Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Science and Engineering

Elixir Comp. Sci. & Engg. 51A (2012) 11066-11068

ISSN: 2229-712X

Fuzzy based data mining system for E-government

Anjum Mujawar[1] and Vijay Patil[2]

[1]Department of Electronics and Telecommunication Engineering, Vidyalankar Polytechnic, Wadala(E), Mumbai-37.

[2]Department of Computer Engineering, Vidyalankar Polytechnic, Wadala(E), Mumbai-37.

ARTICLE INFO

Article history:

Received: 3 January 2012;

Received in revised form:

13 October 2012;

Accepted: 27 October 2012;

Keywords

Data mining,

Fuzzy logic.

ABSTRACT

As the modernization of government is growing, E-Government is a grand new domain in recent years. More and more data is sent to databases system and large amounts of data are accumulated. Abundant knowledge exists in those Historical data. Meaning is to analyze those historical data and find useful knowledge and rules from the mass of data to provide better decision support and better adjustment guidance. The concept and steps of data mining is introduced in particular. When we use the E-Government system to process data, we need to choose what data is useful and what kind of new information we can get from the log file or from the database. Because of the special characters of knowledge, this paper presents an algorithm of the fuzzy data mining, and put great importance on the steps of the Fuzzy data mining.

Introduction

Data Mining sometimes referred to as knowledge discovery from databases, is a non-trivial process of Identifying implicit, valid, novel, potentially useful and ultimately understandable patterns in data. To improve the decision-making ability, find the abnormal pattern and predict the future trend base on the archived data. The electronic government affairs are the government apparatus applying modern information and the communication, they will manage and carries the service on the integration through the network technology, and they will realize the official organization Structure and the work flow optimized organization on Internet, between the surmounting time and spatial and department's separation limit, In other words, it is the use of IT and e-commerce to provide access to government information and delivery of public services to citizens and business partners.

The different stages required for Transformation Process:

Stage 1: Information publishing dissemination.

Stage 2: “Official” two-way transactions with one Department at a time.

Stage 3: Multipurpose portals.

Stage 4: Portal personalization.

Stage 5: Clustering of common services.

Stage 6: Full integration and enterprise transformation.

The E-government is a systems engineering and integrity System should include the following functions: (long-distance, distributing) information collection, information management (Electronic files, record management), information security; Electronic office; electronic post, electronic document and Decision system of the government and the report forms. Today, e-government is associated with the Internet. However, governments have been using other networks, especially internal ones, to improve government operations, the nation had all levels of government departments to establish his own website, the service content has been richer and richer, the function unceasingly strengthens, the interaction also obtains great enhancement. However facing such complex magnanimous data, we needs to use the new Technical method to come to these data to carry on the analysis, enables its "to recycle waste". The data mining technology can solve this problem.

Data mining and its approach:

Excavation of information from the large-scale database, which is formerly unknown, effective, useful information, and used this information to make decision or rich knowledge. This technology faces the application from the very beginning, now it is used widely in bank, telecommunication, transportation, retail sales (for example supermarket) and so on. In sense of cognition science, data mining mostly uses greatly induction to discover knowledge while uses deduction when appraising the discovered Knowledge, thus the algorithm of the data mining combines induction with deduction. Following parts usually composes the data mining:

(1) Controller-controlling other devices’ operation

(2) Database Interface-creating and processing Database inquiry

(3) Knowledge base-storing special information of Fields

(4) Focus-deciding assignment of analysis data

(5) Pattern extraction-choosing algorithms of pattern Extraction

(6)Evaluation-evaluating whether extracted pattern is Interesting and effective.

The main function of data mining includes automatic Forecasting the trend and behavior function, relationship Analysis functions and cluster analysis function, and this system uses the price automatic forecasting function. There are many methods to realize the automatic forecasting function in the data mining such as time serial, regression, decision-making trees, NN network, rough set arithmetic and the inherit arithmetic ,fuzzy set etc. To make decision or predict automatically traditional data mining method can’t fit for these data, so a new technology –fuzzy data mining [5] emerges. The principle of this theory is to use fuzzy set in data mining. The steps of data mining based on the fuzzy Theory is as the following [2]

Tele:

E-mail addresses: vijay.patil.me@gmail.com

To classify targets and collect the factor data

In all data recording of the data warehouse, at first we set up a categorized sample collection X(x1, x2,…, xn),collecting the numbers of sample to bring sample two-dimensional data lists that classify write down number. To every sample, there is a sample index. There is m a sample index, available m links vector and express samples i, xi=(xi1, xi2,…xin). Because the data that gather in reality are not here [0,1] in the number block, and do not accord with the demands for fuzzy set, so it's time to standardize it first, and then makes mach index of making samples concentrate each sample is here [0,1].

Set up fuzzy similar relation

Let *ij r* [0,1] score the relevant degree between *i x* and *j x* , and let *R*~=(*ij r*)n x n x n x n_ *ij r* = *ji r* , *ij r* =1(i, j=1,2…n). The key problem is how to set up the R ~. The commonly used methods are the accumulates method, coefficient correlation law method, greatest minimum law method, minimum law of arithmetic average method, the minimum law of geometric average method, index law of Absolute value method etc.

Here we introduce the accumulates method:

$$r_{ij} = \begin{cases} 1 & i=j \\ \frac{1}{M}\sum_{k=1}^{m} x_{ik} \cdot x_{ki} & i<>j \end{cases}$$

$$M>0 \quad M>= \max_{i \neq j}\left(\sum_{k=1}^{m} x_{ik} \cdot x_{jk}\right)$$

Cluster analysis:

The cluster analysis has three methods: equivalence close bag law, most great number method and weave network law, the most commonly used is the biggest tree. This method is utilized when n is very big. When the work load is presented under the state that the index multiple increases, it makes use of fuzzy matrix to carry on a kind of method of the cluster directly, and the concrete measure is the following:

(1) For the summit pinnacle classified, when *ij r* <>0,xi and xj can link a side.

(2) Let *ij r* permutation from small to large_a1>a2>…> al, and ak(k=1,2…l) is a certain *ij r*

(3) Link the objects which relational degree is a1_and Indicate a1 on the corresponding line segment, if the Loop appears while joining some two targets, this line will not be drawn.

(4) To a2…al in proper order, repeats the measure 3, until all targets feed through, and then we can get the greatest tree at this moment, but this biggest tree is not the only one.

(5) Let α [0,1] , cut the line which upper value is smaller than the line segment, and it have left what has been joined there is targets belong to everyone under level

Forecast

To every mode that is received during cluster analysis, try to achieve the average index of this mode according to the lower type:

Mod eij=_*u p ki* / _i=1,2…,s;j=1,2…m

S shows that all modes are counted, k shows that this mode has in data warehouses several records are put out, P show that introduces the total amount of records of this mode. The sample waiting to be predicted Y(y1, y2,…yn) is N a fuzzy sub collection in talking about land X of sample and compares with mode which data classify in the warehouse and ask and publish their pressing close to degree:

(X, Mod ei)=(1/2)[X•Mod ei+(1-X Mod ei)

The Applying Of The Fuzzy Data Mining On The E-Government

The content of the E-Government is broad and its applying technology is much more than others.

The electron-duty means that you can finish duty register, duty declare tax transfer and revenue query at home. This system can not only offer convenience to the enterprise but also decrease the expense of the government. However this sounds good in thought, in realities, there are so many data that you cannot distinguish what is useful to you what is no. How to distil useful information from the data house is urgent Problem that need to solve.

In a tax data house of Maharashtra tax bureau, there is a datasheet as table1.

The time granularity is divided into three layers: year, Season and month. The department granularity is divided into four layers: province, terra, country and town. The Economy granularity is divided into two layers: foreign capital and national capital. The vocation granularity is divided into two layers: Metal industry and metal Manufacturing industry. In practice applying, we often fall across such question: the real tax of a certain time, a certain department, a certain economy and a certain vocation belongs to what kinds of levels. A certain time, a certain department, a certain economy and a certain vocation are some known data if we have a datasheet. The real tax is a fuzzy variable and we should use the technology of fuzzy data mining to get the number of it. According to the former arithmetic we can get the *R*~:

$$\tilde{R} = \begin{bmatrix} 1 & 0.4 & 0.8 & 0.5 & 0.5 \\ 0.4 & 1 & 0.4 & 0.4 & 0.4 \\ 0.8 & 0.4. & 1 & 0.5 & 0.5 \\ 0.5 & 0.4 & 0.5 & 1 & 0.6 \\ 0.5 & 0.4 & 0.5 & 0.6 & 1 \end{bmatrix}$$

Adopt the max-tree and get different value , Assume different value of α . The process is as the fig. 2. According to the experiment result, when a[0.5,0.6] ,the five record in the datasheet can be divided into three glass. This classify is reasonable according with other Classification.

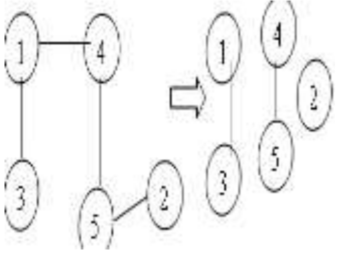

Fig.2 The process of the max-tree

From the fig.2(b), we know that 1 and 3 belong to the same level while 4 and 5 belong to the same level.2 belongs to the other level. The last step is to predict. According to the step 3.5 we get the inner product, the outer product and fuzzy near-tude of the A, B, C node of the data. We can know which tax money level this data does belong to by Looking, which kind of this empirical datum does draw close to, and most approaches.

Conclusion:

An emerging decision-making analysis method of data mining based on the data Warehouse. It can distill concealed, latent and unknown useful information or the pattern from the mass data in order to assist the policy-maker to carry on the decision-making. This article uses the fuzzy data mining to excavate the useful information in the electronic government affairs system, thus the help government policy maker makes the right decision-making.

References

[1] Zhang Junhua. Will the government 'serve the people? The development

[2] D. H. Hong. A note on correlation of interval-valued intuitionistic fuzzy of Chinese e-government. "New Media & Society 4.2 (June): 163-184,2002.

[3] Y.C.Hu, R.S.Chen and G.H Tzeng," Generating ;learning sequences for decision makers though data mining and competence set expansion:, IEEEE Transactions on Systems, Man and Cybernetics (2002)(to appear)

[4] K.Atanasov, More on intuitionistic fuzzy sets, fuzzy sets and systems 33(1989)37-46 sets, Fuzzy Sets and Systems 95 (1998)

Table-1 Datasheet of A Tax Data Warehouse

Id no.	Duration	Department	Capital	Profession
100	5	5	3	2
101	2	3	4	5
102	5	5	2	4
103	1	5	1	5
104	2	4	3	2