Emerging Schools

A guide for post-secondary investment

January 31, 2016

1 Introduction

1.1 The Problem

The Goodgrant Foundation is a charitable organization with a mission to improve educational performance of students attending colleges and universities in the United States. The foundation intends to donate a total of 100,000,000 USD to a group of schools each year, for five years. They also wish to avoid investing in universities that have already received grants from major charitable organizations such as the Gates and Lumina Foundation.

We present a mathematical model to outline an investment strategy to provide the maximal return on investment(ROI) appropriate for an education charity. The model is split into two sections and is applicable in choosing $1 \cdots N$ institutions.

1.2 Assumptions and Rationale/Justification

- **The Gates Foundation and Lumina Foundation will continue to donate to the same school for at least two years:** If first year donation shows promise results then they will continue to donate for a second time or more. If first year donation shows inadequate result, then they will increase fund for a second donation in hope for better result (i.e. give another chance). This allow us to avoid duplicate investment by reading their donation data from 2015 to see a list of schools the two foundations donated to.

- **Students who received grant will stay in the program and successfully graduate:** There are many factors that could cause students to not be able to graduate, but we will focus on financial support because it is the root of other factors. However, we will provide a note on how to improve the model with the assumption that receiving a grant simply reduces the likelihood of not completing a degree.

-

2 Model Design and Approach

2.1 Notations and Definitions

1. $U_j^{[i]}$ = university j of class i
2. $U_o^{[i]}$ = benchmark university for class i
3. $A_j^{[i]}$ = amount invested in university j for class i
4. $T_j^{[i]}$ = Tuition in university j for class i
5. $n_j^{[i]}$ = number of students receving our grant in university j for class i
6. $IP_j^{[i]}$ = poverty rate of students entering university j for class i
7. $OP_j^{[i]}$ = poverty rate of students exiting university j for class i
8. $\phi_j^{[i]}$ = number of students in university j for class i
9. $g_j^{[i]}$ = maximum amout of grants to be given to university j for class i
10. Group, refers to the Carnegie Classification of the university
11. Class, refers to which benchmark within the group, the university is closest to.

2.2 Summary of our Approach

- Create a benchmark setting: top (...) schools that we consider to have good performance. Each school represent a class.
- Identify universities that have similar characteristics to the benchmark list.
- Prioritize universities to be invested in by the amount of grants they received from other sources.
- Determine the amount of investment and allocation to maximize the return on investment.

3 Flattening Out the Approach

3.1 Benchmark Setting

First we omitted any below medium size universities or communal universities such as racial or religious schools. Then we partition all universities by number of years required for graduation. Lastly, we rank the universities by the amount of their poverty elevation standards:

$$R_j^{[i]} = \frac{IP_j^{[i]} - OP_j^{[i]}}{IP_j^{[i]}}$$

Next we choose the top, highest ranking universities per 'years of graduation'-group. We set these chosen universities as our benchmarks for each class in each group.

i.e.

$$o^{[i]} = \underset{j}{\operatorname{argmax}} R_j^{[i]}$$

Where o, is to represent the benchmark index of the class i.

3.2 Identifying Matches and Schools to be Invested

We aim to classify each university into some class i. This is done via the kNN algorithm: a non-parametric method used in classification supervised learning problems. The output is determined by a majority vote, where the object is assigned to the class most common with it's neighbours. [5]

We set up the following model: Assuming $j \in [1, n]$,where n = amount of variables. For a detailed selection list please see Appendix 1. Let $y_j^{[i]}$ be the j-th variable for the i-th class benchmark. Let x_j be the j-th variable of a particular university. A university u, is considered a member of the k-th class if the variables x_i of that university is 'close' to a particular variable of a given class in euclidean-space.

That is, if u belongs to the k-th class then

$$k = \underset{i}{\operatorname{argmin}} \sqrt{\sum_{j=1}^{n} (x_j - y_j^{[i]})^2}$$

3.3 Determine Investment

3.3.1 Number of student-Helped

Let $A_j^{[i]}$ be the amount we invest in university j of the class i, and $T_j^{[i]}$ be the tuition of low-income students. Then we can compute the number of students who we help as:

$$n_j^{[i]} = \lfloor \frac{A_j^{[i]}}{T_j^{[i]}} \rfloor$$

$$n^{[i]} = \sum_{j=1}^{\infty} n_j^{[i]}$$

$$n = \sum_{i=1}^{\infty} n^{[i]}$$

Please note that the number of universities is finite, we use infinity here in order to avoid the introduction of more variables.

3.3.2 Adjusted Poverty

Next we computed the adjusted poverty rate. We make a crucial assumption here about the correlation between grant-money and retention rates. We assume that every student who receive the grant will stay in the program and university (this assumption can be adjusted by a future study on the subject matter).

Let $\phi_j^{[i]}$ be the total number of students in university j for class i.
The adjusted rate of existing poor students based on our assumptions is:

$$OP_j^{[i]*} = \frac{OP_j^{[i]}\phi_j^{[i]} - n_j^{[i]}}{\phi_j^{[i]}}$$

We can then provide a rank to this university, in the similar fashion we computed the rank for the benchmark universities:

$$R_j^{[i]*} = \frac{IP_j^{[i]} - OP_j^{[i]*}}{IP_j^{[i]}}$$

We use this to compute the increase to the social well-being of students going to this university.

3.3.3 Return on Investment

The return on investment for our model will be the change in the level of alleviation of poverty provided by our investment.
We do this as follows:

$$R_j^{[i]}(1 + r_j^{[i]}) = R_j^{[i]*}$$

where $r_j^{[i]}$ is the return on investment for a particular school.
We also note that this can be rewritten as $r_j^{[i]} = \frac{OP_j^{[i]} - OP_j^{[i]*}}{IP_j^{[i]} - OP_j^{[i]}}$. This will later be used for simplification purpose of the maximization problem.

We then average these rates, to get the approximated return:

$$\bar{r} = \frac{\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} r_j^{[i]}}{\#(\{\forall i, j : U_j^{[i]}\})}$$

For simplification, the total number universities, or $\#(\{\forall i, j : U_j^{[i]}\})$ will be written as Φ, so

$$\bar{r} = \frac{\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} r_j^{[i]}}{\Phi}$$

3.3.4 Maximization of return

Now that we have a measure of return on investment, we wish to maximize this measure based on the physical restriction that are given to us.
Since we only have a limit of $100M, and would not want to overshoot our investment by giving any particular university more than what the top-tier schools are given. We put the following restriction forth.

Let $g_j^{[i]} = \frac{G_o^{[i]}}{\phi_o^{[i]}} - \frac{G_j^{[i]}}{\phi_j^{[i]}}$, be the restriction on the per-student investment.
We do this, in order to make sure that our investment has a reasonable bound that does not exceed the higher-tiered school.

So now our problem can be formulated as follows:

$$\begin{aligned} \underset{A}{\text{maximize}} \quad & \bar{r} \\ \text{subject to} \quad & \sum \sum A_j^{[i]} = 100,000,000 \\ & \forall i, j \quad 0 \leq A_j^{[i]} \leq g_j^{[i]} \phi_j^{[i]} \end{aligned}$$

3.3.5 Simplifying the complexity

In order to simplify the complexity of the previous problem we outlined before. We reformulated the problem as follows:
First recall, results:
(1) $r_j^{[i]} = \frac{OP_j^{[i]} - OP_j^{[i]*}}{IP_j^{[i]} - OP_j^{[i]}}$
(2) $OP_j^{[i]*} = \frac{OP_j^{[i]}\phi_j^{[i]} - n_j^{[i]}}{\phi_j^{[i]}}$
, this can be further simplified:

$$\begin{aligned} r_j^{[i]} &= \frac{1}{\phi_j^{[i]}(IP_j^{[i]} - OP_j^{[i]})} n_j^{[i]} = \frac{1}{\phi_j^{[i]}(IP_j^{[i]} - OP_j^{[i]})} \lfloor \frac{A_j^{[i]}}{T_j^{[i]}} \rfloor \\ &\approx \frac{1}{\phi_j^{[i]}(IP_j^{[i]} - OP_j^{[i]})T_j^{[i]}} A_j^{[i]} = \Delta_j^{[i]} A_j^{[i]} \end{aligned}$$

We note that $\Delta_j^{[i]}$ is the information given by our data, whereas $A_j^{[i]}$ is our investment, which is a variable that can be controlled by us. Hence the mean return can be written as:

$$\bar{r} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \frac{1}{\Phi} r_j^{[i]} = \frac{1}{\Phi} \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \Delta_j^{[i]} A_j^{[i]}$$

So our new maximization problem can be formulated as follows:

$$\begin{aligned} \underset{A}{\text{maximize}} \quad & \frac{1}{\Phi} \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \Delta_j^{[i]} A_j^{[i]} \\ \text{subject to} \quad & \sum \sum A_j^{[i]} = 100,000,000 \\ & \forall i, j \quad 0 \leq A_j^{[i]} \leq g_j^{[i]} \phi_j^{[i]} \end{aligned}$$

4 Model Testing on Big Data

5 Sensitivity Analysis

6 Conclusions

References

[1] Trombitas, Kate. *Financial Stress: An Everyday Reality for College Students.* Inceptia, July 2012. PDF.

[2] Carnevale, Anthony, Ban Cheah, and Martin Van Der Werf. *Ranking Your College.* Georgetown University Center on Education and the Workforce, Dec. 2015. PDF.

[3] "Bill & Melinda Gates Foundation." *Bill & Melinda Gates Foundation.* Web. 29 Jan. 2016.

[4] "Grants Database." *Grants Database.* Web. 29 Jan. 2016. <https://www.luminafoundation.org/grants-database/strategy/student-financial-supports>.

[5] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. "Prototype Methods and Nearest-Neighbors". *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. New York: Springer, 2009. Print.

[6] "IPEDS Data Center." *IPEDS Data Center.* Web. 30 Jan. 2016. <https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx>.

[7] Hanushek, Eric and Margaret Raymond. 2005. "Does School Accountability Lead to Improved School Performance? *Journal of Policy Analysis and Management* 24(2): 297-329.

[8] Winston, Wayne L. "Linear Programming." *Operations Research: Applications and Algorithms.* 4th ed. Belmont, Calif.: Duxbury, 2003. Print.

[9] "Using Federal Data To Measure And Improve The Performance Of U.S. Institutions of Higher Education" Sept 2015. <https://collegescorecard.ed.gov/assets/UsingFederalDataToMeasureAndImprovePerformance.pdf>.