The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence
Expanding the Boundaries of Health Informatics Using AI:
Technical Report WS-16-08

A Novel Method for Mining Semantics from Patterns over ECG Data

Zhen Qiu, Feifei Li, Shenda Hong, Hongyan Li*
School of Electronics Engineering and Computer Science, Peking University, Beijing, China
{qiuzhen, liff, hongshenda, lihy}@cis.pku.edu.cn

Abstract

In intensive care units (ICU), electrocardiogram (ECG) waveforms show diverse variations under different patients' physical conditions. In general, physicians can diagnose patients efficiently by detecting any disorder of heart rate or rhythm and any change in the morphological pattern of ECG data, which contain underlying semantics. To help physicians better analyze ECG data in a fairly short time, it is essential to develop a novel method for mining semantics from ECG patterns. This paper is the very first time to characterize ECG patterns by using Prefix Scalable Pattern Tree (PSP-Tree). Comparing with similar currently existing methods, PSP-Tree can mine significant semantics, such as scalability, temporality and hierarchy over ECG patterns. We conduct extensive experiments on real ECG data set which are obtained from PhysioBank Community and Beijing No.3 People Hospital. The experiment results show that our method performs more feasibly and effectively than other related work.

Introduction

One of the most active hotspots in biomedical informatics is emergency medicine. It is critical that patients in intensive care units (ICU) can get timely diagnosis and treatment according to the massive medical data collected by the advanced medical technologies. Among all recorded medical data in ICU, ECG waveform data, which characters the heart electrical depolarization and repolarization patterns, have attracted a significant amount of attention (Eom, Kim, and Zhang 2008; Srinivas, Rani, and Govrdhan 2010).

ECG waveforms can be used by physicians to diagnose the patient intuitively and accurately. As time elapse, ECG waveforms show significant variations under different patients' physical conditions. Different ECG waveform patterns characterized by diverse morphological and temporal features can be sufficiently complicated. Any change in ECG waveform may sign underlying pathology (Goldberger 2012; Banerjee and Mitra 2014), such as missing or repeating waveforms may indicate some pathology symptoms (Kim et al. 2007; Chinnasami, Rathore, and Duncan 2013). It's important for physicians detecting the change in the morphological and temporal pattern. For this, mining useful semantics over ECG pattern takes the prominent position. However, the current high-throughput measurement systems-produced massive data make it difficult for physicians to parse through the so much information for timely diagnoses (Belle, Kon, and Najarian 2013). Therefore, data mining methods play important role for providing computer-aided solutions for ECG semantics mining for timely diagnosis.

Currently, the related studies can be categorized into frequent pattern mining and specific pattern mining.

Firstly, frequent pattern mining methods aim to identify patterns that frequently occur in the ECG waveforms. (Porta et al. 2001) define frequent deterministic pattern in short heart period variability series, then group all possible patterns in four categories characterized by different frequency contents. Besides, (Noh et al. 2006) propose the CAD-AC, an associative classifier, to build up a classification technique by using Frequent Pattern Growth (FP-Growth). CAD-AC operates on ECG patterns and clinical investigation of Heart Rate Variability (HRV), can automatically diagnose Coronary Artery Disease. However, apart from the frequent occurred pattern which can be detected by frequent pattern mining methods, there are some patterns which appear infrequently may even contain significant semantics in medical domain, such as ventricular fibrillation, which loses some waves, can cause sudden death (Othman and Safri 2012). It is essential for researchers to provide new method for mining the loses in semantics for physicians to discover the significant infrequent ECG pattern.

Secondly, specific pattern mining methods are usually developed to extract features of ECG waveform for representing ECG classification. The most common work is QRS detection, which is necessary to determine the heart rate and can be used as the reference point for beat alignment (Mitra, Mitra, and Chaudhuri 2009; Chatterjee, Gupta, and Mitra 2011). Besides, (Ghaffari et al. 2010) use two innovative modified Hilbert transform-based algorithms to extracted QRS complexes and end-systolic end-diastolic pulses for detecting acute hypotensive episodes and mean arterial pressure dropping regimes. However, specific pattern mining can only give effective actions for classification. In fact, some combination of the specific patterns can also express significant semantics in medicine domain beyond classification. For example, if QRS complex repeat many times indicates

*Corresponding author

ventricular tachycardia (Simson 1981). Therefore, developing new mining method, which can not only be used for classification but can also be used for medical diagnose, has become a more and more urgently solved problem.

In order to make up for the inadequacy of previous mining method, we systematically analyze the ECG patterns, characterization of ECG pattern by using Prefix Scalable Pattern Tree (PSP-Tree). It is proposed for the first time in the literature for mining semantics from ECG patterns, which can find out the majority of patterns and related information stored in the complex medical ECG waveforms.

Methods

Observation

As shown in Figure 1(a), a single normal cycle of the ECG represents the successive arterial depolarization and ventricular repolarization, and can be approximately associated with the peaks and other ECG waveforms, which labeled P, Q, R, S, T, U (McSharry et al. 2003). Recently, different characteristic segments, such as PQ, QRS, ST segments, are also used for diagnostics (Chiang et al. 2014). Our method begins with a brief observation on ECG patterns to summarize some useful semantics based on pathophysiological mechanisms.

Scalability: We define the pattern formed by gain and loss of basis units as scalable pattern. For example, the waveforms in Figure 1(b) and Figure 1(c) are so different from the normal ones. However, they are consist of the waves which are basically unchanged over the ECG cycles. QRS complex, and T wave are gained in both Figure 1(b) and Figure 1(c) as the same form. But U wave is loss in Figure 1(b) and P as well as U wave is loss in Figure 1(c). The loss and gain of the basic units represent scalability of patterns.

Temporality: The scalable ECG pattern follows linear temporal logic as shown in Figure 1(a). The sequence of basic waves can indicate the mechanism of heart rate, such as P wave indicates the start of a new cardiac cycle. Once the linear temporal logic turns to be inverted, a new cycle begins (Shibahara 1985).

Hierarchy: Further, some patterns are formed by gain and loss of scalable pattern as a whole, which represents hierarchy of scalable patterns. As shown in Figure 1(b), T wave repeats many times. In the medical domain, this phenomenon can be considered as myocardial infarction (Thygesen, Alpert, and White 2007). QRS complex repeat many times in Figure 1(d), which reflect that patient may have experienced episodes of sustained ventricular tachycardia (Simson 1981). Therefore, the hierarchy of scalable patterns has significant semantics in medicine domain.

Overall approach

Our proposed framework for mining semantics over ECG patterns can be describe as follow, which is illustrated in Figure2. First, with the help of our previous work (Li et al. 2010), the base pattern matcher is applied to generate the pattern stream based on the ECG original data. Then, we present a data structure called PSP-Tree, which is used to characterize ECG patterns. After that we illustrate the method of mining semantics from ECG patterns with PSP-Tree. Finally, the mining semantics will be returned to the physicians.

Preliminary Definition

In order to introduce how the scalable pattern is formalized, we take the pattern in Figure 1(b) for example: T pattern repeats, U pattern loses, the scalable pattern can be expressed as below:

$$sp_1 = P_1^1 Q_2^1 R_3^1 S_4^1 T_5^3 U_6^0 = P_1 Q_2 R_3 S_4 T_5^3$$

where the superscript represents the number of the pattern appear times, if the pattern appears one time, the superscript can be omitted, and if the pattern disappears, neither the letter of wave pattern nor the superscript is shown (e.g. U wave lost and can be leave out). The subscript represents the pattern appear order in the normal ECG temporal logic. If the whole scalable pattern sp_1 repeated 5 times, the more complex phenomenon can be expressed by a higher-hierarchy scalable pattern which can be expressed as below:

$$sp_2 = (sp_1)^5 = (P_1 Q_2 R_3 S_4 T_5^4)^5$$

Prefix Scalable Pattern Tree

In order to express semantics of scalable pattern over ECG data, we present Prefix Scalable Pattern Tree (PSP-Tree) which can organize the scalable patterns.

Definition 1. *Let PSP-Tree stands for the prefix scalable pattern tree, it can be formalized as follow,*

$$PSP - Tree = ((N_{root}, N_{inner}, N_{leaf}), E)$$

where N_{root} is the root node, N_{inner} is the internal node, N_{leaf} is a complex leaf node, E stands for the edges that links the nodes.

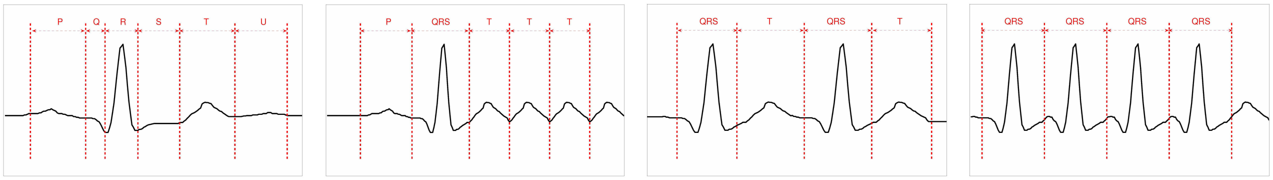
The information in N_{inner} N_{leaf} supports the semantics mentioned in the observation section, which is explained in detail as follow.

(1) **Root node**: the initial node with no actual information;

(2) **Internal node**: it represents a base pattern. The appearance or disappearance of base pattern in the inter node represents the semantics of gain or loss, namely scalability, all the nodes in the paths comply with the temporality of ECG;

(3) **Leaf node**: it indicates the ending of a scalable pattern and store abundant information, which contains Base Pattern Table (BPT) and Pattern Frequency and Navigation Linked List (PFNLL). The numbers in the BPT represents the times of scalable patterns, which explain the semantics of hierarchy.

- **Base Pattern Table (BPT)**: each line is corresponding to a path from root node to the last internal node, and records the appearance of each base pattern. Then the path with the appearance record will form a scalable pattern. That is to say, the base pattern in the path can form a benchmark with temporal logic, and the BPT will provide the more detailed appearance information;



(a) Normal ECG waveform (b) Myocardial infarction (c) Premature systole (d) Ventricular tachycardia

Figure 1: Normal ECG pattern and some ECG patterns implying symptoms

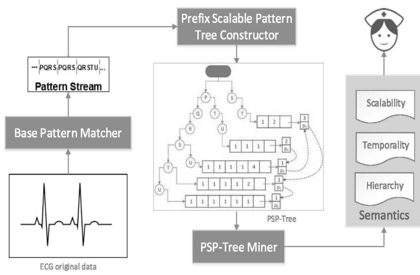


Figure 2: Framework for mining semantics over ECG data

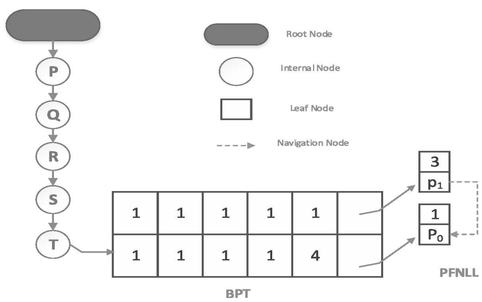


Figure 3: The illustration of PSP-Tree data structure

- **Pattern Frequency and Navigation Linked List (PFNLL)**: each linked list entry represents the continuous appearance times of the scalable pattern with the corresponding line. The entry will store the continuous appearance times, as well as the navigation pointer, which will records the location or index of the former scalable pattern.

As shown in Figure 3, one pattern of "P, Q, R, S, T" appears in the corresponding internal node. In the BPT, the former five columns represent the number of the base pattern appears times along with the temporal logic. Each row stands for a scalable pattern, the first row is sp_1, which can be formalized by $P_1Q_2R_3S_4T_5$, the second row stands for sp_2, which is $P_1Q_2R_3S_4T_5^4$. The last column of BPT points to PFNLL, which contains the appearance number of the corresponding scalable pattern and navigation information. For example, the linked list in the first row shows that the scalable pattern sp_1 appears for 3 times repeatedly, then the navigation pointer points to the entry in the second row, which means the scalable pattern $(P_1Q_2R_3S_4T_5)^3$ follows $P_1Q_2R_3S_4T_5^4$.

Table 1: A Scene with Base Pattern Stream to Construct PSP-Tree

Order	Scalable Pattern	Appearance Times
1	$P_1Q_2R_3S_4T_5U_6$	1
2	$P_1Q_2R_3S_4T_5$	1
3	$S_4T_5^2$	3
4	$P_1Q_2R_3S_4U_6^4$	1
5	$P_1T_5U_6$	2

Mining Semantics from Patterns with PSP-Tree

In this subsection, we will introduce how to characterize ECG patterns and mine semantics with PSP-Tree. At the beginning, PSP-Tree is empty. When a pattern stream comes, N_{root} will be created, which just represents the beginning of the scalable pattern. Then an inner node will be added into PSP-Tree with value of the corresponding matched basic pattern . Let p_i be the current basic pattern and n_i be the current added node. If the order of the previous basic pattern p_{i-1} and p_i do not meet the temporal logic, n_i will be added as the roots child, if the root does not have a child with this value of p_i. Or else, it will be added as the child of n_{i-1}. The times of each basic pattern appears from root to leaf will be added into the BPT. In addition, if the current scalable pattern is the same as the previous one, the times in its corresponding PFNLL will increase by 1, or else a new one will be added to PFNLL by the value of 1.

We take a ECG waveform example, after transforming it into a pattern stream (Li et al. 2010), we get a pattern stream which is "$PQRSTUPQRSTSTTSTTSTTPQRSUUUUPTU PTU$". Table 1 characterize the scalable patterns of the ECG waveform example. We can find that the pattern $P_1Q_2R_3S_4T_5U_6$ appears once firstly, then pattern $P_1Q_2R_3S_4T_5$ arrives. Next, pattern $S_4T_5^2$ appears 3 times repeatedly, which is followed by the scalable pattern $P_1Q_2R_3S_4U_6^4$. The last scalable pattern of this ECG example slice is $P_1T_5U_6$ which repeats twice. With the help of above semantics information, we will get the PSP-Tree which is shown in Figure 4. Algorithm 1 shows the detailed process of building PSP-Tree.

Then we introduce how to use the PSP-Tree to mine semantics from ECG patterns. By the more in-depth observing the PSP-Tree data structure, we can find that the majority of the semantics information is stored in the complex leaf nodes. Algorithm 2 shows the detailed process of mining se-

Algorithm 1 PSP-Tree Builder

Input: pattern (P) ;
$count\_temp \leftarrow 1$
if root doesn't exit **then**
    create root;
end if
$CurNode \leftarrow root$;
for each $i \in [0, P\_length - 1]$ **do**
    if $i! = P\_length - 1$ and $P[i] = P[i+1]$ **then**
        $count\_temp + +$;
        continue;
    end if
end for
AddCounts($count\_temp$);
$count\_temp \leftarrow 1$;
$Children \leftarrow$ getAllChildren($CurNode$);
if $!P[i] \in$ Children **then**
    $temp$ = CreateNode($P[i]$);
    AddChildren($CurNode, temp$);
    $CurNode = temp$;
else
    $CurNode = Children[j]$;
end if
if $CurNode$ has no leaf **then**
    CreateLeaf($CurNode$);
    Add P into leaf;
else
    for each $k \in [0, P\_length]$ **do**
        if $k! = P\_length$ **then**
            if $patterns[k] equals P$ **then**
                Add show times to $patterns[k]$;
            end if
        else
            Add P into leaf;
        end if
    end for
end if
Output: PSPTree;

mantics with PSP-Tree.

- **Scalability Mining**: According to the each path from root node to leaf nodes and the base pattern dictionary, we can easily find the gain and loss information of a scalable pattern. As shown in Figure 4, the shortest path represent scalable pattern $S_4T_5^2$, the corresponding BPT means all the base patterns in "P, Q, R, U" are lost and pattern T gains twice.
- **Temporality Mining**: Each path of the PSP-Tree shows the appearance order of base pattern, the longest path in the Figure 4 shows the ECG temporal logic which is obeyed by the other pathes.
- **Hierarchy Mining**: According to the definition of scalable pattern which is mentioned in the previous preliminary definition section, the existing scalable patterns can be regarded as basic units to generate a new scalable pattern, the related information is recorded in PFNLL.

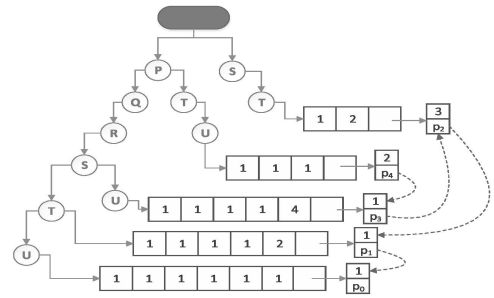


Figure 4: The PSP-Tree of ECG pattern example

As shown in Figure 4, the corresponding PFNLL shows pattern $S_4T_5^2$ appears 3 times repeatedly, which can be means that a whole hierarchical scalable pattern gains 3 times. For the more complex situation, the navigation pointer of scalable pattern $S_4T_5^2$ points to scalable pattern $P_1Q_2R_3S_4T_5^2$, we can find that the suffix of the prior scalable pattern and the prefix of the later one are the same, the navigation pointer links and combines them together to form a more complex scalable pattern, here the scalable pattern $S_4T_5^2$ can be regarded as a basic unit. The new generated scalable pattern can be expressed as $P_1Q_2R_3(S_4T_5^2)^4$.

Algorithm 2 SemanticsMining

Input: PSPTree (T), PSPTree node (n), Base Pattern Dictionary (D);
if n is a leaf **then**
    for each $row \in [0, BPT.row.size - 1]$ **do**
        for each $col \in [0, BPT.row[i].length - 1]$ **do**
            while true **do**
                $j \leftarrow 0$;
                if $P[col]! = D[i]$ **then**
                    $pset \leftarrow (D[j], 0)$
                    $j + +$;
                else
                    $set \leftarrow (D[j], BPT[row][col])$;
                    break;
                end if
            end while
        end for
    end for
end if
Output: set of scalable patterns with semantics set;

Experiment

In this section, we present our experiments to evaluate the flexibility and effectiveness of the proposed method, followed by some discussion about semantics results of extracted ECG patterns in comparison with some existing techniques.

Data set

We used real datasets instead of simulated data, which are obtained from PhysioBank Community (Goldberger et al. 2000) and Beijing No.3 People Hospital. The detail information of experiment data is listed as follow.

- **MIT-BIH normal sinus rhythm database:** We mark this database as *mnsrd*, which includes 18 long-term ECG recordings of subjects referred to the Arrhythmia Laboratory at Boston's Beth Israel Hospital. The 18 long-term ECG recordings were obtained from 5 men, aged 26 to 45, and 13 women, aged 20 to 50. No significant arrhythmias were found in the subjects included in this database.
- **Creighton University Ventricular Tachyarrhythmia Database:** We mark this database as *cuvtd*, which includes 35 eight-minute ECG recordings of human subjects who experienced episodes of sustained ventricular tachycardia. The ECG waveform form of ventricular tachycardia is the repeatability of QRS complex.
- **The MIT-BIH Malignant Ventricular Arrhythmia Database:** We mark this database as *mvad*, which includes 22 half-hour ECG recordings of subjects who experienced episodes of sustained ventricular fibrillation. The ECG waveform of ventricular fibrillation loses P wave, QRS complex and T wave.
- **MIT-BIH Arrhythmia Database Directory:** We mark this database as *madd*, which is the set of over 4000 long-term Holter recordings which obtained from inpatients who have premature beats, their ECG waveform lose P wave.
- **ECG from Beijing No.3 People Hospital:** We mark this database as *bj3*, which is recorded during a six hour period simultaneously from a pediatric patient with traumatic brain injury in ICU. The sample rates of the signals are from 125 Hz to 500 Hz, varying according to the states of illness. The whole dataset includes over 25,000,000 data points.

Among the above data sets, *nsrd* database is applied to generate pattern matcher with our previous work (Li et al. 2010), due to that it contains ECG waveform data of normal human beings. With the help of pattern matcher, we can transfer the rest of ECG database into pattern stream for processing.

Experiment results

This section presents the evaluation results in details. In our framework, the PSP-Tree is builded to mine the ECG patterns with semantics, which is few focused on by the most of the methods mentioned in the introduction section. We will compare our method with the classic FP-Growth (Noh et al. 2006), the most relevant method for evaluating the mining patterns. In addition, we have some discussion about the interesting findings of patterns mined by our method. The PSP-Tree's ability in supporting the semantics of ECG patterns is also discussed in this part.

By execution of mining methods based on PSP-Tree and FP-Growth respectively on ECG datasets, we compared

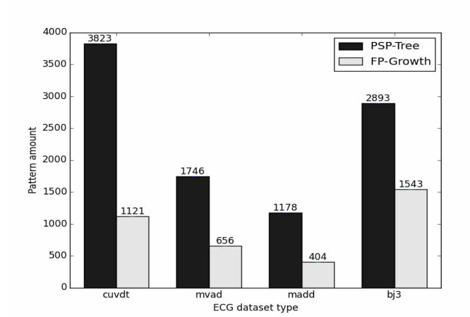


Figure 5: The comparison of mining patterns amount

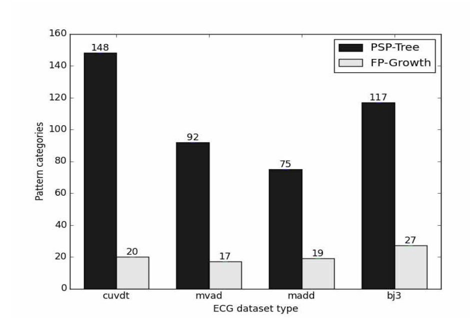


Figure 6: The comparison of mining patterns categories

them in terms of the category and amount of patterns mined out. By conducting experiments on different kinds of ECG data set, we got the results as shown in Figure 5 and Figure 6. From the results above, the method based on PSP-Tree can mine out more categories and a larger amount of patterns from ECG data sets. The main reason is that the method based on FP-Growth does not support the scalability semantics of basic patterns, it only mine patterns which appear frequently, thus it is impossible to mine out the scalable patterns which has gain and loss of basic patterns. In addition, PSP-Tree supports mining out scalable patterns which occur continuously along with the temporal logic of ECG waveform, while the other one does not have this ability.

For further performance evaluation, we use the following important measurements.

- **Sensitivity (True Positive Rate):** The probability that the algorithm can find meaningful patterns over ECG waveform;
- **Selectivity (True Negative Rate):** The probability that the algorithm does not find false patterns.

The two measurements are conflict in a sense that increasing sensitivity to find more meaningful patterns will inevitably finding more false patterns. In an ICU environment, sensitivity is much more important than selectivity as missing a meaningful pattern may cost the patient's life. There are already tags in *cwdt*, *mvad*, *madd* data set, which can represent the meaningful patterns. Besides, we ask three professional doctors to tag out meaningful patterns for *bj3* data set. The

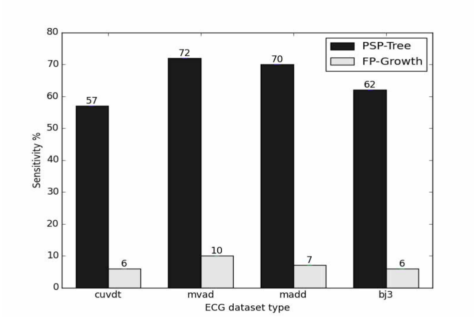


Figure 7: The sensitivity comparison of mining patterns

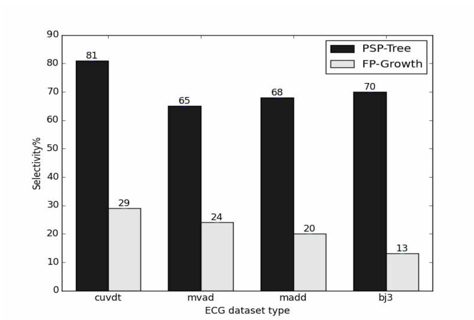


Figure 8: The selectivity comparison of mining patterns

sensitivity is defined as follows:

$$Sensitivity = \frac{\#p_{mp}}{\#p_{tp}} \quad (1)$$

$$Selectivity = \frac{\#p_{mp}}{\#p_{p}} \quad (2)$$

where p_{mp} denotes the patterns mined by methods which are also in the set of tagged meaningful patterns, p_{tp} denotes the set of tagged meaningful patterns, p_p denotes the patterns mined by methods.

Figure 7 and Figure 8 shows the performances of the two algorithms on finding meaningful patterns, along with the numbers of the false patterns. The result indicates that PSP-Tree has the better performance on sensitivity than FP-Growth, the former discovers more meaningful patterns and misses more false patterns. This behavior is caused by the limitation of the FP-Growth method's characteristic. Just as mentioned above, FP-Growth can only mines patterns which appears frequently, however, PSP-Tree can mine more patterns which is formed by the gain and loss of basis patterns.

Further more, we have some interesting findings about the meaningful patterns. We conduct an experiment on *bj3* data set in consideration of its continuity. As the volume of data increases, the changing trends of sensitivity mined by PSP-Tree and FP-Growth can be shown in Figure 9. We can find that sensitivity of FP-Growth decreases when more ECG data comes, which can be also explained by the characteristic of the FP-Growth method. Because it only mines frequent patterns, which leads few in number of patterns category. With the increase of the meaningful patterns, sensitivity of FP-Growth decreases. Due to the scalability of mined patterns, sensitivity of PSP-Tree performs steady.

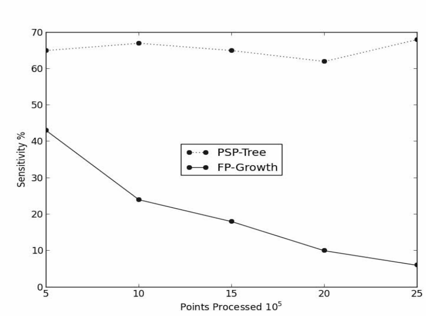


Figure 9: The comparison of meaningful patterns of bj3 data set

As shown in the description of data set, the ECG waveforms of patient who has ventricular tachycardia show the repeatability of QRS complex. In the results of experiment conducted on the *cuvdt* data set, we find a lot of QRS complexes, when the QRS complexes appear repeatability, the corresponding tags in the dataset actually show that the patients are in the risk of ventricular tachycardia. We totally find 134 QRS complex, among them there are 41 QRS repetitions which cover the whole 34 tagged ventricular tachycardia. Therefore, the interesting experiment results show that our method can provide effective help for physicians.

Conclusion

In this paper, we analyze the requirements and challenges of pattern mining over ECG waveform. By in-depth observation, we summarize significant semantics over of ECG waveform. In order to mining semantics, we proposed a new data structure called PSP-Tree for characterizing ECG patterns. With PSP-Tree, we can mine out scalability, temporality and hierarchy from ECG patterns. With extensive experiments on real ECG data sets, compared to the state-of-the-art methods, the experimental results show that our method is more feasible and effective.

In the near future, we plan to develop friendly man-machine interface based on our method for physicians in ICU, so that they can analyze the ECG waveform of patients in real time and provide accurate treatment timely. Besides, we also plan to extend our method on other bioinformatics scenarios, such as mining microsatellite DNA, because DNA sequence is a complex structure composed by basic units, A, T, G, C base. Similarly, microsatellite DNA sequence contains significant semantics which have great value, such as in population genetic study, cancer carcinogenesis analysis and Hemophilia A diagnosis.

Acknowledgement

This work was supported by Natural Science Foundation of China (No.61170003 and No.60973002), and MOE-CMCC Research Fund (MCM20130361).

References

Banerjee, S., and Mitra, M. 2014. Application of cross wavelet transform for ecg pattern analysis and classification. *Instrumentation and Measurement, IEEE Transactions on* 63(2):326–333.

Belle, A.; Kon, M. A.; and Najarian, K. 2013. Biomedical informatics for computer-aided decision support systems: a survey. *The Scientific World Journal* 2013.

Chatterjee, H.; Gupta, R.; and Mitra, M. 2011. A statistical approach for determination of time plane features from digitized ecg. *Computers in biology and medicine* 41(5):278–284.

Chiang, H.-S.; Shih, D.-H.; Lin, B.; and Shih, M.-H. 2014. An apn model for arrhythmic beat classification. *Bioinformatics* 30(12):1739–1746.

Chinnasami, S.; Rathore, C.; and Duncan, J. S. 2013. Sinus node dysfunction: an adverse effect of lacosamide. *Epilepsia* 54(6):e90–e93.

Eom, J.-H.; Kim, S.-C.; and Zhang, B.-T. 2008. Aptacdss-e: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction. *Expert Systems with Applications* 34(4):2465–2479.

Ghaffari, A.; Homaeinezhad, M.; Atarod, M.; and Akraminia, M. 2010. Parallel processing of ecg and blood pressure waveforms for detection of acute hypotensive episodes: a simulation study using a risk scoring model. *Computer methods in biomechanics and biomedical engineering* 13(2):197–213.

Goldberger, A. L.; Amaral, L. A.; Glass, L.; Hausdorff, J. M.; Ivanov, P. C.; Mark, R. G.; Mietus, J. E.; Moody, G. B.; Peng, C.-K.; and Stanley, H. E. 2000. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation* 101(23):e215–e220.

Goldberger, A. L. 2012. *Clinical electrocardiography: a simplified approach*. Elsevier Health Sciences.

Kim, K. K.; Lim, Y. G.; Kim, J. S.; and Park, K. S. 2007. Effect of missing rr-interval data on heart rate variability analysis in the time domain. *Physiological measurement* 28(12):1485.

Li, F.-F.; Li, H.-Y.; Qu, Q.; and Miao, G.-S. 2010. Spq: A scalable pattern query method over data streams. *Jisuanji Xuebao(Chinese Journal of Computers)* 33(8):1481–1491.

McSharry, P. E.; Clifford, G. D.; Tarassenko, L.; Smith, L.; et al. 2003. A dynamical model for generating synthetic electrocardiogram signals. *Biomedical Engineering, IEEE Transactions on* 50(3):289–294.

Mitra, S.; Mitra, M.; and Chaudhuri, B. 2009. Pattern defined heuristic rules and directional histogram based online ecg parameter extraction. *Measurement* 42(1):150–156.

Noh, K.; Lee, H. G.; Shon, H.-S.; Lee, B. J.; and Ryu, K. H. 2006. Associative classification approach for diagnosing cardiovascular disease. In *Intelligent Computing in Signal Processing and Pattern Recognition*. Springer. 721–727.

Othman, M. A., and Safri, N. M. 2012. Characterization of ventricular arrhythmias using a semantic mining algorithm. *Journal of Mechanics in Medicine and Biology* 12(03):1250049.

Porta, A.; Guzzetti, S.; Montano, N.; Furlan, R.; Pagani, M.; Malliani, A.; and Cerutti, S. 2001. Entropy, entropy rate, and pattern classification as tools to typify complexity in short heart period variability series. *Biomedical Engineering, IEEE Transactions on* 48(11):1282–1291.

Shibahara, T. 1985. On using causal knowledge to recognize vital signals: Knowledge-based interpretation of arrhythmias. In *IJCAI*, 307–314. Citeseer.

Simson, M. B. 1981. Use of signals in the terminal qrs complex to identify patients with ventricular tachycardia after myocardial infarction. *Circulation* 64(2):235–242.

Srinivas, K.; Rani, B. K.; and Govrdhan, A. 2010. Applications of data mining techniques in healthcare and prediction of heart attacks. *International Journal on Computer Science and Engineering (IJCSE)* 2(02):250–255.

Thygesen, K.; Alpert, J. S.; and White, H. D. 2007. Universal definition of myocardial infarction. *Journal of the American College of Cardiology* 50(22):2173–2195.