Fuzzy feature for Traditional Chinese Medical Pulse Data

Jian Cui
Department of Computer Science and Technology
Tsinghua University
cuijian.hh@gmail.com

Bin Xu
Department of Computer Science and Technology
Tsinghua University
xubin@tsinghua.edu.cn

Nanyue Wang
Chinese Academy of Chinese Medical Sciences
Experimental Research Center
wangnanyue1981@sohu.com

Guodong Sun
Department of Computer Science and Technology
Tsinghua University
sgdcs@tsinghua.edu.cn

ABSTRACT

This paper proposed a novel method based on Wavelet Transforms which can be easily used in processing the traditional Chinese pulse data in the context of mobile healthcare. Considering the energy limitation and real-time requirements, we make a new structure to describe the pulse data which we call fuzzy feature. The fuzzy feature can extract the hiding information from the pulse units. Pulse data preprocessing and fuzzy feature extraction only operates the wavelet transform coefficients of the original data. The algorithm complexity of the fuzzy feature extraction is about $O(N)$. Through analyzing the clusters from 28 patients' pulse units, the fuzzy feature can extract the hiding information well. The experimental results show that the fuzzy feature can be easily used in mining useful information from patients' data and assisting doctors to make accurate diagnosis.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Theory

Keywords

pulse data processing, Chinese medical diagnosis, Wavelet transform, mobile healthcare

1. INTRODUCTION

With the development of wireless communication and medical sensing technologies, mobile healthcare systems are attracting more and more attention in recent years. These systems can provide personalized, professional healthcare services to the patients, who are even at homes or on workplaces.

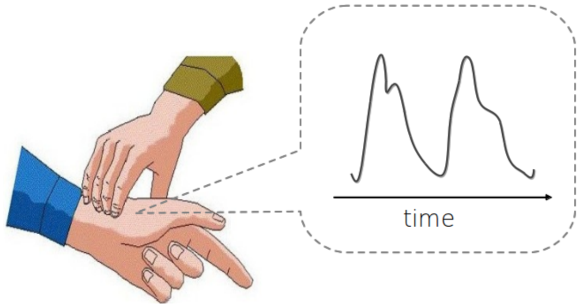

Figure 1: The illustration of traditional Chinese pulse diagnosis

The traditional Chinese pulse diagnosis has been successfully used for clinics in China for more than 2,000 years, and now, is gradually becoming popular in the world. As shown in Fig.1, the traditional Chinese physicians use three fingertips to feel the radial artery of patient at three positions on wrist. Technically, the wrist pulse is a kind of signals, from which a lot of physiological and pathological status of patient can be deduced according to traditional Chinese medicine theories [3, 5, 7].

In the context of mobile healthcare, the traditional Chinese pulse diagnosis is attracting global attention because of its accuracy, convenience, and non-invasiveness. However, there exists two significant challenges in capturing and understanding the wrist-pulse data when integrating the pulse diagnosis into the mobile healthcare system. Firstly, because physicians (or the medical software) need a contiguous feeling (sampling) of the pulse to understand the pulse accurately. The intensive sampling(often 100 $\sim$ 1000Hz) of pulse makes it very hard to transmit samples via wireless channel in a real-time manner.

Secondly, it is difficult for medical software to efficiently understand the pulse data because of the hardness in identifying critical pulse samples accurately and efficiently. To automatically perform pulse diagnosis in mobile health system, therefore, it is very important to feed to the medical software the pulse wave data with appropriate abstraction and granularity.

To address the above two problems that pulse diagnosis experiences in the context of mobile healthcare, we propose a feature description method called fuzzy feature, to process the pulse data such that it can be easily transmitted and understood by the medical software. The primary contributions of our work are: 1)The data processing algorithms

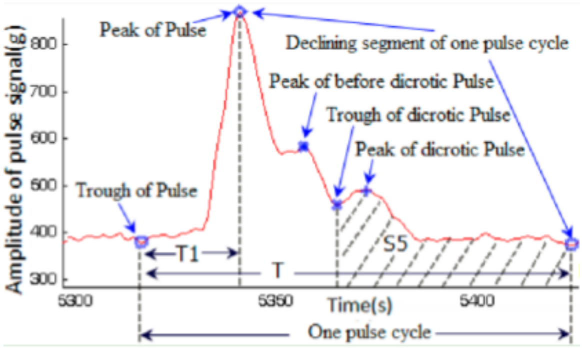

Figure 2: Five most important physiological feature points of human pulse wave data

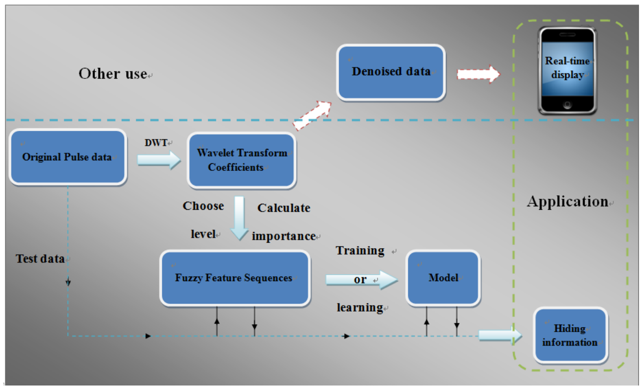

Figure 3: Processing steps

based on the Wavelet Transform are liner complexity, and can be deployed on mobile devices and fit the real-time requirements. 2)The fuzzy feature can reflect the hiding information from the pulse unit which can not be extracted by the traditional methods.

The rest of this paper is organized as follows. Section 2 briefly introduces some related work. Section 3 presents pulse data processing detailedly. In section 4 we conduct some preliminary experiments to evaluate our design. Section 5 concludes our paper.

2. RELATED WORK

The pulse wave data are regularly changed pressure signals. The most important time-domain features of the basic pulse wave unit [4] are shown in Fig.2.

From left to right in the chart above, they are named as point of pulse (T-point), peak point of pulse (P-point), peak point of pulse before dicrotic pulse (PB-point), trough point of dicrotic pulse (TD-point) and peak point of dicrotic pulse (PD-point). PDT [6] said that these features are the inflection points of pulse pressure curves. They are also the transition points from one mechanical process to another in one cardiac cycle. All the feature points have clear physiological meanings, and naturally may reveal much information about human health condition.

Similar with ECG data processing, feature extraction is the first step of the task. The large number of patterns, complexity in waveform, noise and interference in ECG processing also exists. In addition, pulse waveforms have its own mathematics morphology that different from ECG. FEA (feature extraction algorithm) in [4] uses the threshold and slope methods to extract the features. T-point and the P- point, as shown in Fig. 2, can be selected as beacon feature points. Then, FEA generates all the candidate feature points of the PB-point, the TD-Point and the PD-point by analyzing slope variability. It formalizes the slope variability as a trend decision function. Then, the final desirable feature points can be selected from the candidate ones according to some adaptive parameters.

3. PULSE DATA PROCESSING STEPS

The original pulse data flow is a series of voltage signals as the time goes on. It can be seen from Fig.2. The capacity data set may be very large because of the integrity of information. The sample frequency of the pulse diagnosis instrument is required to reach 1000Hz. In traditional Chinese medicine theory, the pulse state reflects at three positions on the wrist. These positions are called inch, bar and cubit, respectively. It is just like the ECG 12-lead, data can reflect the all sided information of the heart state. And the pulse data of both hands should be collected.

In this paper, we analyze the pulse data based on Wavelet Transforms. Pulse data analyzing can be divided into three main steps. First, preprocessing; second, feature extraction; third, pattern training or learning. These steps are shown in Fig.3.

3.1 Wavelet Transform of the Pulse Data

The whole analyzing framework is built based on the Wavelet Transform of the original data. The follow work are all conducted on the wavelet coefficients. We use the discrete wavelet transformation(DWT) to transform the pulse raw data.

The reason for using DWT is, it has fast algorithm and can be easily used in application and real-time scenario. The simple and effective method used to choose the wavelet filter is: the shape of Wavelet function should be similar with the original data shape.

The decomposition level in this paper is 9. And we get that, the most important coefficients for different people's pulse data is in the same level. This will be discussed in section .3.3.

3.2 Pulse data processing

The original pulse data from the pulse diagnosis instruments contain many noises. As proposed in Introduction, processing work should be done on limited energy budget.

In this paper, we use VisuShrink [2] to denoise the data. It uses the global threshold to process the wavelet coefficients. There are many other methods that can get better results. However, the denoising in this paper is not the critical part. We can see in next part that the fuzzy feature can be used without denoising. What we want to say is we can produce the denoised data in this framework if necessary. For example, many mobile devices in mobile healthcare monitoring should provide the displaying of the pulse data in real-time. And the computation complexity of VisuShrink is $O(N)$. The Online display for users can be easily done in this denoising strategy.

3.3 Fuzzy feature extraction

Feature extraction is a common and important step in sensor data processing. Many work have been done in extracting the morphology features in Fig.2. They are either time or frequency domain features. However, what really need are the time-frequency features. And these existed feature extraction work are all focused on how to extract them exactly. Then these features are taken as input for automatic diagnose models. However, this idea is not appropriate for pulse data analyzing in mobile healthcare. The problems are: 1)These morphology feature extraction methods need scan raw data more then one time. And they can hardly fit the real-time requirements. At the same time, the high computation complexity will result serious energy consumption. 2)These methods always heavily depend on the data preprocessing. If data contain more noises or are denoised not enough, the feature extraction will get bad results. 3)Most importantly, these exact feature extraction methods always separate the time and frequency domain features apart. In other words, these features of pulse data are many "name:value" pairs without sequence dependency. However, these features in each pulse unit must have some sequence relation. The traditional feature extraction will lose these sequence information when extracting. 4)Pulse features for different people have different scales. For example: the age, gender and even different phases. This make the feature extraction and model construction more difficult.

Considering above challenges, we proposed the fuzzy feature that can overcome these problems. The fuzzy feature has some key differences with the traditional morphology features: 1)Fuzzy features do not have the definite domain meanings. It means that fuzzy feature can not match any Traditional Chinese medicine metrics. 2)Fuzzy features of one pulse unit are a sequence of features. In this way, the features' dependency can be obtained in the fuzzy feature.

In this paper, original pulse data are transformed into low frequency and high frequency wavelet coefficients. We directly take some high frequency coefficients as the fuzzy feature candidates. Because high frequency coefficients describe the local character of the data, they can be used to distinguish different pulse units.

The algorithm of extracting fuzzy feature is based on one assumption: Important data characters have large values in each high frequency level; on the other hand, characters are important if large coefficients in this level and it's large in next level [1]. In this paper, we look for important coefficients among levels, and take them as the fuzzy feature. Then, the original pulse units are described as a sequence of fuzzy feature.

3.4 Model training

There must be a model to assist diagnosing the pulse data. The model can be an clssifier or some other models that can give an intuitive result for the pulse data.

The data format is the simple 1-D data. There are many algorithms can be used to train the model. In this paper, we use the improved k-means algorithm to cluster the pulse units. And find the hiding information behind the clusters. We add weights for the fuzzy features selected from the wavelet high frequency coefficients. The fuzzy feature sequence FF_i of each pulse unit can be described as Eq.1

$$FF_i = \{ff^1{}_i, ff^2{}_i, ..., ff^k{}_i\} \tag{1}$$

where FF_i is composed of k fuzzy features.

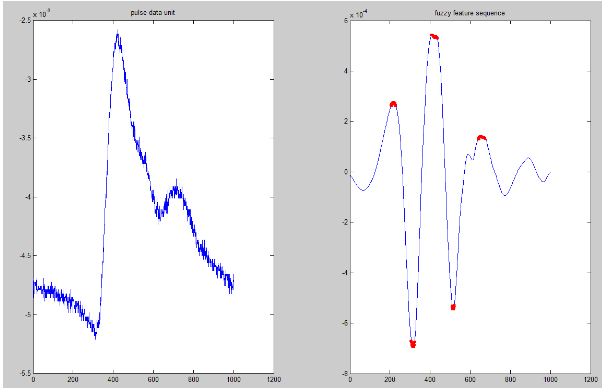

Figure 4: Fuzzy feature and original data

Table 1: One fuzzy feature instance

ff^1	ff^2	ff^3	ff^4	ff^5	position	patientID
0.000250	-0.000367	0.000294	-0.000254	0.000136	1	CFGaD06

4. EVALUATION

We use 28 Cirrhosis patients pulse data to evaluate our work. The sample frequency is 1000Hz. And these data have been labeled by the people's personal information and their symptoms. Using the fuzzy feature to describe the pulse unit and get the hiding information behind the clusters.

4.1 Extract fuzzy feature from pulse data

It has been discussed that we use DWT(Discrete Wavelet Transform) to transform the original pulse data into the wavelet approximation and detail coefficients. The decomposition level used in this paper is 9. Indeed, we only use 6 levels' coefficients. However, we evaluate all levels in this paper for future improvements. We find that level 6 high frequency coefficients can reflect the whole duty character for each pulse unit. In this paper, we take the five obvious peaks separated by some distances as the fuzzy features. The pulse data and its fuzzy feature sequences are shown in Fig.4.

Then, the fuzzy feature sequences are extracted from the original pulse data units. The training set is made up of the fuzzy features sequences for each pulse unit labeled by the patient ID. Each sequence is an instance for the model. For example, the format of the instance is shown in Table.1.

To test the algorithm performance in mobile environments. We make an experiment to run the fuzzy feature extraction algorithm on HTC G8 to test the consumption on the mobile device. We calculate the time consuming of running different data size on different decomposition level. Fig.5 shows the result. It mainly trends on the liner complexity $O(N)$.

4.2 Training Clusters

Now, we have got the fuzzy features training set. Put all the patients instances together into the k-means algorithms. Here we set $k = 5$ to get five clusters. Because the instances are labeled by the patientID, we can find the related information of the patient in the same cluster. Fig.6 shows the results of the clusters.

We can see from the table that each cluster contains some patients' instances. The ground truth is that the instances are labeled by the symptoms. Many patients are caught by

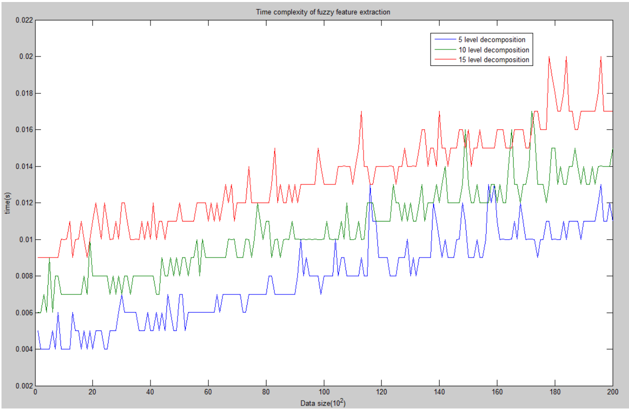

Figure 5: Time consuming on HTC G8

cluster(k)	1	2	3	4	5
patients (ID)	12, 14	4, 7, 9, 11, 15, 17, 18, 23, 25, 27	10, 14, 20, 30, 40	6, 11, 17, 20	10, 14, 20, 33, 40
common symptom	tongue nature is dark	Arrhythmia; bad sleep	have problems in gallbladder	tongue nature is yellowing	being tired and lazy
gender (m:f)	1:1	5:5	3:2	2:2	1:1
age (year)	50, 74	40~50	30~70	40~70	40~50,70+

Figure 6: Fuzzy feature and original data

many different symptoms. From the point of the Chinese Medical theory, these symptoms should have some reflections on the pulse data. In this paper, we take the pulse unit as the data object. And the cluster contains different patients' fuzzy feature of the pulse unit. We think fuzzy feature sequences in the same cluster must have some common or similar attributes. And we found that patients in the same cluster have some other common symptoms besides the cirrhosis in this experiments. Take Cluster 2 as an example, patient 4,7,9,11,15,17,18,23,25 and 27 all have the symptom of arrhythmia and suffers the bad sleep experiences. We can conclude that the fuzzy feature sequence can reflect the hiding information of the pulse unit. Fig.6 shows all the results about the hiding information for each cluster.

Cluster contains different patients' fuzzy feature sequences. However, different patients' data size is not same. We define the contribution of the patient to cluster in Eq.2.

$$contribution(p) = clusterCount(p)/dataCount(p) \qquad (2)$$

$clusterCount(p)$ is the instance number of p contained in this cluster, $dataCount(p)$ is the total instance number of p.

The *contribution* reflects the importance of the patient to this cluster. If $contribution(p)$ is low, patient would not have the common characters of other patients in this cluster. Fig.7 show the contribution of each patient in cluster. We can see that most of the *contribution* in the same cluster are over 50%. This means patients' contributions are very important for this cluster.

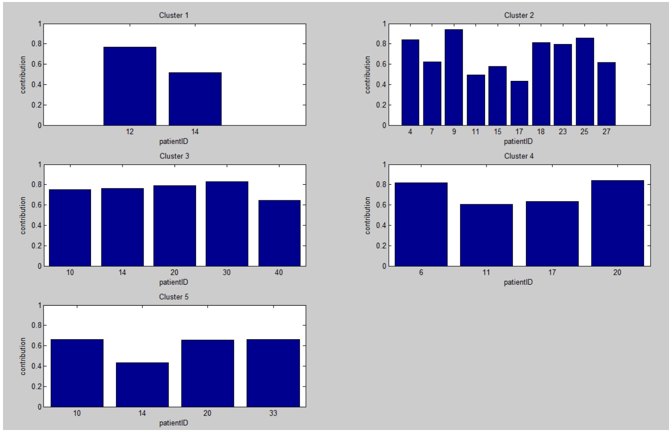

Figure 7: Fuzzy feature and original data

5. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a novel method based on Wavelet Transform to analyze pulse data. We mainly focus on the pulse data management in mHealth environment. The whole process only works on the Wavelet Transform coefficients. The whole process only scans the data once to finish the denoising and feature extraction. And the computing and storage consumption is low by using only the Discrete Wavelet Transform. The experimental results on mobile device show the time complexity is $O(N)$. It can be easily used in mobile mHealth environment where is limited with energy and real-time requirements. The experiment shows the fuzzy feature proposed in this paper can reflect the hiding information well of the pulse units for the cirrhosis patients.

6. ACKNOWLEDGMENT

This work was supported in part by NSF of China, with grant No.61170212.

7. REFERENCES

[1] Lewis A and KnowledgeG. Image compression using the 2-d wavelet transform. *IEEE Transactions on Image Processing*, 1:244–250, 1992.

[2] David L. Donoho. De-noising by soft-thresholding. *Information Theory, IEEE Transactions*, 41:613–627, 1995.

[3] Z. Fei. *Research on Chinese pulse theory*. Shanghai University of Traditional Chinese Medicine Press, 2005.

[4] J. Gong, S. Lu, and R. Wang. Pdhms: Pulse diagnosis via wearable healthcare sensor networks. In *ICC*, 2011.

[5] S. Lukman, Y. He, and S. Hui. Computational methods for traditional chinese medicine: a survey. *Computer methods and Programs in Biomedicine*, pages 283–293, 2007.

[6] L. Xu, K. Wang, D. Zhang, Y. Li, Z. Wan, and J. Wang. Objectifying researches on traditional chinese pulse diagnosis. *Informatica medica Slovenica*, 8(1), 2003.

[7] D. Zhang, W. Zuo, and D. Zhang. Wrist blood flow signal-based computerized pulse diagnosis using spatial and spectrum features. *Biomedical Science and Engineering*, 3:361–366, 2010.