Data processing of lie detector based on fuzzy clustering analysis

Zilong Chen[1, a], Junru Wang[2, b], Zhiwei Gong[2, c], Quanbo Liu[3, d]

[1]School of Electrical Engineering, Chongqing University, Chongqing 400044, China

[2]School of Mechanical Engineering, Chongqing University, Chongqing 400044, China

[3]School of Hong Shen, Chongqing University, Chongqing 400044, China

[a]1764220897@qq.com, [b]392910349@qq.com, [c]312268893@qq.com, [d]905027574@qq.com

Keywords: Data Processing, Fuzzy Clustering, Psychological lie-detection.

Abstract. In this paper, we analyze the problem of many data processing in psychological-detection, then, we deal with the problem based on Fuzzy Clustering, Fuzzy Clustering can use less variable to replace previous variable while the quantity of information is invariant, so, we can eliminate redundancy of information. Last, we did an experiment by our homemade lie detector, then we analyze the data based on Fuzzy Clustering, we found that the previous 6 parameters can be replaced by 4 new parameters, and the information almost didn't lose, so Fuzzy Clustering can be a good choice for data processing.

1. Background of the problem

The traditional lie detector can measure less parameters and accuracy is low, but the data is relatively easy to process. However, with the development of sensor and micro-processing technology, modern lie detector can measure more and more parameters, the complexity of it become more and more high, and some lie detectors can even measure a dozen parameters simultaneously. Although the parameters increase will help improve the accuracy of the polygraph, more parameters will greatly increase the amount of experimental data, it's not conductive to rapidly process and analysis data, in courts and other some occations where polygraph judgements should be made quickly, the massive data is not conducticve to effective and rapid polygraph judgements, thus affecting the progress of the trial.

In order to solve the complicated problems which lie in the modern lie detector's data processing, this paper presnts a new method in which we ruduce the number of parameters by fuzzy clustering. In this method, we replace the strong correlation parameters with a new variable to eliminate relevance and redundancy among the variables, to make maximum use of parameters' information, to prevent the problems of massive parameters caused by repeated information, so we can minimize the within-class similarity and maximize the between-class similarity.

2. Theory of lie directing based on fuzzy clustering

Fuzzy clustering analysis is based on the properties of subjects researched, it classify the subjects researched based on the fuzzy clustering matrix confirmed by membership grade, and according to the different subjects, the classifications can be divided into Q-type and R-type. Q-type classification is to classify the sample, and R-type classification is to classify indicators. Since the paper discusses the correlation among the parameters of polygraph, so we choose R-type classification.

2.1 Determining the measurements of similarity

Assuming the polygraph variable $\mathbf{x}_j$ are $(x_{1j}, x_{2j}, \cdots, x_{nj})^T \in R^n (j = 1,2,\cdots,m)$, then we can regard the sample correlation coefficient between variables $\mathbf{x}_j$ and $\mathbf{x}_k$ as the measurements of similarity, and written as:

$$r_{jk}=\frac{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)(x_{ik}-\bar{x}_k)}{[\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2]^{\frac{1}{2}}} \quad (1)$$

The formula above is called the correlation coefficient matrix between variables. We can clearly get the degree of similarity between the different variables, so as to provide the following variable cluster analysis with bases. Moreover, compared to other method of measuring similarity, such as included angle cosine method, it's effective and simple to use correlation coefficient matrix method, and this method is most commonly used variable similarity analysis method.

It should be noted that the similarity measurement in above formula should be comply with the following properties:

(1) $|r_{jk}|\le 1, \quad j,k\in R$

(2) $r_{jk}=r_{kj}, \quad j,k\in R$

Wherein, $|r_{jk}|$ is closer to one, the more relevant or similar $\mathbf{x}_j$ and $\mathbf{x}_k$ are. On the contrary, $|r_{jk}|$ is closer to zero, the less similar $\mathbf{x}_j$ and $\mathbf{x}_k$ are.

2.2 Determining the method of clustering variables

Variable clustering methods contain the longest distance method, the shortest distance method and other commonly used methods, from the perspective of the complexity of the problem, this paper choose the shortest distance method to carry out variable clustering. In the shortest distance method, we should firstly find the smallest gap parameters in a correlation coefficient matrix of n order, classify these parameters as a class, then calculate distance between the new and old classes on the basis of original method of calculating the distance among parameters, so we can create a new correlation coefficient matrix of m-1order; Then we select new class from the new correlation coefficient matrix, calculate the distance between the new class and the original class, this action goes on until the objects are classified into one kind. Wherein, the formula of the distance between the two types of variables is as follows:

$$R(G_1,G_2)=\max_{\substack{x_j\in G_1\\ x_k\in G_2}}\{d_{jk}\} \quad (2)$$

Wherein, $d_{jk}=1-|r_{jk}|$ or $d_{jk}^2=1-r_{jk}^2$, it can be seen that $R(G_1,G_2)$ is related to the value of similarity measurement between the two variables which have the smallest similarity among the two types.

3. Polygraph test analysis

Use self-made multi-parameter physiological and psychological polygraph measure the heart rate, breathing, skin resistance, blood pressure, body temperature, pulse of the people tested simultaneously, select 140 volunteers to have standard polygraph test in control question test method, then analyze the experimental data by matlab programming, and we can get correlation coefficient matrix of six indicators as follows:

As can be seen from the table above, there is a strong correlation between some parameters, and the correlation coefficients among them are big, the redundant information among them are large, so we can consider to select a representative new parameters to have clustering analysis from the these highly correlated parameters.

Firstly, we can standardize every polygraph parameter, eliminate the influence of dimension, use the correlation coefficient method to determine the similarity measurement between the variables,

then use the average linkage method to determine the similarity measurement between the classes. Finally, we can obtain a following figure of index clustering:

Table 1 Correlation coefficient matrix of polygraph parameter

polygraph parameter	heartbeat	breathing	skin resistance	blood pressure	temperature	pulse
heartbeat	1.0000	0.9334	0.9628	0.9600	0.9743	0.9699
breathing	0.9334	1.0000	0.9835	0.9957	0.9643	0.9602
skin resistance	0.9628	0.9835	1.00000	0.9969	0.9831	0.9807
blood pressure	0.9600	0.9957	0.9969	1.0000	0.9873	0.9856
temperature	0.9743	0.9643	0.9831	0.9873	1.0000	0.9934
pulse	0.9699	0.9602	0.9807	0.9856	0.9934	1.0000

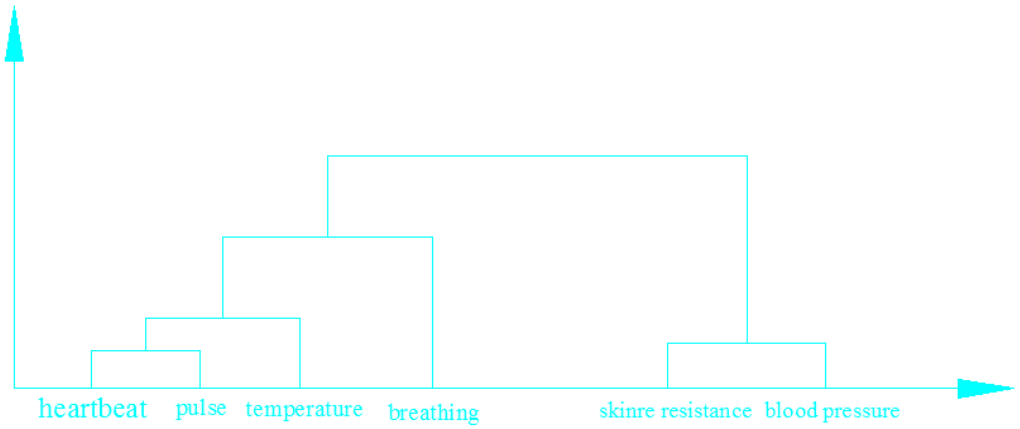

Fig.1 index clustering

From the above figure we can clearly see that heartbeat and breathing can be seen as one kind variable which has a strong correlation, this is consistent with the actual situation, usually human's heartbeat is directly proportional to their respiratory intensity; Skin resistance and blood pressure can also be regarded as one kind variable which has a strong correlation, when people's blood pressure changes, the blood flow rate in vessels changes simultaneously, and skin resistance and velocity of blood flow have a certain relationship, so the two variables are of great correlation, and we can see them as one kind variable. However, the correlation between body temperature and respiratory is not very strong, so they can be considered as independent variables.

In the above analysis, under the premise of ensuring that the information quantity is approximately constant, the original six parameters can be simplified to four parameters. Although the number of parameters reduces, the total amount of information did not reduce too much, only the correlation between different variables reduces, that's advantageous to data processing. In the experiment involved with a dozen parameters, the method of simplifying parameters will greatly reduce the difficulty and complexity of data processing in later period, so it can facilitate processing data quickly and effectively.

Reference

[1] Si shoukui,Sun Xiqing, Application of mathematical model, National Defence industry press, Beijing, 2001.8.

[2] Liang Xiaoru,Yin Chunwu,"An Method to determine the index of polygraph testspolygraph tests," *Journal of Xi an University of Arts & Science* , vol. 13,No.1,2010 .

[3] Zhang Jijun,"Fuzzy Analytical Hierarchy Process," *Fuzzy System and Mathematics*, vol. 14,No.2,2000 .

[4] Lin Wenjuan,"The Overview of advanced psychology," *Advances in Psychological Science*, vol. 16(3),2008 .

[5] You Jian,“An Analysis on Practical Application of Interrogation Applied to Investigation and Trial after Psychology Test,” *Journal of Fujian Public Security College*, vol. 4,2008 .