ISSN 2616-7433 Vol. 5, Issue 16: 40-45, DOI: 10.25236/FSST.2023.051608

Causal Inference-based Study Abroad Program Recommendation System

Qianyi Hou[1,a,#], Shuyuan Bao[2,b,#], Longxi Wang[3,c,#], Zhengxi Hou[4,d,#], Tianyuan Zhang[5,e,#]

[1]Department of Applied Physics, The Hong Kong Polytechnic University, Hong Kong, China
[2]Beijing No. 55 High School, Beijing, China
[3]Beijing Huijia Private School, Beijing, China
[4]Tianjin WeiMing School, Xiqing, Tianjin, China
[5]Beijing New Essence, Beijing, China
[a]hqysophiauk@outlook.com, [b]bao.shuyuan@qq.com, [c]Wanglongxo@outlook.com,
[d]3434994013@qq.com, [e]13126613535@qq.com
[#]Co-first author

Abstract: *In recent years, studying abroad has become an important way for more and more people to achieve their personal development goals. However, with a wide range of study abroad programs and varying standards, choosing a suitable study abroad program has become a challenge. This website proposes a recommendation system for studying abroad projects based on causal reasoning. This website conducts causal analysis and inference of user basic information to generate appropriate decisions. Predicting and inferring conclusions based on causal reasoning will lead to more precise planning actions, enabling international students to choose their study abroad projects more accurately, thereby improving the effectiveness of studying abroad. This website is based on a causal reasoning algorithm for career/future planning goals and has designed a corresponding recommendation system for study abroad projects, providing a scientific and reliable way for study abroad planners to choose their study abroad projects.*

Keywords: *causal inference, recommendation algorithm, SCM*

1. Introduction

When international students choose majors and courses for studying abroad, they often face many choices. Choosing a major and courses for studying abroad is a major decision that every international student must face before embarking on their study abroad journey. In this process, they often face many choices, including various majors and courses offered by different schools. These choices, like a maze, can be confusing and daunting. At this time, an effective study abroad content recommendation can play a crucial role. Recommendations can help students understand the pros and cons of different majors and courses, as well as detailed information about them, thereby helping them make choices that better suit their circumstances. Moreover, international students often face significant cultural and linguistic differences during their studies abroad. These differences can be troubling, making them feel lonely and confused. Study abroad content recommendations can help them integrate more easily into local culture and living environments, enhancing their confidence and improving their quality of life. The necessity of content for international students also lies in helping them alleviate the pressure of learning. With excellent study abroad content recommendations, students can acquire many outstanding knowledge and skills, thereby increasing their learning efficiency and reducing their learning pressure. From this, we can see that the recommendation of study abroad content is an essential powerful tool for international students, which can help them have a smoother and more pleasant study abroad journey.

2. Related Works

For the issue of recommendation, many researchers have conducted related studies. For example, in [1], the author proposed a reference model for product recommendation based on user interest. This model can be applied to product recommendations on e-commerce websites, attracting customer attention and satisfying user preferences, allowing customers to experience more personalized services.

ISSN 2616-7433 Vol. 5, Issue 16: 40-45, DOI: 10.25236/FSST.2023.051608

In [2], the author proposed an algorithm for product recommendation based on customer consumption character analysis of shopping records. [3] introduced a model for electronic product recommendation based on product satisfaction, focusing on the application of fuzzy logic to handle product satisfaction and rules for product recommendation based on comprehensive product satisfaction, and provided simulation examples. [4] proposed a real-time product recommendation method based on browsing preference mining after analyzing the advantages and disadvantages of various algorithms in current recommendation technology. In addition to product recommendation issues, there are many researchers who have conducted related research on personalized recommendation algorithms and job recommendation systems. For example, [5] introduced a job recommendation system designed and implemented based on Mahout, discussing the development process of the recommendation system in actual applications and the optimization of recommendation results. While [6] introduced a job recommendation model based on data mining of employee resignation status data, providing internal job recommendation services for internal staff of the company. At the same time, [7] introduced a job recommendation system based on user social relationships. Research on personalized recommendation algorithms has [8], which mainly introduced social tags providing a new data source for the recommendation system to obtain user preference information, and also posed new challenges to the traditional recommendation technology based on two-dimensional data. This article mainly focused on the project recommendation method based on tags. Then [9] introduced that e-commerce personalized recommendation has become an important part of customer relationship management, and collaborative filtering algorithms are the most widely used personalized recommendation technology. However, traditional collaborative filtering recommendation algorithms are not suitable for personalized recommendations under the situation of multiple user interests. The article combined user-based collaborative filtering and item-based collaborative filtering algorithms to solve the problem. Finally, [10] introduced that collaborative filtering recommendation algorithms are the key technology of personalized recommendation service systems. Due to the extreme sparsity of user rating data in the item space, the user similarity measurement algorithm in traditional recommendation systems is expensive and cannot guarantee item recommendation accuracy. The article constructed different item evaluations based on the attributes of the user's similarity, designed an optimized collaborative filtering recommendation algorithm, and conducted simulation experiments on the EachMovie database. The experiment showed that the algorithm had a higher accuracy rate.

While the above methods have solved many problems, they still have shortcomings, such as being based on user or product characteristics, rather than on objectives, and cannot provide references for the user's long-term planning. Based on the deficiencies of existing methods, we plan to design a recommendation algorithm for study abroad programs that recommends the most suitable programs in line with requirements and serves the user's objectives, based on users' long-term planning.

3. Causal inference-based study abroad program recommendation method

The recommendation algorithm based on causal inference can provide users with more accurate and personalized recommendation results. Causal inference is a method of reasoning based on causation, which deduces causal relationships between specific events or actions by observing and analyzing causal relationships. There are three levels of causal inference, namely association, intervention and counterfactual inference. Association is the simplest type that expresses a correlation between two variables and can be inferred directly from the data based on conditions and expectations. After the introduction of new factors, the correlation between the two variables cannot be directly derived from the statistical data. Counterfactual inference is the inference of cause from effect. Since the recommendation based on the association level has been unable to meet the needs of users, the recommendation based on the intervention level and the counterfactual inference level are also applied in the website.

First, we compare users with similar interests to find out what research projects they are involved in. Then we can recommend similar projects to other users based on the results of these projects and user reviews. Second, constructing a cause-and-effect diagram can visualize the cause-and-effect relationship between project characteristics and user characteristics. For example, the difficulty of the project, the topic, the instructor and other factors may affect the user's interest and engagement. By analyzing these cause-and-effect relationships, it is possible to determine which projects are more suitable for specific types of users and make recommendations. In addition, we recommend suitable research directions and projects according to the user's expectation of salary and job position.

ISSN 2616-7433 Vol. 5, Issue 16: 40-45, DOI: 10.25236/FSST.2023.051608

3.1. Basic Data Collection

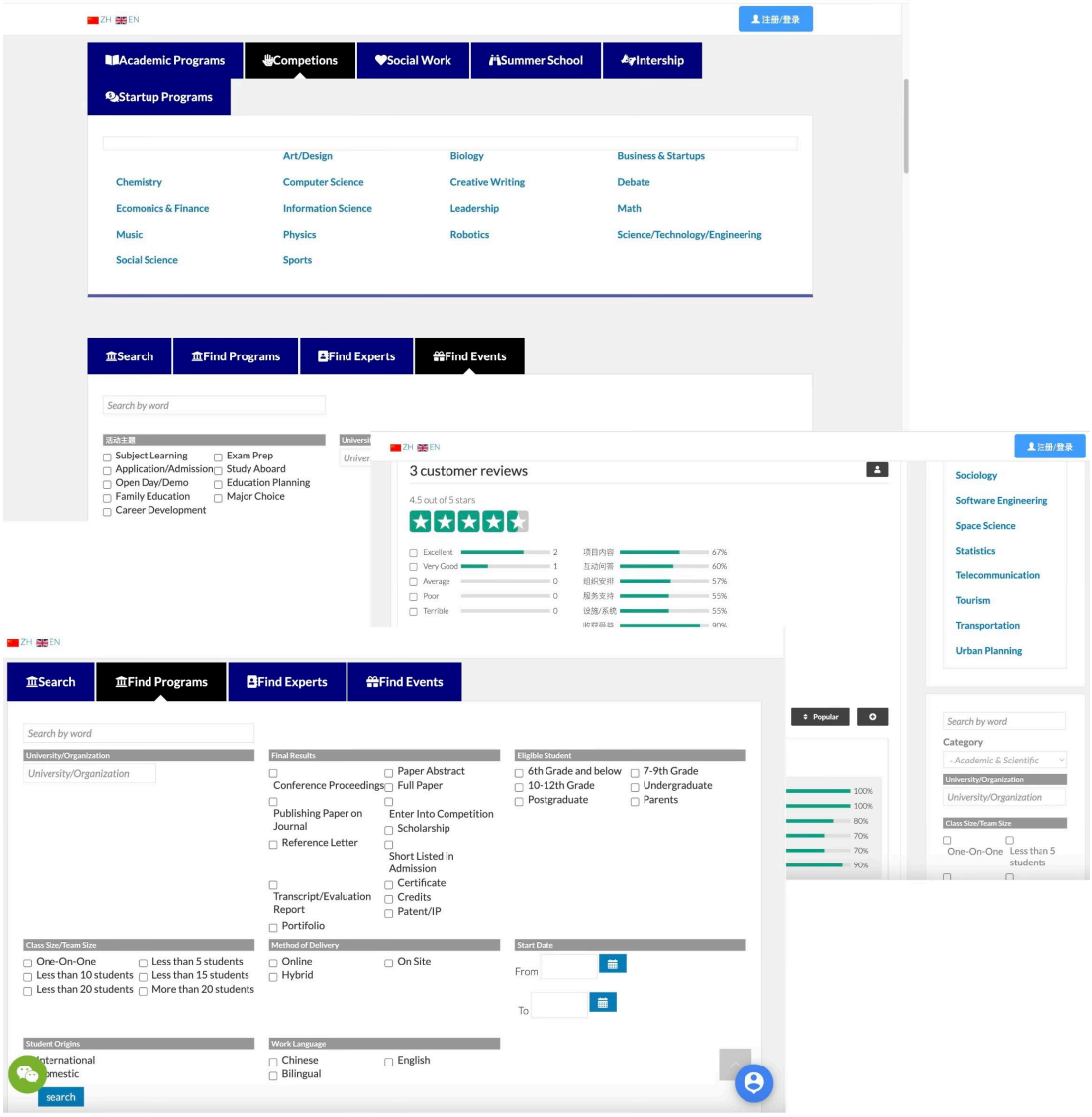

Figure 1: The website interface used to collect user information

In order to implement the study abroad program recommendation based on causal inference, we need to collect user's data. Then we need to analyze the causal relationship between the recommended project and the user's future development to establish a mapping model between the project and the user's expectations on the basis of this. In this way, we can recommend the best project for the user based on their expectations. According to the above objectives, the user data to be collected includes:

(1) User's personal information: including the user's name, gender, age, education level, interests and other basic information, which can be used for basic user analysis and positioning. This information may be collected from user's registration on the website;

(2) Study abroad programs accepted by the user;

(3) User's subsequent development status and user's evaluation of the project.

In addition to collecting user data, it is also necessary to collect information on all study abroad programs available in the system, as well as information on different fields, including:

(1) The majors offered by each university in each country, as well as the majors and strong disciplines that are highly ranked in the world for each university;

(2) The researchers, politicians or industry elites produced by each school and the employment rate

ISSN 2616-7433 Vol. 5, Issue 16: 40-45, DOI: 10.25236/FSST.2023.051608

(3) Employment and income levels of different majors in different countries;

(4) What are the employment prospects of different majors and the income levels of these employment opportunities;

(5) Whether the lifestyle of different fields and industries is slow or fast paced, and the living conditions and standards of the employees under those working state.

The above information can be collected through the user interface of this recommendation system. The typical interface is shown in Figure 1.

3.2. Recommendation algorithm design

3.2.1. Causality analysis of recommendation factors

Through the collection work in Section 3.1, we have obtained basic data for analyzing the potential relationship between study abroad programs and future career development. After collating, we can obtain data covering the following attributes: name, gender, age, interests, family status, project name, project corresponding major, industry, country/region, employer, position, salary, work status, living situation, self-evaluation.

Aiming at the ultimate goal of recommending study abroad programs for students, that is, maximizing their self-satisfaction, we can extract relevant influencing variables and outcome variables from the above data. Among them, self-satisfaction is the outcome variable. Name is obviously not necessarily associated with self-satisfaction, so this variable is eliminated. The other 13 attributes serve as influence variables. The following table 1 shows the meanings of each variable.

Table 1: Elements of project recommendation related variables

Gender	G	Gender attributes of students
Age	A	Student's age at the time of application
Hobby	H	Student's preference and interest in the project and university
Family	F	Student's family members, relationship, home address, income status and political status
Project	P	Briefly summarizes the main content of the research project provided on the website
Major	M	Majors to which the research program can be matched
Business	B	Industries that can be pursued later in different majors
Region	R	The country or region where the student wants to live
Company	C	Units where students want to work after finishing their studies
Job	J	Positions that students want to serve in
Income	I	Ideal future salary for students
Work	W	Students' expectations of work environment, work stress level, etc.
Live	L	Students' desired pace of life, living environment, etc.
Evaluation	E	Students' judgments and evaluations of their thoughts, desires, behaviors, and personal characteristics

Based on the relevant principles of causal reasoning, it is empirically possible to construct a causal diagram containing the above variables which shows the dependence and causality between the variables. The causal diagram is as follows in Figure 2:

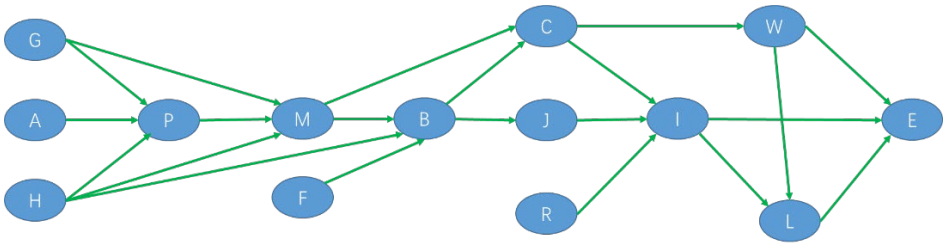

Figure 2: Causality diagram for study abroad program oriented recommendations

The above causal diagram can be interpreted as follows: students' gender, age, and interests will affect the choice of study abroad programs, study abroad programs will affect the choice of students' future majors, students' gender and interests will also affect the choice of their future majors, students'

ISSN 2616-7433 Vol. 5, Issue 16: 40-45, DOI: 10.25236/FSST.2023.051608

majors as well as their family status will affect the industries they will enter after work, etc., and the relationship between the variables is analogous. According to the theoretical basis of causal inference, the corresponding structured causal equation can be constructed for the above causal diagram as follows:

$$\begin{aligned} G &= u_G \\ A &= u_A \\ H &= u_H \\ P &= f_p(G, A, H, u_p) \\ M &= f_M(G, P, H, u_M) \\ B &= f_B(M, H, F, u_B) \\ C &= f_C(M, B, u_c) \\ J &= f_J(B, u_J) \\ I &= f_I(C, J, R, u_I) \\ W &= f_w(C, J, u_W) \\ L &= f_L(I, W, u_L) \\ E &= f_E(W, I, L, u_E) \end{aligned} \tag{1}$$

In the structured causal equation constructed above, the quantitative relationship between the variables is defined through functions such as f_P, f_M, and f_B. The specific form of each function depends on the relationship between each variable exhibited by the underlying data. From the point of view of simplifying the calculation, it can be assumed that the variables satisfy the linear relationship between them. Based on this assumption, the above structured equation can be converted into the following form:

$$\begin{aligned} G &= u_G \\ A &= u_A \\ H &= u_H \\ P &= k_P^1 \cdot G + k_P^2 \cdot A + k_P^3 \cdot H + u_P \\ M &= k_M^1 \cdot G + k_M^2 \cdot P + k_M^3 \cdot H + u_M \\ B &= k_B^1 \cdot M + k_B^2 \cdot H + k_B^3 \cdot F + u_B \\ C &= k_C^1 \cdot M + k_C^2 \cdot B + u_C \\ J &= k_J^1 \cdot B + u_J \\ I &= k_I^1 \cdot C + k_I^2 \cdot J + k_I^3 \cdot R + u_I \\ W &= k_W^1 \cdot C + k_W^2 \cdot J + u_W \\ L &= k_L^1 \cdot I + k_L^2 \cdot W + u_L \\ E &= k_E^1 \cdot W + k_E^2 \cdot I + k_E^3 \cdot L + u_E \end{aligned} \tag{2}$$

where $u_G, u_A, u_H, u_P, u_M, u_B, u_C, u_J, u_I, u_W, u_L, u_E$ is the confounding variable. Based on the above transformed structural equation, by substitution, the quantitative expression between variable E and other variables can be obtained as follows:

$$E = f(G, A, H, P, M, B, C, J, I, W, L, u_G, u_A, u_H, u_P, u_M, u_B, u_C, u_J, u_I, u_W, u_L, u_E) \tag{3}$$

By substituting the underlying data, the coefficients and confounding variables in the equation can be solved to obtain a specific quantitative relationship between the variable E and the other variables. Based on the solved function $f(\ )$, the quantitative influence of each variable on the outcome variable E can be clarified.

3.2.2. Causality-based recommendation algorithm for study abroad programs

Based on the functional relationship between E and each variable obtained from the solution in the

ISSN 2616-7433 Vol. 5, Issue 16: 40-45, DOI: 10.25236/FSST.2023.051608

previous section, the relationship between the study abroad program that the student needs to choose and the other variables can be obtained through the transformation:

$$P = f'(G, A, H, P, M, B, C, J, I, W, L, uG, uA, uH, uP, uM, uB, uC, uJ, uI, uW, uL, uE) \quad (4)$$

Based on the function$f'(\,)$, the recommendation system calculates the study abroad programs that need to be recommended to the student to achieve the specific goal by simply inputting the student's age, gender, hobbies and interests, family situation, the future major they want to study, the industry they want to work in, the organization and position they will be employed in, the country or region they will be living in, and their self-assessment of their living and working conditions, as well as their expectations to be achieved.

4. Conclusion

This war provides a reliable decision-making tool for international students to study abroad. This website is based on causal reasoning analysis to help international students choose the optimal project. Causal reasoning will extract appropriate variables from important information such as age and interest to predict and recommend the most suitable project for international students, while also providing more personalized choices for international students. Helping international students make more informed choices about their study abroad projects, and recommending study abroad content can help them better understand their situation and find suitable projects, thereby maximizing their ability to find their strengths. To better achieve one's career development and future planning goals.

References

[1] Huang Guangqiu, Jin Feng, Peng Xuyou. A collaborative filtering commodity recommendation system model based on interest level [J]. Microelectronics & Computer, 2005(3):5-8.
[2] Zhang Guangqian, Bai Xue. A new product recommendation method based on consumption personality [J]. Management Science, 2015(2):60-68.
[3] Huang Hong, Yang Zhuojun, Wang Ben. Application of fuzzy logic in e-commerce product recommendation system [J]. Computer System Applications, 2012(3):171-175.
[4] Xie Yi, Chen Deren, Gan Honghua. A real-time product recommendation method based on browsing preference mining [J]. Computer Applications, 2011(1):89-92.
[5] Huang Hongkun, Tang Jihua, Tong Wencan. Design and implementation of a job recommendation system based on Mahout [J]. Journal of Longyan University, 2019(5):21-26.
[6] Ren Ranran. Research on AI company job recommendation based on data mining [J]. Value Engineering, 2017(34):42-44.
[7] Wang Chao. Architecture and implementation of a job recommendation system based on social relations [J]. Digital Technology & Application, 2013(11):123-125, 127.
[8] Zhang Fuguo. Overview of personalized project recommendation system research based on tags [J]. Journal of Information, 2012(9):963-972.
[9] Yu Li, Liu Lu, Li Xuefeng. Research on personalized recommendation algorithms under multiple interests of users [J]. Computer Integrated Manufacturing System, 2004(12):1610-1615.
[10] Huang Guoyan, Li Youchao, Gao Jianpei, et al. A user clustering collaborative filtering recommendation algorithm based on item attributes [J]. Computer Engineering & Design, 2010(5): 1038-1041.