MOLEC: Anywhere and at Any Time Arrhythmia Classification

J Rodriguez, A Goñi, L Dranca, A Burgos, A Illarramendi

University of the Basque Country. Donostia - San Sebastián, Spain

Abstract

The new advances in sensor technology, PDAs and wireless communications favor the development of a new type of monitoring systems that can provide patients with assistance anywhere and at any time. Of particular interest are the monitoring systems designed for people that suffer from heart arrhythmias, due to the increasing number of people with cardiovascular diseases. In this paper, we present MOLEC: a PDA-based system that performs local real-time classification and detects the ECG anomalies in situ. In the actual implementation of MOLEC, the signal is acquired by some ECG sensors (ActiveECG sensors) with a 360Hz frequency. For the preprocessing of the signal the ECGPUWAVE tool and an automata developed by us that identifies the beats are used. We have also developed a beat and rhythm classifiers that determine if there has been an anomalous rhythm, and in that case, an alarm is sent to a hospital via wireless communications. The rhythm detection delay of the MOLEC system is of 6.66 seconds.

1. Introduction

Innovation in the fields of PDA, wireless communication and vital parameter sensors enables the development of revolutionary monitoring systems, which strikingly improve the lifestyle of patients, offering them security even outside the hospital. Focusing on electrocardiogram (ECG) sensors new ECG monitoring systems are appearing that outperform traditional holters. Although holters present the advantage that patients can continue living a normal life in their houses, they also present a serious drawback: if the patient suffers from a serious rhythm irregularity, the holters only record it, i.e. they do not react to it. So, in order to overcome the previous restriction we have implemented a system called MOLEC: a monitoring system that performs a local real-time ECG classification in a PDA. That PDA acquires, records and analyzes the ECG signals in order to find high-risk arrhythmias, and in case of detecting them, the PDA sends them to a hospital so that cardiologists can determine the adequate intervention. In that way a real anywhere and at any time assistance can be provided. In [1] we present a complete beat and rhythm classifier that can run in real-time into a PDA.

In this paper we show first the framework of the system, then in section 3 the process that we followed in order to select a beat and rhythm classifier that provides an accuracy result and in section 4 we show some performance aspects considering PDAs resources.

2. Framework of the MOLEC system

In this section we explain the four components that form the global architecture of MOLEC (see figure 1): the ECG sensor, the Monitor Molec, Molec Center and the users of the system (hospital and relative computers).

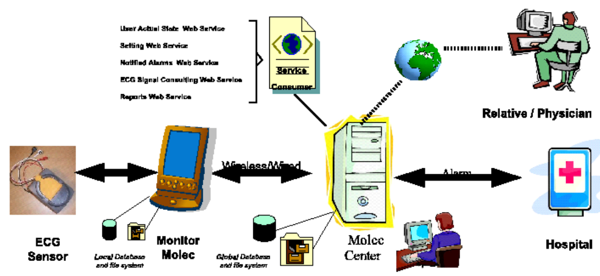

Figure 1. System Architecture.

1. **ECG sensor.** Specialized sensors capture the ECG signals and sends them to the PDA. Due to the diversity of ECG sensors available on the market, the MOLEC system has been designed in such a way that any ECG sensor can be easily incorporated into the system. At this moment, the sensors commercialized by ActiveECG are used in the prototype.
2. **Molec Monitor.** The Molec Monitor is a PDA that is carryied by the user, and that contains the software able of acquiring the ECG data signals, recording them, detecting abnormalities and notifying them immediately in case they are considered serious. That software is constituted by several modules (see figure 2): 1) *Acquisition module.* The raw data sent by the ECG sensors are interpreted by one specialized software module at the PDA which makes those ECG data available to the rest of the application as fixed size packages of samples (we use packages corresponding to 2 seconds of ECG signal) defined in a way

that are independent of the ECG sensor type. This module acts as a mediator between the ECG sensor and the rest of the system and permits to incorporate new ECG sensors dynamically using XML configuration files.

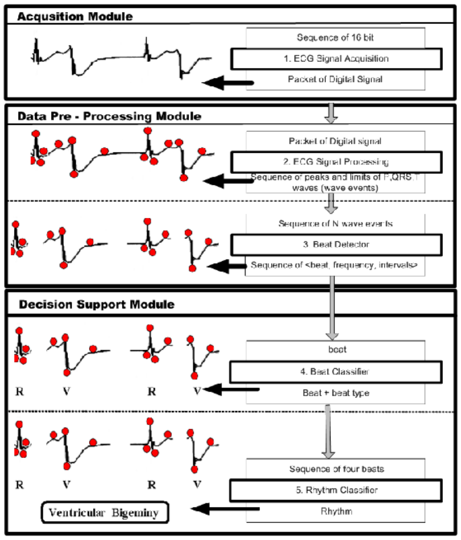

Figure 2. MOLEC modules

2) *Data preprocessing module.* This module receives the packages of an ECG signal and divides them into a sequence of beats. Every beat is usually formed by several waves: the P wave, the QRS complex and the T wave. For the arrhythmia detection it is significant the identification of the wave events: the points where a wave starts, ends and the picks of it, and also if the wave events are absent. The Data Preprocessing Module is constituted by two submodules: the ECG Signal Processing and the Beat Detector. For the implementation of the ECG Signal Processing we have used the ECGPUWAVE tool [2], a Fortran implementation of an on-line detection algorithm developed by Pan and Tompkins [3], that is provided by PhysioNet. This tool was not specifically designed to be run in small computers like PDAs and in general is used to process long duration ECG signals available also in PhysioNet. However, we have adapted it to run successfully in the PDA by using as input short duration ECG signals (2 sec.). Changes related to memory management were made in the "open source" of ECGPUWAVE with the goal of increasing the processing speed. The Beat Detector module has been implemented as an automata that transforms the sequence of wave events into a sequence of beats and computes the length of the relevant intervals and segments determined by two wave events. 3) *Decision support module.* The key element of the decision support system is our ECG classifier that is constituted by a set of rules that first classify beats (Beat Classifier) and then identify rhythms (Rhythm Classifier) by analyzing sequences of four beats. In section 3 more details about the ECG classifier are given.

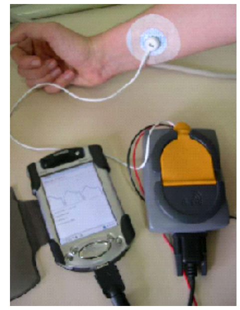

Figure 3. MOLEC system

4) *Database and communication module.* PDAs are small computers with storage and processing capacity limited. In our context, on the one hand ECG sensors generate a great amount of data and on the other hand the classification process also generate aditional data: wave events, beat and rhythm types, etc. The goal of the Database and Communication module is to store, manage and send these data in an efficient way, trying to optimize the limited resources of the PDA and the use of communication links.
3. **Molec Center.** The Molec Center is an intermediate computer between the Monitor Molec PDAs and the client computers (located at the hospital and relative homes). Molec Center receives alarms that are immediately communicated to the adequate hospital.
4. **Hospital and Relative Computers** Physicians in the hospital and relatives can query about the results of the ECG classification and can visualize the signal itself, through a set of Web Service deployed at Molec Center. In [4] we show a set of web services that access ECG data stored at the PDA and that can be technically deployed at the PDA.

3. Real time ECG classification

In order to design a heart beat classifier that provides competitive results, we have built and evaluated a set of experiments, in the area of machine learning, that use some available tools, which apply some methods over a known ECG data source (the MIT-BIH Arrhythmia Database[5]). The data were preprocessed by using the ECGPUWAVE tool and our Beat Detector. Once we got the data in the right format where one raw represents one beat, the data set was divided into two random groups: 66% of the data for training and 33% for validation. The training data was the input data for the tool that allowed the generation of the beat classification model, and the validation data was used to validate that model. We used two well-known machine learning tools in order to perform the experiments: Weka [6] and AnswerTree [6]. The applied methods were typical methods in the machine learning area [7] that can be used to classify beats and rhythms. For each tool-method combination an experiment was run taking the training data set as input, which was automatically selected by the tool. The result of each experiment was a beat classification model, which was later validated against the validation data set also obtained previously.

The goal of running the experiments was to obtain the most accurate beat classifier. Although we tried sixteen experiments (see [1] for more details) only the most accurate beat clasifier is mentionated here: j48.part method that implements the C4.5 algorithm, based on decision tree techniques that obtains a 92.73% accuracy (i.e. the beats identified as x by physician and classified as x by the decision tree).

Once one beat classifier was identified, we continued working with the idea of improving the results of that classifier and we started another training process focusing now on parameters of the scheme j48.part proposed by Quinlan [7]: 1) Determining how deeply to grow a decision tree. 2) Reducing error pruning. 3) Choosing an appropriate attribute selection measurement. 4) Handling training data with missing attribute values. 5) Penalizing bad classifications. With respect to how to penalize bad classification, we fist apply zero/one approach. The 0 and 1 are values used in the zero/one loss approach where the general idea is that in many contexts, the costs of all errors are equal. But in our context to confuse a high risk arrhythmia that requires medical assistance in less than 3 minutes with a low risk arrhythmia may have serious consequences. Therefore to improve the classification *we used information about the specific domain* and we introduced the value 2 in order to penalize bad classifications (see [1] for more details).

Table 1. Beat validation.

===Run information ===	
Scheme:	weka.classifiers.j48.PART -C 0.25 -M 2
Relation: test.txt	Instances: 64260
Attributes: 13	Test mode: evaluation on training data
===Summary===	
Correctly Classified Instances: 21003	96.128%
Incorrectly Classified Instances:846	3.872 %

The results obtained over the validation set are present in table 1 where the scheme is j48.part with the values of two parameters -M 2 that indicates that the minimum number of descendents per node considered is 2; and -C 0.25 that indicates that 0.25 is the threshold of confidence for pruning. The next rows indicate the data source (test.txt), the number of instances (64,260), the number of attributes used (13), the type of test used (evaluation on training data). In the bottom of that table, appear the percentage of correctly classified beats (96.128%), immediately afterwards we show the percentage of incorrectly classified beats (3.872%).

Once the most accurate beat classifier and a set of rules associated with it were obtained, it was necessary to determine an accurate rhythm classifier.

Rhythm classifier. In the specialized cardiologic literature descriptions of arrhythmias can be found. Although, they are not very explicit, it is possible to represent them using a computer language. However, in order to select the most appropriated set of rules we used the following approach: 1) we rewrote the rules, corresponding to the arrhythmia descriptions found in the literature; 2) in parallel, we obtained the arrhythmia rules by using techniques based on decision trees; and last, 3) we used the combination rules that classify arrhythmias and provide competitive results.

Validation of the beat and rhythm classifier: The validation results of the beat and the rhythm classifier are shown in table 2. Considering that MOLEC is a monitoring system that performs a local real-time ECG classification and sends alarms when high-risk arrhythmias are detected, then it is necessary, on the one hand, to classify rhythms depending on the risk of suffering a heart attack, and, on the other hand, to define when an episode is correctly classified.

Table 2. Rhythm validation.

Type	1	2	3	4	5	Total	Correct	Wrong	%
1	8	0	0	0	0	8	8	0	100 %
2	0	48	0	1	0	49	48	1	97.95 %
3	0	0	76	1	3	80	76	4	95 %
4	0	30	23	339	0	392	339	53	86.47 %
5	3	84	40	0	262	389	262	127	67.35 %

In the first column of the table it appears the type of rhythm, according to the level of risk. Rhythms of type 1 (arrhythmias VFL and IVR) require medical assistance in less than 3 minutes. Rhythms of type 2 (arrhythmia VT) require medical assistance in less than one hour. Rhythms of type 3 (arrhythmias T and NOD) are considered predictive arrhythmias, that is, arrhythmias that occur before a heart attack. Rhythms of type 4 are the rest of arrhythmias and rhythms of type 5 are the normal ones. Rhythms of type 1, 2 and 3 are considered high-risk arrhythmias.

An *episode* is a sequence of consecutive beats that appear in a record of the MIT-BIH database, and that are associated with the same rhythm annotation (given by the cardiologists). An episode of type X (for X = 1, 2, 3 or 4) is successfully classified by the system if there is at least one group of four consecutive beats classified as a rhythm of type X. However, a normal episode is successfully classified if all beats [1] in that episode are classified as normal.

In table 2, it can be seen that all episodes of high-risk 1 were correctly classified; 97.95% of episodes of high-risk 2 were correctly classified; 95% of episodes of high-risk 3 were correctly classified. On the contrary, the accuracy percentage for normal sinus rhythms was 67.35% which means that more false alarms would be sent to the hospital.

[1] Except the first and the last beats in that episode, that can be still detecting the previous rhythm or anticipating the next one.

4. Performance of the MOLEC system

Processing ECG signals in the PDA implies running several threads: the thread that performs the preprocessing and classification, the thread that stores the signal and classification results in a local database; the thread that manages the alarms and finally another thread in charge of the communications between the PDA and the Molec Center. The hard restriction is that the thread that preprocesses and classifies the signal has to finish before the signal acquisition obtains the next signal package.

In this section, we answer the next question: how often does the signal processing (preprocessing and clasification) have to be executed? We call processing cycle duration to that time. It is obvious that greater the processing cycle duration is, greater the rhythm detection delay is. The rhythm detection delay grows with the signal package size because at least four beats are needed in order to classify the rhythm, and some beats may delay until the package is completed according to the processing cycle duration. However, the processing cycle duration cannot be very small because the system would get overloaded: the threads of the signal processing have to synchronize with the thread that performs the signal acquisition. And, moreover, it does not have much sense to start a new processing cycle if a new signal package with at least one beat has not yet arrived: no new rhythm can be detected.

Therefore, in order to establish the optimal processing cycle duration we tested the system performance for processing cycles of one and two seconds respectively. The experiment consisted on: 1) running several threads into the device: a signal acquisition thread and all threads involved in the signal processing and classification, storing and visualization; and 2) measuring when the signal package, provided by the signal acquisition, started its processing in the signal processor and classifier threead. Notice that only processing times of those threads into the PDA have an influence on that time and not communication times from the sensors to PDA.

As a result of the experiment we noticed that the processing cycles of one second in the PDA cannot be performed in real time but it was possible with processing cycles of two seconds. Therefore, the optimal processing cycle duration is of 2 seconds and the average rhythm detection delay is of 6.66 seconds (in that period the system performs all the tasks before the next signal package has been arrived).

5. Conclusion

Monitoring systems that perform a complete ECG analysis in a local device near the patients are of great interest because they allow to improve the quality of life of persons that suffer from arrhythmias and reduce communication costs. For an anywhere and at anytime monitoring system, used devices have to be actually mobile. That is why we advocate for using PDAs as the core of these kinds of monitoring systems.

In this paper, we have presented MOLEC: a PDA-based system that performs local real-time classification and detects the ECG anomalies in situ. This solution allows a real-time classification anywhere and at any time where the PDAs can analyze ECG signals, detect anomalies, and make use of wireless communications in order to send those anomalous situations to the control center . The set of rules used to classify the beats have been inferred applying techniques based on decision trees and the set of rules used to classify the rhythms have been obtained in two ways: using rewriting rules found in the medical specialized literature and also techniques based on decision trees. Our ECG classifier provides a very good accuracy for classifying rhythms (100% of accuracy for arrhythmias that require medical assistance in less than 3 minutes -VFL, IVR-, 97.95% for arrhythmias that require medical assistance in less than an hour -VT-, 95% for arrhythmias that usually happen before a worse arrhythmia -NOD, T-, 86.47% for arrhythmias with moderate and low risk of heart attract -AFL, AFIB, PREX, STVA, P, B, SBR- and finally with 67.35% the normal rhythm). Finally, notice that Molec System can run in real-time by using a processing cycle duration of 2 seconds. In that case, the rhythm detection delay is of 6.66 seconds.

References

[1] Rodríguez J, Goñi A, Illarramendi A. Real-time classification of ECG on a PDA. IEEE Transactions of Informations Technology in Biomedicine 2005;9:23–34.

[2] Jané R, Blasi A, García J, Laguna P. Evaluation of an automatic threshold based detector of waveform limits in holter ecg with the qt database. Computers in Cardiology 1997; 24:295–298.

[3] Pan J, Tompkin WJ. A real-time qrs detection algorithm. IEEE Trans Biom Eng 1985;32:230–236.

[4] Rodriguez J, Dranca L, Goñi A, Illarramendi A. A web access to data in a mobile ecg monitoring system. In Duplaga M, Zielinski K, Ingram D (eds.), Transformation of Healthcare with Information Technologies. IOS Press; 2004.

[5] Physionet. The research resource for complex physiologic signals, http://www.physionet.org/.

[6] Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann Publishers, 1999.

[7] Mitchell TM. Machine Learning. Cambridge Univ Press, 1997.

Address for correspondence:

Alfredo Goñi. alfredo@sc.ehu.es