Automated ECG Ventricular Beat Detection with Switching Kalman Filters

Julien Oster[1] and Lionel Tarassenko[1]

[1] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK

Abstract

The exponential rise in availability of clinical data, and especially physiological recordings made using wearables, creates a real need for highly accurate and fully automated analysis techniques. An automated detection of ventricular beat in the ECG is proposed, which is an extension of a recently published switching Kalman filter (skf) approach. The latter technique enables automatic selection of the most likely mode (beat type), and makes novelty detection possible by incorporating a mode for unknown morphologies (X-factor). The previously published technique is semi-supervised and relies on the manual annotation of the different clusters (or modes), thus making it less readily applicable in Big Data scenarios. Here we propose to extend the switching Kalman filter technique by automating the labelling of the modes. Each heartbeat in a mode was classified individually using a feature-based approach, and the cluster was assigned a given type by majority voting. Two different feature-based classifications were tested. First, ecgkit, a state-of-the-art toolkit recently made available online provides a heartbeat classification based on clustering and Linear Discriminant Analysis. Second, a Support Vector Machine (svm) approach was used with the same features as ecgkit. Therefore two different automated switching Kalman filter techniques were tested, ecgkit-skf and svm-skf, that differed only in the way the modes were classified. Both approaches were assessed on an independent subset of the MIT-BIH arrhythmia database (22 individual subjects, 30-minute recordings), and were compared to the semi-supervised switching Kalman filter approach (skf), as well as to the classification techniques, ecgkit and svm. F1 varied from 81.2% *for ecgkit,* 85.4% *for svm,* 91.8% *for ecgkit-skf,* 92.3% *for svm-skf, and* 98.6% *for skf. The proposed combined techniques demonstrated improved automatic beat classification, compared to state-of-the-art fully automated techniques (ecgkit). Performances were however still lower than what was achieved with semi-supervised techniques (skf) highlighting the fact that some clusters were mislabeled.*

1. Introduction

With the exponential rise in the acquisition of physiological data, often for phenotyping purposes, there is an increased importance for the extraction of meaningful information from this vast quantity of data. Cardiac aplications are no exception, and it is widely accepted that big data analytics in cardiology will lead to improved patient outcomes in cardiovacular disease.

Several applications will require the development of robust and fully automated data analysis techniques, among them:(i) telemedicine and in particular mHealth applications as the tool for reaching a wider population and predicting the advent of serious pathologies [1] or (ii) the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses. These applications are supported by the large databases such as Physionet [2], and longitudinal studies such as the UK Biobank[3].

When it comes to the analysis of ECG data, one usually starts with the detection of the QRS complexes. Further feature extraction, such as QT segment or ST levels, can then be performed but that has to be accompanied with heartbeat classification in order to analyse these values on "normal" heartbeats only. Many heartbeat classifiers have been proposed, mainly based on a machine learning approach and relying on the extraction of features representing the temporal evolution of the rhythm, and the morphological changes of the QRS complexes [4].

More recently, Bayesian filtering has been proposed for Ventricular beat detection [5]. This new approach is based on a Switching Kalman Filter (SKF), which allows the automated switching between different modes, and the selection of the mode that best represents the morphology of the heartbeat. Unfortunately, this approach is semi-supervised and relies on an expert, who manually assigns a beat type to the different modes. It is therefore not suitable for analysis of big datasets or of continuous recordings.

In this study, we propose an extension of the SKF based Ventricular beat detection technique by adding an automated cluster type assignment, and thereby fully automating Ventricular beat detection.

Computing in Cardiology 2016; VOL 43 ISSN: 2325-887X DOI:10.22489/CinC.2016.012-314

2. Materials and methods

This section will begin with a short introduction to the recently-proposed SKF approach for ventricular beat detection. In a second subsection, the new automated techniques for assigning a beat type to the clusters will be presented. Lastly, the quality assessment will be described.

2.1. Switching Kalman Filter

A semi-supervised ventricular beat detection has recently been proposed [5]. The technique is based on a Bayesian filtering approach, namely the SKF.

Bayesian filtering is a paradigm aimed at estimating the hidden parameters that govern a given system. This system has therefore to be derivable in a so-called state-space model, which characterises the evolution of the hidden parameters, called the state vector, and links this state vector with the observations being made. Bayesian filtering has been applied to ECG analysis [6], where the state space is derived following the ECG model suggested by McSharry *et al.* [7]. Each heartbeat is modelled as the sum of Gaussian waves, each wave characterising the deflations of the ECG signals, and being represented by three variables (position, amplitude and scale). Bayesian filtering allows for recursively estimating the evolution of these variables with time, and has been applied to a wide range of applications such as denoising, delineation or source separation.

SKF is an extension of Bayesian filtering, which takes avantage of the fact that with Bayesian filtering the level of confidence in the parameter estimation is also freely provided. Monitoring this level of confidence (by monitoring the innovation, which is the difference between a new observation and the value expected by the state-space model) allows for the selection of the most likely mode, in cases for which several modes can explain the behaviour of a given system. In the case of pathological ventricular rhythm, the ECG signal is a “random” succesion of normal and ventricular beats. These beats have different morpholgy, due to the different origins of the electrical excitation. They can therefore be modelled with different parameters. Interested readers are referred to the original paper [5] for further explanations on the SKF approach. It should be noted that an extra mode, called the X-factor, was introduced in order to account for noisy segments but also rare heartbeats, whose morphology cannot be explained by the existing modes.

It is important to highlight that the SKF, as suggested in [5], has a main drawback, which is the need for an expert to annotate the relevant clusters of heartbeats. All heartbeats are gathered in clusters according to their morphology, and these clusters are then assigned a type (Ventricular or Normal) according to the manual annotation of an expert. Such an expert intervention cannot be performed when analysing big datasets, or for near-continuous recordings. The next subsection will therefore be devoted to the presentation of automated ways of assigning the type of the clusters.

2.2. Cluster type assignment

During the initialisation of the SKF, the heartbeats are clustered and the morphology parameters of the relevant modes are estimated. A class needs to be assigned to each of these relevant clusters (or modes). Instead of relying on an expert for manual annotation of the modes, we propose to perform a pre-classification of all the heartbeats of the mode, and assign the class to the mode by majority voting.

Two approaches have been attempted for the pre-classification task:

1. ecgkit: The first approach consists in applying a state-of-the-art method [4], whose code has been recently open-sourced in a large ECG analysis toolbox [8]. The classification is based on machine learning, and so several features (both temporal and morphological) are extracted. Each heartbeat is then assigned a cluster, and assigned a class based on Linear Discriminant Analysis. Finally, all the heartbeats in a given cluster are assigned their final class by majority voting within the cluster. This technique is able to perform a full classification, that is detecting supraventicular beats as well. This property is not of interest for this paper, as our focus is on ventricular beats. Supraventricular beats were therefore considered as normal. Moreover, as the classifier has been trained for a multiclass problem, performance of the detection for the ventricular beat detection might be slightly sub-optimal.
2. svm: The features extracted by ecgkit are re-used in order to train a nonlinear svm to detect ventricular beats only, by using libsm [9]. This new classifier is therefore not as complete as the previous, as it does not offer a full classification, but might have better performance for Ventricular beats detection. As in [4], this classifier has been trained using cross-folds on half of the MIT-BIH arrhythmia database DS1.

Figure 1 summarises the different steps for the automated SKF. Note that the extra step, which consists of the automated cluster type assignment is represented at the bottom of the figure.

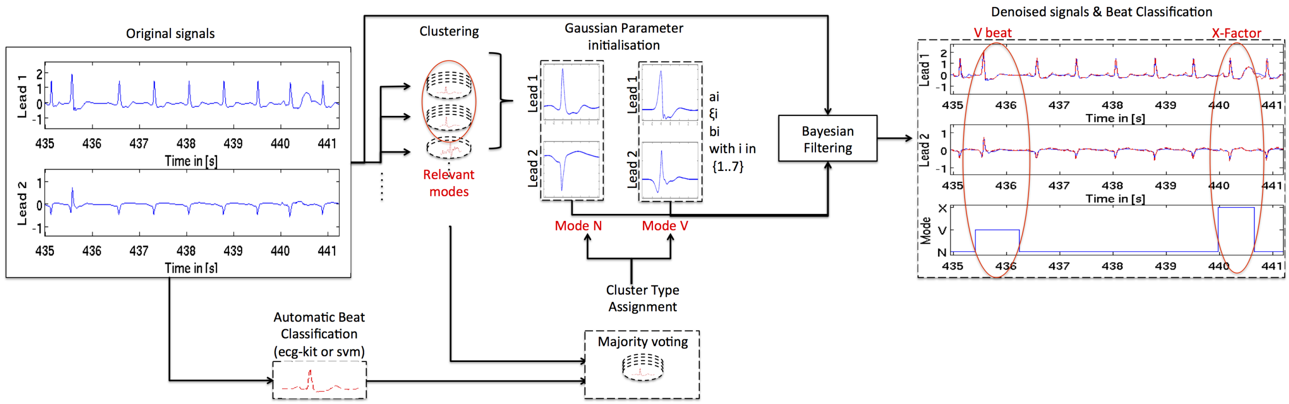

Figure 1. Logic flow of the proposed method. The original ECG signals are depicted on the left. The beats are then clustered, and the Gaussian parameters for the relevant modes are initialised. At this point, each mode can be assigned a class following majority voting, each heartbeat in the mode having been previsouly classified either with ecgkit or svm. The original signals are then filtered using the Bayesian approach and the knowledge provided by the morphological model. The output of the method (right) is the filtered signal and beat classification. An example of V beat detection and of X-factor beat detection are circled in red. This procedure is performed for each new recording.

2.3. Quality assessment

The performance of the new automated SKF was assessed on half of the MIT BIH arrhythmia database [2], on DS2. DS2 consists of 22 recordings lasting 30-minutes each, and contains more than 46,000 normal beats, and more than 3,000 ventricular ones. This ensures that the evaluation is performed on data entirely independent from the training data. The two proposed approaches, called "ecgkit-skf" or "svm-skf" according to the cluster type assginment used, were compared with the automated "ecgkit" approach [4], the ventricular beat detection "svm", and the the semi-supervised technique "skf".

The proposed technique was assessed using sensitivity (Se) and positive predictive value (PPV) as suggested in[10], but also in terms of F1 (which is the harmonic mean of Se and PPV, and penalizes False Positives and False Negatives equally).

3. Results

The results are summarised in the table 1.

The best results are obtained by the semi-supervised technique, which benefits from the expertise of the cardiologists to properly assign the cluster type.

Among the automated techniques, the skf approaches perform better than the other two approaches. Svm-skf is the best performing technique in terms of F1 and PPV, but not for Se. Ecgkit approaches yield the best Se of any automatic techniques, but this comes at the cost of a relatively low PPV.

Finally, it has to be noted that combining an automated heartbeat classifier with the SKF seems to be working effictively, as both the "ecgkit-skf" and the "svm-skf" are outperforming the direct approach, that is "ecgkit" and "svm", with a gain of 10% and 7% respectively.

Metric	Automated				Semi-supervised
	ecgkit	svm	ecgkit-skf	svm-skf	skf
Se	**93.1**	83.6	92.9	88.4	**97.3**
PPV	72.0	87.3	90.7	**96.5**	**99.9**
F1	81.2	85.4	91.8	**92.3**	**98.6**

Table 1. Performance of the different automatic beat classification (ecgkit, svm, ecgkit-skf, svm-skf) and the supervised approach (skf) on the testset of the MIT-BIH arrhythmia database.

4. Discussion & Conclusion

The results presented in the previous subsection show that the automated SKF approach leads to an increased performance compared to state-of-the-art automated techniques. This highlights the capability of the SKF to correctly follow morphological changes. The svm-based cluster type assignment method gives the best results, mainly due to the fact that this classifier only focuses on Ventricular beat detection, as both ecgkit and svm are based on the same features.

Nevertheless, the semi-supervised technique is still outperforming the automated approaches with at least a 6% higher value of F1. This shows that the cluster type assignment can still be improved. One can for example consider a Bayesian voting approach, which has been proven to be more robust than simple majority voting [11]. Another improvement might come from better feature extraction. Data-based feature extraction techniques in Big Data is an interesting avenue of research[12].

This study assumed perfect beat (or QRS) detection. It would be interesting to assess how imperfect beat detection will affect the overal performance. One can assume that wrongly detected beats might be assigned to the X-Factor class, as the morphology of the overall beat is likely to be atypical and not ressemble any of the existing modes.

Approximatively 2.5% of the heartbeats have been classified as X Factor for the svm-skf technique, which is in the same range as for the semi-supervised skf (approximatively 3%). Automatically excluding heartbeats, for which there is large uncertainty in the analysis, is one of the key feature of the SKF approach. Analysing big datasets (acquired in uncontrolled environments, and not having been manually annotated) requires to be able the automated exclusion of poor-quality segments for which analysis is not reliable.

One of the main limitations of this approach comes from the fact that it is limited to Ventricular beat detection. It is therefore not possible to detect supraventricular rhythm. Several approaches could be considered for accounting for such rhythms, (i) either with a two-stage approach, the first one for the detection of Ventricular beats and the second one based on rhythm features only; (ii) or more elegantly by incorporating rhythm based state variables in the state-space of the Bayesian filter and modelling their evolution for the different types of possible pathologies.

A novel automated ventricular beat detection has been presented, based on Bayesian filtering, which switches automatically between different modes according to the morphology of the heartbeat. The type of this mode has been initialised automatically by applying an existing classifer on all the heartbeats belonging to a given cluster, whose type is assigned by majority voting. Such an approach allows high performance while being fully automated and therefore being suitable for the analysis of big datasets.

Acknowledgments

This study has been supported through the NIHR Oxford Biomedical Research Centre, *Biomedical Informatics and Technology*,

References

[1] Clifford GD, Clifton D. Wireless technology in disease management and medicine. Annual Review of Medicine 2012;63:479–492.

[2] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

[3] Collins R. What makes uk biobank special? The Lancet 2012;379(9822):1173–1174.

[4] Llamedo M, Martínez JP. An Automatic Patient-Adapted ECG Heartbeat Classifier Allowing Expert Assistance. Biomedical Engineering IEEE Transactions on 2012; 59(8):2312–2320.

[5] Oster J, Behar J, Johnson AEW, Sayadi O, Nemati S, Clifford GD. Semi-supervised ECG beat Classification and Novelty Detection based on Switching Kalman Filters. Biomedical Engineering IEEE Transactions on 2015; 62:2125–2134.

[6] Sameni R, Shamsollahi MB, Jutten C, Clifford GD. A nonlinear bayesian filtering framework for ecg denoising. IEEE Transactions on Biomedical Engineering 2007; 54(12):2172–2185.

[7] McSharry P, Clifford GD, Tarassenko L, Smith LA. A Dynamical Model for Generating Synthetic Electrocardiogram Signals. Biomedical Engineering IEEE Transactions on 2003;50:289–294.

[8] Demski A, Soria ML. ecg-kit: a matlab toolbox for cardiovascular signal processing. Journal of Open Research Software 2016;4(1).

[9] Chang CC, Lin CJ. Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology TIST 2011;2(3):27.

[10] ANSI/AAMI:EC57. Testing and reporting performance results of cardiac rhythm and ST-segment measurement algorithms, 1998.

[11] Zhu T, Dunkley N, Behar J, Clifton DA, Clifford GD. Fusing continuous-valued medical labels using a bayesian model. Annals of biomedical engineering 2015; 43(12):2892–2902.

[12] Bengio Y, Courville AC, Vincent P. Unsupervised feature learning and deep learning: A review and new perspectives. CoRR abs12065538 2012;1.