Applied Mechanics and Materials Vols 380-384 (2013) pp 1274-1277 Online: 2013-08-30
© (2013) Trans Tech Publications, Switzerland
doi:10.4028/www.scientific.net/AMM.380-384.1274

Analysis and Implementation of Multifactor System Estimation

Mao Tiezheng[a], Mao Aihua[b*]
School of Computer Science & Engineering, South China University of Technology, Guangzhou, 510001, China
[a]email:1150344295@qq.com, [b] email:ahmao@scut.edu.cn

Keywords：Multifactor; AGNES; Section polynomial fitting

Abstract: How to find most relative factor in multifactor system and how to find the function between relative factors and result are very important. In this paper, Bayes is used to find the most relative factors, then the training dataset is clustered by AGNES(Agglomerative Nesting) , then the section polynomial fitting is used to formulate multiple function for those most relative factors. Finally, this method is applied in the eye sight analysis system, and satisfying results can be obtained.

Introduction

Multifactor system is featured as there are many factors correlated with the system. How to find the most relative factors and how to find the relationship between these factors and results is very important for multifactor system.

Zhou used concepts, including information quantity, unites information quantity, conditional information quantity and mutual information quantity to describe information relation, and gave equations of these concepts [1]. In [2] decision information is intuitionistic fuzzy, and the relation based on a decision-making method was solved by intuitionistic fuzzy entropy and score function. XIE studies relation described by continuous attributes, and proposed a data discretization method based on statistical correlation coefficient [3]. However, there is no method thinking of seeking the help from artificial intelligence which can make the process more gathering and precise. In addition, it can be useful when identifying the odd data as well as dealing with raw data to improve accuracy in the next work.

In this paper, Bayes is used to find the most important factors, and then AGNES is used in the training set, then the sections polynomial fitting is used to find the function for those important factors. Finally, we use this method in eye sight analysis system, and obtained satisfying results.

Theoretical Methods

1. Find the Most Important Factors

We use data set to represent multifactor. The set is defined as followed:

Definition 1:The element of data set is （$x_1,x_2,\ldots x_i,\ldots x_n,y$）, in which x_i is i-th factor, y is result. The count of train set is s, elements in dataset are $\{\{x_1^{(1)},x_2^{(1)},\ldots x_i^{(1)},\ldots x_n^{(1)},y^{(1)}\}, \{x_1^{(2)},x_2^{(2)},\ldots x_i^{(2)},\ldots x_n^{(2)},y^{(2)}\}, \ldots,\{x_1^{(s)},x_2^{(s)},\ldots x_i^{(s)},\ldots x_n^{(s)},y^{(s)}\}\}$, where $x_i^{(j)}$ represent the value of factor x_i of the j-th sample.

Definition 2:

$$\vartriangle y^{(i)} = y^{(i)} - y^{(i-1)} \tag{1}$$

$$\forall j, \vartriangle x_j^{(i)} = x_j^{(i)} - x_j^{(i-1)} \tag{2}$$

After this , data set is changed to $\eta=\{\eta_1, \eta_2,\ldots \eta_s\}$, $\eta_1=\{\{\vartriangle x_1^{(1)},\vartriangle x_2^{(1)},\ldots \vartriangle x_i^{(1)},\ldots \vartriangle x_n^{(1)},\vartriangle y^{(1)}\}$, $\eta_2=\{\vartriangle x_1^{(2)},\vartriangle x_2^{(2)},\ldots \vartriangle x_i^{(2)},\ldots \vartriangle x_n^{(2)},\vartriangle y^{(2)}\},\ldots$, $\eta_s=\{\vartriangle x_1^{(s)},\vartriangle x_2^{(s)},\ldots \vartriangle x_i^{(s)},\ldots \vartriangle x_n^{(s)},\vartriangle y^{(s)}\}\}$

Definition 3: $\Re(x) = \{\eta_i \mid \eta_i \in \eta$ and $\eta_i(x)=1$, x is Boolean variable$\}$; $\Re() = \eta$.

Definition 4:

$$P(\triangle y_i) = |\Re(\triangle y^{(i)} > 0)| / |\Re()| \quad (3)$$

$$P(\triangle x_{ji}) = |\Re(\triangle x_j^{(i)} > 0)| / |\Re()| \quad (4)$$

$$P(\triangle y^{(i)}, \triangle x_{ji}) = \frac{|\Re(\triangle x_j^{(i)} > 0 \text{ and } \triangle y^{(i)} > 0)|}{|\Re()|} \quad (5)$$

Definition5: by Bayes, $P(B/A)=\frac{P(AB)}{P(A)}$, if P(B/A)>P(B), then we can get the conclusion that if A happens, B has bigger possibility to happens. So we use Eqs. (6) to calculate.

$$P(\triangle y^{(i)} / \triangle x_{ji}) = \frac{P(\triangle y^{(i)}, \triangle x_{ji})}{P(\triangle x_{ji})} \quad (6)$$

if $\frac{P(\triangle y^{(i)}, \triangle x_{ji})}{P(\triangle x_{ji})} > P(\triangle y_i)$, then x_j is relative to y.

Definition6: j=arg max $\{\frac{P(\triangle y^{(i)}, \triangle x_{ji})}{P(\triangle x_{ji})}\}$, so x_j will be the most relative factor.

2. Use AGNES to cluster dataset

Next, we must find the function for those relative factors. There are large sum of data in a dataset. Using polynomial fitting to get a polynomial usually can not fit the data exactly, but little num data can get better result. Here AGNES is use to cluster dataset into several sub dataset in which data is similar to each other.

AGNES (Agglomerative Nesting) algorithm firstly considers each element as one cluster, then those clusters are merged step by step. Similarity of the two clusters is defined by the distance between the nearest nodes in the clusters. Cluster step repeats until the cluster's count is suitable.

The AGNES algorithm step is listed as followed:

Input: dataset, stop condition num: ω

Output: clusters.

1) Set each element of dataset as a cluster
2) Calculate the nearest cluster
3) Merge the nearest cluster
4) Repeat 2)-4) Until cluster's num is ω

3. Sections Polynomial Fitting

After steps above, we find the most relative factors are $\{x_1,x_2,, x_k,y\}$, the data set is as followed $\{\{x_1^{(1)},x_2^{(1)},\ldots x_k^{(1)}, y^{(1)}\}, \{x_1^{(2)},x_2^{(2)},\ldots x_k^{(2)}, y^{(2)}\}, \{x_1^{(3)},x_2^{(3)},\ldots x_k^{(3)}, y^{(3)}\}, \{x_1^{(\upsilon)},x_2^{(\upsilon)},\ldots x_k^{(\upsilon)}, y^{(\upsilon)}\}\}$,which we called it W',($\upsilon$ is the num of one cluster, $\sum_{i=1}^{w} \upsilon = s$).. Then , n times Newton interpolation will be used to find the function describing relationship between the result y and parameter x_i, which as $N_n(x_i) = y$. $N_n(x_i)$ is as follows:

$$\begin{aligned} N_n(x_i) &= f(x_i^{(1)}) + f(x_i^{(1)}, x_i^{(2)})(x - x_i^{(1)}) + \\ & f(x_i^{(1)}, x_i^{(2)}, x_i^{(3)})(x - x_i^{(1)})(x - x_i^{(2)}) + \\ & f(x_i^{(1)}, x_i^{(2)}, x_i^{(3)}, x_i^{(4)})(x - x_i^{(1)})(x - x_i^{(2)})(x - x_i^{(3)}) + \\ & f(x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(n)})(x - x_i^{(1)})(x - x_i^{(2)}) \ldots (x - x_i^{(n-1)}) \end{aligned} \quad (7)$$

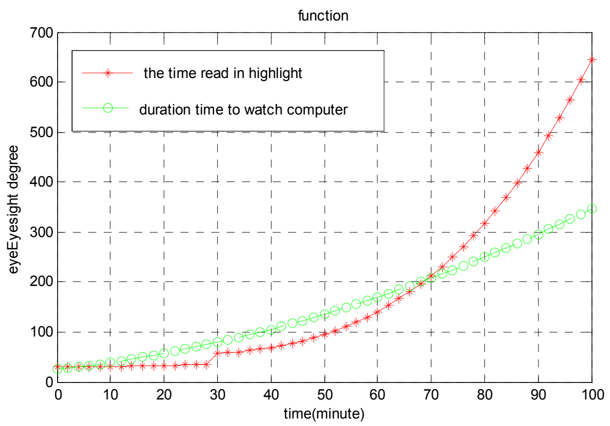
Besides , $f(x_i^{(1)},x_i^{(2)},...,x_i^{(j)})$ can be calculated as follows:

$$f(x_i^{(1)},x_i^{(2)},...,x_i^{(j)})=\frac{f(x_i^{(1)},x_i^{(2)},...,x_i^{(j-1)})-f(x_i^{(2)},x_i^{(3)},...,x_i^{(j)})}{x_i^{(1)}-x_i^{(j)}}$$

Thus, a section polynomial for x_i in a cluster can be got and repeating in this way all sections polynomial of x_i can be obtained, then, the relation between the multifactor and consequence can be described from the function numerically.

Test results

We use this method to find which factor is most relative to the eye sight analysis system. The data set contains the following factors: duration of watching TV, duration of watching computer, duration of sport , heredity, eye-exercises times, duration of e reading in highlight, height for reading, weight, vitamin eaten every day, protein eat everyday etc. After analysis, the most relative factor is ‘duration of watching computer’ and ‘duration of reading in highlight’ , the two factors has less influence are duration of watching TV and duration of reading while moving (like walking or on the bus). Fig.1 shows the function we have learned from the proposed method above. The duration of watching computer has one section polynomial, and the duration of reading in highlight has two sections polynomial. Duration of watching TV has two sections polynomial too.



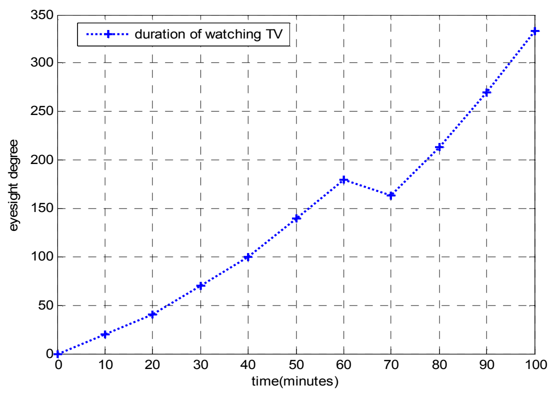


Fig.1. The relationship between factors and eyesight degree

Conclusion

This paper discusses how to find most important factors in the multifactor system and how to create mathematic model for those important factor. Firstly ,the problem is defined in dataset, then Bayes is used to estimate the value of importance. Then, the dataset is clustered to reduce the problem scale. Finally, sections polynomial fitting is used to get function model. Experiments in the eye sight analysis system show that the system can obtain satisfying results.

Acknowledgement

This paper is financially supported by the MOE(Ministry of Education in China)Project of Humanities and Social Sciences(No.13YJC890027), the National Natural Science Foundations of China (No.61003173),the Fundamental Research Funds for the Central Universities (No.2012ZZ0063)and the Science and Technology Project of Guangzhou City (No.2012J4100002).

References

[1]. Zhou Tong,Zhang Jia-lu. Relativity of attributes in information systems and its application[J].Computer Engineering and Design, 2012(33):1192-1196

[2]. WANG Cui-cui,YAO Deng-bao,MAO Jun-jun,SUN Li. Intuitionistic fuzzy multiple attributes decision making method based on entropy and correlation coefficient[J]. Journal of Computer Applications,2012(32):3002-3004

[3]. XIE Ya-ping. Data discretization method based on statistical correlation coefficient[J]. Journal of Computer Applications ,2011(31):1409-1412

[4]. ZHENG Wei, HSU Hou-Tse,ZHONG Min,YUN Mei-Juan. Impacts of interpolation formula,correlation coefficient and sampling interval on the accuracy of GRACE Follow-On intersatellite range-acceleration[J]. Chinese Journal of Geophysics ,2012(55):822-832

[5]. Hao Tao .Some Properties of g-Covariance and g-Correlation Coefficient[J]. Chinese Journal of Applied Probability and Statisties,2012(28):637-646

Vehicle, Mechatronics and Information Technologies
10.4028/www.scientific.net/AMM.380-384

Analysis and Implementation of Multifactor System Estimation
10.4028/www.scientific.net/AMM.380-384.1274