Research on the Investment Return of Study Abroad Based on Machine Learning

Zihao Yang

Wuhan British China School, Wuhan, Hubei Province, China

1187515722@qq.com

Abstract: Sending family teens to study abroad has become an educational option for middle-class families in China's big cities. Due to the high economic cost and time cost of studying abroad, it is necessary to analyze the human resources investment to provide the family with a basis for decision-making on whether to make this investment. This paper proposes a method for calculating the return on investment in studying abroad based on the GBDT model. By selecting multiple micro variables and macro variables as input, the ResNet-GBDT regression model is used to predict the annual salary of students in different regions and industries after graduation and estimate the actual situation of studying abroad. The experimental results show that the model can predict the actual return of study abroad in different regions to a certain extent.

1. Introduction

With the development of globalization and the deepening of China's integration into the world,education abroad from the secondary and undergraduate/graduate levels has become an educational option for Chinese families with teenagers. Compared with the average annual tuition fee of about 1,000 to 2,000 US dollars in the middle school and undergraduate/postgraduate stages of domestic public schools, the first problem encountered in overseas study is the cost problem. Most families are based on the assumption that students can earn more wages after graduating from undergraduate or postgraduate studies than if they did not receive overseas education. From this perspective, the cost of studying abroad can be regarded as a long-term invest on human resources. According to statistics, the average annual income of employees working in non-private organizations in urban China was about $13,600 last year. Considering the tuition and living expenses of nearly $10,000 to $100,000 per year, it is an investment worth careful consideration for most ordinary families, not to mention that studying abroad is an act that lasts for several years.

As the develop of Chinese middle classes and develop of social awareness, more and more middle-class families choose to invest for their children to study aboard [1]. It is not only as a substitute of "gao kao" for future career and competition in labor market, but also personality development like to be more adaptable to new environment, better with manage a team and board horizon. For middle class, it is an opportunity to reproducing social structure through invest education. As they travelling aboard and have more channel to gain information, they will be more likely to choose to study aboard as Chinese ordinary "gao kao" is quite strict and less opportunity to give entrance to top universities and have to learn by rote and discourage students to be creative and have critical thinking.

In addition to the cost of studying abroad, families need to consider the rate of return of studying abroad. That is, whether a student can find a satisfactory job after studying abroad for 4 to 10 years, and can meet the requirements of personal development and personal income at the same time. Behind these requirements is whether there is a strong national demand for a workforce with a background of studying abroad, and whether students of a similar age have received the same level of education. Similar to the labor supply and demand model, the return on study abroad also needs to be forecast for the long-term investment cost and the ratio of price to future investment return. Not only that, but the state of development in the country where students will be employed in the future is also closely

related to the rate of return on this investment.

Growth rate of student's study aboard is twice the growth of world GDP, it could mean more and more financial resources are allocated on educational import, but still some students choose to return to their home land for high income, families, or incentives from government. For the model in paper [2] consider household possessing ability, pay for daily life might be more expensive in aboard than home country and higher education quality in host country, then home country will gain effective labor study aboard (both education quality and time are vital, which are highly relative to time and finance invested in). For the economy the final good depends on two factors (capital and effective labor) for two generations young and old indicate the change or growth in labor effectiveness also different level of time of working and workforce they can produce. Also, different level of household and income level as labor will have higher utility in home country than host country.

At present, researchers have begun to take adolescents from middle-class and higher-class families around the world to study in other countries as research objects, and explore this phenomenon from various dimensions such as politics, culture, and economy. On the other hand, as a way to invest in human resources, studying abroad needs to be like other investment methods, using data and models to analyze in depth, how to select sufficient and necessary data vectors, and how to construct different cultural backgrounds in different regions. The study abroad investment model is still a new field of research. This paper proposes a machine learning-based investment return model for studying abroad. By selecting micro-variables vectors such as the city where teenagers belong to their families, average annual income, social status, and parents' educational background and macro-variables vectors including future employment prospects, the GDP of the country, the average income and consumption level of the city to be employed, and the average education level of the industry etc. as the input of the model, and the ResNet-GBDT model is used to predict the average annual income of students in the year of undergraduate/graduate graduation. Based on this, the rate of return for young people studying abroad under different conditions is given. The results of the data collected through the questionnaire survey show that this model can predict the return rate of overseas study to a certain extent, and provide suggestions on whether to invest in overseas study to families in different backgrounds.

The rest of this paper is structured as follows: "related work" provides the methodology and specific practice for study abroad investment evaluation; in the "Study Abroad Investment GBDT regression model" section describes how to use the GBDT regression model for micro data vectors and macro data vectors to predict the return on study abroad investment, and the experimental results and conclusions are given at the end of the paper.

2. Related Work

Increase in demand for skilled labor raise international labor market migration. The mobility change could be the determinant of such policies. In the paper [3], the ERASMUS program, a academic prove and financial reward based on motivation and academic performance was introduced for encourage students and education exchange in Europe. Increase in number of student's study aboard increase probability of finding a job aboard about 50% increase.

Based on human investment model, it can consider gain and potential loss of society in terms of cost- benefit frame work and evaluate. First, if the core aim is to maximize the social welfare, the immigrant to aboard seems to have low effect because people go to foreign country seems not bring others income or benefit away and have less loss. However, if we considered the country or governments own benefit as their core aim is economy development or military power. In this case, the investment on educates these students or future person with ability might be considered as loss if they just flow into foreign country. In addition, it may lead to the promotion on let students develop aboard would be less effective to country in this way. It could be considered as loss in their potential contribution. To calculation the actual contribution, the probability the loss in labor after study and immigration after calculation the result seems still indicate the study aboard is benefit to the country overall. The benefit and potential loss of benefit not only limit in personal value but also value of interact with outsiders like business and group of individuals and other social resources [4].

In 2016, about 5 million students go for study aboard, contain both short-term and long term. But whether the experience of study aboard is helpful or not require further research [5]. This is because first, some students go to study aboard not only for develop their academic ability but be attracted by foreign culture and different life style. By the way they may also pursue what is known as personality and global awareness development. In this way, it might be hard for business to judge whether the experience of study aboard is more valuable or not if there are no certification for those abstract ability development, even the language depends on the depth of how much students involved in host country environment and how long. Education aboard seems to develop global awareness adaptability and labor mobility [6].

In the background of expansion in trend of study aboard, and the contents before, it is important to give certification or agreement to prove the value of experience of study aboard, and avoid low quality of foreign study experience affect overall image. This accountability specially directed toward students learning is pushed by increased competition for student's enrollment. And better to ensure equivalent to knowing what skills those students acquired or fail to acquire. However, this seems to increase base academic and language ability requirement to students especially for those top universities, to ensure their quality of education effort and quality. So, they can have better image and fame in not only businesses but it could also be a advantage for those applicants and students, a stronger education export status in the field in the multi- environment and students have different cultural language and classes background. And they need a common criterion to measure academic performance like GPA because things like gender revenue and culture background also different time and program cannot well represent effort and advantage bring by study aboard experiences [7].

3. Investment ResNet-GBDT Model

The vast majority of research examining the effects of study abroad has primarily used quantitative methodologies to explore personal, social, academic, and career development, it has become more and more popular as a way of personal investment. However, the proportion took by different coloring population is highly balance especially for Asian, black, Hispanic. The core reason could be lack in family awareness and rigid program structures, lack of course requirements, length of study, lack of family and community support. Such problems are common in lots of developing country that have too much domestic educational resources needs for large population like Chinese GAO KAO and in India. For development study program aboard, the level of participant is important [8]. As study aboard become a more and more famous way to develop labor skill and productivity, more and more government tend to adjust their structure of education and give financial aid as promotion to encourage students to study aboard, but this investment could be a loss if students choose to work for foreign country. It seems it is still important to determine the demand of what type of labor skills and relative utility labor can gain through study aboard that seems more reasonable and friendly, or if students' study to work at aboard first, the working experience would increase the probability to keep working aboard. In this way, the length of staying seems to be the matter but the benefit government and environment can offer seems effect more as there still lots of labor choose to return with experience of working for foreign business which could also be recognize as a process of learning [9].

From an investment point of view, investment in the stock market can be compared with investment in overseas study. Stock index forecasts mainly include fundamental forecasts and technical analysis forecasts [10]. The former needs to focus on the value of stocks, industries, politics, the general economy and the market, while the latter can describe future patterns and trends by studying market statistics and constructing patterns and trends. A great task and challenge in stock market investing is predicting the direction of daily changes in stock indices. In this type of forecast, it is a popular trend to accurately predict stock index and implement investment decisions through machine learning and big data processing technology. In predicting stock indices, Gradient Boosting Decision Tree (GBDT) models are widely used, for example, to describe the characteristics of stock indices based on raw stock price data and other indicators, in addition to GBDT structures, models such as support vector machines and neural networks are also widely used use. Under the same test conditions, the GBDT-based model outperformed the other models in a statistical sense.

A GBDT model can be viewed as a combination of a basic set of decision tree classifiers through gradient boosting techniques. Different from classical linear models, GBDT has the ability to generate nonlinear cross-transformed input features, and since the GBDT model is not sensitive to outliers, it is suitable for the representation of nonlinear decision boundaries, and the GBDT model has good scalability, which is widely applicable to the processing of numerical features and other categorical features. A typical GBDT is shown in Eq.1. Given a sample, GBDT uses this function for the output:

$$\hat{y}_{GBDT} = \sum_{m=1}^{M} \eta_m b(X; \gamma_m) \tag{1}$$

Where, M is the maximum number of iterations in training the model, η is the weight used to combine individual trees in updates to prevent overfitting, γ_m is the minimum loss reduction required for further partitioning on the leaf nodes of the tree. The main role of Gradient Boosting is to descend in the direction of the steepest gradient, in order to further reduce the loss. Compared with AdaBoost's binary classification, Gradient Boosting deals with multiple categories by decomposing the loss function. Gradient Boosting in GBDT relies on negative gradients, while basis functions employ decision trees.
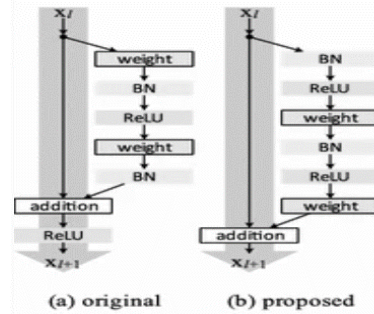

Figure 1. The residual unit in ResNets [11].

With the advent of deep neural networks, many problems in the field of artificial intelligence have been greatly improved in performance. The depth of the deep neural network, that is, the increase in the number of network layers, is very beneficial to the performance improvement [11]. The reason is that deeper networks can obtain more complex features, but as the depth increases, it brings the Degradation problem, that is, the performance decreases. Residual learning is an effective way to solve the Degradation problem, which structure is shown in Fig.1, because the stacked layer can learn new features based on the input features.
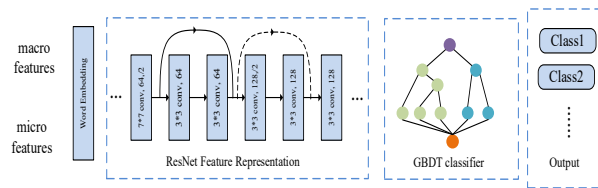

Figure 2. The proposed ResNet-GBDT structure.

Fig.2 shows the investment model for studying abroad based on the ResNet-GBDT structure. The input data are micro-features, including the city to which the youth belong, average annual income, social status, parents' educational background, etc., and macro-features, including future employment prospects, the income and consumption levels of the employment city, and the average education level of the employment industry. These features are mainly from the natural language descriptions in the questionnaires. First, the word vectors are extracted from these natural languages through Word Embedding, and then the output of the word vector layer is extracted through ResNet as features and the last layer of output vectors are extracted to obtain the feature representation processed by ResNet. Then input GBDT to further prevent over-fitting, and improve the accuracy by increasing the weight of the wrong instance.

4. Experimental results

The way to achieve prediction in the market is to look for the information embedded in the price and then to be interpreted by the market participants, our goal is to predict whether it is worthwhile to invest in studying abroad, i.e. whether studying abroad can be obtained after the student graduates not studying abroad. High-paying returns that crowds can't get. To this end, we propose the learning architecture ResNet-GBDT and use multiple macro- and micro-features as input features, and finally compare the results with those obtained with SVM and decision tree models using the same data.

The input dataset comes from natural language responses to 20 questions in 80 questionnaires from families in Beijing, Shanghai, and Shenzhen with future study abroad destinations including the United States, United Kingdom, Australia, and Singapore. The questions are divided into two categories, one is micro-features, including 10 questions such as the city to which the teenager belongs, average annual income, social status, and parental educational background; The other category is macro characteristics, including 10 questions such as future employment prospects, income and consumption levels in employment cities, and average education levels in employment industries. The output is the ratio of the student's annual average income in the year of undergraduate/graduate graduation to the total cost of studying abroad, and on this basis, the rate of return for youth studying abroad under different conditions is given. The purpose is to advise families of different backgrounds on whether to invest in studying abroad.

The definition of Accuracy and other index is shown in Eq.2 to Eq.5.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{4}$$

$$F\text{measure} = 2*\frac{(\text{Precision}*\text{Recall})}{(\text{Precision}+\text{Recall})} \tag{5}$$

The results of the algorithm comparison on the micro data are shown in Fig.3. As can be seen from the figure, compared with other methods, the accuracy and F-measure of the Resnet-GBDT classifier proposed in this paper are at a better level on microscopic data.
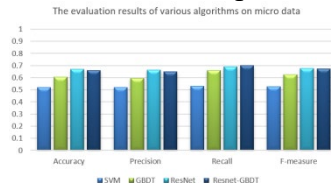

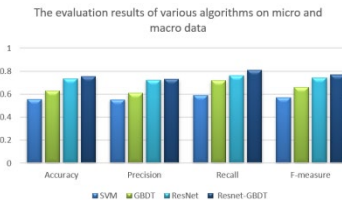Figure 3.The evaluation results of various algorithms on micro data.

Figure 4.The evaluation results of various algorithms on micro and macro data.

The results of the algorithm comparison on the micro and macro input data are shown in Fig.4. Compared with the standard SVM, GBDT and ResNet structure, the Resnet-GBDT classifier has better results in various indicators. In addition, compared with the algorithms based on micro-feature data, each algorithm has a certain degree of improvement in indicators. One of the reasons for this result may be that macro characteristics have a greater impact on annual income than micro characteristics.

5. Conclusions

Studying abroad has become one of the educational choices of many middle-class families, and the cost of capital investment and time investment is significant for a family. Therefore, studying abroad can be regarded as a way to invest in human resources in the future. From an economic perspective, analyze its advantages and disadvantages to help families make decisions about whether to study abroad. This paper proposes a method for calculating the return on investment of studying abroad based on the ResNet-GBDT model. By selecting micro variables and macro variables as input, the ResNet-GBDT regression model is used to predict the post-graduation income level of students to calculate the actual return on overseas study. The experimental results show that, to a certain extent, the ResNet-GBDT model can predict the economic return family obtained after youths graduate from different regions.

References

[1] Wang X. Capital, habitus, and education in contemporary China: Understanding motivations of middle-class families in pursuing studying abroad in the United States[J]. Educational Philosophy and Theory, 2020, 52(12): 1314-1328.

[2] Naito T, Zhao L. Capital accumulation through studying abroad and return migration[J]. Economic Modelling, 2020, 87: 185-196.

[3] Parey M, Waldinger F. Studying abroad and the effect on international labour market mobility: Evidence from the introduction of ERASMUS[J]. The economic journal, 2011, 121(551): 194-222.

[4] Comay Y. The benefits and costs of study abroad and migration[J]. The Canadian Journal of Economics/Revue canadienne d'Economique, 1970, 3(2): 300-308.

[5] Nerlich S. Outcomes-focused evaluation of study abroad experiences[J]. Journal of Higher Education Policy and Management, 2021, 43(2): 166-181.

[6] Hubbard A, Rexeisen R J. Investing in Study Abroad and Cultural Engagement: A Win-Win for Career Development[J]. Frontiers: The Interdisciplinary Journal of Study Abroad, 2020, 32(3): 3-21.

[7] Sutton R C, Rubin D L. The GLOSSARI project: Initial findings from a system-wide research initiative on study abroad learning outcomes[J]. Frontiers: The Interdisciplinary Journal of Study Abroad, 2004, 10(1): 65-82.

[8] Lu C, Reddick R, Dean D, et al. Coloring up study abroad: Exploring Black students' decision to study in China[J]. Journal of Student Affairs Research and Practice, 2015, 52(4): 440-451.

[9] Oosterbeek H, Webbink D. Does studying abroad induce a brain drain?[J]. Economica, 2011, 78(310): 347-366.

[10] Zhou F, Zhang Q, Sornette D, et al. Cascading logistic regression onto gradient boosted decision trees for forecasting and trading stock indices[J]. Applied Soft Computing, 2019, 84: 105747.

[11] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks[C]. European conference on computer vision. Springer, Cham, 2016: 630-645.