35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Classification of Oscillometric Envelope Shape Using Frequent Sequence Mining

Hung-Wen Diao, Weichih Hu, Gong-Yau Lan, and Liang-Yu Shyu, *Member, IEEE*

Abstract— **The shape of the oscillometric envelope is known to affect the accuracy of automatic noninvasive blood pressure (NIBP) measurement devices that use the oscillometric principle to determine systolic and diastolic blood pressures. This study proposes a novel shape classification method that uses data mining techniques to determine the characteristic sequences for different envelope shapes. The results indicate that the proposed method effectively determines the characteristic sequences for different subject groups. Subjects in the high- score group and in the low- score group tend to have an envelope with a broader plateau and are bell-shaped, respectively. This information about shape can be used for future determination of the correct algorithm for systolic and diastolic blood pressures determination in NIBP devices.**

I. Introduction

Because of an imbalance of diet and a change of life style, cardiovascular disease has become the leading cause of mortality in developed countries. On the other hand, blood pressure is a good indicator of cardiovascular condition. Hypertension is also correlated with several life-threatening diseases such as atrial fibrillation [1] and renal disease [2]; so, numerous studies have suggested the long term control and monitoring of blood pressure [1, 2].

The use of automatic noninvasive blood pressure (NIBP) measurement devices for blood pressure monitoring at home has become widespread. Most of the NIBP devices use the oscillometric method which utilizes the oscillometric waveform from the pressure cuff in conjunction with specific characteristic ratios to determine the systolic blood pressure (SBP) and diastolic blood pressure (DBP). However, the accuracy of NIBP devices is not assured [3]. One of the possible sources of inaccuracy is the different characteristic ratios that are used by different manufacturers [3]. It is also found that the shape of oscillometric waveform envelopes contribute to this difference [4]. When the oscillometric envelopes are smooth and have well-defined peaks, most NIBP devices generate a reliable SBP and DBP. However, when the envelopes have complex shapes and a broad plateau, most devices significantly overestimate or underestimate SBP and DBP [4]. In order to improve the accuracy of NIBP devices, a different characteristic ratio for different envelope shapes is required.

Several known clinical factors, such as arterial hardness, affect the shape of the oscillometric waveform envelope. It is also found that coronary artery calcification is correlated with arterial stiffness [5], so this study uses oscillometric waveforms from patients with different degrees of coronary artery calcification. The study aims to develop a new method, using sequence data mining technique, to classify the pattern of oscillometric envelopes. The result of this classification will improve the future accuracy of NIBP devices.

II. Methods

A. Subject

Sixty patients recruited for this study under-went coronary artery calcification examination, using an 64-slices CT scan in the Cheng-Hsin Rehabilitation and Medical Center. There were 27 male and 33 female subjects. The patients were divided into three groups: low-score (0~11), mid-score (11~100) and high-score (101~), according to their CAC score. The demographic information is summarized in Table 1. There are 30 low-score patients and the numbers of patient for mid-score and high-score are 17 and 13, respectively. The highest CAC score is 646. After a short orientation and rest, blood pressure was measured using a custom-built device from the Chung-Yuan Christian University. The oscillometric pulses were acquired from the recorded cuff pressure using the custom-built software. This study received ethical permission from the Cheng-Hsin Rehabilitation and Medical Center, and all subjects gave their informed consent before the task.

B. Direct Sequence Comparison (DISC)

Mining frequent sequences or patterns from a large database or over data streams is a common problem in the data mining community. This study aims to classify the shape of oscillometric envelopes which is similar to finding the specified pattern for each envelope. Therefore, this study

TABLE I. Demographic Information

CAD score	*Male/Female*	*Age, y*	*SBP*	*DBP*
low-score (0~11)	8/22	50±7	121±10	92±20
mid-score (11~100)	10/7	56±6	129±11	86±10
high-score (101~)	9/4	57±3	125±14	81±9

*Resrach supported by The National Science Council, Taiwan (NSC100-2221-E-033-038-).

H. W. Diao was with Chung Yuan Christian University. She is now with the Enzytek Inc., Hsinchu, Taiwan (e-mail: marten@cycu.org.tw).

G. Y. Lan, is with the Dept. of Medical Image, Cheng Hsin Rehabilitation and Medical Center, Taipei, Taiwan (e-mail: ch5697@chgh.org.tw).

W. C. Hu is with the Department of Biomedical Engineering, Chung Yuan Christian University, Chung Li, 32023 Taiwan (e-mail: weichih@be.cycu.edu.tw).

L. Y. Shyu is with the Department of Biomedical Engineering, Chung Yuan Christian University, Chung Li, 32023 Taiwan (phone: 886-3-265-4515; fax: 886-3-265-4599; e-mail: liangyu2@cycu.edu.tw).

uses sequence data mining technique to obtain the frequent characteristic sequence from the oscillometric envelopes.

Using a transaction database as an example, a transaction contains the items purchased by a customer at a certain time. Sorting the purchase list in ascending order of time to form a sequence and generate the customer sequence. The frequent sequences for each customer sequence represent the consumer patterns of particular customers. In general, in order to min the frequent sequences, all candidate sequences are generated and the support counts of each candidate are computed.

For efficiency, candidate pruning, database partitioning and sequence shortening are commonly used strategies in the mining processes. By using the locative AVL-tree, the direct sequence comparison (DISC) strategy prunes candidate sequences with the same length directly and increases efficiency [6]. The DISC strategy finds all of the frequent k-sequences for any specified length, k, and sorts them in ascending order. A sequence is frequent if and only if at least δ customer sequences support it. The DISC compares the minimum of all the k-subsequences with the δth position of the sorted list. If these are not equal, then the minimum k-subsequence must be non-frequent and sequences between the minimum and the $(\delta-1)$th position are all non-frequent. The dynamic DISC-all algorithm incorporates a database partitioning strategy and DISC to further improve the performance. This study uses DISC strategy to mine the frequent sequence from the oscillometric envelopes [7]. To order to apply the DISC strategy, two parameters, the sequence length k and the minimum required count, must be determined.

C. Pre-Process of Oscillometric envelope

In order to ensure consistency with the transaction database example, this study treats every oscillometric envelope as a customer sequence. It is assumed that a particular customer (subject) group with a similar consumer pattern (clinical condition) results in a similar customer sequence (oscillometric envelope). Therefore, after mining, the frequent sequences from each customer group can be used to represent the pattern of this group. However, the original oscillometric envelopes are not suitable for frequent sequences mining, because the amplitudes of the envelope are continuous. In order to convert oscillometric envelopes into the customer sequences, this study quantizes the normalized oscillometric envelopes. Firstly, the oscillometric envelope is normalized to unity according to its maximum and minimum. After normalization, the envelope is then divided and rounded into seven levels representing seven purchasing activities. Fig. 1 shows three typical oscillometric envelopes and the results of quantization.

D. Characteristic Sequences

In general, the mining process produces multiple frequent sequences. However, not all of the frequent sequences are meaningful and some are repeated in different groups. Further analysis is required to screen out the unwanted

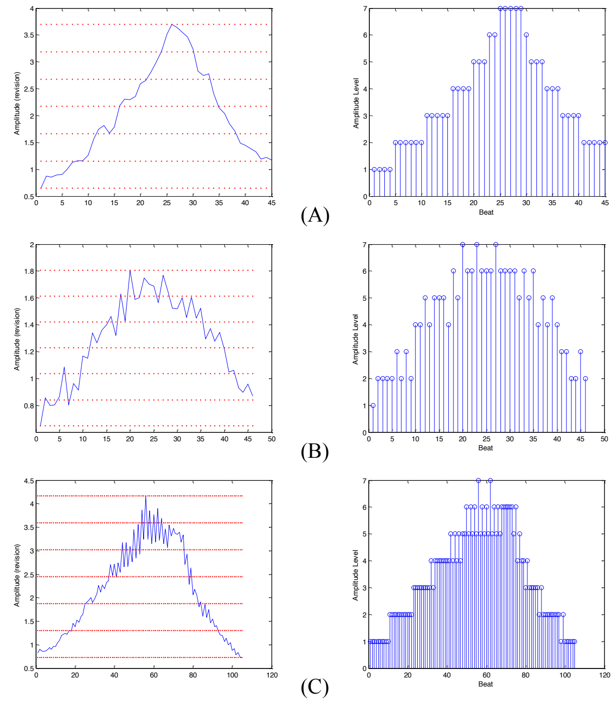

Figure 1. Quantization results for three typical oscillometric envelopes. Where (A) has a well-defined peaks, and (B) and (C) have a broader plateau and complex shapes, respectively.

sequences in order to identify the most representative sequences for a particular group of subjects.

In order to emphasize the uniqueness of the oscillometric envelope pattern for different groups, any repeated frequent sequences are removed from the final frequent sequences. Additionally, only the frequent sequences with sufficient supports are preserved for the final analysis and these sequences are the so-called characteristic sequences.

Characteristic sequences with good quality can represent a particular shape, distinctively. In order to assess the quality of extracted characteristic sequences, this study uses the concept of coverage, as follows:

$$\text{coverage} = \frac{N_c}{N_t} \times 100 \qquad (1)$$

Where N_c is the number of supported customer of this particular set of characteristic sequence and N_t is the total number of customer in the particular group. A higher coverage means that these sequences are supported by more customers. In other words, these sequences appear in more customer sequences.

III. Results

A. Effect of Sequence Length, k

During frequent sequence mining, several factors, such as the number of customer sequences, the sequence length, and the minimum support, affect the complexity of the mining process and the time required to complete it. For the sake of simplicity, the minimum support is fixed at 0.4, in

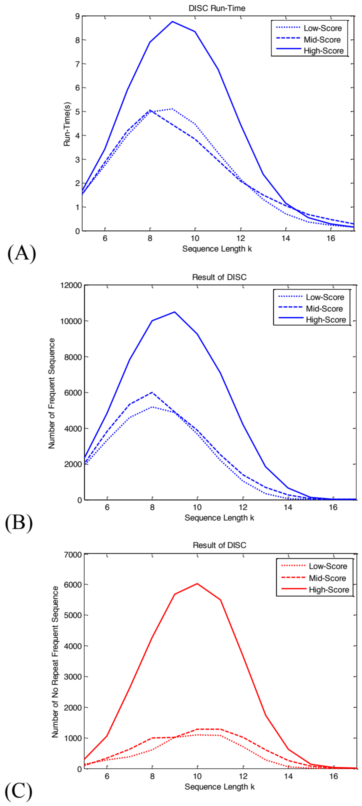

Figure 2. The effect of sequence length, k. (A) illustrates that the time for frequent sequence mining increases and then decreases as k increases. (B) and (C) show the total number of frequent sequences before and after removal of the repeated sequences, respectively.

this study, so with 13 subjects in high-score group, the required support for a frequent sequence is, 0.4*13 = 5.2, 5. That is, in the high-score group, a frequent sequence must have support of at least five subjects.

In order to determine the best parameter setting and the best characteristic sequences for each group, this study determines the effect of sequence length, k, ranging from 5 to 17. Fig. 2 shows that the time for frequent sequence mining increases and then decreases as k increases. The longest mining is required when the length is nine. There is also a maximum number of frequent sequences when the sequence length is nine. The total numbers of frequent sequences are 27177, 31181 and 58648, for the low-, mid- and high-score groups, respectively. However, the total number of frequent sequences reduces dramatically to 5608, 7613 and 31597, for the low-, mid- and high-score groups, respectively, after removal of the repeated sequences. However, this is still too great a number for efficient computation.

In order to further reduce the number of non-repeated frequent sequences, the supporting customers for each sequence are recorded. The sequences are then sorted according to their supporting number. Sequences with a maximum supporting number are collected and designated as the characteristic sequences for that particular group. It is found that the number as well as the coverage of each characteristic sequences vary dramatically for different sequence length. When the sequence length is lower than 7

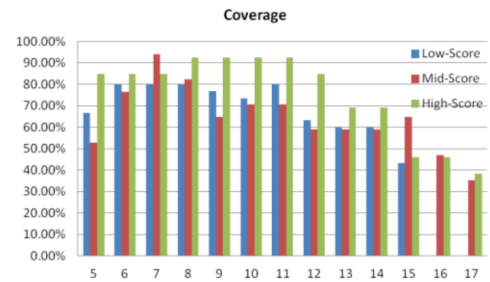

Figure 3. The coverage of characteristic sequence. The overall coverage reduced when the sequence length is lower than 7 or higher than 12.

or higher than 12, the overall coverage rates were reduced (Fig. 3). Thus, table II summarizes results of mining for sequence length equals to 9, 10 and 11 only. The coverage rates for high-score group are as high as 92.31% indicating subjects with high CAD score have very similar envelope pattern. On the other hand, subjects in low and mid-score group have much more diverse pattern.

B. *Characteristic Sequence*

The results indicate that for oscillometric envelopes with different shapes, the characteristic sequences reflect the difference in shape. For example, the oscillometric envelope for subject JC015, a subject in the high-score group, has a broad plateau and matches one of the high-score characteristic sequences: sequence 1122367643 (Fig. 4a). On the other hand, a bell-shape envelope of subject JC025, a low-score group subject, matches with the sequence 1112237222 in the low-score characteristic sequences (Fig. 4b).

On further examination, it is found that the characteristic sequences have multiple characteristic sequences that resemble each other and that all of these have the typical shape of an oscillometric envelopes (Fig. 5). This

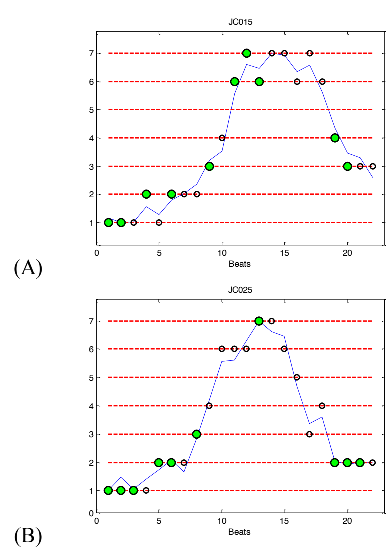

Figure 4. Typical results for characteristic sequences of (A) an oscillometric envelope with broad plateau and (B) an oscillometric envelope with a well-defined peak. Where the blue lines are the original oscillometric envelopes, the hollow circles are the results of envelope quantization and the green dots are the characteristic sequences.

TABLE II. THE MINING RESULTS

	k	*Low*	*Mid*	*High*
Number of Frequent Sequence	9	4873	4900	10486
	10	3678	3882	9252
	11	2183	2526	7082
Number of Non-repeated Frequent Sequence	9	1006	1019	5681
	10	1089	1282	6029
	11	1083	1272	5486
Maximum Support	9	21	11	12
	10	20	12	12
	11	19	12	12
Number of Characteristic Sequence	9	23	11	12
	10	22	12	12
	11	24	12	12
Coverage	9	76.67%	64.71%	92.31%
	10	73.33%	70.59%	92.31%
	11	80.0%	70.59%	92.31%

phenomenon demonstrates the effectiveness of the mining and screening processes, which determine the representative characteristic, regardless of the type of oscillometric envelope shape.

Additionally, because the characteristic sequences in each group are exclusive to that group, the coverage ratios can also be used for group identification. For example, when the oscillometric envelope of a new subject matches the one of characteristic sequences in the high-score group, this subject has, more than a 90% probability of being a high-score subject (Table 2).

IV. CONCLUSION

This study proposes a novel technique for the classification of the shape of an oscillometric envelope. Using the DISC strategy, the characteristic sequences are extracted and the shape of envelope is represented faithfully. The high coverage ratio indicates that these characteristic sequences are very common pattern in the particular group. The information for oscillometric envelope classification can be used to apply a specific algorithm for a particular envelope shape, in the NIBP measurement devices. This additional information can improve the accuracy of NIBP devices in the future. This information can also be used to distinguish subjects with different diseases.

REFERENCES

[1] N. M. Van Popele, W. J. S. Bos, N. A. M. De Beer, D. A. M. Van der Kuip, A. Hofman, D. E. Grobbee, and J. C. M. Witteman, "Arterial stiffness as underlying mechanism of disagreemen between an oscillometric blood pressure monitor and a sphygmomanometer." *Hypertension*, 36:484-488, 2000.

[2] M. Ursino, C. Cristalli, "A mathematical study of some biomechanical factors affecting the oscillometric blood pressure measurement," *IEEE Trans Biomed Engng*, 43; 761-778, 1996.

[3] J. N. Amoore, Y. Lemesre, I. C. Murray, S. Mieke, S. T. King, F. E. Smith and A. Murray, "Automatic blood pressure measurement: the oscillometric waveform shape is a potential contributor to differences between oswcillometric and auscultatory pressure measurements," *J. Hypertension*, 26: 35-43, 2008.

[4] D. Zheng, H. N. Amoore, S. Mieke, and A. Murray, "How impotant is the recommended slow cuff pressure deflation rate for blood pressure measurement?" *Amals of Biomed. Eng.*, 39:10, 2584-2591, 2005.

[5] Y.-C. Chen, L.-Y. Shyu, "Investigate the relation between coronary calcification and peripheral arterial compliance," in *Proc. Conf. Biomed. Eng. Technology*, Taiwan, 2008.

[6] D.-Y. Chiu, Y.-H. Wu, A. L. P. Chen, "Efficient frequent sequence mining by a dynamic strategy switching algorithm," *The VLDB J.*, 18:303-327, 2009.

[7] H. W.Daio, "The investigation of disease classification by analyzing the blood pressure waveform and waveform envelope," M.S. thesis, Dept. Biomed. Eng., Chung Yuan Christian Univ., Chung Li, Taiwan, 2012.

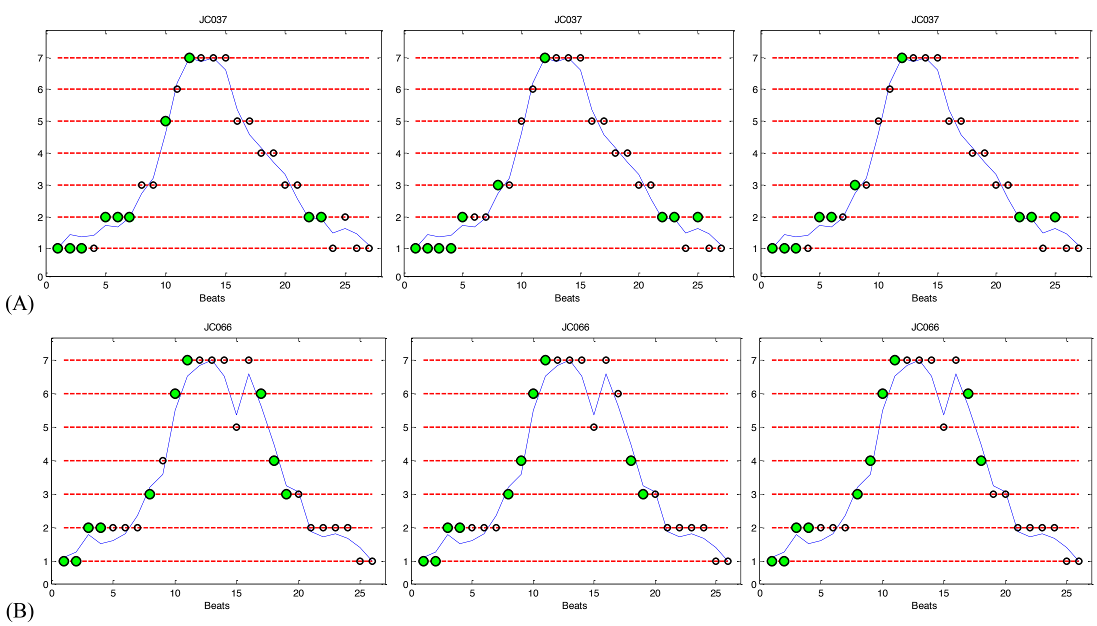

Figure 5. Multiple resemble characteristic sequences for different type of shape. (A) an oscillometric envelope with a well-defined peak and (B) an oscillometric envelope with broad and complex plateau. Where the blue lines are the original oscillometric envelopes, the hollow circles are the results of envelope quantization and the green dots are the characteristic sequences.