ATRIAL FIBRILLATION INTELLIGENT DIAGNOSIS BASED ON ROUGH SET APPROACH

Xiaomin Xu and Ying Liu

Department of Electronic & Computer Engineering
University of Portsmouth, UK

Xiaomin.xu@port.ac.uk

Abstract: This paper presents an artificial intelligent technique for atrial fibrillation prediction in ECG analysis and tele-cardiologic monitoring application. A model based on rough set theory is established for the purpose of mining decision rules in training span and classifying new data in test span. In this process, a novel discretization algorithm is proposed to generate efficient cuts which may be used for data discretization. As the result of this work, 32 decision rules for AF prediction have been mined out with eight attributes which relate to ECG parameters such as heart rate, heart rate variability, P-wave form, T-wave form, and QRS complex. The test accuracy of average 96.33% demonstrates that rough set is an excellent approach for AF intelligent prediction.

Introduction

Atrial fibrillation (AF) is an extremely common cardiac arrhythmia among elder population. It can cause or exacerbate heart failure [1] and affect quality of life. Predicting AF is one of the most important tasks in tele-cardiologic monitoring.

In the last decades several studies have focused on finding algorithms able to predict AF by the analysis of surface electrocardiography records, such as, twelve-lead electrocardiograms [2]; signal-averaged P-wave morphology [3]; R-R interval dynamics [4, 5]; and neural network [6]. But all the methods are mainly research matters and their clinical application is still limited [7].

In this paper, we present an approach based on rough set theory to make intelligent ECG diagnosis for AF prediction. The objectives are to investigate the possible clinical predictors for AF diagnosis, establish predicting model by means of machine learning technique, and extract AF predicting decision rules for applying in tele-cardiologic monitoring domain.

Materials and Methods

For the AF intelligent diagnosis and prediction challenges, ECG signals from MIT-BIH Atrial Fibrillation Database (afdb) [8], which are real-time recorded and reliable ECG data accepted in many academic articles, are used in our study. Sampling frequency is set to 250Hz. Twenty representative segments from different patient's records are random divided into two groups, as shown in Table 1. Each of the two categories includes 10 segments. Each segment includes some fragments extracted randomly from the same patient's record, and each fragment has 4500 points of digital ECG data. The total data in each segment is ranged including 1500 to 3500 QRS complexes.

Table 1: Training and testing sets in this work

	Training		Test
Tr-1	04126	Te-1	08455
Tr-2	04746	Te-2	08434
Tr-3	04908	Te-3	08405
Tr-4	04936	Te-4	08378
Tr-5	05261	Te-5	08219
Tr-6	06462	Te-6	08215
Tr-7	06453	Te-7	04015
Tr-8	07162	Te-8	04043
Tr-9	07879	Te-9	04048
Tr-10	07910	Te-10	05091

Using rough set tool package-ROSETTA, which allows for the analysis of tabular data, we discrete attributes extracted from raw ECG data, and classify them according to various algorithms. In our study, six kinds of discretization algorithms and two classification methods are considered and compared.

A.. Rough Set Approach

Rough set theory [9] was introduced as a method to model static uncertain multi-variety relationships in 1980's. In recent years, some efforts in data mining have focused on applying the rough set method to knowledge discovery. By now, RS becomes one of the most important technologies for data mining, and is widely applied in rule generation and decision-making fields.

The general scheme for AF artificial intelligent diagnosis based on rough set approach is shown in Figure 1. The process of rule generation according to some definite knowledge is called the training span; the process of recognizing new and unknown data depending on these rules is called the test span. Since rough set method can only handle discrete data, the first step in RS approach is to calculate cuts for

discretization. There are a lot of articles talking about this issue, but no one is effective for AF prediction. In this paper, we propose a novel algorithm, denote interaction algorithm, to do this job. The main ideal is that, for any two sets of objects $R \subseteq C$, and $O \subseteq D$ where C and D are condition and decision attributes, respectively, in the information system $S = \{U, C \bigcup D, V, f\}$, the relationship function may be expressed as (see figure 2):

$$f(R, O) = \{R \bigcap O \,|\, 0 : R \not\subseteq O \;\; ; 1 : R \subseteq O, or \;\; \alpha : else\}$$

where α is in the value between 0 and 1. The possible cuts are written as:

$$C_{R,O} = \{\langle r_o^{min}, o^{max} \rangle, or \langle r_o^{max}, o^{min} \rangle \,|\, r \in R, o \in O\}$$
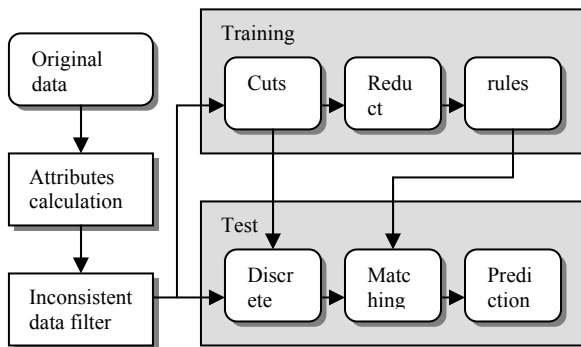

Figure 1: General scheme for AF prediction based on rough set approach
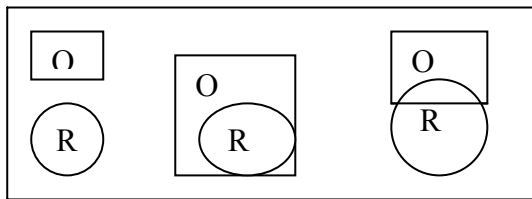

Figure 2. Relationship between sets O and R

The process of cuts calculation is described as follows:

Input: one of condition attribute, and relative decision attribute;
Output: Cuts of intersection discretization.
 Step1: Calculate decision indiscernibility of attribute $O \subseteq D$, and sort them in
$$o_R^1 < \cdots < o_R^j < \cdots < o_R^{V_D} ;$$
Step 2: Calculate indiscernibility of attribute $R \subseteq C$, and sort them in
$$r_O^1 < \cdots < r_O^k < \cdots < r_O^{V_C} ;$$

 Step 3: For each decision indisceribility o_R^j, find out the two sorts of indiscernibility r_O^k and r_O^{k+1} (or r_O^{k-1}) which are nearest to o_R^j;
Step 4: The possible cuts of discretization for decision indiscrnibility o_R^j are $\langle r_O^k, o_R^j \rangle$ and $\langle o_R^j, r_O^{k+1} \rangle$.

Sometimes, not whole knowledge is always necessary to define some categories available in the knowledge considered. Or in other words, some sets of attribute are dependent for making decision. The knowledge reduct is introduced after process of cut calculation, for the purpose of removing of superfluous partitions and/or superfluous basic categories, and forming a minimal set of attributes that preserve the indiscernibility relation.

Once the reducts have been computed, they are used to generate rules which are normally in the form of IF-THEN, or $\alpha \rightarrow \beta$, where α is a conjunction of descriptors and β consists of a decision descriptor. If the decision system contains objects that match the rule's antecedent α, but that have a different value for the decision attribute than the one indicated by the rule's consequent β, the quality accuracy is needed to measure of how trustworthy the rule is in drawing conclusion β on the basis of evidence α:

$$accuracy(\alpha \rightarrow \beta) = \frac{card\{[\alpha] \cap [\beta]\}}{card\{[\alpha]\}}$$

When rules derived from the reducts are used to classify new and unknown data, the set of rules is referred to as a classifier. One popular classifying method is the voting system which consists the following phases:

a). For a test object x, the rule set is scanned to find applicable rules;
b). If a rule is matched, this rule is said to fire for object x;
c). If a rule is not found, the most frequently occurring outcome or decision class is chosen, or alternatives include reverting to a nearest neighbor method and instead consider a collection of rules close to object x in some sense [10];
d). If more than one rule fires, this indicates that there are more than one possible outcomes. The most possible outcome will be the one which has the biggest accuracy value.

B. Attributes

Since AF identification depends mainly on ECG characteristics such as P-wave form and heart rate, in our study, eight parameters from ECG analysis are selected as attributes in rough set processing.

1. Heart rate (HR);
2. Difference between Heart rate and dominant Heart rate (DHR):

$$DHR = \frac{HR - \overline{DHR}}{HR} \times 100\%$$

3. Difference between Heart rate and neighbour heart rate (DHRN):

$$DHRN = \frac{(HR - HRB) + (HR - HRA)}{HR} \times 100\%$$

where HRB and HRA are the heart rates before and after the calculating one;

4. Difference between Heart rate and the next one (DHRA):

$$DHRA = \frac{HR - HRA}{HR} \times 100\%$$

5. The rate of ST segment in R-R interval (PT):

$$PT = \frac{R_2 - T}{T - R_1} \times 100\%$$

where R_1 is the position of offset of the first QRS complex, R_2 is the position of onset of the second QRS complex, and T is the position of the calculating T-wave form.
6. QRS complex (QRS);
7. P-wave form (P);
8. T-wave form (T).

The dominant heart rate in our study is defined as the average heart rate in special region in which most heart rates are concentrated. Comparing with the statistical average in whole domain, dominant heart rate has more ability to represent the standard characteristic of heart. For example, for a set of heart rate $HR = \{60,60,60,100,60\}$, the statistical average is 68, but the dominant heart rate is 60.

The dominant heart rate ($\overline{DHR}$) can be calculated by suing moving window function algorithm. The window function is defined as:

$$F(f) = \begin{cases} 1, & (f_d \leq f \leq f_u) \\ 0, & (f < f_d, f > f_u) \end{cases}$$

where f_d and f_u are the up and down limitation of moving windows.

Step 1. Calculate maximum and minimum heart rates ($f_{\max}$ and $f_{\min}$) in whole domain;
Step 2. Set window width (Δf) equals 2;
Step 3. Count the number of QRS complexes (n_1) in the window of $\{f_{\max}, f_{\max} - \Delta f\}$;
Step 4. Move window by 1 unit, count the number of QRS complexes (n_2) in the new window;
Step 5. Repeat step 4 till to $f_{\min}$, a set of QRS numbers are obtained as $N = \{n_1, n_2, \cdots n_i\}$. Denote the maximum number in this set as n_m, and compare it with special threshold T_h. If n_m is les than T_h, go back to step 2, and plus 1 to Δf;
Step 6. Repeat step 2 to step 5, till n_m is equal to or bigger than T_h, the average heart rate in this region is dominant heart rate:

$$\overline{DHR} = \frac{\sum_{i=1}^{n_m} HR_i}{\Delta f_m}$$

In this algorithm, the value of threshold determines the final results. According to our experiment, in our study , we take 65% of total QRS complexes in whole domain as the threshold value.

Results and discussion

In the training span, six algorithms are calculated for tabular data disertization for eight condition attributes. The created cuts for each attribute are shown in Table 2. Comparing with all six algorithms, the cuts only in intersection algorithm and equal frequency algorithm have less sections for all condition attributes. This result indicates that these two discretization algorithms are not sensitive for noise data. In addition, the obvious difference between of them is in QRS attributes. There are 3 cuts in equal frequency, but only 1 cut in intersection algorithm. According to the previous knowledge of ECG diagnosis, QRS complex has no contribution for AF prediction. Therefore, the intersection algorithm is more reasonable and more efficient to mine AF decision rules.

Because not all attributes are always necessary for rule mining, reduction operation is introduced into rough set approach to find those attributes that are necessary and sufficient to final decision. In our work, two kinds of reduct algorithms, Genetic and Johnson's algorithms, are used and computed. The final decision rules generated by these two algorithms are shown in Table 3.

The mathematical notations for some rule samples in Table 3 can be translated as follows:

Table 2: Discretization cuts with different kinds of algorithms for HR, DHR, DHRN, DHRA, PT, QRS, P and T

Algorithms	HR		DHR		DHRN		DHRA	
	Value range	%	Value range	%	Value range	%	Value range	%
Intersection	[*, 92) [92, 104) [104, *)	34.87 27.30 37.83	[*, 18) [18, 27) [27, *)	37.33 24.84 27.83	[*, -39) [-39, 9) [9, *)	2.96 84.38 12.66	[*, -46) [-46, 43) [43, *)	6.74 89.31 3.95
Boolean reasoning	[*, 92) [92, 95) [95, 98) [98, 101) [101, *)	34.87 8.22 8.55 7.57 40.79	[*, -50) [-50, -42) … [55, 56) [56, *)	0.16 0.16 … 0.49 0.33	[*, -28) [-28, *)	12.34 87.67	[*, 7) [7, *)	74.51 25.49
Entropy/ MDL	[*, 105) [105, *)	62.17 37.83	[*, 28) [28, *)	62.17 37.83	[*, -5) [-5, -4) … [44, *)	29.28 0.66 … 3.95	[*, -1) [-1, 15) … [29, *)	31.41 50.33 … 9.21
Equal frequency	[*, 90) [90, 111) [111, *)	33.55 33.22 33.22	[*, 16) [16, 32) [32, *)	33.55 32.89 33.55	[*, -2) [-2, 1) [1, *)	34.54 30.43 35.03	[*, -1) [-1, 3) [3, *)	31.41 35.69 32.89
Naive	[*, 92) [92, 93) … [104, 105) [105, *)	34.87 2.47 … 0.82 37.83	[*, 18) [18, 19) … [27, 28) [28, *)	34.87 2.47 … 1.15 37.83	[*, -46) [-46, -45) … [43, 44) [44, *)	6.25 0.49 … 0.16 3.95	[*, -39) [-39, -38) … [28, 29) [29, *)	5.59 0.33 … 0.82 9.21
Semi-Naive	[*, 101) [101, *)	59.21 40.79	[*, 26) [26, *)	60.03 39.97	[*, -40) [-40, -39) … [43, 44) [44, *)	7.73 0.49 … 0.16 3.95	[*, -36) [-36, -28) … [5, 7) [7, *)	6.41 4.28 … 4.61 25.49

Algorithms	PT		QRS *		P **		T ***	
	Value range	%	Value range	%	Value range	%	Value range	%
Intersection	[*, 160) [160, 280) [280, *)	15.79 55.26 28.95	1	100	1 4	41.61 58.39	1	100
Boolean reasoning	[*, 232) [232, 275) [275, *)	59.05 9.87 31.08	[*, 78) [78, *)	66.61 33.39	[*, 3) [3, *)	41.61 58.39	1	100
Entropy/ MDL	[*, 203) [203, 205) … [274, 276) [276, *)	35.20 1.81 … 0.66 30.76	[*, 126) [126,130) [130, *)	97.53 1.15 1.32	1 4	41.61 58.39	1	100
Equal frequency	[*, 201) [201, 266) [266, *)	33.55 33.39 33.06	[*, 76) [76, 78) [78, *)	0.49 66.12 33.39	1 4	41.61 58.39	1	100
Naïve	[*, 160) [160, 161) …	14.14 1.64 …	[*, 76) [76, 78) …	0.49 66.12 …	[*, 3) [3, *)	41.61 58.39	1	100

	[275, 276)	0.33	[126, 130)	1.15				
	[276, *)	30.76	[130, *)	1.32				
Semi-Naïve	[*, 184)	24.67	[*, 78)	66.61	[*, 3)	41.61	1	100
	[184, 185)	0.16	[78, 82]	8.06	[3, *)	58.39		
	…	…	...	...				
	[273, 274)	0.49	[82, 86)	6.58				
	[274, *)	31.41	[86, *)	18.75				

*   QRS=1 means normal;
**   P=1 means normal, P=4 means absent;
*** T=1 means normal.

Table 3: Rules for AF prediction generated by using Intersection discretization and genetic reduct algorithms

Number of rules	AF Decision Rules *	RHS Accuracy
1	DHR(1)&DHRN(3)&DHRA(3)&ST(2)&P(1)→D(0)	1.0
2	DHR(1)&DHRN(3)&DHRA(3)&ST(3)&P(1)→D(0)	1.0
3	DHR(1)&DHRN(3)&DHRA(3)&ST(2)&P(4)→D(0)	1.0
4	DHR(1)&DHRN(1)&DHRA(1)&ST(3)&P(4)→D(0)	1.0
5	DHR(3)&DHRN(3)&DHRA(3)&ST(3)&P(4)→D(0)	0.25
	DHR(3)&DHRN(3)&DHRA(3)&ST(3)&P(4)→D(1)	0.75
6	DHR(1)&DHRN(1)&DHRA(1)&ST(2)&P(4)→D(0)	1.0
…	…	…
31	DHR(2)&DHRN(3)&DHRA(2)&ST(1)&P(4)→D(1)	1.0
32	DHR(2)&DHRN(3)&DHRA(2)&ST(3)&P(4)→D(1)	1.0

Rule 1: If DHR is less than 18 and DHRN falls in the region of (-39,9) and DHRA is in the range of (-46,43) and PT is more than 280 and P is normal, the decision of non-AF with 100% accuracy will be resulted.

Rule 31: If DHR is more than 27 and DHRN falls in the range of (-39,9) and DHRA is more than 43 and PT is less than 160 and P is abnormal, the decision of AF with 100% accuracy will be resulted.

But in many cases, not all created rules can make decision with 100% accuracy for classifying new object. For example:

Rule 5: If DHR falls in the region of (18,27) and DHRN falls in the region of (-39,9) and DHRA is in the region of (-46,43) and PT is in the region of (160,280) and P is absent, the decision will be made as non-AF with 25% accuracy, or as AF with 75 accuracy.

This kind of result may be predicted because not all previous knowledge are distinct for decision result. The condition attributes in Rule 5 are falling in vague region and led indistinct decision results. In the other hand, this result demonstrates that the most advantage of rough set

approach is that, it can give numeral results for vague decision problems. According to the principles of data classification in rough set approach, the final decision will be made for outcome with higher support or accuracy, i.e., in Rule 5, the final decision is AF.

Experiments in our work mined out 32 final decision rules for AF prediction with good decision accuracy (if the decision accuracy is 50/50, it will be poor result). The rule number of 32 is the smallest among the six kinds of algorithms in Table 2. It expresses that the intersection discretization algorithm is efficient and reliable for rough set approach in AF diagnosis. Programming using this algorithm should have higher efficiency for real-time AF monitoring application.

Experimental results in test span are shown in Figure 3. The average accuracy of AF prediction achieved 96.33%. Comparing with some other work [11,12], this approach is good enough in AF intelligent prediction application.

Besides of the discovered decision rules, another contribution of this approach is that, it has proved in theory that heart rate and heart rate variability (HRV) are the key attributes for AF prediction. In addition, this approach has mined out relative attributes such as

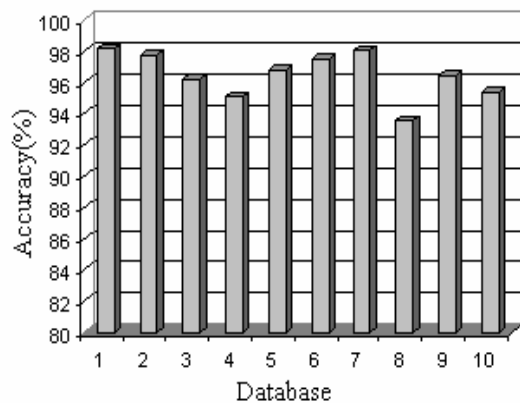DHRN and DHRA. This work enriches principles of AF intelligent diagnosis.

Figure 3. Test results of AF prediction in precentage accuracy for 10 ECG sets of data

Because the condition attributes are characterized depending on object features extracted by some methods, it needs to emphasize that, various feature extraction methods may result in the change of final decision rules.

Conclusions

Rough set theory is a very useful mathematic tool for data mining. In this work, we have established a model based on rough set approach for atrial fibrillation prediction, and mined out 32 decision rules. Experiment results demonstrate that this approach is reliable and efficient in AF intelligent prediction. Also, it presents a good example of rough set data analysis in biomedical application.

References

[1] WAKTARE, J.E., AND CAMM, A.J., (1998): 'Acute treatment of atrial fibrillation: why and when to maintain sinus rhythm', *Am J Cardiol.*, **81**, pp. 3C-15C

[2] DILAVERIS, P.E., GIALAFOS, E.J., SIDERIS, S.K., THEOPISTOU, A.M., ANDRIKOPOULOS, G.K., KYRIAKIDIS, M., GIALAFOS, J.E., AND TOUTOUZAS, P.K., (1998): 'Simple electrocardiographic markers for the prediction of paroxysmal idiopathic atrial fibrillation', *Am Heart J.*, **135**, pp. 733-738

[3] ISHIMOTO, N., ITO, M., AND KINOSHITA, M., (2000): 'Signal-averaged P-wave abnormalities and atrial size in patients with and without idiopathic paroxysmal atrial fibrillation', *Am Heart J* , **139**, pp. 684-689

[4] VIKMAN, S., MAKIKALLIO, T.H., YLI-MAYRY, S., PIKKUJAMSA, S., KOIVISTO, A.M., REINIKAINEN, P., AIRAKSINEN, K.F., AND HUIKURI, H.V., (1999): 'Altered complexity and correlation properties of R-R interval dynamics before the spontaneous onset of paroxysmal atrial fibrillation', *Circulation*, **16**, pp. 2079-2084

[5] TRAN, T., JAMES, M.N., MATEO, A., AND BRAHM, G., (2004): 'Prediction of paroxysmal atrial fibrillation by analysis of atrial premature complexes', *IEEE Tran. On Biomed. Eng.,* **51(4)**, pp. 561-569

[6] YANG, T.F., DEVINE, B., AND MACFARLANE, P.W., (1994): 'Atrificial neural networks for the diagnosis of atrial fibrillation', *Med. Biol. Eng. Comput.*, **32(6)**, pp. 615-619

[7] POLI, S., BARBARO, V., BARTOLINI, P., CALCAGNINI, G., AND CENSI, F., (2003): 'Prediction of atrialf fibrillation from surface ECG: review of methods and algorithms', Proc. of Ann 1st Super Sanita 2003, **39(2)**, pp. 195-203

[8] MARK, R., AND MOODY, G., (1998): '*MIT-BIH Arrhythmia Data Base Directory*', Cambridge: Massachusetts Inst. Technology

[9] PAWLAK, Z., (1982): 'Rough Sets', *International J computer and Information Sciences*, **11(5),** pp. 341-356

[10] ROMAN, S., (1993): 'Rough set learning of preferential attitude in multi-criteria decision making', Proc. International Symposium on Methodologies for Intelligent Systems (ISMIS'93), **689**, PP. 642-651

[11] ROSSI, P., CASALEGGIO, A., ET AL., (2000): 'Automatic atrial tachyarrhythmia detection from intracardiac electrograms', *Ital Heart J.*, **1(6),** pp. 412-419

[12] KOLB, C., NURNBERGER, S., NDREPEPA, G., SCHOMIC, A., AND SCHMITT, C., (2001): 'Modes of initiation of paroxysmal atrial fibrillation from analysis of spontaneously occurring episodes using a 12-lead Holter monitoring system', *Am. J. Cardiol..*, **88(8),** pp. 853-857