

Learning Decision Tree for Selecting QRS Detectors for Cardiac Monitoring

François Portet¹, René Quiniou², Marie-Odile Cordier², and Guy Carrault³

¹Department of Computing Science, University of Aberdeen, Aberdeen, AB24 3UE, UK
fportet@csd.abdn.ac.uk

²Irisa, INRIA, Université de rennes 1, Campus de Beaulieu, 35042, Rennes, France
{quiniou,cordier}@irisa.fr

³LTSI, Université de rennes 1, Campus de Beaulieu, 35042, Rennes, France
guy.carrault@univ-rennes1.fr

Abstract. The QRS complex is the main wave of the ECG. It is widely used for diagnosing many cardiac diseases. Automatic QRS detection is an essential task of cardiac monitoring and many detection algorithms have been proposed in the literature. Although most of the algorithms perform satisfactorily in normal situations, there are contexts, in the presence of noise or a specific pathology, where one algorithm performs better than the others. We propose a combination method that selects, on line, the detector that is the most adapted to the current context. The selection is done by a decision tree that has been learnt from the performance measures of 7 algorithms in various instances of 130 combinations of arrhythmias and noises. The decision tree is compared to expert rules tested in the framework of the cardiac monitoring system IP-CALICOT.

1 Introduction

The QRS complex is the main wave in the ECG as it reflects the ventricular activity of the heart. Its automatic detection is an essential task for cardiac monitoring systems that has been studied for several decades and has resulted in a large number of methods [1-5]. But, each method has situations where it fails as each QRS detector reacts differently to the large number of different QRS waveforms and noises.

In this paper, we propose to combine the strength of several algorithms to detect the QRS complex even in difficult situations. The approach is not to fuse the detectors outputs but to select, on-line, the best detector from among a set of algorithms according to an evaluation of the current context of the chunk of ECG to process. The selection is done by decision tree (DT) which is learnt to select 7 QRS detectors according to various situations, called *contexts*, representing 130 combinations of arrhythmia and clinical noise. The learning is detailed in section 2 and its results is analysed in section 3. This method is then compared with expert rules previously acquired in section 4. Finally, the paper ends with a short discussion.

2 Learning Method

Selecting the QRS detector the most suited to the specific context of some ECG chunk is a difficult task. We advocate the use of selection rules but these rules must be acquired. In a previous experiment [6], we described an expert acquisition method. This experiment emphasized the complexity of the task. That is why an automatic approach is considered. Many methods could be used to learn selection rules but decision tree learning presents several advantages in our application: (1) the ECG contexts are composed of both nominal and categorical data that are easily handled by the learning method; (2) rules derived from decision tree are explicit and checkable by human experts; and (3) the learned decision tree can process large volume of data in a short time, which is mandatory in ICU monitoring.

Succinctly, a decision tree (DT) consists of several test nodes and class (or decision) leaves. It classifies an input by executing the tests in the tree beginning at the root and going down the tree until a leaf is reached which gives the class of the input (or the decision to be taken). The C4.5 algorithm of Quinlan [7] has been used to learn the tree. In our application, the DT input is an ECG context described by a set of attributes (i.e. properties of an ECG chunk) and the output is the algorithm to apply to the input context (i.e. the decision). To learn the DT, a training data set has been created. This has been achieved by (1) generating all possible contexts that can be found in an ECG, (2) applying all the QRS detectors to these contexts and (3) deciding what the best detector to apply to a given context is.

2.1 Definition of the Context

A context is defined as the combination of a *rhythm context* and a *noise context*. Indeed, in clinical practice, an ECG is composed of the original ECG—the rhythm context—which is usually corrupted by noise—the noise context.

An ECG is composed of different QRS waveforms and the variation of waveforms inside an ECG signal disturbs the detection. To the best of our knowledge, the influence of QRS waveform variation on the QRS detection has been studied only in very few papers [6]. The rhythm contexts have been extracted from the MIT-BIH Arrhythmia Database [8]. 10 rhythms that are representative of normal rhythms and arrhythmic situations have been chosen to assess the detectors on sequences of identical QRSs as well as on sequences of non identical QRSs.

The noise used to corrupt the ECG comes from several sources. Few studies have analyzed the influence of noise on QRS detectors. Most of them used composite noise that is not representative of real clinical situation. In our study, clinical additive noise was extracted from the MIT-BIH Noise Stress Test Database[9], which contains three noise records, lasting 30 mins each, predominantly composed of baseline wander (*bw*), muscle artifact (*ma*), and electrode motion artifact (*em*). We used these three types of noise at four Signal-to-Noise Ratios (SNR). Thus the attributes of a context are: rhythm context type, noise context type, and the SNR of the noise context.

2.2 Selected QRS Detectors

Many QRS detection schemes have been described in the literature for the last 30 years; however, since the 90s the performance has improved slowly in non noisy situations. For example, the Pan and Tompkins detector (1985) [1] (ER=0.68%) performs slightly less than the Christov's one (2004) [5] (ER=0.44%) in uncorrupted situations. That is why selection rules are learnt in noisy situations in order to emphasize the difference between detectors. Seven algorithms were selected. They were those used in [6] (*pan*, *gritzali*, *af2*, and *df2*) plus those proposed by Benitez *et al.*[4] (*benitez*), Suppappola and Sun[2] (*mobd*), and Kadambe *et al.*[3] (*kadambe*).

2.3 Computation of the Training Set for Decision Tree Learning

The decision rules for selecting some algorithm in a specific context are given by a decision tree previously learned. The training set was computed by the algorithm given Fig. 1.

```

Let  $Q$  be a set of ECGs;
for  $r = 1$  to  $R$  rhythm contexts do
  for  $t = 1$  to  $T$  trials do
    Select randomly from  $Q$  a chunk  $S$  containing  $B$  QRSs of the rhythm  $r$ ;
    for  $n = 1$  to  $N$  noise (context) types do
      Corrupt the chunk  $S$  with the noise  $n$  to obtain  $S'$ ;
      Filter  $S'$  to obtain  $S^*$ ;
      for  $d = 1$  to  $D$  detectors do
        compute  $p = performance(d, S^*)$ ;
         $M(t, r, n, d) = p$ ;
        /* save the result in the matrix  $M$  */
      end
    end
  end
end
end

```

Fig. 1. Algorithm used for the computation of the results

This algorithm computes a matrix \mathbf{M} where the performance p of each QRS detector is related to the context and the trial. p is composed of 3 values: True Positives (TP -- correct detections), False Negatives (FN -- missed detections) and False Positives (FP -- false alarms). These values are then used to compute: the Error Ratio, $ER=(FN+FP)/(TP+FN)$, the Sensitivity, $Se=TP/(TP+FN)$, the Positive Predictivity, $PP=TP/(TP+FP)$ and the F-Measure, $FM=2*PP*Se/(PP+Se)$.

The training set was composed using the following parameters: chunk length $\mathbf{B=10}$ beats (QRSs), $\mathbf{N=13}$ noise values, $\mathbf{R=10}$ rhythm types, $\mathbf{T=100}$ instances of each context type, $\mathbf{D=7}$ detectors. In all, 9,500 QRSs were used representing 8.70% of the whole database and leading to a training set composed of $T*N*R$ ($100*13*10$) =13,000 individuals. The class (i.e. algorithm that should be chosen in this context) of each individual was found by selecting the QRS detector with the best FM value, for each instance of context.

3 Learning Results

A tree of size 69 for 48 leaves was obtained from the training set. *benitez* was used in a majority of the contexts, particularly in the contexts *no_noise* and *bw*. These two contexts do not perturb the signal as *bw* can be easily removed by the input filter of the detectors. Then, for the *ma* noise it alternates between *kadambe* and *benitez*. The choice of *kadambe* comes from its wavelet filtering, which appears to be the most able to deal with high-frequency noise. For the *em* context, the choice alternates between all the detectors of the set except *kadambe*. The *em* noise is composed of high and low frequency components which can affect the detectors very much. Moreover, in this branch, the rhythm type rather than the level of noise is used to distinguish the different cases. This demonstrates the value of using the rhythm information as a factor which influences the detectors performance.

4 Experiments and Comparison with Expert Rules

A test set of 11 uncorrupted ECG records including 10 different rhythm contexts and representing 5 hours and 30 mins, was extracted from the MIT-BIH Arrhythmia database. The clinical noise (*bw*, *ma*, *em* at 4 SNRs) has been added randomly to the uncorrupted ECGs in order to control the SNR. The learned decision tree was translated into production rules and loaded into the cardiac monitoring system IP-CALICOT [10]. IP-CALICOT is a piloting system which enables the selection of signal processing algorithms on line, to treat chunks of ECG according to a context analysis. For comparison, expert rules acquired following the method described in Portet *et al.* [6] have also been tested on the dataset. To assess the maximum performance reachable with the selection rules, the best detector performance (the detector with maximum FM) for each chunk of ECG has also been retained. The results collected are used as gold standard and are grouped under the name *idealSelection*.

Table 1. Performance of the selected detectors

detector	ER ^{±STD} (%)	Se ^{±STD} (%)	PP ^{±STD} (%)	FM ^{±STD} (%)	nb of switch
<i>af2</i>	51.82 ^{±22.17}	92.16 ^{±3.19}	67.69 ^{±8.67}	78.05 ^{±6.62}	-
<i>benitez</i>	27.87 ^{±8.60}	96.46 ^{±1.62}	79.85 ^{±4.85}	87.38 ^{±3.37}	-
<i>df2</i>	37.20 ^{±23.78}	78.93 ^{±13.43}	83.03 ^{±11.66}	80.93 ^{±11.94}	-
<i>gritzali</i>	52.10 ^{±10.68}	86.66 ^{±4.64}	69.09 ^{±4.95}	76.89 ^{±4.12}	-
<i>kadambe</i>	22.24 ^{±7.77}	93.13 ^{±2.99}	85.83 ^{±4.80}	89.33 ^{±3.37}	-
<i>mobd</i>	55.70 ^{±13.69}	95.56 ^{±2.51}	65.09 ^{±5.37}	77.43 ^{±4.26}	-
<i>pan</i>	34.90 ^{±12.52}	78.20 ^{±11.06}	85.65 ^{±5.48}	81.76 ^{±7.38}	-
<i>expert rules</i>	20.60 ^{±6.62}	93.72 ^{±2.62}	86.75 ^{±4.12}	90.10 ^{±3.05}	623
DT	22.68 ^{±7.95}	94.44 ^{±2.48}	84.66 ^{±4.81}	89.28 ^{±3.46}	542
<i>idealSelection</i>	14.38 ^{±4.86}	94.85 ^{±1.93}	91.13 ^{±2.92}	92.96 ^{±2.32}	2119

The result synthesized in Table 1 shows that, according to the FM, expert rules outperform all the other methods (*idealSelection* is used as gold standard). The best

algorithm is *kadambe* with FM=89.33% and ER=22.24%. However, *benitez* shows a Se = 96.46% superior to *kadambe*. These two algorithms outperform the others with an FM greater by 5.6%. Among the algorithm selection methods, only expert rules (ER=20.60%) outperform *kadambe* (ER=22.24%) reducing ER by 1.64%. Moreover, expert rules obtain the lowest standard deviation for FM. This shows that the selection method by expert rules is more stable than the DT and, thereby is more reliable. DT is slightly below the results of *kadambe*. *idealSelection* shows that the upper bound for FM is 92.96%. Thus, expert rules contribute to fill 21.2% of the gap between the best algorithm *kadambe* and the gold standard. This shows that the algorithm selection strategy can be greatly improved with more accurate rules. Selection rules are also interesting because they switch algorithms fewer times than *idealSelection* (2119).

5 Discussion

This experience showed that the selection of algorithms rests mainly on the acquisition of good selection rules. The experience undertaken with QRS detection algorithms shows that there remains some room for improvement. According to the gold standard, the maximum reachable FM is 92.96%. Expert rules reached 90.10%, improving the best algorithm by 0.77% in noisy contexts. This is a good score according to the current literature studies in which the sensitivity is improved typically by less than 1% [5] even in non noisy situation. This method will be applied to the other kinds of signal processing algorithms used in cardiac monitoring, such as QRS classification and P wave detection, which are less developed fields than the QRS detection and for which more significant results are expected.

References

1. Pan, J., Tompkins, W.J.: A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* 32(3), 230–236 (1985)
2. Suppappola, S., Sun, Y.: Nonlinear transforms of ECG signals for digital QRS detection: a quantitative analysis. *IEEE Trans. Biomed. Eng.* 41(4), 397–400 (1994)
3. Kadambe, S., Murray, R., Boudreaux-Bartels, F.: Wavelet transform-based QRS complex detector. *IEEE Trans. Biomed. Eng.* 47(7), 838–848 (1999)
4. Benitez, D., et al.: The use of the Hilbert transform in ECG signal analysis. *Comput. Biol. Med.* 31, 399–406 (2001)
5. Christov, I.: Real time electrocardiogram QRS detection using combined adaptive threshold. *Biomed. Eng. Online.* 3(28), 1–9 (2004)
6. Portet, F., Hernández, A., Carrault, G.: Evaluation of real-time QRS detection algorithms in variable contexts. *Med. Biol. Eng. Comput.* 43(3), 381–387 (2005)
7. Quinlan, J.R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Francisco (1993)
8. Mark, R., Moody, G.: *MIT-BIH arrhythmia data base directory*. MIT Press, Cambridge (1988)
9. Moody, G., Muldrow, W., Mark, R.: A noise stress test for arrhythmia detectors. *Comput. Cardiol.* (1984)
10. Portet, F., et al.: Piloting signal processing algorithms in a cardiac monitoring context. In: Miksch, S., Hunter, J., Keravnou, E.T. (eds.) *AIME 2005. LNCS (LNAI)*, vol. 3581, Springer, Heidelberg (2005)