

## Design method of data acquisition in intelligent sensor based on web data mining clustering technology

Ting Zhong Wang<sup>1, a</sup> and Hui Juan Sun<sup>2</sup>

<sup>1</sup>College of Information Technology, Luoyang Normal University, Luoyang, 471022, China

<sup>2</sup>Dept. of Information Engineering, Henan College of Finance & Taxation, Henan Zhengzhou, China

<sup>a</sup>wangtingzhongedu@163.com

**Keywords:** Intelligent sensor; Data acquisition; Web data mining; Association rule; K-Mean clustering

**Abstract.** Data mining mainly has two functions, namely prediction / validation function and function description. The main goal of Web mining technology is from the Web access record extraction interest pattern. The intelligent sensor has been widely applied in various fields of spaceflight, aviation, national defense, science and technology. The paper presents design method of data acquisition in intelligent sensor based on web data mining clustering technology. The experimental results show that the improved clustering method has good data collection effect.

### Introduction

The collection of the information is passed through the sensor technology to achieve, so the sensor detection technology is essentially information collection technology. Obviously, in the three major links of the modern information technology, however, with the rapid development of computer technology, information processing technology is constantly updated to improve. But as the element of the sensor to provide information, its development with respect to the information processing function of computer is lagging behind [1]. This allows the automatic detection technology is affected, but also directly affects the further development of various technologies. Based on the above factors, more and more scientific and technological workers attach importance to sensor technology, prompting the sensor technology to accelerate development, to meet the needs of the information processing technology.

The intelligent sensor has been widely applied in various fields of spaceflight, aviation, national defense, science and technology and the production of industry and agriculture etc.. For example, it has a broad application prospect in the field of robot, intelligent sensor make robots with human like facial features and functions of the brain, can sense various phenomena, to complete various actions.

In data mining, clustering analysis as an independent tool to get the data distribution characteristics, through the observation of each cluster, focus on the further analysis to specific cluster. In addition, the cluster analysis can also be used as a preprocessing step for other algorithms, the algorithm further treatment in the generated clusters.

Data, information and knowledge can be regarded as the different forms of generalized data expression. It is no exaggeration to say that people have to be greedy for data; especially the development of computer storage technology and network technology has accelerated the scope of people collecting data and capacity [2]. This greedy as a result of the "data rich and information poor (Data Rich & Information Poor) produce" phenomenon. Enterprise data volume is very large, which really has very little valuable information from the massive data, so after a deep analysis, be conducive to commercial operation, to improve the competitiveness of the information, like the gold from ore, data mining and hence the name.

Smart sensor is people no longer meet the detection sensor information of simple function, but hope that the sensor can be massive information analysis optimization, filter out useless data, will be the most useful information transfer for the upper actuator or controller. This is an important background of the birth of intelligent sensor. Web data mining is the mining technology in the Web

document and the server to automatically discover and extract interesting, useful pattern and hidden information data. The paper presents design method of data acquisition in intelligent sensor based on web data mining clustering technology.

## Web data mining and clustering analysis method

Web content mining is from the Web document content or its description for knowledge extraction process. Text mining based on Web document (Text Mining) is the main research content of Web content mining. Web structure mining is mining on web framework, that is, network traffic analysis, link distribution analysis, for improving the structure of network and enhances the website performance.

Association rules are initially proposed motivation for basket analysis questions. The ideal assumption is that stores more understand the customer's shopping habits [3]. In particular, wanted to know which products the customer may purchase while shopping at a time when? To answer this question, the number of retail customers cans things on store shopping basket analysis. The process is by finding customers into "association between different shopping baskets" in the analysis of customer's shopping habits.

The root node is to create the FP- tree, with "null" mark. For each transaction in D Trans, the executive: the selection of Trans frequent, and sort by L sequence. A frequent table after sorting is [p|P], where p is the first element, and P is the rest of the table of the elements. The following implementation of the process: if T have children N such that N.item-name = p.item-name, count of N increased by 1; otherwise, create a new node N, the count is set to 1, link to its parent T, and through the node chain structure linking it to the node with the same item-name, as is shown by equation(1).

$$Var[X] = E[(X - \mu_x)(X - \mu_x)^T] = C_x \quad (1)$$

Because of the organic combination of various new technologies and the database technology, the new content, new applications, new technologies in the database field emerge in an endless stream, forming a huge database of family. However, the application of these databases is based on real-time query processing technology as the foundation. In essence, the query is passive use of database. Due to the selective output of simple query are the contents of a database, analysis of it and expected there is still great distance Senior Application of prediction and decision support.

In order to reach the global optimum, partitioning clustering will require the enumeration of all possible partition based on. Most applications using the following two heuristic methods: K average algorithm, in this algorithm, each cluster using average object representing the cluster; k center algorithm, in this algorithm, each cluster with an object close to the cluster center to represent.

Often the work of data analysis is the aggregation of data, such as count, sum, average, and max. Since a lot of aggregation functions need to be frequently repeated calculation, and the calculation quantity of this kind of operation in general and particularly big, therefore, a natural idea is, put these summary operating result, pre computing and storing up, in order to use advanced data analysis. The most popular method of collection of data storage class is a multidimensional database, as is shown by equation (2) [4].

$$E_{jB}^{\xi}(m, n) = \sum_{m' \in J, n' \in K} w_B^{\xi}(m', n') [D_{jB}^{\xi}(m + m', n + n')]^2 \quad (2)$$

Effect of porosity is the ratio of credibility and confidence. Much of the impact mechanism degree to describe the items set A appear to the items set B. Because the probability in B items appear in all affairs is expected confidence; and the items set B in probability in A items appear in the transaction is credibility, credibility of confidence by the ratio reflects the join "the condition of A appeared" item set, sets the emergence of B changed how much probability.

Web log mining has been successfully applied to the personalized recommendation, system improvement and business intelligence etc.. Especially in the e-commerce site in the accumulation of a large number of users' access log data, businesses can provide personalized service for the users to use these data to improve the customer trust and service quality.

The Web server to collect user information through interaction process, according to the information of cutting a user requests a page, the page content and style according to the user requirements, and the customized pages returned to users as a result of the request. Such as: user registration to provide customized services website, can provide column and favorite option selected interest from the server page style, layout, color and so on, after the server in response to the user's request, will be returned to the user page revised according to the user's customized information.

The average k- algorithm is a classical clustering methods in statistics, but it is only in the cluster mean is predefined to good conditions of use, in addition to the noise data sensitive, which makes the poor adaptability for data mining, therefore, there have been some improved algorithm.

In fact, an important content of Web link structure is Web. In addition to link information, Web is the main content contains text, sound, pictures and other documents and information. Obviously, this information is key Association in-depth understanding of the Web site location. At the same time, this kind of mining also has a more challenging. Web content is rich, and the composition is complex (semi structure without structure, etc.), describe the semantic analysis of content can not be separated from the specific details of the words.

### **Design method of data acquisition in intelligent sensor based on web data mining clustering technology**

Intelligent sensor based on the measured sample value learning, new processor measured value by using the approximate formula and iterative algorithm can be cognitive, namely re learning ability. At the same time, based on the measured and the influence of the amount of learning, adaptive processor criteria by the reconstruction of structure and reset the parameters.

Node data output signal through the A/D converter to convert the intelligent sensor, digital signal obtained generally can not directly enter the micro processor for applications to use, must according to the need for processing, such as scale transformation, nonlinear compensation, temperature compensation, digital filtering, these processes are also known as processing software, as is shown in equation (3) [5].

$$C_1 = \frac{1}{\sum_{i=1}^n k \left( \left\| \frac{y_0 - X_i}{h_1} \right\|^2 \right) (1 + d(X_i))} \quad (3)$$

Because the extension of this kind of intelligent sensor is any functional needs corresponding space in hardware. Sensor system of multi function space limitation, and the performance of the same subject to subsequent processing circuit (filter, zero, compensation) control. Overall, the hardware constraints, makes use of a single sensor system quantity is limited, is unable to realize the measurement of some related quantities [6]. In order to guarantee the performance of the system, the need to increase the processing circuits more. Such as temperature compensation, need in the development of the sensor through many experiments to guarantee, and the circuit is special, it can not be directly transplanted.

Clustering analysis is used to find the distribution of data and models is an important research direction in data mining [7]. The clustering problem can be described as follows: the set of data points into several categories (clusters, cluster), between each cluster of data points in the maximum similarity, and the different clusters of data points maximally different.

Web access to the IP address, time stamp, method contains the user logs in (GET, POST), the requested URL, hypertext transfer protocol (HTTP) version number, return code (the requested state, success or error code), transmission bytes, has the characteristics of large amount of data, data distribution sparse, as is shown by equation (4).

$$p' = W(p, A, D) = Ap + D = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} p + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \quad (4)$$

Data and its associated always can come to understand it in a number of different conceptual levels. The problem of mining multi-level generalized knowledge relationship we described earlier, in certain background knowledge, a relational database can be in multiple concept level mining related knowledge.

The minimum support threshold  $\text{minsupport}$  represents a data item sets minimum major in statistics. The minimum reliability minimum confidence threshold  $\text{minconfidence}$  said rules. If the data set  $X$  satisfy  $X.\text{support} \geq \text{minsupport}$ , then  $X$  is a large set of data items.

Because clustering objects may be multidimensional data or high dimensional data clustering, a lot of research and index research unifies, is committed to the clustering method based on effective index structure, such as: CURE based on k-D tree; DBSCAN can make full use of the  $R^*$  tree or  $X$  tree; BIRCH based on CF tree foundation of STING; in the class four tree; and DENCLUE is using B ten trees [8].

In order to meet the exchange between the multi sensor information, design of intelligent sensor software occupies the main position, through the software will be contacted within each sensor will total sensitive unit or with the outside of the intelligent sensor units together. A software object is no longer the object before the individual, but the entire system, the digital signal output is in line with a certain protocol format [9]. Thus to realize sensor and sensor, sensor and actuator, and it is sensor between the system and data exchange and sharing. So the intelligent network is the future development direction of sensor.

The Web server through the analysis, judgement of users A and B belong to the same group, assuming the user profile in the group  $\{a.html, b.html, c.html\}$ , user A has access to the page contains a.html pages, then the real-time recommendation module will be intelligent Web site to the user A recommendation b.html and c.html this two page [10]. If the user has access to the B page contains b.html pages, then the real-time recommendation module should recommend the a.html page and c.html page to the user B.

Clustering is in Web log on: there are two kinds of user and page clustering. User to user session clustering is analyzed, according to the user access behavior mode of action, to find similar users. That is user proximity, visit the page with more overlap, these users are divided into a group, group users can share a user configuration file (User Profile), the configuration file for the group of users to access the high frequency page set. The paper presents design method of data acquisition in intelligent sensor based on web data mining clustering technology, as is shown by figure1. Finally, it is to cluster the users using clustering techniques, the experiment using the clustering of the data after the personalized recommendation. Improved K-Mean clustering algorithm is to reduce the outlier, improving the accuracy of recommendation, as shown by Fig.1.

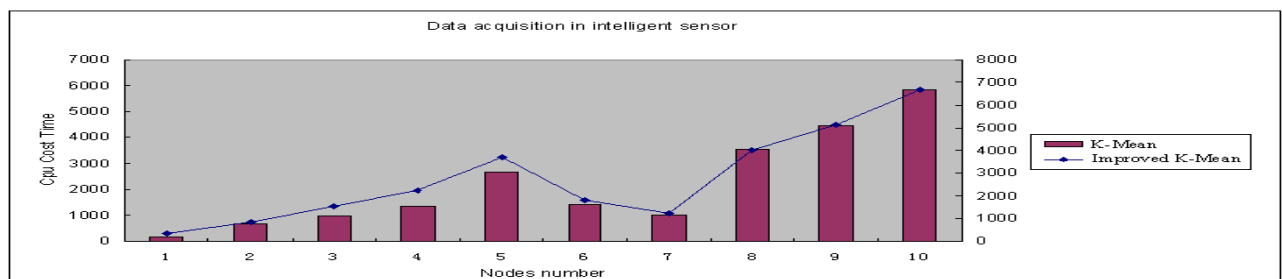


Figure. 1 Design method of data acquisition in intelligent sensor based on web data mining clustering technology

In most cases, the proposed intelligent sensor in this text is actually refers to the three categories of devices: non electrical input parameter is converted into electromagnetic sensor signal output; the electrical signal is converted into the actuator output non electrical parameters; and can be used as a sensor and can be used as an actuator, which is more an electromagnetism parameter form into

another form of electromagnetic parameters of transmitter output. That is to say, about the intelligent sensor data acquisition experiment can expand the analogy to the micro actuator.

## Summary

Data mining mainly has two functions, namely prediction / validation function and function description. The former refers to predict or verify other unknown properties with the number of known properties of the database, and can automatically predict the trend and behavior; the latter means finding comprehensible description of the data model. Although not constituting method intelligence sensor in all occasions is reasonable, but in many applications, the relative advantages of traditional sensors and is irresistible. In most cases, the intelligent sensor cheap, convenient use, superior performance, simple maintenance, easy expansion of functions has the advantages of the traditional sensor can not match. Especially in some of the more sensor application occasions undoubtedly intelligent sensor is the most reasonable choice.

## References

- [1] Mobasher B, Cooley R, Srivastava J. Automatic personalization based on web usage mining. *Communications of the ACM*, 2000, 43(8):142-151.
- [2] Xuchong Liu, Dafang Zhang, Qiang Fang, Wei Zhao, Ningjia Zhu, "Convergence Optimization of K-means Clustering Algorithm in Network Intrusion Detection", *JDCTA*, Vol. 8, No. 6, pp. 89 ~ 98, 2014.
- [3] Srivastava J, Cooley R, Deshpande M, et al. Web usage mining:discovery and application of usage patterns from web data. *SIGKDD Explorations*, 2000,1(2):12-23.
- [4] DanDan Cui, CaiQian Zhang, "The Novel Model of Fusion Temperature Error Compensation System by Neural Network and Smart Sensors", *JCIT*, Vol. 8, No. 5, pp. 379 ~ 387, 2013.
- [5] DanDan Cui, CaiQian Zhang, "The Novel Model of Fusion Temperature Error Compensation System by Neural Network and Smart Sensors", *JCIT*, Vol. 8, No. 5, pp. 379 ~ 387, 2013.
- [6] Jao-Hong Cheng, Chih-Ming Lee, Ting-En Lee, Liang-Chien Lee, "An Integrated Business Bankruptcy Prediction Model based on K-mean Clustering, Rough Sets and Support Vector Machines", *AISS*, Vol. 5, No. 17, pp. 15 ~ 31, 2013.
- [7] Jingyong Li, Lan Huang, Tian Bai, Zhe Wang, "A Novel High-Quality Community Detection Algorithm Based On Modified K-Means Clustering", *IJACT*, Vol. 4, No. 11, pp. 248 ~ 256, 2012.
- [8] Schafer J B, Konstan J A and Riedl J. E-Commerce recommendation applications. *Data Mining and Knowledge Discovery*, 2001,5(1-2):115-153.
- [9] Da-meng DAI, De-jun MU, "A Fast Approach to K-means Clustering for Time Series Based on Symbolic Representation", *IJACT*, Vol. 4, No. 5, pp. 233 ~ 239, 2012.
- [10] Ling Zhu, Jie Lin, Shuai Zhang, Yanan Liu, Weiwei Fu, "The Niche Hybrid Genetic Algorithm with K-means for Text Clustering", *AISS*, Vol. 5, No. 3, pp. 717 ~ 725, 2013.