Study of Hybrid Strategy for Ambulatory ECG Waveform Clustering

Gang Zheng*
Laboratory of Biological signal and Intelligent Processing, Tianjin University of Technology, Tianjin, China
Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin, China
School of Computer and Communication Engineering, Tianjin University of Technology, Tianjin, China
Email: kenneth_zheng@vip.163.com

Tian Yu
Laboratory of Biological signal and Intelligent Processing, Tianjin University of Technology, Tianjin, China
Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin, China
Email: qq190254094@hotmail.com

***Abstract*—A hybrid strategy has been proposed to reduce the wrong clustering on Ambulatory ECG (electrocardiogram). Since Ambulatory ECG is usually composed by 24 hours data, the number of individual ECG waveform can reach to 100,000, the request for accurate clustering result is highly required. The proposed strategy adopted some intelligent algorithms to solve the above problem. It clusters ECG waveform sample (selected from Ambulatory ECG) synchronously by Max-Min distance clustering algorithm, *K*-means algorithm and Simulated annealing algorithm first. And then, it adopted all three outputs from the above three algorithms as input on Back Propagation Artificial Neural Network (BP ANN). In the end, we got more accurate clustering result from the output of ANN. For testing the results, data of MIT/BIH arrhythmia database were used for experiments. After the controlled trial on MIT/BIH data, it can be safely concluded that the clustering result achieved by improved strategy can got more accurate than that by the traditional clustering algorithm. An average accuracy ratio is about 94.6%, 1.6% higher than *k*-means algorithm averagely and 1.3% higher than Simulated Annealing algorithm averagely.**

***Index Terms*—ambulatory electrocardiogram, *k*-means, artificial neural networks, clustering algorithm, Simulated Annealing algorithm**

I. Introduction

Ambulatory ECG has been used widely in early detecting and preventing some kinds of heart disease since it was invented. Recently, lots of portable ECG recording system are produced in relative small size and equipped with powerful analysis function. Therefore, a number of contribution studies emerge, which leads to propose some analysis structure and algorithms in long time ECG waveform recording system[1]. Considering the individual characteristics of the sample, clustering algorithm has been used. This paper analyses traditional clustering algorithms and also points out the disadvantages in those clustering algorithms. Then another improved strategy based on Max-min Distance clustering algorithm, *k*-means algorithm and Simulated Annealing algorithm is proposed. It is proved that the proposed strategy is more suitable to process electrocardiogram data.

In the paper, Section One is the introduction of this paper; Section Two analyzes strengths and weaknesses of traditional clustering algorithms; Section Three introduces the improved strategy and analyses the significant parameters used in proposed strategy; Section Four analyzes the results of the controlled trial; Section Five gives the conclusions and future directions of the research.

There are two main reasons for analyzing ambulatory electrocardiogram waveforms by clustering algorithm.

First of all, the intention of using the clustering algorithm is to obtain the individual congenital waveforms. In the practical application, the most significant waveforms are the abnormal waveform rather than congenital waveforms. If these congenital waveforms can be achieved, the remaining meaningless waveforms, about 90% in whole data, can be reduced. The abnormal waveforms can be discriminated to identify the abnormity of heart. Through clustering the original data and reducing the meaningless waveform, the workload of analysis system can be reduced at the same time.

Secondly, the electrocardiogram analysis systems used nowadays mostly base on some template-matching algorithms, such as the algorithms are proposed in [2] [3]. Unfortunately, constructing an ideal set of all-purpose templates is a hardship because of the difference of

*Gang Zheng is with the School of Computer and Communication Engineering, Tianjin University of Technology
Tian Yu is with the same university as Gang Zheng
The paper is supported by Tianjin Natural Science Foundation contracts 10JCYBJC00700 and Science and Development Funds of Tianjin Higher Education, contracts SB20080044

doi:10.4304/jsw.6.7.1257-1264

individual characteristics of test samples. This difference has twofold meanings. One is the difference between the different states of one test sample. The waveforms of one sample present different shapes when the sample is involved in different activity states, such as walking, running and sleeping. The other difference occurs between the different test samples. Because of the differences in their body, such as heart volume and coronary artery, every test sample has the individual characteristics. Weighing and describing anybody's waveforms by one or several uniform criterion(s) is apparently impossible.

Considering all above, the clustering algorithm is an available solution which can attain the individual's congenital waveforms and several types of abnormal waveforms. These representative waveforms are significant for further examination.

II. Related Works

A. Traditional clustering algorithm

Since the target of cluster and classification are all intend to category the data into several partitions, the number of class need to be known in classification analysis. But, in clustering, supervised and unsupervised methods are exist concurrently. If the number of class is finite, data will be categorized by finite partitions. If the number is uncertain, it can be decided by other parameters in clustering procedure [4][7][10]. Through the survey on clustering algorithm, since the end of 1990s [19][20][25][34][35][41], we can found that unsupervised learning methods have been broadly developed, which have become significant methods and have been used in more and more modern application domains, such as image processing [17]and signal processing [6] [32]. The clustering algorithms cluster a collection of data objects into several groups based on the similarity of all objects in the space. Except this, density based cluster [39], cluster validity methods are also widely studied [40] on clustering study. Sometimes, it combined with other method [21].

Optimization problem were also introduced into cluster study. In the paper, simulated annealing algorithm was adopted. Since simulated annealing algorithm was broadly applied on optimization algorithm [3][5][9][27], it has also been used to solve the problem in cluster analysis. Some contributions support this [11][24][26].

In this paper, two typical clustering algorithms have been adopted, one is Max-min Distance clustering algorithm, the other is *k*-means algorithm.

K-means algorithm is a prototype-based, clustering technique. This traditional clustering algorithm attempts to find *k* clusters and the *k* is assigned by the expert. A number of modifies have been studied on it to improve the clustering quality of *k*-means. Chen [8] proposed a combined *k*-means and hierarchical clustering method, which can improve the efficiency of microarray clustering to research's demands. It was also used on nonlinear component analysis by kernel function method [16]. Unsupervised *k*-means do not need pre-assigning the exact number of cluster. Krista [38] presented a method based on minimizing a suggested cost-function. The cost-function extends the mean-square-error cost-function of k-means. Another method proposed by Bafirov [43] dynamically add a new clustering center on large data set. He computes the starting point for the k^{th} cluster center by minimizing the auxiliary cluster function. In Tsai's [42] contribution, he presented a feature weights self adjustment mechanism for *k*-means. The problem of initiated center was also studied [37]. *K*-means was also used on medical data [14]. Through adjusting the margin of feature weights, a good clustering quality was gotten. Being a traditional cluster algorithm, *k*-means has its advantage on relative low time-space complexity, and modifiable.

The flow chart of *k*-means algorithm is illustrated in Figure1.

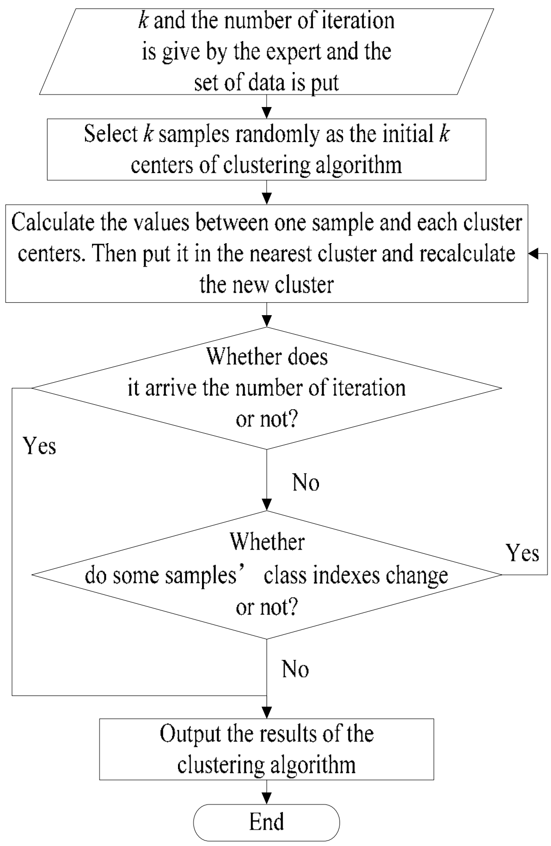

Figure 1. The flow chart of *k*-means algorithm

Max-min Distance clustering algorithm is a probing-based clustering algorithm. This clustering algorithm attempts to find the furthest object to be one of the centers, because it wants to avoid that several initial centers are too close to lead the poor performance in the clustering [5]. The flow chart of Max-min Distance clustering algorithm is illustrated in Figure 2.

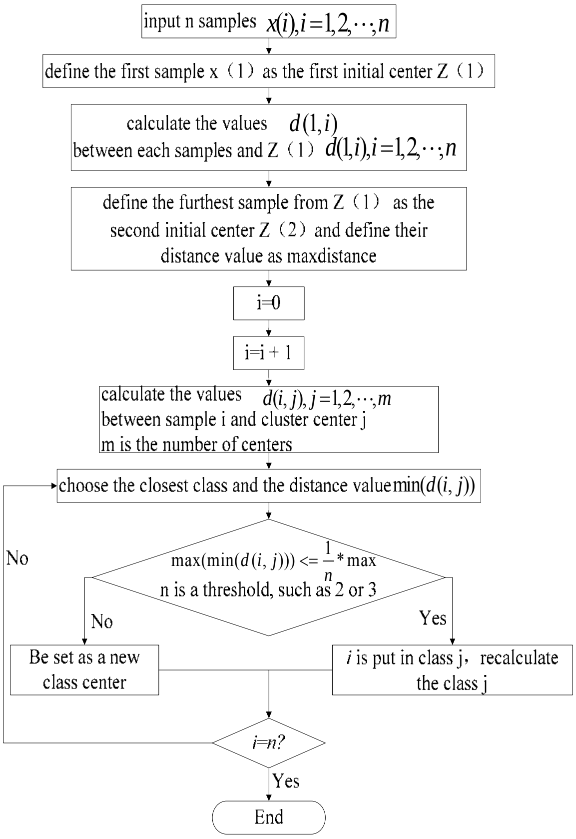

Figure 2. The flow chart of Max-min Distance clustering algorithm

B. Study on ECG waveform

A number of intelligent algorithms are used to analyze ECG waveform. As one of the bio signal, ECG waveform has been broadly studied in waveform detection, classification, and clustering In the detection, linear prediction [29], syntactic recognition [30], Wavelet transform [44][45][46], Morphology[12], Hidden Markov Model [36]etc. were applied. In ECG classification, the correction rate of Principal Component Analysis reached 98.37%[13], 98%[15], and 98.5%[18] ; By Super Vector Machine, the correction rate reached 99.5%[31]; By Artificial Neural Network, it reached from 84.9% to 92.95%, and from 95.53% to 97.82%[32] on different dataset; By Wavelet transform, it reached 90% [33].

C. The problem in traditional clustering algorithm

In the study of Ambulatory electrocardiogram, Max-min Distance clustering algorithm and *k*-means algorithm has been introduced in [5]. But both algorithms have their drawbacks.

As well known, Max-min Distance clustering algorithm and *k*-means algorithm both have two strict prerequisites in practice: one is that the intra-group of the data is similar and the inter-group is dissimilar. In view of the electrocardiogram data, the boundaries of different classes are dim and many waveforms distribute loosely in the whole high-dimension space.

These inappropriate applications of algorithms lead to several drawbacks.

To Max-min Distance clustering algorithm, the distribution of electrocardiogram data is so unsatisfactory that many objects belonged to several different clusters might be merged into one cluster.

To *k*-means algorithm, there are two poor drawbacks. The first one is that a good result depends seriously on the selection of the initial clustering centers. This drawback leads to the improper initial waveform clustering centers. And the second one is the influence of “noise point” and “isolated point” in the class boundary. This drawback leads to the wrong clustering results of these points. Consequently, the abnormal waveforms can be considered as congenital waveforms.

Considering these drawbacks of the traditional algorithms, an improved strategy based on Max-min Distance clustering algorithm, *k*-means algorithm and Simulated Annealing algorithm [9] is introduced in this paper. The aim is to eliminate the adverse effects brought by the traditional clustering algorithms.

III. PROPOSED CLUSTERING STRATEGY

A. Description the strategy

The flow chart of the strategy proposed in this paper is shown as Figure 3.

In the step ①, Max-min Distance clustering algorithm is used to cluster the waveforms. The clustering result of Max-min Distance clustering algorithm is regarded as one of the input layer nodes of ANN in the step ④.

In the step ②, k-means algorithm is used to cluster the waveform. The clustering result of *k*-means algorithm is not only regarded as the initial result of Simulated Annealing algorithm in the step ③, but also regarded as one of the input layer nodes of ANN in the step ④.

In the step ③, Simulated Annealing algorithm is used to get more accurate categorization. The clustering result of Simulated Annealing algorithm is regarded as another of the input layer nodes of ANN in the step ④.

In the step ④, ANN is used to get more proper categorization. The clustering results by Max-min Distance clustering algorithm, by *k*-means algorithm and by Simulated Annealing algorithm are regarded as the input layer nodes of ANN in the step ④.

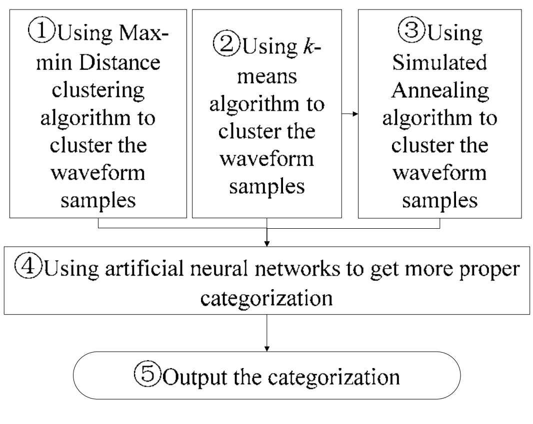

Figure 3. The flow chart of the strategy proposed in this paper

In this paper, BP neural networks is chosen in the step ④. And it is trained as supervised learning.

In the step ⑤, the category result achieved by ANN is output.

B. Description of Simulated Annealing

Simulated Annealing algorithm is a heuristic and stochastic searching algorithm. In theory, it has been proved that it can converge at the optimal result by Kirkpatrick, Gelatt and Vecchi [3]. In Figure 4, the flow chart Simulated Annealing algorithm is illustrated.

The clustering result by *k*-means algorithm is regarded as the initial result of Simulated Annealing algorithm. The initial temperature "T_0" can be numerated by the objective function expression (1) and the initial result. The Simulated Annealing algorithm carries an iterative procedure which is explained in part c of Section 3.

Through the iteration, the value of T is decreased gradually. And when the procession is over, the solution is the optimum theoretically. Simulated Annealing reaches the optimal result just through perturbing continuously some current results at random.

Benefiting from this continuous perturbation at random, Simulated Annealing algorithm has chances to change the improper categories of "noise points" and "isolated points" into proper categories. Moreover, these proper clustering results of "noise points" and "isolated points" can further adjust the clustering centers to reach the optimal result.

C. The parameters in Simulated Annealing algorithm

Firstly, the objective function has been defined. Choose the between-class scatter of the current clustering partition as the objective function, which is defined by:

$$J_\omega = \sum_{i=1}^{M} \sum_{X \in \omega_i} d(X, \overline{X^{(\omega_i)}}) \quad (1)$$

with X one electrocardiogram waveform vector which covers corresponding R-R period; ω clustering partition; $\overline{X^{(\omega_i)}}$ the class central vector of the X sample; $d(X, \overline{X^{(\omega_i)}})$ the distance between the sample and the corresponding class center; M the number of clustering. Objective function J_ω, which is also called cost function, is the summation of the distances between the samples and their corresponding centers.

Secondly, this paper chooses the clustering result achieved by *k*-means algorithm as the initial result, which the initial temperature is $T_0 = J_\omega$.

Thirdly, in Simulated Annealing algorithm, a proper stochastic perturbation measure must be applied. The new result used for every iterative procedure in Simulated Annealing algorithm is achieved by perturbing the current result randomly. Through altering the category of one sample randomly each time, a new partition can be produced. That measure is the most important step to make sure that the result can escape the wrong clustering result.

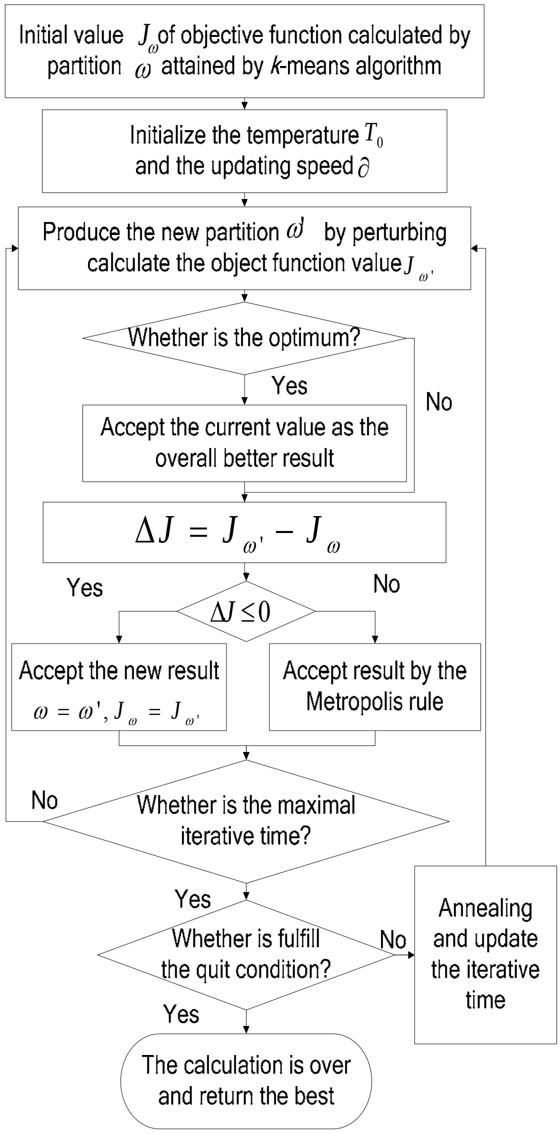

Figure 4. The flow chart Simulated Annealing algorithm

Fourthly, it has been proved that Simulated Annealing algorithm can converge at the optimal result by Kirkpatrick and Gelatt. However, considering the time-consuming in practice, infinitely stochastic perturbation is unavailable. If not enough points are perturbed, it cannot be ensured that the result of Simulated Annealing algorithm is the optimal result. The time of perturbation is chosen as 1000 in the proposed algorithm.

Finally, the annealing mode is significant in Simulated Annealing algorithm. To avoid too long Markov chain made by the algorithm, the control parameter should be decreased slowly. An expression which used for updating the parameters is that:

$$T_{k+1} = \alpha \bullet T_k \quad (2)$$

The equation (2) is updating equation and it was firstly proposed by Kirkpatrick, α is the annealing speed parameter which controls the falling speed of the temperature. In experiment, α is chosen as 0.99.

D. Implementation of Simulated Annealing algorithm

1) Cluster the samples by *k*-means algorithm and the Simulated Annealing algorithm's initial result ω has been achieved. Objective function value J_ω is calculated by equation (1).
2) Initial temperature $T_0 = J_\omega$. Initialize the annealing speed α and the maximum annealing time.
3) For the input temperature T, iterative procedure was processed as step 4） - step 7) till maximum annealing time is reached and then jumps into step 8）.
4) Stochastic perturbation produces new partition ω' by changing some sample's category each time and calculating the new objective function value $J_{\omega'}$.
5) Judge whether the new objective function value $J_{\omega'}$ is a smaller value. If it is, the clustering partition ω' is preserved as the better partition and $J_{\omega'}$ is preserved as the better objective function; else go to the next step.
6) Calculate the difference ΔJ between the new objective function value and current objective function value.
7) Judge whether ΔJ is less than 0. If it is, the new result is accepted as current result; else it is accepted as the probability $p(p = e^{\Delta J / KT})$ according to the Metropolis rule. K is a constant and T is the current temperature.
8) Judge whether the maximal annealing time has reached. If it has, go to the next step; else anneal the temperature according to the annealing expression (1) and then return the step 3) to continue the iterative procedure.
9) Judge the quit condition is same as judging whether the clustering result is not changed. If it is not, output the better partition as the best partition; else anneal according to expression (2), update the iterative time and return the step 3).

E. Description of ANN

To get the more accurate result, In this paper, BP neural networks is chosen in the step ④. And it is trained as supervised learning by experts.

Sigmoid function is chosen as the activation function of the intermediate layer. And the learning rate is set as 0.9. The learning samples are 500 waveforms from MIT/BIH Arrhythmia database.

IV. Experiment and Results

The MIT/BIH Arrhythmia database provided by Massachusetts Institute of Technology and Boston's Beth Israel Hospital is used for evaluation of the proposed strategy. The software used in the experiment is VC++6.0.

A complete data recording is composed by three files which are signals file, annotations file and header file. The recordings from MIT-BIH Arrhythmia Database were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. And these digitized signals are conserved in the file which filename extension is ".dat". Two or more cardiologists independently annotated each record; disagreements were resolved to obtain the computer-readable reference annotations for each beat (approximately 110,000 annotations in all) included with the database. And these annotations are conserved in the file which filename extension is ".atr". The information about the tested people and the format of the signals file are conserved in the file which filename extension is ".hea".

The samples collected from II lead and V lead are used in the experiment. The electrocardiogram data has been preprocessed in order to get a vector group. Every vector has been chosen according to the R-wave positions which are recognized by experts in annotations file. According to the sampling rate of the data and the time interval which is significant to analyze the waveforms, 145 eigenvalues are chosen to constitute an eigenvector. These eigenvectors are used in the experiment.

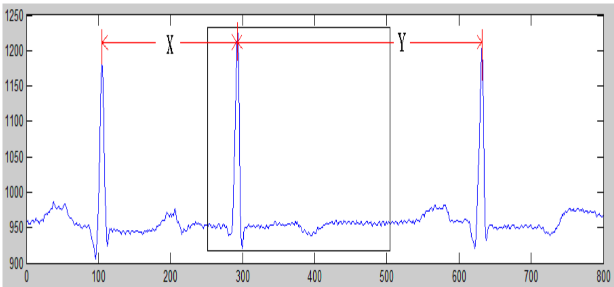

Figure 5. The examples of APC (Atrial Premature Contraction) from signals file 100

Considering the particular attributes of the electrocardiogram waveforms, the waveforms in some certain categories cannot be recognized only by their shapes, because these waveforms are same as the congenital waveforms. As a case in point, in Figure 5, APC (Atrial Premature Contraction) is illustrated with black frame and the other two waveforms are normal heart rate. However, the shape of APC waveform has little difference with normal waveforms, so this type of arrhythmias can not be identified by its shape. Analyzing the time interval X and Y in the Figure 3 is a more proper method to identify this type of arrhythmias. X is the time interval between current R-wave and the R-wave before and Y is the time interval between current R-wave and the R-wave after. So those waveforms unrecognized in shape are not chosen in the experiment.

TABLE II.
THE DATA USED IN THE EXPERIMENT

Sample No.	The category annotations in the original file	The number of waveforms in the original file	The categories which is used in the experiment	The number of waveforms used in the experiment
100_1	NORMAL, APC, PVC	2273	NORMAL, PVC	2240
102_1	NORMAL, PVC, PFUS, PACE	2187	NORMAL, PVC, PFUS, PACE	2187
104_1	NORMAL, PVC, PACE, PFUS, UNKNOWN	2229	NORMAL, PVC, PACE, PFUS	2211
118_1	RBBB, APC, PVC ,NAPC	2278	RBBB, PVC	2182
119_1	NORMAL, PVC	1987	NORMAL, PVC	1987
121_1	NORMAL, APC,PVC	1863	NORMAL, PVC	1862

TABLE I.
THE ACCURACY NUMBERS OF THREE ALGORITHMS IN THE EXPERIMENT

Sample No.	No.1 algorithm	No.2 algorithm	No.3 algorithm
100_1	2239	2238	2237
102_1	2028	2036	2050
104_1	1915	1934	1960
118_1	2084	2082	2166
119_1	1656	1666	1714
121_1	1861	1859	1861

In the Table Ⅰ, the recognizable categories and corresponding number of waveforms used in the experiment are listed. And the number of all waveforms and all categories appeared in the original file are also listed.

In the controlled trial on MIT/BIH data, the results of *k*-means algorithm (which is called No.1 algorithm for short), the results of Simulated Annealing algorithm (which is called No.2 algorithm for short) and the results of proposed algorithm in the paper (which is called No.3 algorithm for short) are listed in Table 2 and Table 3.

From Table Ⅱ and Table Ⅲ, 100_1 sample's results by Simulated Annealing algorithm and by proposed strategy are not better than that by *k*-means algorithm. That is caused by the inappropriate stochastic perturbation. The perturbation leads to movements of some class centers, and these movements cause some improper classification in the next iteration. But through contrasting and analyzing other samples' results and corresponding annotations files, it can be concluded that if the intra-group similarity and inter-group dissimilarity of the data are large enough in high dimensional space, the stochastic perturbation may lead improper results. By contrast, if the intra-group similarity and inter-group dissimilarity of the data are not large enough, the negative influence of "noise point" and "isolated point" in the class boundary can be eliminated orderly by the stochastic perturbation.

TABLE III.
THE ACCURACY PERCENTAGE TWO ALGORITHMS IN THE EXPERIMENT

Sample No.	No.1 algorithm	No.2 algorithm	No.3 algorithm	The contrast between No.2 algorithm and No.3 algorithm	The contrast between No.1 algorithm and No.3 algorithm
100_1	99.9554 %	99.9107 %	99.8661 %	0.0446 %	0.0893 %
102_1	92.7298 %	93.0956 %	93.7357 %	0.6401 %	1.0059 %
104_1	86.6124 %	87.4717 %	88.6477 %	1.1760 %	2.0353 %
118_1	95.50871 %	95.4170 %	99.2667 %	3.8497 %	3.7580 %
119_1	83.3417 %	83.8450 %	86.2607 %	2.4157 %	2.9190 %
121_1	99.8926 %	99.8389 %	99.8926 %	0.0537 %	0%

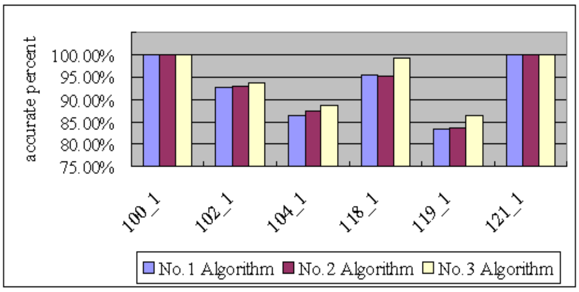

Figure 6. Column diagram of analysis accurate rate

By contrasting the accurate numbers in 118_1 and 121_1, it can be easily found that the results of Simulated Annealing algorithm are little worse than the results of k-means algorithm, but after adjusting by the ANN of proposed algorithm in this paper the accurate rate has been improved. From the results of 102_1, 104_1 and 119_1, it can be learned that in these samples the performance of Simulated Annealing algorithm is better than that of k-means algorithm. And the performance of proposed algorithm is better than two others. After different clustering results obtained by different clustering algorithms have been input as the initial values of BP neutral net, the neutral net weights these results repeatedly and finally obtains more accurate samples' categories.

Through analyzing all results in the experiment, it can be safely concluded that the results attained by proposed strategy are more accurate than the clustering attained by two other algorithms and the accuracy ratio has been increase by about 1.6% averagely than *k*-means algorithm and about 1.3% averagely than Simulated Annealing algorithm.

V. CONCLUSION

Through the controlled trial in the Section 4, it spontaneously comes to the conclusion that using the new strategy improves the accuracy. In most samples, the misclassification times are decreased. Reaping the benefit of proposed strategy, the more qualitative and reliable data are used to support the further professional analysis and diagnosis. However, the new strategy has a drawback in time-consuming. This problem becomes the main bottleneck to restrict the application of Ambulatory electrocardiogram auto-analysis. The authors will focus on this problem in the future research.

ACKNOWLEDGMENT

This work was carried out based on academic achievement of Laboratory of biological signal and intelligent processing, Tianjin University of Technology. During the study, the work got highly supporting from Tianjin Key Laboratory of intelligence computing and novel software technology on its hardware environment. The work was supported by Tianjin Natural Science Foundation, contracts 10JCYBJC00700 and Science and Development Funds of Tianjin Higher Education SB20080044 on studying consuming. The authors would like to acknowledge Dr. Bingrang Zhao and Dr. QinQin (Tianjin Chest Hospital)'s team in medical knowledge supporting. Thoughtful comments and suggestions from Prof. Yalou Huang, Dr. Shusong Xing from NanKai University, Tianjin, and Prof. Pengtao Wang, Prof. Junqing Sun from School of computer and communication engineering, Tianjin University of Technology, are all greatly appreciated.

REFERENCES

[1] D. Yu. Larionov, A. A. Shelupanov. "Analysis structure and algorithms used in personal ECG recorders," *Biomedical Engineering*, 2007.12, pp.150-153.

[2] Ping Zhou, Lowery, M.M.Dewald, J.P.A. Kuiken. T.A. "Towards Improved Myoelectric Prosthesis Control: High Density Surface EMG Recording After Targeted Muscle Reinnervation," *Engineering in Medicine and Biology Society 27th Annual International Conference,* 2006.4, pp.4064-4067.

[3] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi. "Optimization by Simulated Annealing," *Science*, 1983.5, pp.671-680

[4] C.Bishop, Neural Networks for Pattern Recognition. NewYork: Oxford Univ. Press, 1995

[5] Kirkpatrick. "Optimization by simulated annealing," *Science*, 1983.220, pp.671-680.

[6] Qinghua Yue. "A study of Ambulatory Electrocardiogram waveform clustering strategy," *Journal of Tianjin University of Technology*, 2008.2, pp.49-52.

[7] V. Cherkassky and F. Mulier, Learning From Datra: Concepts, Theory, and Methods. New York: Wiley,1998

[8] Chen Tung-shou, Tsai Tzu-hsin, and Chen Yi-tzu. "A combined *k*-means and hierarchical clustering method for improving the clustering efficiency of micro-array," (2006-02-01),http://ieeexplore.ieee.org/xpls/abs-all.jsp?arnumber=1595432.

[9] Jeffcoat D E, and Bulfin R. "Simulated annealing for resource constrained scheduling." *European Journal of Operational Research*, 1993.70, pp.43-51.

[10] R. Duda, P. Hart, and D. Stork, Pattern Classification, 2nd Ed. NewYork: Wiely, 2001

[11] Shokri Z. Selim, K. Alsultan. "A simulated annealing algorithm for the clustering problem," *Pattern Recognition*, 1991, pp.1003-1008.

[12] Dickhaus H, Gittinger J."Classification of QRS morphology in holter monitoring". In: *Proceedings of the 1st Joint BMES/EMBS conference*. *Atlanta: Serving Humanity, Advancing Technology*, 1999. 270-270.

[13] Sung-Nien Yu, Kuan-To Chou. Combining Independent Component Analysis and Backpropagation Neural Network for ECG Beat Classification[C]. IN: *Proceedings of the 28th Annual International Conference of IEEE Engineering in Medicine and Biology Society, New York, USA,* 2006, 3090-3093

[14] Zheng Gang, Huang Yalou, etc. "Study of Classficication Rules on Weighted Coronary Heart Disease Data. "*Proceedings of the Fourth International conference on Machine Learning and Cybernetics*. Guangzhou: 2005, 1845~1850.

[15] Osowski, S. Linh Tran Hoai. Analysis of features for efficient ECG signal classification using neuro-fuzzy network[C]. IN: *Proceedings of Neural Networks of IEEE International Joint Conference,* 2004 , 3: 2443 – 2448

[16] SCHOLKOPFB, SMOLA A J, MULLER K R. "Nonlinear component analysis as a kernel eigenvalue problem". Neural Computation, 1998, 10(5):1299-1319.

[17] Jain AK, Flynn PJ. "Image segmentation using clustering." In: Ahuja N, Bowyer K, eds. Advances in Image Understanding: A Festchrift for Azriel Rosenfeld. Piscataway: IEEE Press, 1996.

[18] Zhao Yong, Hong Wenxue, Xu Yonghong, et al. ECG Beats Classification Based on Ensemble Feature Composed of Independent Components and QRS Complex Width[C]. IN: *Proceedings of the International Conference on Computer Science and Software Engineering, Wuhan, China*, 2008, 868 – 871

[19] Jain AK, Murty MN, Flynn PJ. "Data clustering": A review. ACM Computing Surveys, 1999.

[20] Jain AK, Dubes RC. "Algorithms for Clustering Data." Prentice-Hall Advanced Reference Series, 1988.

[21] Fred ALN, Leitão JMN. "Partitional vs hierarchical clustering using a minimum grammar complexity approach." In: *Proc. of the SSPR&SPR 2000*. LNCS 1876, 2000. 193−202. http://www.sigmod.org/dblp/db/conf/sspr/sspr2000.html.

[22] E. Aarts and Jan Korst, "Simulated Annealing and Boltzman Machines", *A Stochastic Approach to Combinatorial Optimization and Neural Computing*. John Wiley & Sons Ltd. 1989.

[23] V. Cemy, "Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm". *Journal of Optimization Theory and Applications*, Vol.45, No. 1, 1985.

[24] Shokri Z. Selim and K. Alsultan, "A Simulated Annealing Algorithm for The Clustering Problem", *Pattern Recognition*, Vol. 24, No. 10, 1991, pp. 1003-1008.

[25] S. P.Ma, Y.X.Zhu, Artificial intelligence, Beijing: Tsinghua University Press, 1993, pp. 12-15.

[26] S. Bandyopadhyay, U. Maulik, and M. K. Pakhira, "Clustering using simulated annealing with probabilistic redistribution," *Int. J. Pattern Recognition. Artificial.*, vol. 15, no. 2, pp. 269–285, 2001.

[27] D. K. Nam and C. Park, "Multiobjective simulated annealing: A comparative study to evolutionary algorithms," *Int. J. Fuzzy Syst.*, vol. 2, no. 2, pp. 87–97, 2000.

[28] P. Laguna, R. Jane, and P. Caminal. "Adaptive feature extraction for QRS classification and ectopic beat detection". *Proc. Computers in Cardiology*, 1991, pp. 613-616.

[29] K.P. Lin, and W.H. Chang. "QRS feature extraction using linear prediction". *IEEE Trans. on BME*, 1989, 36(10), pp. 1050-1055.

[30] A.I. Rasiah, and Y. Attikiouzel. "Syntactic recognition of common cardiac arrhythmias". *Proc. Engineering in Medicine* and Biology Society, 1994, 1, pp. 155-156.

[31] Ghongade, Rajesh Ghatol, Ashok. A robust and reliable ECG pattern classification using QRS morphological features and ANN[C]. IN: Proceedings of TENCON of IEEE Region 10 Conference, Hyderabad, India, 2008, 1 – 6

[32] Tadejko, P.，Rakowski, W. Hybrid wavelet-mathematical morphology feature extraction for heartbeat classification[C]. IN: *Proceedings of the International Conference on Computer as a Tool, Warsaw, Poland,* 2007, 127 – 132

[33] Al-Naima, F.M. Ali, A.H. Mahdi, S.S. Data acquisition for myocardial infarction classification based on wavelets and Neural Networks[C]. IN: *Proceedings of 5th International Multi-Conference of IEEE Systems Signals and Devices, Amman , Jordan, 2008, 1-6*

[34] A.T.Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 24, no. 3, pp. 381-396, March 2002.

[35] Xu R, Wunsch D II, "Survey of clustering algorithms", *IEEE Transaction on Neural Networks,* Vol. 16, Issue 3, May 2005, pp.645-678.

[36] Jana Bardonova, Ivo Provaznik, et al. Detection of Myocardial Ischemia Using Hidden Markov Models[C], IN: *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Cancun, Mexico*, 2003, 2869-2872

[37] Shehroz S. Khan, Amir Ahmad, "Cluster Center Initialization Algorithm for K-Means Clustering", Pattern Recognition Letters, Vol. 25, 2004, pp.1293-1302.

[38] Krista Rizman Zalik, "An Efficient K'means Clustering Algorithm",*Pattern Recognition Letters*, Vol. 29, 2008, pp. 1385-1391.

[39] Januzaj E, Kriegel Hans P, Pfeifle M., "DBDC: Density Based Distributed Clustering", *Advances in Databases Technology – EDBT 2004, Springer Berlin / Heidelberg,* Vol. 2992, February 2004, pp.529-530.

[40] M. Halkidi, Y. Batistakis, M. Vazirgiannis, "Cluster Validity Methods", *ACM SIGMOD Record,* Vol. 31, No. 3, 2002, pp. 19-27.

[41] Daxin Jiang, Chun Tang and Aidong Zhang, "Cluster Analysis for Gene Expression Data: A Survery", *IEEE Transactions on Knowledge and Data Engineering,* Vol. 16, No. 11, November 2004, pp. 1370-1386

[42] Chieh-Yuan Tsai, Chuang-Cheng Chiu, "Developing a Feature Weight self-adjustment Mechanism for a K-Means Clustering Algorithm", *Computational Statistics and Data Analysis*, Vol. 52, 2008, pp. 4658-4672.

[43] Adil M. Bagirov, "Modified Global K-Means Algorithm for Minimum Sum-of-squares Clustering Problems", *Pattern Recognition*, Vol. 41, 2008, 3192-3199.

[44] Liang Yu Shyu,Ying Hsuan Wu,Hu,W. Using wavelet transform and fuzzy neural network for VPC detection from the holter ECG[J]. *IEEE Transactions on Biomedical Engineering. Taiwan* , 2004, 51(7) :1269 – 1273

[45] Mrs.S .N. Shivappriya, Mrs.R. ShanthaSelvaKumari. ECG Delineation Using Stationary Wavelet Transform[C]. IN:*Proceedings of International Conference on Advanced Computing and Communications. Surathkal* ,2006,271-274

[46] XIUYU ZHENG, ZHEN LI. Detection Of QRS Complexes Based On Biorthogonal Spline Wavelet[C]. In: *Proceedings of International Symposium on Information Science and Engieering. Shanghai, china*, 2008, 502-506

Gang Zheng received the B.S. Degree in Computer Software from Tianjin University of Technology (TUT) in 1993. He received M.S and Ph.D. Degrees in Computer Application Technology from NanKai University, Tianjin, China in 2003, and 2007, respectively. His Ph.D. degree thesis was on the AECG (Ambulatory Electrocardiogram) waveform analysis based on machine learning theory.

From 1993 to now, he worked as a faculty in TUT, China. He is a professor with school of computer and communication engineering in the same university. He Joined in Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology as a research in 2006. In 2007, he established Laboratory of Biological Signal and Intelligent Processing under controlled by TUT. He studied on ECG (Electrocardiogram) signal since 2001, and proposed some publications about ECG waveform analysis based on machine learning techniques. Meanwhile, he started to design ECG signal collecting system from the beginning of 2003, and he had presented one lead 24 hours ECG signal collecting system, 3-leads 24 hours synchronic ECG signal collecting system, 24 hours ECG waveform morph analysis software system. He was also a Visiting Professor at KanSai University, Japan, and University of Aizu, Japan for cooperation on bioinformatics research.

Dr. Zheng is a Member of IEEE.

Tian Yu received the B.S. Degree in Material Engineering from Soochow University in 2006. He received M.S Degree in Computer Application Technology from Tianjin University of Technology (TUT) in March, 2010. His M.S degree thesis was "A Clustering Analysis Strategy for On-line Dynamic Electrocardiogram Waveform".

From March, 2010 to now, he join in NanKai University, Tianjin, China as a Ph.D. candidate for his further study in data mining.