Establishment and Analysis of University Enrollment Promotion Strategy Prediction Model Based on Big Data Analysis Model and Market Response Prediction

Xin Zhao[1*], Yiyu Chen[2], Wenlong Wang[3]

[1] Dean's Office Undergraduate admissions office, University of Electronic Science and Technology of China, Chengdu
[2] Enrollment and Employment Division, Southwest Petroleum University, Chengdu
[3] Admissions office, Sichuan Agricultural University, Chengdu

zx@uestc.edu.cn, 616475853@qq.com, 46913907@qq.com

Abstract. The article analyzes and studies the problem of college enrollment publicity strategy prediction, uses the market reaction prediction theory in marketing, and uses the brand selection model, Cox proportional risk model, Buzz model, big data analysis model, and the concept of spot testing to build a college enrollment publicity prediction model, which provides a scientific and reasonable solution for colleges and universities to formulate enrollment publicity strategies. The conclusions of this study can better guide colleges and universities to formulate enrollment publicity strategies, and provide methods for colleges and universities to optimize the structure of students.

Keywords: College Admissions, Market Forecasting, Nowcasting, Big Data Analysis Model

1 Introduction

With the deepening of the comprehensive reform of the college entrance examination, the competition among college students has become increasingly fierce. In order to ensure the quality of students and optimize the structure of students, the recruitment publicity of colleges and universities has gradually become the core component of the recruitment work. With the continuous upgrading and deepening of enrollment publicity, the enrollment publicity strategy of colleges and universities has become particularly critical. However, there has been a lack of scientific and reasonable prediction models for the formulated enrollment publicity strategy to play a role, produce results and optimize the impact on the student source structure.

H. Wu et al. (Eds.): CSIEDE 2022, ACSR 103, pp. 872–878, 2022.
https://doi.org/10.2991/978-94-6463-108-1_96

2 Establishment of the Prediction Model of College Enrollment Publicity Strategy

In the context of big data, Hadoop distributed storage and computing platform should be used to build the prediction model, which is mainly composed of HDFS distributed file system and MapReduce parallel computing as shown in Figure 1.

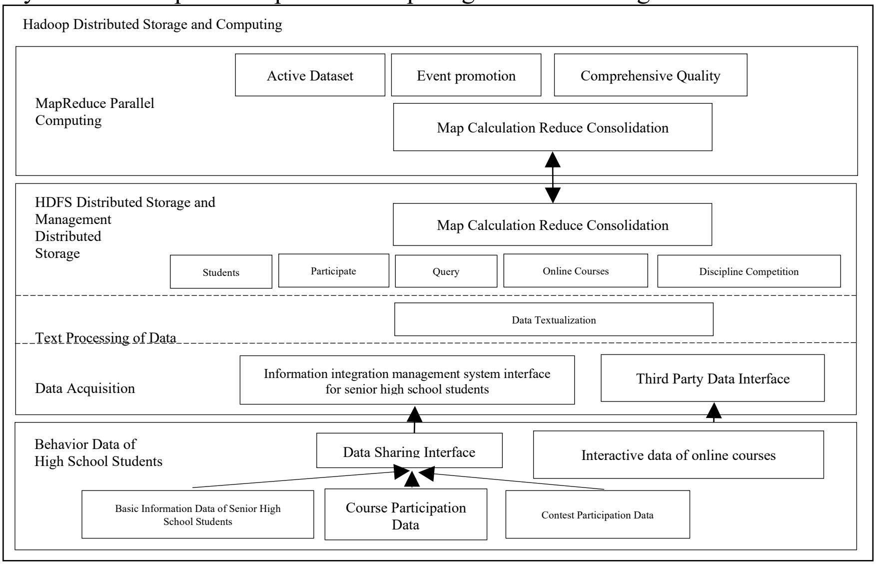

Fig. 1. Big Data Analysis Model of College Enrollment Behavior [Owner-draw]

In order to help colleges and universities make scientific and reasonable predictions on the enrollment promotion strategy, this paper uses the brand selection model, Cox proportional risk model and Buzz model in the market reaction prediction in marketing to build a prediction model for the enrollment promotion strategy of colleges and universities as a whole, as shown in Figure 2.

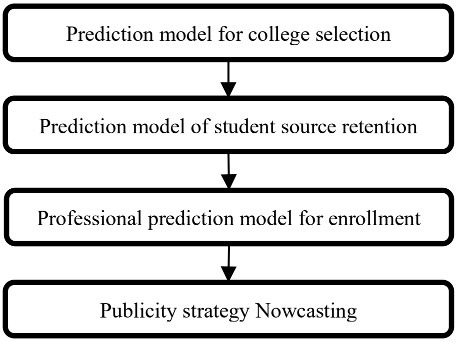

Fig. 2. Prediction Model of College Enrollment Publicity Strategy [Owner-draw]

2.1 Prediction Model for College Selection

For examinees, the greater the utility of a university, the higher the probability of examinees applying for the examination. The utility of colleges and universities includes two parts, one is the solid and effective use of colleges and universities, and the other is the impact of external factors such as college enrollment scores, enrollment strategies, and the quality of candidates. The specific model is as follows:

First, suppose that K observable variables jointly determine the utility, and use a linear equation to express the relationship between these variables and utility, as shown in equation (1):

$$V_{nj} = a_j + \sum_{k=1}^{K} b_k x_{njk}\, ; k = 1, \cdots, K; j \in C; j = 1, \cdots, J \tag{1}$$

Among them, a_j is the inherent utility of each university. Each university has its own unique a_j value, and J universities to be selected have a total of J such parameters. Usually these parameters are interpreted as the value of college entrance examination after controlling other variables. According to the estimated demand of the model, one of the J parameters can be limited to 0. Therefore, only J-1 such parameters need to be estimated.

x_{njk} is a common variable that can be observed in every university. Some of these variables are related to examinees and some are related to voluntary reporting. They will affect the effectiveness of each university. b_k is the parameter or weight corresponding to the k common variable. Each variable has a parameter corresponding to it, but different candidates with the same variable share the same parameter. Therefore, the candidate logo n is omitted from the subscript of the parameter.

It can be seen that although examinees share the same parameters on the same variable, the utility of the same institution may vary among different examinees due to different variable observations [1]. Taking the common Logit selection model as an example, the random part of its assumed utility follows Weibull distribution [2].

$$P_{nj} = \frac{e^{v_{nj}}}{\sum_{j=1}^{J} e^{v_{nj}}}; i, j \in C; j = 1, \cdots, J \tag{2}$$

In the formula (2), the numerator is the power function of the partial utility determined by College i, and the denominator is the sum of the power functions of the partial utility determined by all colleges in the college library. The random part of the utility no longer exists, which greatly simplifies the calculation process of the selection probability. By introducing equation (1) into equation (2), we get equation (3):

$$P_{nj} = \frac{e^{a_j + b_k x_{nik}}}{\sum_{j=1}^{J} e^{a_j + b_k x_{nik}}}; i, j \in C; j = 1, \cdots, J \tag{3}$$

In the above formula, x_{njk} is a known observation value, and a_j and b_k are unknown parameters that need to be estimated. We don't know the selection probability P_{nj}, but we know which university the examinee has applied from the university library. We use y_{nj} to express the result of n selection. If the examinee n chooses the university j, $y_{nj} = 1$, else $y_{nj} = 0$.

On the basis of the prediction model for college selection, the maximum likelihood method can be used to estimate the parameters of the model, and based on the estimated parameters, the selection probability of the future college and the student occupation can be predicted.

2.2 Candidate Retention Prediction Model

On the basis of reference to the survival analysis model, the examinee retention analysis model regards the process of the examinee from becoming the target student source of the university to entering the university as a complete life cycle [3]. For examinees, the time from the first contact with the university to the time of entering or giving up the university is called "survival time", which is a random variable and can be described by four functions: survival function s(t), probability function f(t), distribution function F(t) and risk function h(t).

According to whether to make assumptions about the distribution of parameters, the commonly used survival analysis method for examinees is the semi parametric estimation method, among which the Cox proportional hazard model is the most widely used, which can study the impact of each covariate on the survival status [4].

2.2.1. Risk function.

Let h (t) be the risk rate, which is the probability that the examinee will remain at time t and leave the university after time t.

$$h(t) = \frac{\text{Number of customers lost in the interval}(t, t+\Delta t)}{\text{Number of customers retained at time t} \times \Delta t} \tag{4}$$

Among them, the risk rate at time t without the influence of factor $X_1, X_2, \cdots, X_p$ is, and the risk rate at time t with the influence of factor $X_1, X_2, \cdots, X_p$ is $h_0(t)$. With the help of the construction idea of Logit model, Cox proportional risk model can be constructed as (5):

$$ln \frac{h(t|X)}{h_0(t)} = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \tag{5}$$

Therefore

$$h(t|X) = h_0(t)\, exp(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p) \tag{6}$$

Since $h_0(t)$ is unknown, it is also called semi parametric model. In this model, the partial likelihood function is established and the maximum likelihood method is used to estimate the parameters. The meaning of the parameter is: β_i means that the risk rate of death in the presence of X_i factor is β_i times that in the absence of X_i factor. $\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$ is the hazard index (HI), also known as the prognosis index or prognosis score. HI=0 means the risk is at an average level, HI<0 means the risk is at a low level, and HI>0 means the risk is at a high level.

2.2.2 Survival Function.

The survival function $S(t_i)$ represents the probability that the examinee will remain until time t, which is consistent with the construction idea of the above risk function. When there is no factor $X_1, X_2, \cdots, X_p$, the probability that the examinee will remain until time t is $S_0(t)$, while when there is factor $X_1, X_2, \cdots, X_p$, the probability that the examinee will remain until time t is $S(t_i)$, which can be expressed as Formula (7):

$$S(t_i) = S_0(t_i)\, exp(\, b_1 X_1 + b_2 X_2 + \cdots + b_p X_p) \tag{7}$$

2.3 Professional Prediction Model for Enrollment

After the major of enrollment is launched, colleges and universities are very concerned about the willingness of candidates to apply for the examination and their attraction to candidates. In this regard, the Bath model can be used for prediction. Based on the Bath model, we divide the candidates in the professional examination into the following two categories: (1) Innovative candidates. When innovative candidates apply for the major, they are only affected by external factors, especially the propaganda caliber. The innovation coefficient is set in the Bath model to reflect the possibility of these candidates applying for the major due to external factors. (2) Imitative candidates. Imitative candidates are only affected by internal factors, especially word of mouth, when they apply for the major. When former candidates apply and spread good word of mouth, they will follow up. Based on this, we build a major prediction model for enrollment according to the Bath model. The model formula is as follows [5]:

$$S_t = \left[p + \frac{q}{N} N_{t-1}\right] [N - N_{t-1}]$$

Among them, St represents the number of students applying for admission at time t; N represents the number of candidates who finally apply for admission; Nt-1 refers to the number of candidates applying for admission before time t; P is the innovation coefficient, which is also the external influence coefficient corresponding to innovative candidates; Q is the imitation coefficient, which is also the internal influence coefficient corresponding to the imitation candidates.

2.4 Field test of Publicity Strategy

The so-called spot test refers to the quick induction and presentation of changes in key influencing factors of the strategy, rather than the prediction of external factors. The previous prediction was based on the analysis of the impact and effect of the enrollment strategy, while the current measurement, because of the use of real-time and massive high-frequency data for analysis, can analyze and judge the entire process of the decision in the impact of the publicity strategy [6]. At the same time, colleges and universities can also analyze the inertia trend of candidates through their portraits, draw the needs reflected in the inertia trend in specific scenarios.

3 Conclusion

With the help of brand selection model, Cox proportional risk model, Buzz model and the concept of spot test, this paper constructs a prediction model for college enrollment promotion strategy, and makes a scientific and reasonable solution for colleges and universities to formulate enrollment promotion strategy. The conclusions of this study can better guide colleges and universities to formulate enrollment publicity strategies, and provide methods for colleges and universities to optimize the structure of students.

References

1. Eve J. Higginbotham A Singular Focus on Insurance Coverage Will Not Lead to Eye Health Equity [J] Ophthalmology, 2017, 124(10)
2. Sailesh Konda, Joseph Francis, Kiran Motaparthi et al. Future considerations for clinical dermatology in the setting of 21st century American policy reform: Corporatization and the rise of private equity in dermatology[J] Journal of the American Academy of Dermatology, 2019, 81(1)
3. The Lancet Abortion bans in the USA harm health equity[J] The Lancet, 2021, 398(10310)
4. Sara Cherri, Silvia Noventa, Alberto Zaniboni Is the oncologic patient more susceptible to covid19 but perhaps less likely to undergo severe infection-related complications due to fewer cytokines storm as a consequent of the associated immunodeficiency? [J] Medical Hypotheses, 2020, 141(prepublish)
5. Kenneth H. Cohn, Thomas R. Allyn, Robert H. Rosenfield et al. Overview of physician-hospital ventures[J] The American Journal of Surgery, 2004, 189(1)
6. Chen Chi-Hsien, Condello Giancarlo, Chen Wei-Han et al. Effects of a novel inclined-adaptive footwear on change-of-direction performance in male athletes[J] Gait & Posture, 2022, 94