The contribution of this project is mainly is these points:

- Dataset built-up: here I decide to use data from the BBS GTER¹, one of the most popular graduate application BBSs in China. Many Chinese students post their admission decision here. I decide to write crawler to collection data of admission and rejection. Also due to the low data quality, much work on data cleaning has to be done. Such data can be used for related purpose in the future research

¹<http://bbs.gter.net/>

- Model training: here I decide to train several popular machine learning models on such data, including neural networks, decision tree, naive bayes, etc, find better fit model and make optimization.

Clarification

This project is finished by myself alone. With a self-defined dataset, some related work may continues on it afterwards.

2 Data Summary

Chinese students take a large portation in graduate application, however, there is no avaiable dataset about it. So, for the first part of this project, we intent to biult a dataset about Chinese student graduate application. We wrote web crawler to collect application data from GTER BBS, both admission and rejection application in the range of 2012-2016. Since web data usually holds low quality, some data cleaning skills are applied on the dataset to make it easy for model training.

Current dataset contains 11056 cases. Due to natural language-based expression, the raw data is hard to catch related feature. Here, we only extract information about decision, target school, degree, year of application, TOEFL² score, GRE³ score, GPA, GPA ranking. Such information is transformed into 14 features and well normalized for training. Summary of the dataset is shown in Table 1.

| General Information | |
|----------------------|--------------------------|
| data point amount | 11056 |
| feature amount | 14 |
| Features for dataset | |
| result | admission:reject=4.48 |
| most populat school | Columbia University |
| TOEFL average | 84.8 25.2 22.1 23.1 26.6 |
| GRE average | 319.0 |
| GPA average | 3.2 |

Table 1: Data summary

References

²<https://www.ets.org/toefl>

³<https://www.ets.org/gre>