

Research on Early Warning Control of University Budget Process Risk Based on Naive Bayes*

Lingfang Zhu

School of Economics and Management

Beijing Jiaotong University

Beijing, China 100044

Abstract—With the implementation of budget management in universities, some good results have been achieved. However, there are some problems in the course of operation, such as how to find out the risks in the course of budget project implementation and judge the risk level, and then take action effectively. This paper constructs the relevant indicators of budget risk early warning in universities. The indicators are mainly classified from the inside and outside the risk factors of the universities. Then the classification of the indicators is carried out by using naive Bayes in R language. It proved that the construction of these indicators can be used in universities budget process control in early warning, so that managers can timely detect the risk factors and take appropriate measures.

Keywords—budget process risk; early warning; naive Bayes; R language

I. INTRODUCTION

In 2012, the Ministry of Finance issued the "Financial System for Universities", which pointed out that the budget of a university consists of two parts: income budget and expenditure budget. The income budget consists of financial subsidy income, business income, superior subsidy income, subordinate units' paid-in income, operating income, and other income. The expenditure budget consists of business expenses, operating expenses, subsidies for subsidized units, higher-level payments, and other expenses. The budget of universities follows the principle of "reconciliation of income and expenditure, balance of payments". This confirms that universities implement budget management and set up budget-related systems to provide guidance. However, there are some problems in the implementation of the budget management process in universities, such as how to control the risk in the implementation of the budget and tap the potential source of risk to achieve the purpose of prevention, it still needs to be improved.

II. RESEARCH STATUS AND LITERATURE REVIEW

The earliest research on financial early-warning model was the univariate bankruptcy prediction model made by Fitzpatrick. Then Beaver selected 30 the financial ratio was studied. On the premise of excluding industry factors and corporate asset size factors, through a single test of 30 ratios,

*The Project Supported by China Railway Corporation Science and Technology Research and Development Program (2017X006-B).

the study found that the following financial ratios are effective in predicting financial crisis: cash flow/total debt, net income/asset total, total debt / total assets, of which cash flow / total debt indicators perform best [1]. However, there are also some problems. The established early warning mechanism cannot effectively serve the real society. The first warning mechanism is not complicated, but there is no consensus on whether the warning indicators are scientific, comprehensive, and the effectiveness and correctness of the early warning decision mechanism. Secondly, due to the extreme complexity and systemic nature of the external environment, it is difficult to obtain comprehensive information on the police situation. The establishment of a long-term effective, full-process, dynamic and intelligent warning system exists [2].

Early warning and control of financial risks are currently related to research focused on the relationship between corporate financial risk management and internal controls. Vanhorne et al [3] defined corporate financial risk as the risk of insolvency and the change in earnings per share due to the use of financial leverage. This view provides the basis for expanding the scope of financial risk.

The current methods for studying risk early warning models at home and abroad are shown in "Table I" below [4]:

TABLE I. COMPARISON OF RISK EARLY WARNING MODELS

Country	model
domestic	Single variable discriminant model
	Multiple discriminant analysis model (Z-score model, ZETA model, Logistic model, DEA model, neural network model, Black-Scholes option pricing model)
foreign	Single variable and multivariable analysis model
	Logistic regression model
	Neural network model
	Other models

III. INTRODUCTION OF FINANCIAL BUDGET MANAGEMENT IN M UNIVERSITIES

M University used traditional budget management methods in the past. Whether they used manual methods, spreadsheet methods, or financial software methods, they did not achieve the integration of business and finance, they could not report real-time budget execution status and they had insufficient control over business processes, which often result in the separation of business process and budget control, therefore, M

university adopted the process of budget-based management based on process collaboration and business finance integration, thus the purpose of business finance integration, flexibility

enhancement, horizontal coordination decision-making model change. A model of M university budget management is as follows:

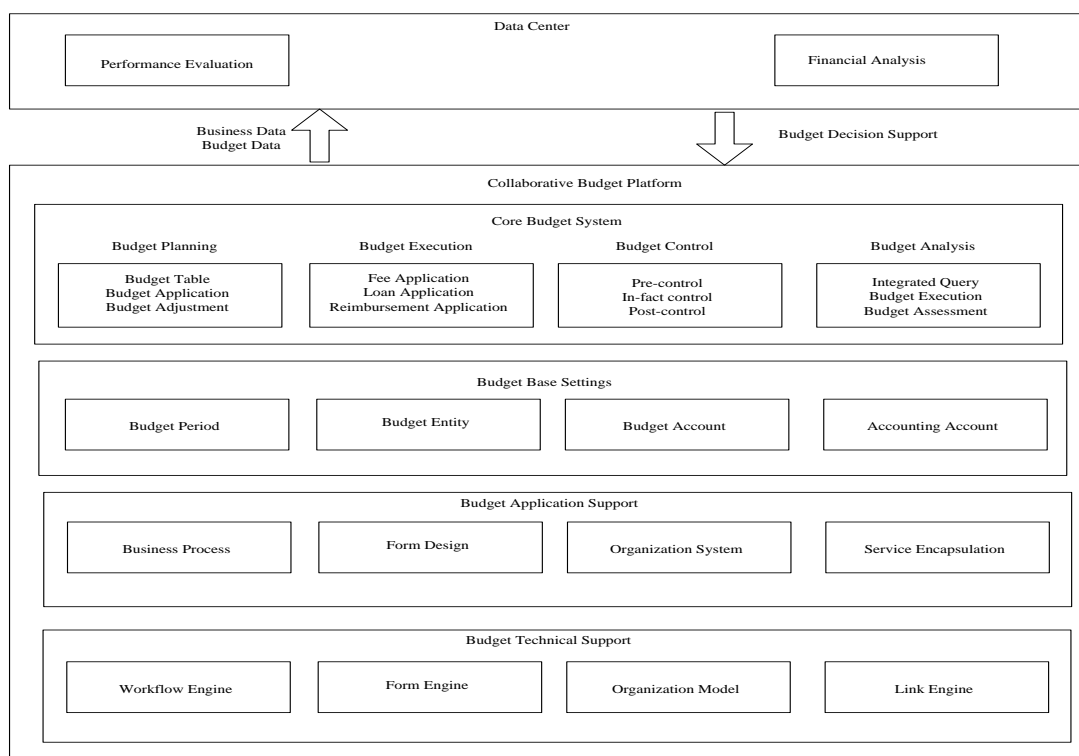


Fig. 1. Budget Management Model.

From the "Fig. 1", we can see that the collaborative budget management functional framework consists of two parts. The first part is the collaborative budget platform and the second part is the data center. The collaborative budgeting platform has good scalability, which can realize integration and exchange the data center and supports the economic activity analysis and business finance integration management of the M University. Integrate business data and budget data into the data center, use the data center's powerful data integration and data processing capabilities to achieve financial analysis and performance evaluation, and provide strong data support for budgetary decisions.

The collaborative budgeting platform realizes the basic functions of budget management and divides it into four levels in the application of overall planning. The first layer is the budget technical support layer. The key technologies involved include the workflow engine, form engine, organization model, and connection engine. It provides technical support for collaborative budget management. The second layer is the budget application support layer, which involves business process construction, form design, organization system, and business encapsulation. The third level is the basic budget setting, which relates to the setting of the budget period, budget entity, budget account, and accounting account. The fourth level is the core budget system, including budgeting, budget execution, budget control, and budget evaluation. The above-mentioned architecture system forms the current budget management mode of M universities.

IV. DATA ANALYSIS AND APPLICATION

A. Construction of Process Budget Control Index System

According to the relevant indicators related to early warning of the financial budget and the related interviews conducted by relevant experts and college personnel, the process budget indicator system obtained is shown in "Table II" below.

TABLE II. EARLY WARNING MODEL INDICATOR SYSTEM

Early warning indicator level	Indicators reflect the risks	Indicator name	Indicator code
Internal risk control	Department staff performance indicators	The role of the conference	S4
		The number of documents created	S7
	financial indicator	Executive rate of department budget	S8
		The difference between the Department's application budget and the actual reimbursement	S10
External risk coordination control	Staff coordination ability index	Interdepartmental communication and coordination	S11

The meaning of each indicator is as follows:

The role of the conference: The number of conferences attended by individuals in different roles can be counted to facilitate individuals and leaders to understand the participation of employees in meetings, reflecting the workload and value of personnel, and also reflecting the working ability of business personnel.

The number of documents created: The number of various interactions that statisticians have within the knowledge community, which reflects the degree of personal or team involvement in the knowledge community, and also reflects the degree of knowledge of the business staff.

Executive rate of department budget: It reflects the ratio of business department personnel to execution and budget application. The greater the ratio, the smaller the financial risk.

The difference between the Department's application budget and the actual reimbursement: It reflects the ability of department personnel to control the ability and process of budget execution.

Interdepartmental communication and coordination: It is an indicator of external risk coordination. The higher the score, the stronger the ability of personnel to handle emergencies and perceived risks.

B. Naive Bayes Model

Naive Bayes learning model is a very practical machine learning model in Bayesian learning. In the literature [5], Mitchell introduces the Bayes learning method in detail. Michie and others have studied the simple Bayes classification model in the literature, and compared the simple Bayes classification model and other learning algorithms, such as neural network and decision tree. The early text classification using naive Bayesian model is proposed by Lewis[6], Dumais[7], Joachims[8] and so on. It has become the most effective class of text document classification algorithms at present.

According to Bayes' theorem, for a classification problem, given the sample feature x , the probability that the sample belongs to class y is:

$$P(y/x) = \frac{P(x/y)p(y)}{P(x)} \quad (1)$$

Where x is a feature vector, assuming the x dimension is M . Because of the simple assumption that the feature conditions are independent, according to the full-probability formula, formula (1) can be expressed as:

$$P(y = c_k / x) = \frac{\prod_{i=1}^M P(x_i / y = c_k) p(y = c_k)}{\sum_k p(y = c_k) \prod_{i=1}^M P(x_i / y = c_k)} \quad (2)$$

Step 1: learning (parameter estimation) is to learn Naive Bayesian classification model from the given data.

The training set
TrainingSet = $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ contains N training data, where $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(m)})^T$ is an M -

dimensional vector, $y_i \in \{c_1, c_2, \dots, c_k\}$ belongs to One of the K classes.

Study 1: First use formula (2)

$$P(y = c_k) = \frac{\sum_{i=1}^N I(y_i = c_k)}{N} \quad (3)$$

Where $I(x)$ is the indicator function, count 1 if parentheses are true, 0 otherwise.

Study 2: the conditional probability in the calculation molecule, the j dimension of the M dimension characteristic has L values, then a certain value a_{jl} of a certain dimension, the conditional probability under a given classification c_k is:

$$P(x^j = a_{jl}) = \frac{\sum_{i=1}^N I(x_i^j = a_{jl}, y_i = c_k)}{\sum_{i=1}^N I(y_i = c_k)} \quad (4)$$

After the above steps, the basic probability of the model can be obtained and the data can be learned.

Step 2: classification. By learning the probability, given the unclassified new sample X , we can calculate the probabilities introduced above to obtain the posterior probability that the sample belongs to each class $P(y = c_k / x)$, because for all classes, the denominator value in formula (2) is The same, so only the molecular part can be calculated as follows:

Classification 1: calculate the probability of the instance belonging to the class $P(y = c_k / X)$

$$P(y = c_k / X) = p(y = c_k) \prod_{j=1}^n p(X^{(j)} = x^{(j)} / y = c_k) \quad (5)$$

Classification 2: determine the classification y of the instance.

$$y = \arg \max_{c_k} p(y = c_k / X) \quad (6)$$

This results in the classification of the new sample.

C. M University Budget Process Risk Early Warning Model

The data collection of this paper is mainly extracted from the M university budget management system, calculated and analyzed, analyzed the indexes of the 36 departments in Colleges and universities, and divided the risk grades by expert scoring methods, and a-c indicated that the risk grade increased in turn. The final analysis of the data is as shown in the following chart.

TABLE III. SAMPLE DATA (PART)

type	S4	S7	S8	S10	S11
a	3	65	33.65%	89	5
a	4	25	27.50%	50	4
a	5	53	22.80%	25	2
a	8	53	20.51%	19	6
a	7	32	18.34%	127	8
a	6	13	14.33%	98	5
a	6	50	7.60%	89	6
a	8	70	5.50%	77	9
a	6	53	5.36%	43	8
a	6	86	4.55%	71	7
a	8	40	2.33%	90	7

Using R language to analyze the above data. The proportion between a, b, and c in the original data is relatively close. Through the selection of the training set and the test set, it is found that the proportion is approximately the same as the overall proportion, so the sampled can be determined. The results can reflect the overall situation and can be further related to modeling and testing.

Through the naive Bayes model, the accuracy of the model is about 84% in the test set, and the AUC value is 81%. Generally, it is more than 0.8 that the model is ideal. The model established by this process is relatively satisfactory.

V. CONCLUSION

Based on the naive Bayes classifier, the indicators of the M university budget process control system is established. The accuracy rate of the establishment of variable is more than 84%. To a certain extent, the indicators of the system is found and established, which provides guidance for the management and control of the budget process in Universities.

REFERENCES

- [1] Dong Yaqin, Wang Yanan. A review of financial warning research literature [J], Hebei Business, 2013 (6): 68-68 2011(4):155-156
- [2] Shi Zhiwu. Research on the construction and evaluation model of safety risk early warning index system for large-scale projects[D]. Beijing University of Posts and Telecommunications, 2014
- [3] Xiong Tao. Research on China's Listed Companies Credit Risk Early Warning Based on Bayesian Network[D]. Zhejiang Normal University, 2013
- [4] Vanhorne JC, Wachowicz JM. Fundamentals of financial management[M]. New Jersey: Pearson Education, 2008
- [5] Mitchell T M. Machine Learning [M]. New York: McGrawHill, 1997: 112-143
- [6] Lewis D. Representation and learning in information retrieval[R]. Amherst, Massachusetts: University of Massachusetts Amherst, 1991
- [7] Dumais S, Platt J, Heckerman D, et al. Inductive learning algorithms and representations for text categorization [C] Proc of the 7th IntConf on Information and Knowledge Management. New York: ACM, 1998:148-155
- [8] Wu Yongliang, Zhao Shuliang, Li Changjing, et al. Text classification method based on TF-IDF and cosine similarity [J]. Journal of Chinese Information Processing, 2017, 31 (5) :138-145.