

An Optimal Strategy of Donation for Educational Purpose

Summary

To determine the best donation strategy, this paper presents a data motivation model based on the original definition of return on investment (ROI) for charitable organizations.

In order to evaluate the value of a school, we describe metrics in given data files. We selected two main species vector. First, We used the imputation method to complete and clean the data. By breaking the knowledge, we gave it the initial score and gave the corresponding weight through post processing. The initial \hat{y} is given by numpy and then the results are homogenized. Finally, the Sigmoid method is used to output a reasonable distribution and the high-segmentation is found to be relatively significant.

Secondly, in order to solve the problems caused by high-dimensional data, we use principal component analysis (PCA) to reduce the data and obtain new feature vectors, that is, new performance indicators.

Next, we developed a composite index called the Performance Index to quantify the student's educational performance. The performance index is a linear combination of the performance indicators analyzed by the PCA in the previous step, and the coefficients are obtained by stepwise fitting through the cyclic neural network (CNN).

Finally, we calculate the ROI, defined as an increase in the performance index rather than an increase in donations. This return on investment is institution-specific and depends on the increase in donations. By adopting a two-step ROI maximization algorithm, we determine the optimal investment strategy.

-
- [An Optimal Strategy of Donation for Educational Purpose](#)
 - [Summary](#)
 - [Letter](#)
 - [Below is the recommended investment strategy](#)
 - [1. Introduction](#)
 - [1.1 Statement of the Problem](#)
 - [2. Symbols, Definitions and assumptions](#)
 - [2.1 Detailed Definitions](#)
 - [2.2 Assumption](#)
 - [3. Articulate our metrics](#)
 - [3.1 Problem with too high data dimension](#)
 - [3.2 principal component analysis \(PCA\)](#)
 - [4. Determining the Performance Index](#)
 - [5. Determining Investment Strategy based on ROI](#)

- [6. Calculation process](#)
 - [6.1 Data Cleaning and Processing](#)
- [7. Conclusions and Discussion](#)
 - [7.1 Conclusions](#)
 - [7.2 Limitation and extensions](#)

Letter

Dear Chiang,

Our team proposes a performance indicator that quantifies the educational performance of each institution of the student and appropriately defines the return on investment (ROI) for charities such as the Goodgrant Foundation. Mathematical models are built to help predict return on investment after determining the mechanism by which donations affect performance.

The optimal investment strategy is determined by maximizing the estimated return on investment. More specifically, the Composite Performance Index is developed after considering all possible performance indicators, such as graduation rates and graduate income. The performance index is built to represent the performance of the school and the positive impact the school has on students and the community. From this perspective, our definition seeks to capture the social benefits of donations. Then we use principal component analysis to find performance contribution variables that strongly influence performance metrics. We fit the relationship between these variables and the donation amount to predict the change in the value of each performance contribution variable relative to the donation amount. We calculate the ROI (defined as the increase in performance indicators relative to the amount of donations) by multiplying the value of each performance contribution variable by the amount of the donation and the impact of each performance contribution variable on the performance index, and then adding the products for all performances. Contribution variable. After maximizing the return on investment based on the selection algorithm, the optimal investment strategy is determined.

In short, our model has successfully developed an investment strategy that includes a list of target institutions and the amount of investment for each institution. (The list for the first year is attached to the end of the letter). The duration of the investment can also be determined based on our model. Since the model and assessment method are completely data-driven and do not contain any standards, it is suitable for solving future philanthropic education investment issues. We firmly believe that our model can effectively improve the efficiency of philanthropic education investment and provide an appropriate and feasible way to best improve students' educational performance.

Below is the recommended investment strategy

NO	INSTNM	UNITID	SCORE	INVESTMENT
1	Medical College of Wisconsin	239169	0.999459574	\$5,086,407.39
2	Albany Medical College	188580	0.997669333	\$5,077,296.57
3	A T Still University of Health Sciences	177834	0.997564058	\$5,076,760.81
4	West Virginia School of Osteopathic Medicine	237880	0.997476675	\$5,076,316.10
5	University of Massachusetts Medical School Worcester	166708	0.996238404	\$5,070,014.35
6	New York Medical College	193830	0.99406238	\$5,058,940.22
7	Rosalind Franklin University of Medicine and Science	145558	0.993442022	\$5,055,783.12
8	Philadelphia College of Osteopathic Medicine	215123	0.988850119	\$5,032,414.20
9	University of North Texas Health Science Center	228909	0.982072607	\$4,997,922.37
10	Baylor College of Medicine	223223	0.980199221	\$4,988,388.42
11	Meharry Medical College	220792	0.978832453	\$4,981,432.71
12	Western University of Health Sciences	112525	0.976554211	\$4,969,838.38
13	SUNY Downstate Medical Center	196255	0.975377605	\$4,963,850.45
14	University of California-San Francisco	110699	0.973446364	\$4,954,022.06
15	Midwestern University-Downers Grove	143853	0.971981321	\$4,946,566.23
16	Midwestern University-Glendale	423643	0.971949743	\$4,946,405.52
17	MCPHS University	166656	0.971513982	\$4,944,187.87
18	Upstate Medical University	196307	0.968459056	\$4,928,640.87
19	St Louis College of Pharmacy	179265	0.967265721	\$4,922,567.79
20	Samuel Merritt University	122296	0.967202212	\$4,922,244.59
				\$100,000,000.00

1. Introduction

1.1 Statement of the Problem

There exists no doubt in the significance of postsecondary education to the development of society, especially with the ascending need for skilled employees capable of complex work. Nevertheless, U.S. ranks only 11th in the higher education attachment worldwide, which makes the financial support from large charitable organizations necessary.

As it's essential for charitable organizations to maximize the effectiveness of donations, an objective and systematic assessment model is in demand to develop appropriate investment strategies. To achieve this goal, several large foundations like Gates Foundation and Lumina Foundation have developed different evaluation approaches, where they mainly focus on specific indexes like attendance and graduation rate. In other empirical literature, a Forbes approach (Shifrin and Chen, 2015) proposes a new indicator called the Grateful Graduates Index, using the median amount of private donations per student over a 10-year period to measure the return on investment. Also, performance funding indicators (Burke, 2002, Cave, 1997, Serban and Burke, 1998, Banta et al, 1996), which include but are not limited to external indicators like graduates' employment rate and internal indicators like teaching quality, are one of the most prevailing methods to evaluate effectiveness of educational donations.

However, those methods also arise with widely acknowledged concerns (Burke, 1998). Most of them require subjective choice of indexes and are rather arbitrary than data-based. And they perform badly in a data environment where there is miscellaneous cross-section data but scarce time-series data. Besides, they lack quantified analysis in precisely predicting or measuring the social benefits and the positive effect that the investment can generate, which serves as one of the targets for the Goodgrant Foundation.

In accordance with Goodgrant Foundation's request, this paper provides a prudent definition of return on investment (ROI) for charitable organizations, and develops an original data-motivated model, which is feasible even faced with tangled cross-section data and absent time-series data, to determine the optimal strategy for funding. The strategy contains selection of institutions and distribution of investment across institutions, time and regions.

2. Symbols, Definitions and assumptions

2.1 Detailed Definitions

name	definition	denotation
index	The data value of a school, a total of p data values	$x_1, x_2, x_3 \dots x_p$
school_num	number of candidate schools	n
performance index	performance index of i th school	Z_i
simplify index	The main data of a school obtained after principal component analysis, a total of j data values	$y_1, y_2, y_3 \dots y_j$
donation	Amount of donation to i th school	F_i

2.2 Assumption

1. Stability. We assume data of any institution should be stable without the impact from outside. To be specific, the key factors like the donation amount and the performance index should remain unchanged if the college does not receive new donations.
2. The performance index is a linear composition of all given performance indicators.
3. Performance contributing variables linearly affect the performance index.
4. Increase in donation amount affects the performance index through performance contributing variables.
5. The impact of increase in donation amount on performance contributing variables contains 2 parts: homogenous one and heterogenous one. The homogenous influence is represented by a smooth function from donation amount to performance contributing variables. And the heterogenous one is represented by deviation from the function.

6. 1.3 The Advantages of Our Model

Our model has many advantages in the application: • The evaluation model is based entirely on data and there are few subjective or arbitrary decision rules. • Our model successfully identifies potential mechanisms, not just indicators such as graduation rates. • Our model makes full use of cross-section data and does not require time series data to produce reasonable results

3. Articture our metrics

3.1 Problem with too high data dimension

Because there are too many related factors to consider, the results are too much, and the correlation between variables is high, which brings inconvenience to modeling. It is therefore desirable to be able to interpret most of the variation in the data with fewer variables and to convert highly relevant variables into variables that are independent or unrelated to each other. Therefore, we finally choose Principal Component Analysis.

3.2 principal component analysis (PCA)

x_1, x_2, \dots, x_p is the original feature describing the school, and c_1, c_2, \dots, c_p represents the weight of each variable. Add a weight to each eigenvalue and sum it to get s:

$$s = c_1 x_1 + c_2 x_2 + \dots + c_p x_p$$

We hope that choosing the right weights can better reflect the school's effectiveness. Each school corresponds to a comprehensive score, which is recorded as s_1, s_2, \dots, s_n , where n is the number of schools. If these values are very scattered, it means that it is well distinguished, that is, looking for such a series of parameters c, so that $s_1, s_2, s_3, \dots, s_n$ are scattered as much as possible. Its statistical definition is described as follows:

Let X_1, X_2, \dots, X_p denote a random variable with x_1, x_2, \dots, x_p as the sample observation, if you can find c_1, c_2, \dots, c_p

$$Var(c_1 X_1 + c_2 X_2 + \dots + c_p X_p) \quad (3.1)$$

The value of equation (3.1) is maximized. Since the variance reflects the degree of data difference, it shows that we have captured the maximum variation of the p variables. The same expression must be added with a limit, otherwise the weight may choose infinity and no meaning, here we specify

$$c_1^2 + c_2^2 + \dots + c_p^2 = 1 \quad (3.2)$$

So far we get a principal component direction $\vec{a}=[c_1, c_2, \dots, c_p]$, which is a unit vector of a p -dimensional space. But a principal component is not enough to represent the original p variables, so it is necessary to find multiple principal components, and the second principal component should not contain the information of the first principal component, that is, the covariance of the two principal components is 0, the direction is orthogonal.

Let Z_i denote the i -th principal component, $i=1, 2, \dots, p$, we can assume

$$\begin{cases} Z_1 = c_{11}X_1 + c_{12}X_2 + \dots + c_{1p}X_p, \\ Z_2 = c_{21}X_1 + c_{22}X_2 + \dots + c_{2p}X_p, \\ \dots \\ Z_p = c_{p1}X_1 + c_{p2}X_2 + \dots + c_{pp}X_p, \end{cases} \quad (3.3)$$

For each i , the equation (3.2) is satisfied. Then you need to determine the value of $j(j < p)$, which is the number of indicators after the dimension is lowered.

The information contribution rate and the cumulative contribution rate of the feature value $\lambda_j(j=1, 2, \dots, p)$ corresponding to each principal component direction are calculated.

$$b_j = \frac{\lambda_j}{\sum_{k=1}^p \lambda_k} \quad (3.4)$$

Equation (3.4) is the information contribution rate of the main component y_i .

$$\alpha_j = \frac{\sum_{k=1}^j \lambda_k}{\sum_{k=1}^p \lambda_k} \quad (3.5)$$

Equation (3.4) is the cumulative contribution rate of the main components y_1, y_2, \dots, y_j . When α_j is close to 1, the first j indicator variables are selected instead of the original p indicators.

4. Determining the Performance Index

From the above principal component analysis method, j main variables can be obtained, but they are all standardized values and do not have the original precise meaning. At this point, the preparation of the component evaluation index has been completed, and the comprehensive score is calculated.

$$Z = \sum_{j=1}^p b_j y_j \quad (4.1)$$

$$Z = 0.566426x_1 + 0.433574x_2 \quad (4.2)$$

Equation (4.2) is a linear calculation formula of the evaluation index obtained after the above processing.

The final calculated evaluation index is the ordinate, the school is the abscissa, and the distribution shown in the following figure is obtained by sorting the evaluation indicators from high to low.

It can be seen that the evaluation indicators can distinguish the performance of each school very well. The index of the top schools is very high, and the subsequent data gradients are gradually flattened.

5. Determining Investment Strategy based on ROI

$$ROI = \frac{Z}{F}$$

Z represents the evaluation criteria obtained above. In order to maximize the ROI, the most intuitive idea is to put all the money to the school with the highest rate of return. However, it is obvious that this is not advisable. The significance of investing is to sponsor more students and help more people in need. Considering the meaning of rewards, no matter what kind of investment has its risks, you should not put your eggs in the same basket. Therefore, it is obviously unreasonable to simply take the most worthwhile investment method through ROI.

Then, it is necessary to ensure the number of schools invested, but also to ensure a relatively high rate of return. We need to calculate based on the number of schools invested and the rate of return, and meet the following conditions

1. There are two parameters, one is the number of schools invested, and the other is the amount of investment for each school. At the same time, the above evaluation indicators are utilized.
2. The increase in the rate of return decreases as the number of schools decreases, or decreases
3. There is a maximum value that makes the point satisfy the optimal investment situation

Get the following formula

$$ROI_1 = \frac{n}{N} \sum_{i=1}^n \frac{Z_i}{F_i}$$

n indicates the number of schools invested, N indicates the total number of schools, Z_i indicates the performance indicators of the i-th school, and F_i indicates the investment amount of the i-th school.

6. Calculation process

6.1 Data Cleaning and Processing

Because the number of schools given by charities is much smaller than the number of schools in the dataset, the first step is to screen out the data of these alternative schools, and exclude the unrelated factors such as the school's official website and the number of graduate students, and finally select the graduation rate, the repayment rate after graduation, the number of students who won the prize, and other indicators related to the return index.

- Problem: We found that there are a large number of missing values in the data file. Variables with missing items in the data set become incomplete variables. From the missing distribution we think it is a missing random (MAR). For random deletions, deletions are not appropriate for recording, and random deletions can be used to estimate missing values by known variables. Simply discarding missing items can lead to model misalignment and loss of information.

- **Necessity:** Data loss is a complex issue in many research areas. For data mining, the existence of default values has the following effects:
 - The system has lost a lot of useful information;
 - The uncertainty shown in the system is more significant, and the deterministic components contained in the system are more difficult to grasp;
 - Data with null values can confuse the mining process, resulting in unreliable output.
 - The process is designed to avoid over-fitting the model to the data, a feature that makes it difficult to handle incomplete data well with its own algorithms. Therefore, the default value needs to be deduced, filled, etc. through a special method to reduce the gap between the data mining algorithm and the actual application.
- **Processing method:** There are three main methods for processing incomplete data sets: deleting tuples, data completion, and no processing.
- **Delete tuples:** that is, delete objects (tuples, records) with missing information attribute values, and get a complete information table. This method is simple and easy. It is very effective when the object has multiple attribute missing values and the deleted object with missing values is very small compared to the initial data set. This method is usually used when the class label is missing. However, this method has great limitations. It replaces historical data in exchange for complete information and discards a large amount of information hidden in these objects. In the case where the initial data set contains few objects, deleting a small number of objects is enough to seriously affect the objectivity of the information and the correctness of the results; therefore, when the proportion of missing data is large, especially when the missing data is not randomly distributed, this This approach can lead to deviations in the data, leading to erroneous conclusions. Description: Deleting a tuple, or directly deleting the column feature can sometimes result in performance degradation.
- **Data Completion:** This method fills in null values with a certain value to complete the information table. Usually based on the statistical principle, a missing value is filled according to the distribution of the values of the remaining objects in the initial data set. The following methods are commonly used in data mining:
 - **Treating Missing Attribute values as Special values** Handling null values as a special attribute value is different from any other attribute value. If all null values are filled with unknown. This will create another interesting concept that can lead to serious data deviations and is generally not recommended.
 - **Mean/Mode Completer** The attributes in the initial data set are divided into numeric attributes and non-numeric attributes to be processed separately. If the null value is numeric, the missing attribute value is filled according to the average value of the value of all other objects according to the attribute; if the null value is non-numeric, the general value principle is used according to statistics. The attribute takes the most value of all other objects (that is, the most frequently occurring value) to fill in the missing attribute value. Another method similar to this is called Conditional Mean Completer. In this method, the value used for averaging is not taken from all objects in the dataset, but from objects that have the same decision attribute value as the object. The basic starting point of the two methods of complementing the data is the same, and the missing attribute values are supplemented with the maximum possible value, but the specific method is a little different. Compared to other methods, it uses most of the information of existing data to estimate missing values.

- Hot deck imputation For an object with a null value, the hot card fill method finds an object that is most similar to it in the full data, and then populates it with the value of this similar object. Different questions may use different criteria to determine similarity. This method is conceptually simple and utilizes the relationship between data for null estimation. The disadvantage of this method is that it is difficult to define similar standards and there are many subjective factors.
- K-means clustering First, the K samples with the closest missing data samples are determined according to the Euclidean distance or correlation analysis, and the K values are weighted and averaged to estimate the missing data of the sample.
- Assigning All Possible values of the Attribute Filling in all the possible attribute values of the vacancy attribute value can get a better completion effect. However, when the amount of data is large or the value of the missing attribute is large, the calculation is costly, and there are many possible test scenarios.
- Combinatorial Completer Try all the possible attribute values of the vacancy attribute value, and select the best one from the final attribute's reduction result as the filled attribute value. This is a data filling method for the purpose of reduction, which can obtain a good reduction result; however, when the amount of data is large or the value of the missing attribute is large, the calculation is costly.
- Regression Based on the complete data set, a regression equation is established. For objects that contain null values, the known attribute values are substituted into the equation to estimate the unknown attribute values, which are populated with this estimate. A biased estimate is produced when the variables are not linearly related.
- Expectation maximization (EM) The EM algorithm is an iterative algorithm that computes a maximum likelihood estimate or a posterior distribution in the case of incomplete data. Two steps are alternately executed during each iteration loop: E step (Expectation step), and the log likelihood function corresponding to the complete data is calculated given the full data and the parameter estimates obtained in the previous iteration. The conditional expectation; M step (maximization step), maximizes the log likelihood function to determine the value of the parameter, and is used for the iteration of the next step. The algorithm iterates between E and M steps until convergence, ie when the parameter change between two iterations is less than a predetermined threshold. This method may fall into local extremum, the convergence speed is not very fast, and the calculation is very complicated.
- Multiple Imputation, MI The multiple padding method is divided into three steps: Generate a set of possible padding values for each null value that reflect the uncertainty of the non-responsive model; each value is used to fill the missing values in the data set, resulting in Several complete data sets. Each filled data set is statistically analyzed using statistical methods for the complete data set. The results from each of the filled data sets are combined to produce a final statistical inference that takes into account the uncertainty created by the data padding. This method treats the vacancy value as a random sample, so that the calculated statistical inference may be affected by the uncertainty of the vacancy value. The calculation of this method is also complicated.
- C4.5 Fill in missing values by looking for relationships between attributes. It looks for two attributes that have the greatest correlation between them, one of which is not a proxy attribute, and the other is called the original attribute, which uses the proxy attribute to determine the missing value in the original attribute. This method of rule-based induction can only deal with noun-type attributes with a small base.

- For several statistical-based methods, the delete tuple method and the mean method are different from the hot card filling method, the expectation value maximization method, and the multiple filling method; regression is a better method, but still not hot Deck and EM; EM lacks the uncertainties contained in MI. It is worth noting that these methods deal directly with the estimation of model parameters rather than the vacancy prediction itself. They are suitable for dealing with unsupervised learning, and for supervised learning, the situation is different. For example, you can delete objects that contain null values and train them with a complete data set, but you can't ignore objects that contain null values when predicting. In addition, C4.5 and the use of all possible value filling methods also have a good complement effect, manual filling and special value filling is generally not recommended.
- Select:
 - Due to the lack of data missing items, deleting the tuple will cause a lot of valid data to be invalidated. If it is not processed, the reliability of the calculation will be low. We have chosen to fill in the elements;
 - Due to the lack of data,
 - filling manually is not realistic, let go.
 - Treating Missing Attribute values as Special values will cause serious data deviation and will not be used.
- Calculation: In order to process a large amount of data, we used tensorflow as a calculation tool, and built a shallow network to calculate the Widget according to the algorithm model. In order to control the output value within a limited range while ensuring a reasonable distribution, we use a log probability regression model to constrain the output.
- Data standardization processing There are known p indicators. In order to solve the impact of different dimensions, we need to standardize the raw data and convert x_i to the standard value \hat{x}_i :

$$\hat{x}_{ij} = \frac{x_{ij} - \mu_j}{s_j}, i = 1, 2, 3, \dots, n, j = 1, 2, 3, \dots, p \quad (2.1)$$

$$\mu_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

$$S_j = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \mu_j)^2}, j = 1, 2, \dots, p$$

μ_j, s_j is the sample mean and sample standard deviation of the ith indicator.

7. Conclusions and Discussion

7.1 Conclusions

This paper manages to develop a fully data-based model to produce a provi- dent investment strategy that maximizing ROI. Our model exhibits a great potential in drawing the conclusions below:

- We formulate a performance index for each institution with principal component analysis and develop an appropriate concept for return of investment (ROI) for the charitable foundations like Goodgrant Foundation.
- We derive the relation between the performance contributing variables and donation amount from a GAM fitting model to predict ROI of performance contributing variables.

- The final list of selected institutions and appropriate amount of donation is determined by a two-step selection algorithm.

7.2 Limitation and extensions

Though our model successfully produced an investment strategy, it can be improved in several ways:

- Since only cross-section data is available, our results of time duration of donation can be improved if we have access to time-series.
- Our model only allows for a relatively simple linear model. A more general selection method is needed when applying to a complicated model.