Research on Data Optimization of Communication Resource Management System Based on Data Mining

Wenzhong Zhu, Hualong Jiang, Shuwen Zhou

Sichuan Provincial Academician (Expert) Workstation, Sichuan University of Science and Engineering, China

Abstract — In this paper, the author studies on the data optimization of communication resource management system based on data mining. Network extracts implicit process mode characteristics of input function to self-organize. The improved competition learning algorithm is given. Clustering result of UCI datasets shows that the proposed approach can be applied to timing clustering effectively.

Keywords - Keywords: Data optimization; Communication resource management system; Data mining

I. INTRODUCTION

Software engineering data refers to the data generated in the whole life cycle of software, including source codes, specification documents, and bug reports and so on. In most situations, software engineering data is the only information source for developers. With the rapid enlargement of software scale, to acquire manually the relevant information of software development and maintenance is becoming more and more difficult. Data mining technology can help to discover useful information from software engineering data automatically, which thus speeds up the process of software development.

Wu’s [1] paper investigates some data mining problems in software engineering data. Identifiers play an important role in software maintenance. A good method name explicitly indicates the function of the method to the developers, while bad names mislead them. However, except for developers’ individual experiences and some naming rules, the developers lack effective tools to guide them to name methods. We consider the problem of automatically choosing proper action names, or verbs, for methods as a classification problem. We build an automated method name selection tool, automatically choosing action names based on text categorization. Experiments show that our approach can achieve up to 70% naming accuracy for Java methods.

In Fu’s [2] paper, process neural networks is introduced into time series data mining to deeply studying on clustering, classification and prediction problems, combining with wavelet multi-resolution analysis and related technologies. A time series clustering method is proposed based on wavelet and improved self-organization process neural network for time series clustering problem. Original time series data is decomposed by wavelet. Under the principle of preserving clustering characteristics, the signal is reconstructed. And then reconstructed signal which has been fitted into time-varying functions is used as the input of process neural network. Self-organization PNN is trained by improved competition algorithm. Making use of time-varying input characteristic of PNN, the timing signal characteristics processed by wavelet is considered adequately in clustering analysis. Network extracts implicit process mode characteristics of input function to self-organize. The improved competition learning algorithm is given. Clustering result of UCI datasets shows that the proposed approach can be applied to timing clustering effectively.

A time series classifier is proposed in Huang’s paper [3] based on competitive radial basis process neural network for time series classification. The topology of competitive radial basis process neural network is given. The compound competition process neuron hidden layer is added to network. Discrete data in continues time points are fitted to time-varied functions as network input. The classifier breaks through time series data unequal length restriction. Pattern matching and temporal aggregation operation of time-varied input is achieved by competition process neuron units. The linear connection weights calculation in the output layer is omitted to simplify network structure and training process. Then generalization ability of classifier is improved by using different clustering coefficient for each clustering sizes and network training results. Learning algorithm is given. Finally performance and effectiveness of classifier are proved by multivariable time series classification simulation data.

Two process neural network models are proposed for time series prediction. One is genetic process neural network model. After input data are represented as a set of orthogonal basis expansions, using improved genetic algorithm to optimize genetic process neural network training process by introducing immigration operator. Prediction accuracy and generalization performance of the model is verified by per capita GDP prediction problem and contrast analysis. The second model is an improved feed forward process neural network, giving combined process neural network algorithm and applying to CPI prediction. The result shows that this model’s prediction accuracy is obviously higher than traditional neural network model.

A time series control and system identification method based on PNN is proposed in Dai’s paper [4] for control problem in time series context. Based on time-varying feature and complex nonlinear characteristic of time series control problem, double hidden layer process neural network is introduced into time series control and system
identification. Nonlinear and time-varying characteristics are taken into the consideration of time series control for the advantage of processing time-varying problem using PNN, and analyzed the model advantage used in time series control. Then, two wood drying control models are built by PNN trained by improved learning algorithm for the characteristics of wood drying process control. Two models’ analysis results compared to traditional neural network models show that process neural network model has better prediction control precision and generalization ability, the performance is superior to the traditional neural network model, PNN applied in time series control and nonlinear system identification problem is feasible. Lai’s paper [5] is mainly carried on some problems of time series data mining, which involves in some methods and models of clustering, classification and prediction. The performance and effectiveness of these methods and models are verified by experiments, example validations and comparative analysis. The research on time series clustering and time series classifier provides a good theoretical basis and ideas for further time series data mining research.

II. THE FRAMEWORK OF DYNAMIC VOLTAGE RESTORER

In contrast to the rapid development of software industry, especially open source projects, we lack effective and efficient plagiarism detection tools to protect them from piracy. A less self-disciplined developer, who committed software plagiarism, may elude punishment by seriously disguising the original open source projects. Existing Program Dependence Graph (PDG) based approaches can identify a few kinds of intentional disguises. However, they are rather slow and can be defeated by PDG affecting disguises. We develop a Type Dependence Graph (PDG) based plagiarism detection tool called TPLAG. TPLAG reduces the time complexity of graph based plagiarism detection from exponential to polynomial, and can see through several disguises that confuse existing tools.

Figure 1 shows the composition of data mining technology.
ignored. We propose a topic model based approach to analyze bug report and quantify the information in bug reports, i.e., bug report quality. Understanding WHAT has gone wrong and WHEN did it happen with an open source software project can provide high-level and useful information for software maintenance. We try to acquire such information from the bug repository of the project. The LDA model is applied to answer the WHAT question, and an event detection tool is developed to answer the WHEN question. Experiments validate the proposed methods in Liu’s [8] paper. Figure 2 shows the basic procedure of data mining technology.

III. THE DATA MINING ALGORITHM

A less self-disciplined developer, who committed software plagiarism, may elude punishment by seriously disguising the original open source projects.

The basic model for online error identification as follows:  

\[
\begin{align*}
TSP_1 & = 0 \leq t \leq \Delta t \\
TSP_2 & = \Delta t \leq t \leq 2\Delta t \\
& \cdots \\
TSP_n & = (n-1)\Delta t \leq t \leq n\Delta t \\
\min d(T) & = \sum_{j=1}^{n} \sum_{k=1}^{N} c_{0,j}(k\Delta t) \\
\text{s.t.} & \quad \Delta t = \frac{T}{n} \\
& \quad \Delta c_{i,j} \geq 0 
\end{align*}
\]

We may get the calculating method for the main index in the following equation (4)-(5):

\[
M_{ij} = \exp \left\{ \frac{k_i - x_j}{\sigma^2} \right\} \\
L = \begin{bmatrix} L_1 & \cdots & L_q \end{bmatrix}
\]

Their matching eigenvectors matrix is shown in the following equation (6):

\[
H = [h_1, h_2, \ldots, h_j] = A^{1/2} E
\]

So, we can get:

\[
U_y = \frac{H_i}{\sqrt{\sum_{i=1}^{n} H_i^2}}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, k
\]

\[
P = I - A^{1/2} MA^{-1/2}
\]

According to the equation (6), the calculating formula can be obtained in equation (7)-(10).

\[
\mathcal{G}(x, \omega) = \frac{1}{(2\pi)^{d/2}} \int \mathcal{G}(k, \omega) \exp(-i\omega \cdot x) dk
\]

\[
\mathcal{G}(k, \omega) = \begin{bmatrix} G_{\alpha}(k, \omega) \\
\gamma_i(k, \omega) \\
g(k, \omega) \end{bmatrix}
\]

\[
G_{\alpha} = (\Lambda_{\alpha} + \frac{1}{\lambda} h_j h_i^*)^{-1},
\]

\[
g = (-\lambda + h_j h_i^{-1} h_i^*)^{-1},
\]

\[
\gamma_i = \frac{1}{\lambda} h_i^* G_{\alpha}
\]

\[
\Lambda_{\alpha}(k, \omega) = k_j C_{\alpha,j} k_i - \rho^2 \omega^2 \delta_{ij},
\]

\[
h_i(k) = \exp(i \omega k_i),
\]

\[
h_i^* = \exp(-i \omega k_i),
\]

\[
\lambda(k) = \eta_i(k) k_i
\]

\[
1 \int_0^1 e^{-i \omega k_i} dk_i = \delta(k_i)
\]

\[
s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}
\]

\[
c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}
\]

The formula generates labels for each file block.

\[
\text{for } j = 0; j < n-1; j++;
\]

\[
W_j = e^* (j + 1); T
\]

\[
\text{Output}(T_0, T_1, \ldots, T_n);
\]

And local fractional integral of \(f(x)\) defined by Eq.9.

\[
a I^{(a)} f(t) = \frac{1}{\Gamma(1 + \alpha)} \frac{d}{dt} f(t)(dt)^a
\]

\[
= \frac{1}{\Gamma(1 + \alpha)} \lim_{\epsilon \to 0} \sum_{j=0}^{N-1} f(t_j)(\Delta t_j)^a
\]

Its local fractional Hilbert transform, denoted by \(f_x^{H,a}(x)\), is defined by

\[
H_a \{ f(t) \} = \hat{f}_H^{(a)}(x)
\]

\[
= \frac{1}{\Gamma(1 + \alpha)} \int_{a}^{b} f(t)(t-x)^a (dt)^a
\]

DOI 10.5013/IJSSST.a.16.5B.25 25.3 ISSN: 1473-804x online, 1473-8031 print
Where $x$ is real and the integral is treated as a Cauchy principal value, that is,

$$
\frac{1}{\Gamma(1 + \alpha)} \int_{x}^{\infty} \frac{f(t)}{(t-x)^{\alpha}} (dt)^{\alpha} = \lim_{\varepsilon \to 0^+} \frac{1}{\Gamma(1 + \alpha)} \int_{x-\varepsilon}^{\varepsilon} f(t) (dt)^{\alpha} + \int_{\varepsilon}^{\infty} \frac{f(t)}{(t-x)^{\alpha}} (dt)^{\alpha}
$$

(21)

IV. RESULTS AND DISCUSSION

Time series problem involves several fields, such as economy, meteorology, water conservancy, forestry etc. At present, time series data mining has been the focus of data mining, which has strong theoretical and practical significance. Because of some complex characteristics of time series data, for example, time variation, high dimension, noise jamming and volatility, time series data mining is always one of the difficulties in data mining research. Process neural network (PNN) is a development of traditional neural network in the time domain. Process input of PNN relax synchronization instantaneous limit on inputs in traditional neural network models, and is more general artificial neural network model. It has its own unique advantages in dealing with time related problems. Figure 3 shows the learning satisfaction based on data mining in communication resource management system and figure 4 shows data resource usage based on data mining in communication resource management system.

Figure 3. Learning satisfaction based on data mining in communication resource management system.

Figure 4. Data resource usage based on data mining in communication resource management system.

Figure 5 shows the optimization outcomes based on data mining in communication resource management system and figure 6 shows the result for communication resource management system on the data optimization.

Figure 5. Optimization outcomes based on data mining in communication resource management system.

Figure 6. The result for communication resource management system on the data optimization.
CONCLUSIONS

In this paper, the author studies on the data optimization of communication resource management system based on data mining. Network extracts implicit process mode characteristics of input function to self-organize. In order to improve the accuracy of the K-means method of knowledge discovery, the use of genetic algorithms to optimize the space of the initial value and obtaining the most valuable knowledge in the sub-space by weighting methods are separately proposed. This algorithm is defined as Valued K-means Genetic Algorithm. Based on the actual needs of the field of telecommunications knowledge discovery, the Data Mining Grid is an effective means of settlement to the problem that the efficiency of data mining algorithms will continue to fall when the data scale is increasing in geometric patterns. The improved competition learning algorithm is given. Clustering result of UCI datasets shows that the proposed approach can be applied to timing clustering effectively.

ACKNOWLEDGEMENTS

This work is supported by:

1. Fund Project of Sichuan Provincial Academician (Expert) Workstation (2015YSGZZ04);

2. Key Laboratory Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things (No.2014WYJ06, 2015WZJ01);

3. Sichuan Provincial Key Research Base of Intelligent Tourism (No.ZHZ14-02,ZHZ14-03).

REFERENCES


