International Journal of Smart Home
Vol. 10, No. 5 (2016), pp. 321-332
http://dx.doi.org/10.14257/ijsh.2016.10.5.30

A Novel Analysis System for Urban Construction Information Based on Case-Based Reasoning

Guanlin Chen[1,2], Yizhe Lu[1], Qun Liu[3] and Ting Zhang[4 *]

[1]*School of Computer and Computing Science, Zhejiang University City College, Hangzhou, 310015, P.R. China*
[2]*College of Computer Science and Technology, Zhejiang University, Hangzhou, 310027, P.R. China*
[3]*Zhejiang Geography Information Center, Hangzhou, 310012, P.R. China*
[4]*International Design Institute, Zhejiang University, Hangzhou, 310058, P.R. China*
**zhangting5@zju.edu.cn*

Abstract

Urban construction is the important infrastructure of the city, and it is the material basis for the urban survival and development. In this paper, a NASUCI (Novel Analysis System for Urban Construction Information) is proposed. Using data mining technology, case-based reasoning technology, the system focused on providing better decision support for urban construction. The system is composed of user management, enterprise information management, geographical information management, construction information display, construction information management, urban construction information analysis, and comprehensive statistics. NASUCI can help urban planners from a large number of original data mining more effective information, and to provide the chart display, make the analysis of urban construction become more scientific and efficient.

Keywords: *Urban Construction; Data Mining; CBR Technology; J2EE Technology; Baidu Maps API*

1. Introduction

With the economic progress and social development, the composition of the city is increasingly complex, and the rhythm of urban construction is getting faster and faster. Hangzhou capital construction commission accumulated a large number of data and the construction management experience in the construction field, but the relative information system and the business data are independent, therefore, it's unable to reach the requirements for unified decision and information integrated application. As a new task of modern urban construction, data mining technology can help us transform the data into useful information and knowledge [1].

Case-Based Reasoning is a recent approach to problem solving and learning that has got a lot of attention [2]. 2009, Goh Y M *et al.* proposed a case-based reasoning (CBR) approach to construction hazard identification [3]. 2010, Seo Joon-oh *et al* designed a cost model based on the owner's decision making at the early stages of a construction project [4]. 2012, Kim Miseon *et al* proposed an approximate cost estimating model for river facility construction based on case-based reasoning [5]. 2013, Long Le Hoai *et al* proposed a comparison of construction cost estimating models based on regression analysis [6]. 2014, Guanlin Chen *et al* proposed an intelligent flood control decision support system (IFCDSS) using data mining and statistical analysis technology for digital urban management [7].

* Corresponding Author

ISSN: 1975-4094 IJSH

However, we find that there is little research on the urban construction information analysis and mining system.

Urban construction is the important part of urban management, which is the basic work for the creation of good conditions for the management of cities. Urban construction can create a pleasant living environment for the people, protect the normal life, and service the urban economic development.

The main content of this paper is to design and implement a novel analysis system for urban construction information. This system can help urban planners from a large number of original data mining more effective information, and to provide the chart display, make the analysis of urban construction information become more scientific and efficient.

Besides, this system provides the function of construction information display and construction information management. With this system, urban planners can manage construction information and get effective information easily, which can improve their work efficiency and quality.

This system uses the MVC design pattern, the J2EE technology [8], CBR technology, the Hibernate framework, Dhtmlx component, MySQL database and Baidu Maps API plug-in technology to realize the urban construction information analysis.

2. Overall System Design and Analysis

2.1. Overall System Design

The main task of this system is to analyze the new construction information case, and search for the original case similar to the new construction information case. Then provide some useful information for urban construction. Finally, give the appropriate choice suggestion about the construction enterprises. It will help to speed up the progress of urban construction and improve the urban construction efficiency.

The main function modules of the system include user management, enterprise information management, geographical information management, construction information display, construction information management, urban construction information analysis, and comprehensive statistics.

NASUCI is composed of seven modules, and the basic function framework is shown in Figure 1 below.

- User management: including user add, user delete, user modification, used to manage user accounts;
- Enterprise information management: including enterprise information add, enterprise information delete, Enterprise information modification, enterprise information search, used to manage enterprise information and display on the map;
- Geographical information management: including geographical information add, geographical information delete, geographical information modification, geographical information search, used to manage geographical information and other functions;
- Construction information display: including display construction information, used to display urban construction information;
- Construction information management: including import construction information, construction information add, construction information delete, construction information modification, construction information search, used to manage construction information and display on the map, support batch import data from XLS file;
- Urban construction information analysis: including analysis result list, analysis result chart, used to analyze the urban construction information data, and show it in the form or a chart, and help the urban construction planner to make a better decision;
- Comprehensive statistics: including geographical information statistics, construction information statistics, used to show the number of projects in each district and construction information state distribution graphically.

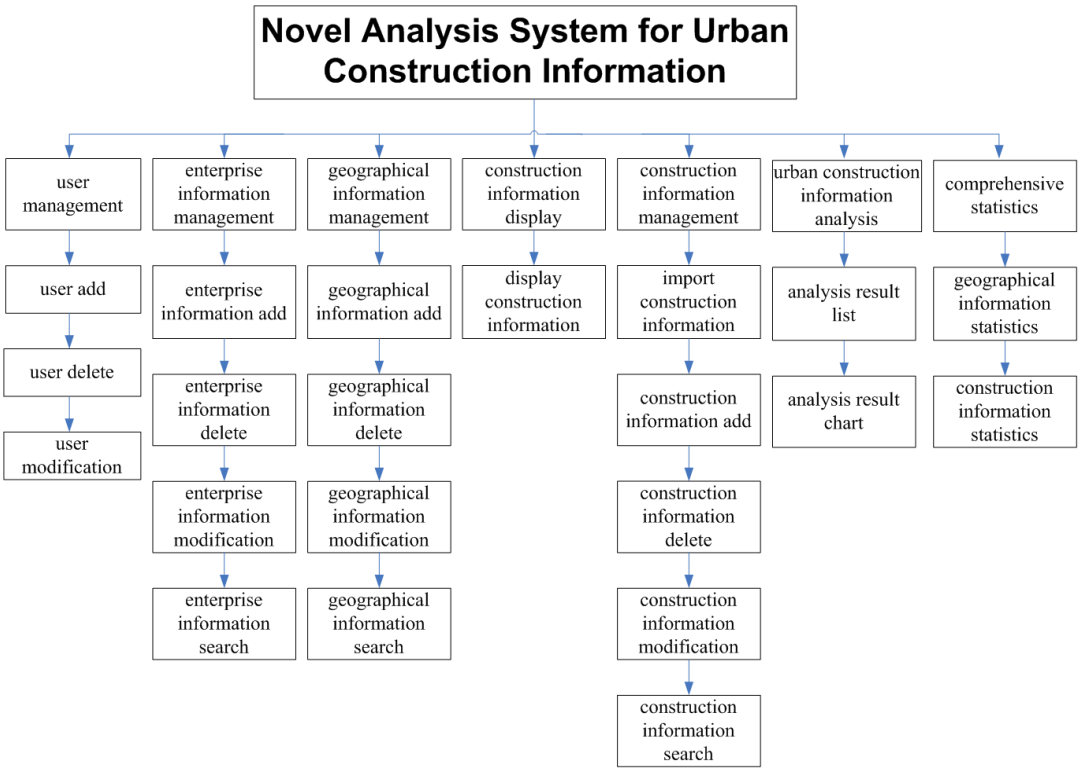

Figure 1. The Basic Framework of NASUCI

2.2. Database Design

This system uses the MySQL relational database, 5 data tables are designed, namely user table, enterprise info table, geographic info table, project info table and project status table.

The database structure of this system is shown in Figure 2 below.

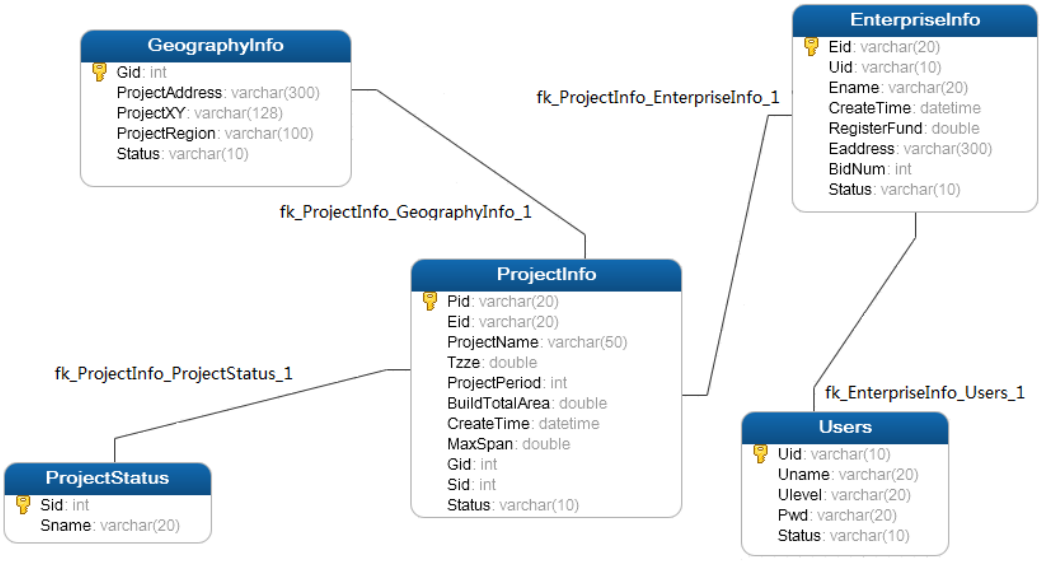

Figure 2. The Design of the System Database

The description of the relationships of these tables is as follows: The user in the user table has a role, and the role permission strings are different. The user table is also referenced by

the enterprise information table as an enterprise account. Besides, enterprise information table, geographic information table and construction information status table are referenced by the construction information table.

3. Detailed System Design and Implementation

NASUCI is designed based on the mainstream of the J2EE platform and the MVC pattern. At the same time, it comprehensively utilizes many kinds of intelligent technologies for construction information analysis and mining. For example, the system integrates with Baidu Maps API technology to realize the combination of construction information and geographic location; Using Case-Based Reasoning technology, through the database of case data analysis, find the matching results and analyze results; using Dhtmlx controls to display statistical results.

In the following, we will give detailed descriptions to some main modules such as construction information display module, urban construction information analysis module and comprehensive statistics module.

3.1. Design and Implementation of Construction Information Display Module

The construction information display module is designed to realize the GIS function in NASUCI. It will display the urban construction information and geographic information on the map. Therefore this function can help urban construction planners to find and view urban construction information more easily. The map display in NASUCI is implemented with the Baidu Maps API, which is easy to be used and open to all developers.

The construction information display module can display from the geographical position and information of the urban construction information in the database, including the exact latitude and longitude.

The JavaScript code of using the Baidu Maps API is as follows:

```
<script type="text/javascript"
    src="http://api.map.baidu.com/api?v=2.0&ak=ZDoQi3HcqAHehdo8MLLZsSPR"><
/script>
```

The system supports map zoom and drag, mini map display and satellite map display function. Specific code is as follows:

```
        map = new BMap.Map("allmap"); // Create map instance
        map.centerAndZoom(new BMap.Point(120.178987, 30.291556), 14); // Initialize the
map, set the center coordinates and map level as 14
        var mapType1 = new BMap.MapTypeControl({
    mapTypes: [BMAP_NORMAL_MAP,BMAP_HYBRID_MAP]
          });
        var overViewOpen = new BMap.OverviewMapControl({
    isOpen:true, anchor: BMAP_ANCHOR_BOTTOM_RIGHT
          });
        var top_left_control = new BMap.ScaleControl({
    anchor: BMAP_ANCHOR_TOP_LEFT});              // upper left corner, adding scale
        var top_left_navigation = new BMap.NavigationControl();  //upper left corner, add
default zoom translation controls
        map.addControl(mapType1);              //2D map and satellite map
        map.addControl(overViewOpen);     //Lower right corner ,set mini map open
        map.addControl(top_left_control);
        map.addControl(top_left_navigation);
        var opts = {      //build an information window
```


width : 250, // Information window width
height : 180, // Information window height
title : "CONSTRUCTION INFORMATION", // Information window title
enableMessage : false //Set the information window to send a short message
};

When initializing the map, we set the longitude and latitude in Hangzhou city and set the map zoom level to 14 in order to make the initial showing range to the city level. The red mark on the initial map is building information, and click on the red marker to see the corresponding construction information. What's more, the system has achieved three levels of building information display, respectively, showing the same company and display a single construction information.

The Baidu Maps interface integrated in NASUCI is shown in Figure 3.

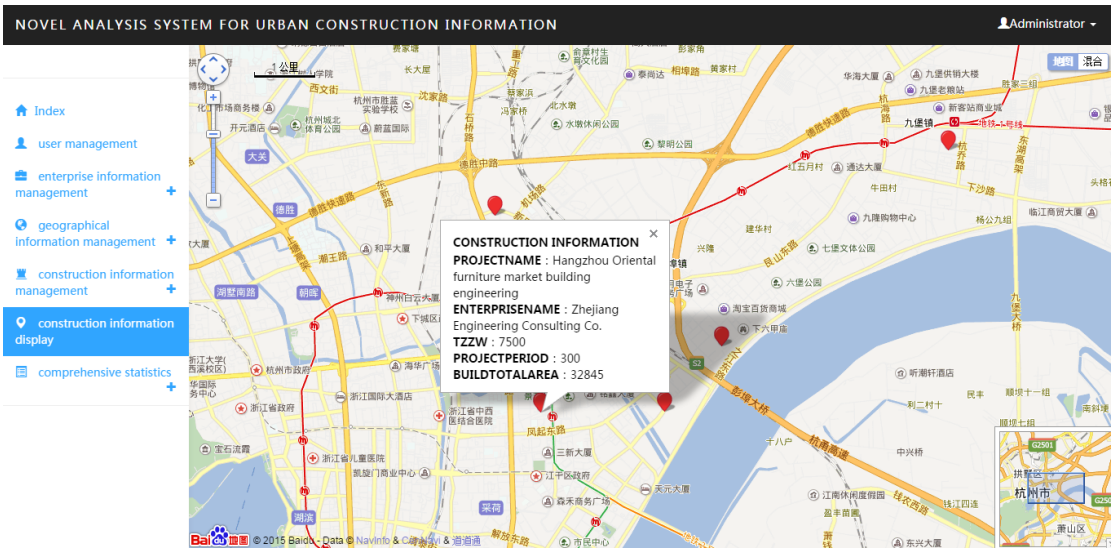

Figure 3. User Interface Using the Baidu Maps API

3.2. Design and Implementation of Urban Construction Information Analysis Module

The data analysis module of urban construction realizes the function of data analysis and data mining, using Case-Based Reasoning technology [9].

The basic idea of solving the problem based on CBR technology is: After modeling the experience (disposal results) and specific case information of the past, the case information is saved in the case database. When there is a new case, search for the most similar case with the new case from the case library. After appropriate modification and adjustment, it can provide a reference solution for the target case. Then put this objective case solution and concrete case information as an original case into the case database. With cases in the case base constantly expanded, and accumulated to a certain amount, for the experience to solve the problem of the field has become more and more abundant, carries on the analysis and prediction has become more accurate.

In this system, it will accord to the conditions of user input to analyze the conclusions and recommendations for the city construction information in the company's decision based on the choice, help city construction planner to make judgments faster and better. Specific process as showed in Figure 4.

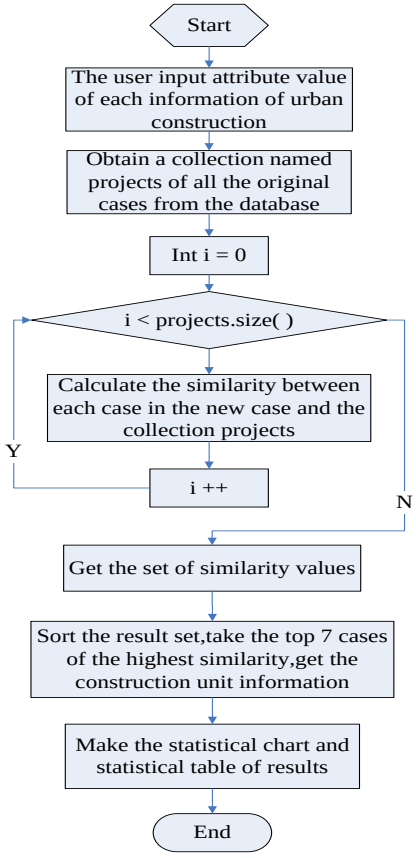

Figure 4. Analysis and Mining Process Flow

We can see from the chart that the most important part in the analysis and mining process is case similarity calculation.

In this system, for different types of properties, we use the different similarity calculation method.

(1) The similarity of continuous numerical attributes: For continuous numerical attributes, such as create time and project period. The similarity definition is defined in Equation 1.

$$Sim\ (s,t) = 1 - \frac{|s - t|}{\max(\ s,t) - \min(\ s,t)} \quad (1)$$

In this equation, s represents the original case attribute, t represents the same attribute of the target case in the case database, max(s, t) represents the maximum value of the attribute value domain, min(s, t) represents the minimum value of the attribute value domain.

(2) The similarity of character data attributes: The similarity of character data attributes can be divided into four cases:

a). Exact matching: If the structure of the two strings are similar, so that the two strings are similar, this method is suitable for the long sentences.

b). Spelling check: Compares the number of the same letters in the two strings. It is only applicable to each string contains only a single word.

c). Character statistics: Similar to the spell checker, the number of characters appearing in the same two strings is proportional to the total number of characters, and the method applies to a string containing multiple characters.

d). Segmentation based matching: Respectively to participle two strings, the similarity equals the result of the same number of words divided by the total number of words.

The accurate matching of the similarity computation not only needs to analyze the syntax of the sentence, but also analyzes the semantics of the sentence, which involves the knowledge of natural language processing, and appears more complex. Considering the ease of use and accuracy, the system is based on the MMSEG Chinese word segmentation matching algorithm [10].

For character data attributes, such as project name and project address, the similarity definition is defined in Equation 2.

$$Sim\ (string\ 1, string\ 2) = \frac{same\ (Stringtoke\ n(string\ 1), Stringtoke\ n(string\ 2))}{\max\ len\ (Stringtoke\ n(string\ 1), Stringtoke\ n(string\ 2))} \quad (2)$$

In this equation, same() function used for calculating the number of synonyms after two strings were segmented, maxlen() function used for calculating the maximum number of words after two strings were segmented.

(3) Case similarity and attribute weight vector: In the construction information, every attribute of each construction case has a different importance. In addition, with the new case continues to increase, the relative weights for each attribute in the case also need to change accordingly. Therefore, the system needs to adjust the weight of case attributes, for adaptation to the different needs.

When calculating the similarity of the case, use the Equation 3.

$$Sim\ (S, T) = \frac{\sum_{k=1}^{m} w_k \times sim\ (a_{ik}, a_{jk})}{\sum_{k=1}^{m} w_k} \quad (3)$$

In this equation, w_k represents the weight of the case feature vector, usually $\sum_{k=1}^{m} w_k = 1$. a_{ik} and a_{jk} represent the values of the k-th features of case S and T, $sim\ (a_{ik}, a_{jk})$ represents the similarity of the k-th features of case S and T.

The concrete implementation of urban construction information analysis and mining is as follows:

The user needs to enter the new construction information about the project name, total investment, project period, total construction area, create time, max span, project address and project region.

The construction data input interface is shown in Figure 5.

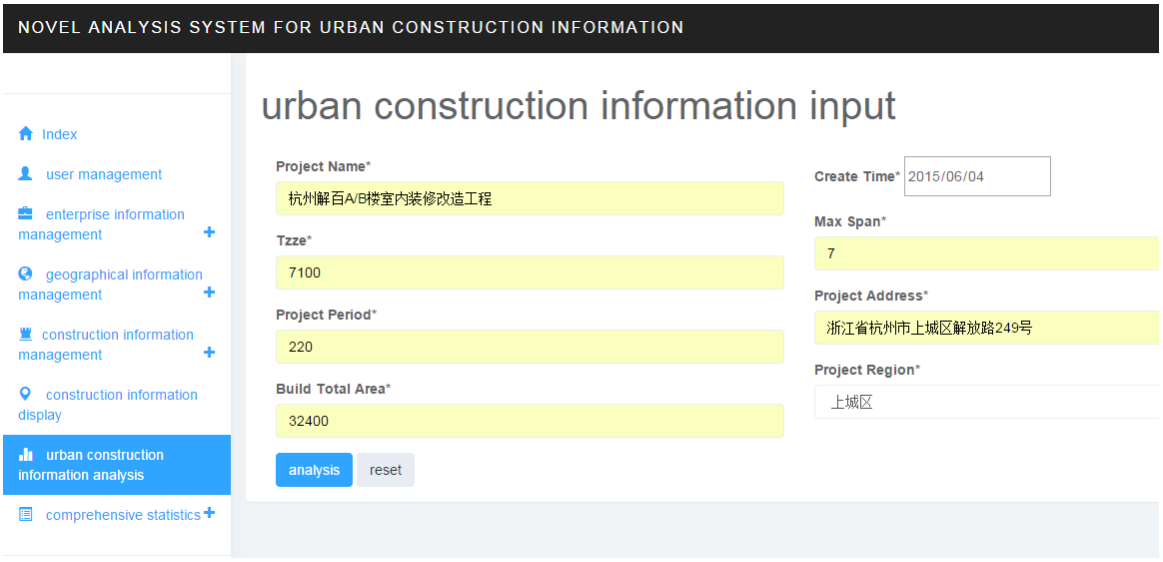

Figure 5. Construction Data Input Interface

After a series of calculations, system will show the inputs of the new case, the top 7 cases of the highest similarity, and the distribution of the construction enterprises. Also, there will be a pie chart and a relevant conclusion as showed in Figure 6.

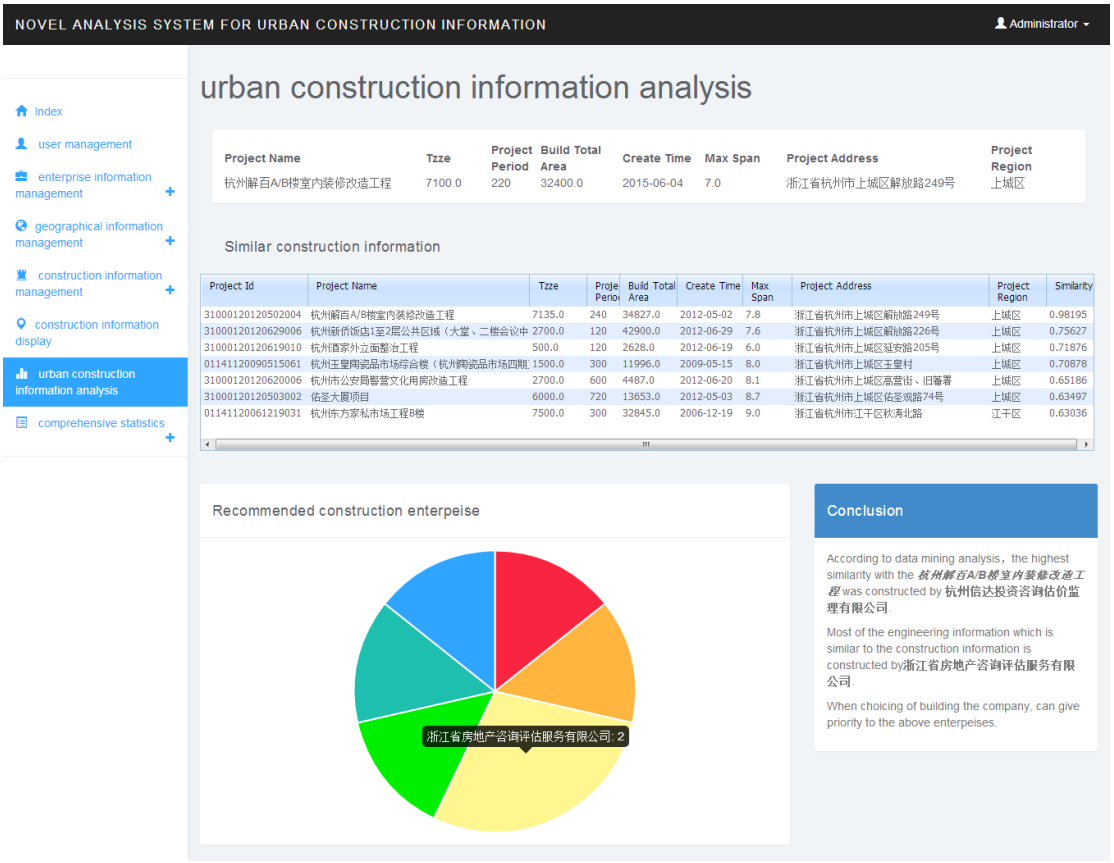

Figure 6. Building Data Analysis Results Interface

In the process of urban construction information analysis, the algorithm of calculating the similarity is most important. This system uses BigDecimal type of data for numerical calculation.

For character data attribute, the following method is used:

```
. public BigDecimal compareName(String e, String r) throws IOException { // The similarity of character attribute
        String example = new MessageSeg().segWords(e, ","); // Chinese word segmentation technology
        String reference =new MessageSeg().segWords(r, ",");
        String[] ee = example.split(",");
        String[] rr = reference.split(",");
        int count = 0;
        for (int i = 0; i < ee.length; i++) { // Calculate the same number of words
                for (int j = 0; j < rr.length; j++) {
                        if (ee[i].equals(rr[j]))
                                count++;
                }
        }
        BigDecimal denominator;
        if (ee.length > rr.length)
                denominator = new BigDecimal(ee.length);
        else
                denominator = new BigDecimal(rr.length);
        BigDecimal numerator = new BigDecimal(count);
        BigDecimal result;
        result = numerator.divide(denominator, 5,
                        BigDecimal.ROUND_HALF_UP);
        return result;
}
```

MessageSeg().segWord() method is based on MMSEG Chinese word segmentation matching algorithm, which separates the segmentation result by comma and puts into a string array. Then statistics the number of identical words in two string arrays as molecule, select the maximum number of words segmentation as the denominator. The similarity of two strings is equal to the ratio of the numerator and denominator.

For continuous numerical attributes, the following method is used:

```
public BigDecimal compareProjectPeriod(int e, int r) { // The similarity of continuous numerical attributes
        BigDecimal ee = new BigDecimal(e);
        BigDecimal rr = new BigDecimal(r);
        BigDecimal numerator = ee.subtract(rr);
        BigDecimal denominator = THOUSAND.subtract(TEN); //codomain[10,1000]
        BigDecimal result = ONE.subtract(numerator.abs().divide(denominator, 5,
                        BigDecimal.ROUND_HALF_UP));
        return result;
}
```

The above code takes the project period as an example. The project period of the new case minus the project period of the original case as the molecule, maximum range 1000 minus the minimum range 10 as the denominator, the similarity of two project periods equal to one minus the value of the absolute value of molecules divided by the denominator.

The similarity of the case is the attribute similarity multiplied by the attribute weight. In this system, due to the comparison of the simple analysis, so the weight values except the project name are 1.

3.3. Design and Implementation of Comprehensive Statistics Module

The comprehensive classification and statistics module realize the geographic information statistics and the construction information statistics, used to show the number of projects in each district and construction information state distribution graphically. The two charts are displayed using the Chart.js library. Chart.js is a simple, clean and engaging chart for designers and developers. Chart.js uses the HTML5 canvas element. It's supported in all modern browsers, and poly fills support for IE7/8. Chart.js is dependency free and super lightweight.

In this module, geographic information statistics mainly count the number of construction information within each region area of Hangzhou City in the form of a histogram display. The module's interface is showed in Figure 7.

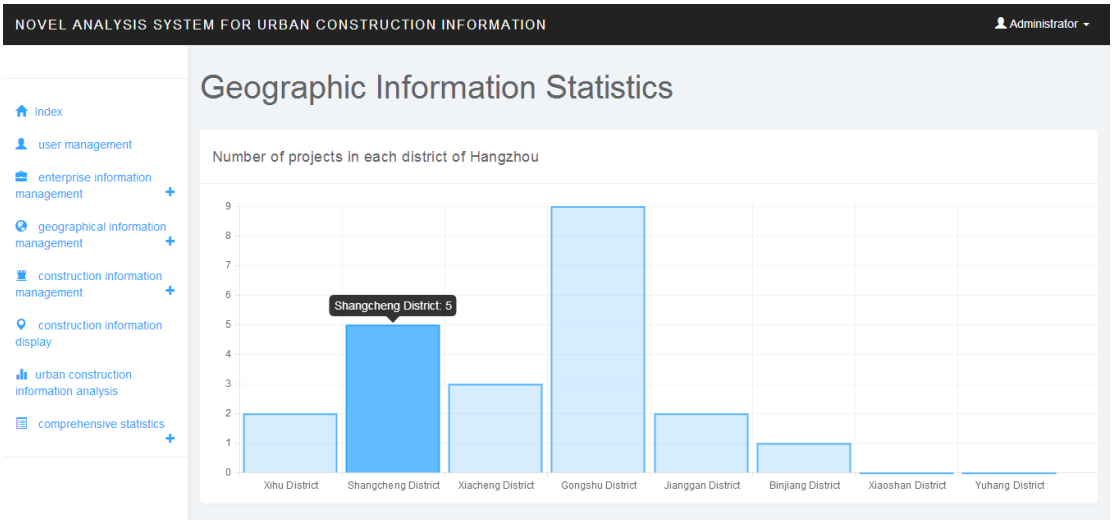

Figure 7. Geographic Information Statistical Interface

The JS code of the histogram realization is as follows.

```
var barChartData = {
        labels : ["Xihu District","Shangcheng District","Xiacheng District","Gongshu District","Jianggan District","Binjiang District","Xiaoshan District","Yuhang District"],
        datasets : [
        {
                fillColor : "rgba(48, 164, 255, 0.2)",
                strokeColor : "rgba(48, 164, 255, 0.8)",
                highlightFill : "rgba(48, 164, 255, 0.75)",
                highlightStroke : "rgba(48, 164, 255, 1)",
data:["<%=datas.get(0)%>","<%=datas.get(1)%>","<%=datas.get(2)%>","<%=datas.get(3)%>","<%=datas.get(4)%>
        ","<%=datas.get(5)%>","<%=datas.get(6)%>","<%=datas.get(7)%>"]}
                ]
}
window.onload = function(){
        var linechartgeo = document.getElementById("bar-chart-geo").getContext("2d");
        window.myBar = new Chart(linechartgeo).Bar(barChartData, {
```

```
                responsive : true
        });
};
```

4. Conclusions

This paper implements a novel analysis system for urban construction information (NASUCI), which is a complex system using data analysis and CBR technology. We concentrate on management, analysis and mining of the urban construction data. At the same time, we can show the results diversely in the system, which supports Dhtmlx component, map information and charts. In the future, we will continue to introduce the big data mining method in the system, enhancing the applied value of urban construction information.

Acknowledgements

This work is supported by the New-Shoot Talented Man Plan Project of Zhejiang Provence, China (No. 2015R401197) and the Public Welfare Technology Project of Zhejiang Province, China (No. 2014C31047).

References

[1] Y. Zheng, L. Capra, O. Wolfson and H. Yang, "Urban Computing: Concepts, Methodologies, and Applications", ACM Trans. Intelligent Systems and Technology, vol. 5, no. 3, **(2014)**.
[2] A. Agnar and E. Plaza, "Case-based reasoning: foundational issues methodological variations and system approaches", Artificial Intelligence Communications, vol. 7, no. 1, **(1994).**
[3] Y. M. Gohand and D. K. H. Chua, "Case-Based Reasoning for Construction Hazard Identification: Case Representation and Retrieval", Journal of Construction Engineering & Management, vol. 135, no. 11, **(2009).**
[4] K. C.-Wan, P. Sang, S. Joon-Oh, H. TaeHoon and H. Chang-Taek, "A Study on the Development of a Cost Model Based on the Owner's Decision Making at the Early Stages of a Construction Project", International Journal of Strategic Property Management, vol. 14, no. 2, **(2010)**
[5] M. Kim, S. Lee, S. Woo and D. H. Shin, "Approximate cost estimating model for river facility construction based on case-based reasoning with genetic algorithms", Ksce Journal of Civil Engineering, vol. 16, no. 3, **(2012).**
[6] L. L. Hoai, Y. D. Lee and A. T. Nguyen, "Estimating time performance for building construction projects in Vietnam", Ksce Journal of Civil Engineering, vol. 17, no.1, **(2013).**
[7] G. Chen, X. Sun, S. Li, J. He and J. Zhang, "An Intelligent Flood Control Decision Support System for Digital Urban Management", International Journal on Smart Sensing and Intelligent Systems, vol.7, no.1, **(2014).**
[8] L. Xue and L. Jie, "Research and implementation of information security emergency plan management system based on J2EE technology", Computer Engineering & Design, vol. 34, no. 4, **(2013).**
[9] D. H. Jonassenand and J. Hernandez-Serrano, "Case-Based Reasoning And Instructional Design: Using Stories To Support Problem Solving", Educational Technology Research & Development, vol. 50, no. 2, **(2002).**
[10] J. H. Jiang, S. Z. Zhao and M. Luo, "Analysis and application of Chinese word segmentation model which consist of dictionary and statistics method", Computer Engineering & Design, vol. 33, no. 1, **(2012).**

Authors

Guanlin Chen, Born in 1978, Ph.D., professor, chenguanlin@zucc.edu.cn. His main research interests include E-government, computer networks and information security.

Ting Zhang, Born in 1980, M.E., lecturer, zhangting5@zju.edu.cn. Her main research interests include innovative design, data management and software engineering.