

# Comparison of Similarity Measures for Clustering Electrocardiogram Complexes

KC Chang<sup>1</sup>, RG Lee<sup>2</sup>, C Wen<sup>1</sup>, MF Yeh<sup>1</sup>

<sup>1</sup>Lunghwa University of Science and Technology, Taoyuan, Taiwan, Republic of China  
<sup>2</sup>National Taipei University of Technology, Taipei, Taiwan, Republic of China

## Abstract

*This paper compares four similarity measures such as the city block ( $L_1$ -norm), the Euclidean ( $L_2$ -norm), the normalized correlation coefficient, and the simplified gray relational grade for clustering QRS complexes. Performances of the measures include classification accuracy, threshold value selection, noise robustness, and execution time. The clustering algorithm used is the so-called two-step unsupervised method. The best out of the 10 independent runs of the clustering algorithm with randomly selected initial template beat for each run is used to compare the performances of each similarity measure. Simulation results show that the simplified gray relational grade outperforms the other measures.*

## 1. Introduction

Generally speaking, in classification or clustering algorithms for the ECG, a similarity measure is used to measure the distances between the query beat and the templates in the database, or adjustable weights in the artificial neural network used, for ranking. The smaller the distance is, the more similar the template (adjustable weight) is to the query. Then, a threshold value on the measurements is used for the help with the decision-making. It is known that the distance or similarity measure used in a clustering or classification algorithm will have large effect on the performance of the classification procedure. In [1], performance of several similarity measures in four types of clustering algorithms was investigated on the practical medical data set containing both numerical and categorical of attributes. In [2], nine image similarity measures were compared for texture image retrieval. In this paper, we compare four similarity measures for clustering of the ECG.

Distance measures widely used in the time domain ECG classification algorithms include  $L_1$ -norm (city block),  $L_2$ -norm (Euclidean distance), normalized cross correlation, likelihood function, and Mahalanobis distance. In Mahalanobis distance, the covariance matrices can be hard to determine accurately, and the memory and the time requirements grow quadratically

rather than linearly with the number of features. These problems may be insignificant when only a few features of the ECG are needed, but they become quite serious when the number of features becomes large. A likelihood function is usually used in an algorithm such as a linear discriminant, where the likelihood function of the data is maximized with respect to the parameters of the model. The model is of suitable functional form with parameters to be adjusted so that it approximates the distribution of the data. The computational load of this approach is very high.

The gray relational grade [3] is also a distance measure whose value is within the interval (0, 1). For a given reference sequence and a given set of comparative sequences, gray relational analysis [4] can be used to determine the gray relational grade between the reference and each comparative sequence in the given set. Then the best comparative sequence is the one having the largest gray relational grade. Recently, an unsupervised algorithm has been developed by using gray relational grade as the distance measure for data clustering and showed that it could achieve a good performance [10]. The original gray relational grade is simplified for the reduction of computation, in this paper.

In this paper, we compare performance of four similarity measures, namely, the  $L_1$ -norm, the  $L_2$ -norm, the normalized cross correlation, and the simplified gray relational grade, that are relatively simple in computation. Performances of the measures include classification accuracy, threshold value selection, noise robustness, and execution time. The four similarity measures are compared in a clustering algorithm for clustering the ECG. The clustering algorithm used in this paper is the so-called two-step unsupervised method [5], which was reported to have a better performance than the hierarchical cluster analysis approach.

## 2. Methods

### 2.1. Similarity measures

First, we introduce the gray relational grade and simplify it. Denote the reference sequence by  $y = \langle y(1), y(2), \dots, y(n) \rangle$ , and the comparative

sequences by  $x_i = \langle x_i(1), x_i(2), \dots, x_i(n) \rangle$ ,  $i = 1, \dots, m$ . Define the gray relational coefficient between  $y$  and  $x_i$  at the  $k$ -th datum as [9]

$$r(y(k), x_i(k)) = \frac{\Delta_{\min} + \xi \cdot \Delta_{\max}}{\Delta_i(k) + \xi \cdot \Delta_{\max}} \quad (1a)$$

where  $\Delta_i(k) = |y(k) - x_i(k)|$ ,  $\Delta_{\max} = \max_i \max_k \Delta_i(k)$ ,  $\Delta_{\min} = \min_i \min_k \Delta_i(k)$ ,  $k = 1, 2, \dots, n$ , and  $\xi \in (0, 1]$  is the distinguishing coefficient controlling the resolution between  $\Delta_{\max}$  and  $\Delta_{\min}$ . The weighted average of the gray relational coefficients, termed by the gray relational grade, can be computed by

$$g(y, x_i) = \sum_{k=1}^n [w(k) \cdot r(y(k), x_i(k))] \quad (2)$$

where  $w(k) \geq 0$  is the weighting function with  $\sum_{k=1}^n w(k) = 1$ . The weighting function can be seen as a window that can be used to emphasize a segment of ECG signal such as P wave. In this paper, we deal only with QRS complex classification problem, and we select the weighting function as  $w(k) = 1/n$  for all  $k$ . From (1a) and (2), it can be seen that  $0 < g(y, x_i) \leq 1$  and that the smaller the absolute value  $\Delta_i(k)$ , the larger the gray relational grade, and vice versa. The best comparative sequence is the one with the largest gray relational grade to the reference. Hence, the gray relational grade can be used as a distance measure.

The gray relational grade in (2) is simplified by using

$$r(y(k), x_i(k)) = \frac{d}{\Delta_i(k) + d} \quad (1b)$$

where  $d$  is a positive real constant. If we use (1b) and (2) as the similarity measure for ranking the comparative sequences, it results in the same rank as that obtained by using (1a) and (2). By this way, the best comparative sequence can be found by using (1b) and (2). The grade obtained by (1b) and (2) is called the simplified gray relational grade, denoted by  $d_{sg}$ , in this paper. The range of the simplified gray relational grade is still in the interval (0, 1). Note that, for a specific sequence in the comparative set, its simplified gray relational grade value and its gray relational grade value obtained by (1a) and (2) may be different, but the rank of the sequence to the reference in that set is always the same by either of the two measures.

The other measures used in this paper are summarized in the following. Denote the city block or L1 distance of two sequences  $x$  and  $y$  by

$$d_{L1}(y, x) = \sum_{k=1}^n |y(k) - x(k)|. \quad (3)$$

The Euclidean distance or L2 distance is defined by

$$d_{L2}(y, x) = \sqrt{\sum_{k=1}^n (y(k) - x(k))^2}. \quad (4)$$

The Euclidean distance is not always the best measure. The fact that the distances in each element of the sequences are squared before summation places great emphasis on those elements for which the dissimilarity is large. Finally, the normalized cross correlation is defined by

$$d_{cc}(y, x) = \frac{\sum_{k=1}^n (y(k) - \bar{y})(x(k) - \bar{x})}{\sqrt{\sum_{k=1}^n (y(k) - \bar{y})^2 (x(k) - \bar{x})^2}} \quad (5)$$

where

$$\bar{y} = \frac{1}{n} \sum_{k=1}^n y(k) \quad \bar{x} = \frac{1}{n} \sum_{k=1}^n x(k).$$

A distance function with a wide range may cause a problem in the selection of the threshold value for decision-making when it is used in a clustering algorithm. From the definition of each distance function, we see that  $d_{L1} \in [0, \infty)$ ,  $d_{L2} \in [0, \infty)$ ,  $d_{cc} \in [-1, 1]$ , and  $d_{sg} \in (0, 1]$ . For the easy selection of the threshold value,  $d_{L1}$  and  $d_{L2}$  are usually normalized, and the absolute value of  $d_{cc}$  is used in clustering algorithms. Let the normalized  $d_{L1}$  and  $d_{L2}$  be denoted by  $\hat{d}_{L1}$  and  $\hat{d}_{L2}$ , respectively. Finally, define the similarity measure for each distance function as

$$s_{L1} = 1 - \hat{d}_{L1} \quad (6)$$

$$s_{L2} = 1 - \hat{d}_{L2} \quad (7)$$

$$s_{cc} = |d_{cc}| \quad (8)$$

$$s_{sg} = d_{sg}. \quad (9)$$

From these definitions, we see that all similarity measures lay within the range [0, 1] except  $s_{sg} \in (0, 1]$  and that, for all similarity measures, the larger the measurement of two sequences, the more similar the two sequences.

## 2.2. Material and clustering method

The MIT/BIH Arrhythmia Database, sampled at 360 Hz with 11-bit resolution, was used in this study. Excluding records containing paced beats (102, 104, 107, and 217), we used the other 44 records of the MIT/BIH Database. Raw data from single channel (channel 1) were

used (with some beats missing). The size of each QRS complex was 73 samples (about 200 ms) with the R wave at the center. This study did not investigate the problem of QRS detection from the ECG, instead the position of annotation labels provided with the MIT/BIH Database was used to identify the R wave. The AAMI recommended practice was used to combine the MIT/BIH beat types into four classes. Class N contains normal, bundle branch block, and supraventricular ectopic beats, class V contains ventricular ectopic beats, class F contains fusion of ventricular and normal beats, and class Q contains unclassifiable beats.

Without feature extraction, the QRS morphology with 73 samples was used in this study. To test the robustness of the similarity measures to high frequency noise and power line interference, we did not use any data-preprocessing step such as filtering in this study. However, the mean value was subtracted from each beat to remove the baseline wander.

The so-called two-step clustering algorithm used in this study was suggested in [5]. The algorithm takes the first beat of the ECG as the template and compares it to all successive beats by one of the similarity measures (6)-(9). The beats showing certain degree of similarity are put together to form a cluster. By this we mean that if the measurement exceeds a threshold value, the compared beat belongs to the same cluster as the template. If no measurement exceeds the threshold value, the template itself forms a cluster. When all successive beats have been compared, the first beat of the residual QRS complexes is taken as a new template, and it is used to iterate the whole procedure until no QRS complex is left unclassified. In the second step of the algorithm, the clusters are merged by the same way used in the first step. However, in this step, we use the cluster center to represent the cluster and select a slightly lower threshold value than that used in step one. Finally, all beats in one-element clusters are collected in a signal cluster, and they are assigned as unclassifiable beats (class Q).

The two-step clustering algorithm described in the above was used to evaluate performances of the four similarity measures (6)-(9). Performances included the accuracy, the execution time, the variation of threshold for the best result for each record, and the noise robustness.

### 3. Results

All programs coded by Matlab® version R12 were executed by a personal computer with Intel® Pentium® 4 1.6 GHz processor and 512 MB RAM. In the two-step clustering algorithm, different initial templates may result in different performances. To have a fair comparison, we randomly selected 10 ECG beats from each test record as the initial templates for that record. Then, for each record,

we used 10 runs, one initial template for a run, to evaluate the performances of each of the 4 similarity measures. By this way, each similarity measure was evaluated 10 times independently under the same clustering method with the same initial template for every test record. Then, for each measure, the best result of the 10 runs was selected for the comparison. In each run, the threshold value used in the first step of the algorithm was selected to make the algorithm have the best performance in accuracy, and in the second step it was selected to minimize the resulting number of clusters while keeping the accuracy obtained by the first step.

Table 1. Results of ECG beats clustering using various measures on 44 MIT/BIH ECG data files.

Record	Number of beats	Accuracy (%)			
		$s_{L1}$	$s_{L2}$	$s_{cc}$	$s_{sg}$
105	2572	98.60	99.61	99.96	99.73
106	2027	98.67	96.35	100	99.95
108	1773	99.38	99.89	99.94	99.89
109	2531	100	99.96	100	100
114	1879	100	99.89	99.89	99.84
116	2411	99.96	99.96	100	99.96
118	2287	99.78	99.96	99.96	99.96
124	1619	99.94	99.82	100	100
200	2600	99.69	99.35	99.92	99.92
201	2000	99.55	99.50	100	100
203	2980	97.68	98.05	99.66	99.66
205	2656	99.74	99.81	99.70	99.74
207	2331	97.43	94.21	99.96	99.96
208	2953	97.97	97.56	97.53	98.04
210	2650	99.58	99.77	99.93	99.55
213	3250	95.94	95.54	95.02	95.54
214	2260	98.85	99.87	99.74	99.87
219	2287	97.60	97.60	97.51	97.55
221	2427	99.67	99.51	100	100
223	2605	99.54	99.62	99.54	99.62
228	2053	99.95	99.95	100	100
231	1573	99.94	99.94	100	100
233	3078	99.77	99.81	99.84	99.84
X	46572	100	100	100	100
Total	101374	99.46	99.38	99.66	99.68
Beats (N)	62710	99.50	99.46	99.82	99.83
Beats(C)	38664	99.40	99.25	99.40	99.44
Beats (c)	35415	99.72	99.59	99.80	99.80
Execution time (sec)		35.73	38.06	50.05	41.50

In respect to the performance in accuracy, the best clustering result of such a 10 runs for each measure for each record is listed in Table 1, in which the term X denotes all records with 100% accuracy by all the

measures, Beats (N) denotes the number of beats in 27 noisy records, Beats (C) denotes the number of beats in the other 17 clean records, and Beats (c) denotes the number of beats in clean records excluding record 213. Table 1 shows that, over total 101,374 test ECG beats,  $s_{L1}$ ,  $s_{L2}$ ,  $s_{cc}$ , and  $s_{sg}$  achieve 99.46%, 99.38%, 99.66%, and 99.68% in accuracy, respectively. The simplified gray relational grade outperforms the other similarity measure functions in classification accuracy.

Table 1 also shows that the performance in accuracy of each measure on noisy records. Over 62,710 beats in 27 noisy records,  $s_{L1}$ ,  $s_{L2}$ ,  $s_{cc}$ , and  $s_{sg}$  achieve 99.50%, 99.46%, 99.82%, and 99.83% in accuracy, respectively. The simplified gray relational grade has the best performance in accuracy on noisy records. It seems that all measures perform better on noisy records than on clean records. This is because all of the worst cases of the measures appear on record 213 that is labeled clean. Excluding record 213, all measures perform better on clean records as shown in Table 1.

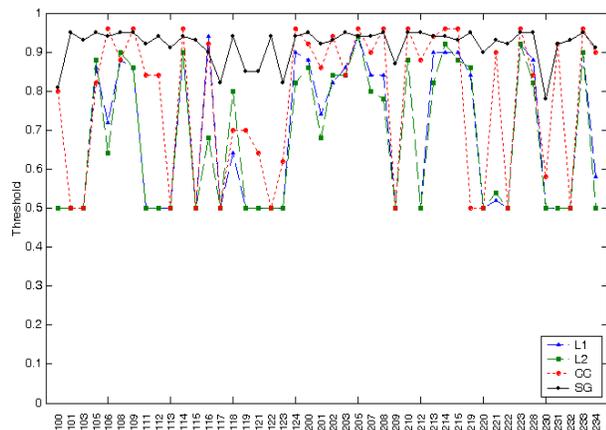


Figure 1. The variation of the threshold.

Fig. 1 shows the variation of the threshold value used in the first step of the algorithm. Note that the threshold value reported is the minimum value for achieving the best performance in accuracy. For values larger than the minimum value, the accuracy can also be achieved but with a larger number of resulting clusters in this step. Though the resulting clusters in the first step will be merged in the second step of the algorithm, a larger number of clusters cause a more computational load in the second step. This figure depicts that, for different records, the minimal threshold value required for the simplified gray relational grade achieving the best result is relatively stable, comparing with the other measures. This is a particularly useful feature since we need not spend much time to decide the threshold, in practice, for

analyzing various ECG recordings. The threshold in the second step of the algorithm is used to minimize the number of clusters by merging them. No significant difference in the number of final resulting clusters was observed for different measures. Once, the threshold values for the first and the second steps were decided for every record, we ran the program again with the fixed threshold values to examine the execution time for each measure. The result is also listed in Table 1. It is seen that, over 101,374 test beats, the clustering algorithm with the  $L_1$  function takes the least time (35.73 sec.). The computational load caused by the additional procedure for normalizing the  $L_1$  or the  $L_2$  function seems that it can be negligible.

#### 4. Conclusions

In this paper, the performances of four similarity measures used in the so-called two-step clustering algorithm for ECG complexes have been investigated. It was found that the simplified gray relational grade outperformed the other measures in classification accuracy, noise robustness, and the selection of thresholds. The best measure in the execution time was the  $L_1$  distance function.

#### Acknowledgement

This work was supported by the National Science Council, Taiwan, under Grant NSC 94-2213-E-027-012.

#### References

- [1] Hirano S, Sun X, Tsumoto S. On similarity measures for cluster analysis in clinical laboratory examination database. In: Proc. 2002 IEEE 26th Ann. Int. Computer Software and Applications Conf. 2002:1170-1175.
- [2] Kokare M, Chatterji BN, Biswas PK. Comparison of similarity metrics for texture image retrieval. In: Proc. 2003 IEEE Conf. on Convergent Technologies for Asia-Pacific Region 2003;2:15-17.
- [3] Deng JL. Introduction to gray system theory. J. Grey Syst. 1989;1:1-24.
- [4] Chang KC, Yeh MF. Gray relational analysis based approach for data clustering. IEE Proc. Vis. Image Signal Process. 2005;152:165-172.
- [5] Maier C, Dickhaus H, Gittinger J. Unsupervised morphological classification of QRS complexes. In: Computers in Cardiology 1999, IEEE Computer Society Press, 1999:683-686.

Address for correspondence

Ren-Guey Lee  
 National Taipei University of Technology  
 Department of Electronic Engineering  
 No. 1, Sec. 3, Chung-Hsiao E. Rd., Taipei 10643, Taiwan  
 evans@ntut.edu.tw