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 Background

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 which graduate school is the best fit, or more possibly offers admission?

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, undergraduate school, work experience, research experience and other supplementary materials. 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, ML techniques are not wildly used 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.

2 TODOs

Work for this project can be divided into two main parts:

- Data collection: here I decide to use data from the BBS 1point3acres[1], the most popular graduate application BBS in China. Many Chinese students post their admission decision here. I decide to write crawler to collection data of admission and rejection.
  Tools to be used: python, python-scrapy, simple NLP skills to deal with plain text description.

[1] http://www.1point3acres.com/bbs/

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

- 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.
  Tools to be used: python, tenserflow.

3 Clarification

This project is going to be finished by myself alone. Several wonderful papers [1] and [2] could help with my project.

References

[1] Narender Gupta, Aman Sawhney, and Dan Roth. Will i get in? modeling the graduate admission process for american universities. In *Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on*, pages 631–638. IEEE, 2016.

[2] Stuart M Keeley and Michael E Doherty. Bayesian and regression modeling of graduate admission policy. *Organizational Behavior and Human Performance*, 8(2):297–323, 1972.