KPI Prediction Algorithm Based on Phase Space Reconstruction and Support Vector Machine

W. Yu, W. Li, H.B. Pan
Department of computer
Kunming Metallurgy College
China

Abstract——**The key performance indicators (KPI) system is of the complexity of setting and data processing. And its effectiveness of the application to the performance evaluation results is questioned in colleges and universities. With sufficiently analyzing on the characteristics of KPI data of colleges and universities, this article putts forward a more effective KPI prediction algorithm to achieve better data processing based on Phase Space Reconstruction and Support Vector Machine, and the pre-process of multi-filtering algorithm.**

Keywords-Key performance indicators (KPI); algorithm; prediction model

I. INTRODUCTION

Key performance indicators (KPI) are a set of target-oriented quantitative management indicators for evaluating the performance of a process by setting, sampling, computing and analyzing the key parameters of the input and output ends of organization flows. KPI extracts and concludes key successful factors in the process of enterprise operations. As the reform of colleges and universities in China develops in depth, the research on the performance appraisal has to focus on both the evaluation results and process. Therefore, more colleges and universities construct KPI system, which has exhibited significant effect in performance appraisal. However, there is still lack of systematically theoretical methods for guiding the set of specific and practicable KPI system that fits the development of colleges and universities. Aiming at this, a new prediction model for constructing KPI system for colleges and universities was proposed and implemented by algorithm by combining advanced theories and methods and summarizing the experience of performance appraisal organization at home and abroad.

II. DATA PROCESSING AND ALGORITHM MODELING

A. Characteristics of KPI Data of Colleges and Universities

At present, Chinese colleges and universities that apply performance appraisal mode mainly adopts KPI method for appraisal[1]. Based on the method, evaluation results are obtained by grading of appraisal subjects for assessment objects. Generally, KPI consists of quantitative and qualitative indicators. The former is accurate as it is based on specific grading standard. While the later fails to reflect the real performance appraisal results owing to the appraisal subjects mark scores depend on their personal judgment, which is influenced by multiple subjective and objective factors; besides, different marking standards cause different marking results of different appraisal subjects.

Based on the characteristics of KPI method for performance appraisal, for the large amount of regular data information of different appraisal indicators, the grade results are analyzed by considering the influence factors, and the data characteristics are as follows:

1) Large data size. According to the characteristics of KPI method, each assessment object presents multi-indicators, each of which has to be graded by multi-appraisal subjects. Therefore, the grade results present large size and are the square of the total number of assessment objects and appraisal subjects.

2) Normal distribution. The grade results of a same indicator show normal distribution characteristic. That is, most of the grades cluster in a close range, and the grades which are far from the center are interference terms.

3) Different grade standards. In different organizations, personal subjective factors cause different reference grades. Therefore, different appraisal results are possibly obtained for assessment objects with same performance, due to the different grade standards of appraisal subjects.

4) Flexible grouping of assessment objects. To reduce the difference of appraisal results, the grades are always sorted to generate assessment coefficient as the final appraisal result. However, owing to not all the groups and the organizations of the assessment objects are consistent; objects of different organizations are likely to be sorted in a same group. In the situation, the different standards show significant influences on the coefficient of appraisal results.

B. Data Preprocessing Method

In general, two times of grouping are conducted in performance appraisal. The first grouping is performed in the appraisal and grade process to divide assessment objects and relevant appraisal subjects according to their organizations; the second grouping is carried out in results computation process to obtain final appraisal results by sorting the results in a same group according to position sequence and ranks.

To make the appraised essential data be fair, eliminate the influences of objective and subjective factors as far as possible and obtain true and effective results, some rectifying deviation methods[2] are applied to adjust the results. Considering the characteristics of appraisal data above, the rectifying deviation process is divided into following two steps:

Filtering interference terms. The small amount of high or low grades were filtered using the normal distribution theory,

and merely the grades in effective coverage were collected and computed;

Processing result smoothing. Composite effect of multi-filtering algorithm was applied to perform data smoothing for the appraisal results in limited scope to obtain appraised essential data at same standard.

The common multi-filtering algorithm is described as follows:

According to the first grouping, the appraisal data are divided into sets A1, A2, …,Anin the number of n. Then filtering algorithm is used to eliminate the interference terms in the sets to obtain the average results of each set B={b1, b2, …, bn}. Afterwards, arithmetic average filtering algorithm is performed for set B to calculate its average value v. After calculating the difference C={c1, c2, …, cn } of the elements in B and v, each difference element in C is utilized for and smoothes sets A1, A2, …, An to compute the final appraised essential data.

Suppose that the filtering algorithm for interference items is $y = F(A)$, then $B = \{b_k | b_k = F(A_k), k \in N\}$. The average value of set B is $v = \sum_{k=1}^{n} b_k / n$, and the set that contains the differences of b and v is $C = \{c_k | c_k = v - b_k, k \in N\}$.

By smoothing, the kth set A_k is

$$y_k = F(A_k) + c_k \tag{1}$$

The theory of filtering algorithm is repeated in the algorithm to optimize the set interior and inter-set. This not only eliminates the influences of interference terms, but also smoothes the appraisal standards. The final results are sorted in a same standard and the appraisal data are more effective.

C. Modeling of KPI Prediction

1) Time series analysis

Based on observed time series data, time series analysis establishes mathematical model by curving fitting and parameter estimation (such as nonlinear least square method). The basic steps of time series modeling are: 1) to obtain dynamic time series data of the observed system by observing, investigating, collecting and sampling; 2) to draw correlogram according to the dynamic data, carry out correlation analysis, and calculate autocorrelation function. Correlogram shows the trend and period of the variation, as well as jump points and inflection points. Jump points refer to the observation points that inconsistent with other data. Jump points of correct observation values have to be considered in modeling; when they are abnormal values, they have to be adjusted to expected values. Inflection points are those points which turn the curve from uptrend to downtrend in the time series. If there is inflection point, piecewise fitting has to be conducted for the time series using different models, such as threshold regressive model, in modeling; 3) to perform curving fitting using suitable stochastic model, that is, to fit observation data of the time series using universal stochastic model. For stationary time series, it can be fitted using universal autoregressive moving average model (ARMA) and its autoregressive model in special situations, moving average model, or combined ARMR model.

2) Phase space reconstruction

Presently, most of the indicator data (taking the satisfaction of the courses on a day as an example) of KPI system of colleges and universities are chaotic time series, which cannot be predicted in normal conditions. However, the introduction of phase space reconstruction solves the prediction of chaotic time series. Takens theorem proved that the chaotic time series $x(t)$ is predictable theoretically, as long as a suitable embedding dimension was found. That is, in the space of the embedding dimension, the regular trajectory (attractor or singular operator) was recovered by the coefficient

$$m \geq 2d - 1 \quad (d \text{ is the dimension of the original system}) \tag{2}$$

of delay coordinate; or the coefficient was differential homeomorphism with the original dynamic system on the trajectory of the reconstructed R^m space [3]. By phase space reconstructing and using suitable parameters m and τ, the time series is

$$\{x(n\tau) | x(n\tau) = [x((n-1)\tau, x((n-2)\tau,..., x((n-m)\tau)], m = 1,2,...\} \tag{3}$$

According to Takens theory, for almost all m and the m meeting certain requirements in certain condition, τ presents a smooth mapping, which satisfies

$$x(n\tau) = f(x((n-1)\tau, x((n-2)\tau,..., x((n-m)\tau) \tag{4}$$

3) Support vector machines

Support vector machines (SVM) are a novel pattern recognition method for supervising learning [3]. Its computation complexity is not associated with the dimension; in contrary, it overcomes the difficulty of dimension. By using a few of support vectors, SVM solves the classification of vectors in high dimensional feature spaces and the linearly inseparable problem of vectors in ordinary two- and three-dimension space. SVM separates a kind of members from other kinds using optimal separating hyperplane. When performs regressive analysis, SVM becomes regression SVM. By introducing proper loss function, the objective of statistical learning is to find a function $f(x, a_0)$ in a function set $f(x,a), a \in \wedge$, in which the function minimizes the following risk function and empirical risk function:

$$R[\alpha] = \int L(y, f(x, \alpha)) P(x, y) dx dy, \alpha \in \wedge \tag{5}$$

$$R_{emp}[\alpha] = \frac{1}{l} \sum_{i=1}^{l} L(y_i, f(x, \alpha)) \tag{6}$$

Therefore the learning function is

$$f_l(x, \alpha_0) = \arg \min_{\alpha \in \wedge} R_{emp}[\alpha] \tag{7}$$

Suppose that the data set is $(x_1, y_1), (x_2, y_2),..., (x_i, y_i)$, where $x_i, y_i \in R^n, i = 1,2,...,l$, then the regression function is

$$f(x,\alpha) = wx + b \qquad (8)$$

That is to find proper w_0 and b_0 to minimize the following empirical risk function with the satisfaction of certain conditions.

$$R_{emp}[w,b] = \frac{1}{l}\sum_{i=1}^{l}|y_i - (w \cdot x) - b| \qquad (9)$$

Afterwards, the problem is transformed to the minimization of the following formula:

$$\Phi(w,\xi) = \frac{1}{2}\|w\|^2 + C\sum_i (\xi_i^- + \xi_i^+) \qquad (10)$$

Based on Lagrange multiply factor method and the inner-product cycle of SVM, the regression function is

$$f(x,v,\beta) = \sum_{i=1}^{l}\beta_i K(x,v_i) + b \qquad (11)$$

Where β_i ($i=1,\ldots l$) is scalar, v_i ($i=1,\ldots l$) is vector, and $K(u,v)$ is a function that meets Mercer condition .

D. Algorithm Implementation

In the performance appraisal system of colleges and universities, KPI indicators for evaluating classroom teaching are significant for students and attended teachers taking part in the appraisal on a day. They are also effective in teaching supervision and education reform. Therefore, the classroom teaching evaluation indicators were used as examples for prediction analysis.

Step 1: Owing to the first evaluation result of each course differs greatly from the data in other appraisal periods, the first appraisal data were extracted and processed independently, and differential process was performed for other data, as expressed by the formula $d(i) = a(i+1) - a(i)$. Where, $a(i)$ is the first appraisal data, and $d(i)$ is the difference of former appraisal data and current ones.

Step 2: Normalization processing,

$$r(i) = \frac{d(i) - mean(d(i))}{dev(d(i))} \qquad (12)$$

It was to transform the original data to the data in the range of (0,1).

Step 3: To calculate time delay by using mutual information method for $\{r(i)\}$.

Step 4: To compute the minimum embedding dimension m using Cao's method based on τ that calculated in step 3.

Step 5: To perform space reconstruction for $\{r(i)\}$ using τ and m, so that

$$r(n\tau) = f(r((n-1)\tau), r((n-2)\tau),\ldots, r((n-m)\tau) \qquad (13)$$

A point range $\{X_i, Y_i\}$, where $X_i \in R^m, y \in R$, was obtained in m-dimension space and utilized as the input of SVM learning machine.

Step 6: Training was conducted using SVM and the support vector set was obtained.

Step 7: X_{test}^m was constructed by using the latest m given data. By inputting X_{test}^m in SVM, the prediction value $Y_{predict}$ of next point was acquired.

Step 8: $Y_{predict}$ was placed in the series $\{d(i)\}$ as a true value and steps 2-8 were repeated until all the points were predicted.

III. CONCLUSIONS

Considering the research on the performance appraisal of colleges and universities is complex, attention has to be paid on multi-aspects. By using multi-filtering algorithm, the appraised data were corrected, the interference terms in the results were eliminated and the appraisal results were smoothed. The algorithm is effective for obtaining more accurate data set. Based on this, a new prediction model for constructing KPI system was proposed by using phase space reconstruction and SVM methods.

REFERENCES

[1] Bartler. The Development of Teacher Appraisal: A Recent History . Blackwell Publishers Ltd.:London, pp.89-91,2003:

[2] A. W.Brown. Model Driven Architecture :Principles and Practice. Software System Model ,2004(3),pp.314-327

[3] Support Vector Machine Based False Alarm Filter of Mechatronic Built-in Test.IEEE Transactions on Power System .R Salat,pp.979-986,2004.