

**Name: Hyomin Seo**

**Date: 29 Sep 2020**

**Paper Title: GRADE : machine learning support for graduate admission**

**Author Names: Austin Waters Risto Miikkulainen**

**Year Published: 2013**

**Open questions:**

Can the historical admission database be reliably used for making prediction of application received multiple years in future?

Should the feature selection and database be updated to reflect the each year's trend?

**The topic areas covered by the paper are:**

The paper introduces and explains the GRADE - a statistical machine learning system developed to support the work of the graduate admissions committee at the University of Texas at Austin, to compensate the increasing number of applications received compared to insufficient number faculties who are responsible for reviewing all. The paper introduces the measurement and statistical computing methods programmed in the system for it to evaluate each application numerically.

**The previous approaches to this problem were:**

The situation problems that gave motivation to this project was the gap between exponentially increasing application for CS Phd and insufficient human resources to review those - in this particular university.

The previously developed system for the similar purposes was decided not applicable for numerous features that those lack. Instead of building a program for deciding the pass or fail of an applicant, the system is aimed to numerically process the large stack of applications to narrow it down to some percentage of applications that are reviewed with higher scores. This clustered group of applications are then subject to human full-review.

**Outline the basic new approach or approaches to this problem:**

The process of GRADE is the following

Application is read based on departmental database and performs preprocessing to standardize the data (into high dimensional feature vectors)

Logistic regression classifier that predicts the admission probability of the application is trained on the feature encoded historical data from step1.

Finally, selects the application that needs a full review, and that can be just quickly checked to verify the model's prediction.

Big percentage of the determination is based on the admission database ( history).

**Critical assumptions made include:**

The most fundamental assumption made for the GRADE is that the historical admission data in fact can be used to make future assumptions.

There is also an assumption that the quality of 'unrepresentable data' should somewhat be the same trend as the other 'representable data' - the unrepresentable data of highly scored application should be in better quality than that of the poorly scored application ( therefore was not included in the group of applications for full review).

That the automation of the admission process is needed to ensure objectivity and fairness even with the exponentially growing number of applications submitted.

**The performance of the techniques discussed in the paper was measured in what manner:**

The paper ensured that the GRADE system provides maximum objectiveness and does not yielded any biased result by presenting that they went through the reliability test on the GRADE for two consecutive years ; showing that the GRADE only made the entire admission process accelerated, saving time and resources simultaneously not losing any critical attribute of the process itself.

It clearly shows how the each features were carefully put into algorithm, even though there weren't much systemic compensation for the unrepresentative data except the statement that this attribute of an highly scored application is 'human' reviewed.

**New background techniques are used in the paper:**

Feature encoding

Probabilistic binary classifiers ( in detail)

Multi-layer perceptrons

Support Vector Machines

historical log odds

**The following terms were defined:**

Feature encoding - for this system, sets three (or four) categories

- Numerical data : GPA GRE

- Categorical data : School reputation

- Text data. : All of recommendation/ Purpose of application. (plan)

- Unrepresentable data : publication, awards, titles.

The paper also defines the logic behind the probabilistic binary classifiers that was generated based on the historical admissions decision taken as labeled training examples. (adapts the l1 regularized logistic regression)

**I rate and justify the value of this paper as:**

This paper is written in a very comprehensible manner, understanding the process and logics behind the system relatively easy.

Though all the reliability test and numbers are presented, it does leave the reader with a bit of a doubt on so just how reliable can the system be when there must be a few applications that was only discoverable by 'human eye'? However, the situational background, where the system is launch, does provide the fair justification for adopting such system into this particular admission process.