Automatic Classification of Arrhythmic Beats Using Gaussian Processes

G Skolidis, RH Clayton, G Sanguinetti

Department of Computer Science, University of Sheffield, Sheffield, United Kingdom

Abstract

We propose a novel approach to the automated discrimination of normal and ventricular arrhythmic beats. The method employs Gaussian Processes, a non-parametric Bayesian technique which is equivalent to a neural network with infinite hidden nodes. The method is shown to perform competitively with other approaches on the MIT-BIH Arrhythmia Database. Furthermore, its probabilistic nature allows to obtain confidence levels on the predictions, which can be very useful to practitioners.

1. Introduction

Cardiac arrhythmias are one of the major causes of morbidity and mortality in the Western world. Their early diagnosis is often reliant on an analysis of electrocardiogram (ECG) traces, generally involving time-consuming manual annotation by expert physicians. Because of this, several automated methods to detect arrhythmic beats have been proposed, often achieving very good levels of performance [1–3].

We present a novel approach for the automatic classification of arrhythmic versus normal beats from ECG signals based on recent developments in Machine Learning. We use the framework of Gaussian Process (GP) classification [4], a non-parametric Bayesian technique which has been shown to be highly accurate on non-linear classification tasks while controlling complexity and avoiding the pitfalls of overfitting. GPs are a natural way to define probability distributions over spaces of functions; they can be viewed as a generalization of Neural Networks where the number of hidden nodes (basis functions) tends to infinity [5]. A key feature of GPs is their probabilistic nature, which means that predictions are always accompanied by an estimate of the associated uncertainty. This is a key advantage over standard non-linear classifiers such as neural networks which generally can only provide a hard assignment.

The method uses as input the spectral or wavelet transform of segmented individual beats from a recording, which requires much less manual annotation than methods based on interval estimation. We use an Automatic Relevance Determination (ARD) kernel for the classifier to automatically reduce dimensionality and extract the most discriminant features by optimising weights.

We test the model on the MIT-BIH arrhythmia data set on the two class problem of discriminating normal and premature ventricular contraction beats (PVC). The results we report show that the method is competitive with the state of the art, obtaining predictive accuracies on test data which are frequently above 90%. This can be further increased by thresholding over posterior probabilities and retaining only predictions with high confidence; the model consistently has a higher accuracy for prediction made with higher posterior probability, indicating that the discriminant obtained from the training data mirrors the structure of the whole data set.

The rest of the paper is organised as follows: in the first section, we briefly review Gaussian Process classification. In the second section, we discuss the beat segmentation algorithm and the feature selection procedure. We then present our results on real ECG data, and conclude the paper with a discussion of the strengths and weaknesses of the method, as well as the possible future extensions.

2. Methods

2.1. Gaussian Processes for classification

In this section we briefly review the statistical foundations of our approach; for a thorough review, the reader is referred to [4]. A Gaussian Process (GP) is a (finite or infinite) collection of random variables any finite subset of which is distributed according to a multivariate normal distribution. As a random function $f(\mathbf{x})$ can be seen as a collection of random variables indexed by its input argument, GPs are a natural way of describing probability distributions over function spaces. A GP is characterised by its *mean function* $\mu(\mathbf{x})$ and *covariance function* $k(\mathbf{x}, \mathbf{x}')$,

a symmetric function of two variables which has to satisfy the Mercer conditions ([4]). In formulae, the definition of GP can be written as

$$\begin{aligned} \mathbf{f} &\sim \mathcal{GP}(\mu, \mathcal{K}) \Leftrightarrow [\mathbf{f}(\mathbf{x_1}), \ldots, \mathbf{f}(\mathbf{x_N})] \\ &\sim \mathcal{N}([\mu(\mathbf{x_1}), \ldots, \mu(\mathbf{x_N})], \mathcal{K}) \end{aligned} \quad (1)$$

for any finite set of inputs $\mathbf{x_1}, \ldots, \mathbf{x_N}$. Here

$$K_{ij} = k(\mathbf{x_i}, \mathbf{x_j}).$$

The choice of mean and covariance functions is largely determined by the problem under consideration. In this paper, we will use a zero mean GP with ARD covariance function, in order to automatically select the most relevant features from a high dimensional input space [6]

$$k(\mathbf{x_i}, \mathbf{x_j}) = exp\left(-\frac{1}{2}(\mathbf{x_i} - \mathbf{x_j})^T \Lambda (\mathbf{x_i} - \mathbf{x_j})\right), \quad (2)$$

where Λ is diagonal, with Λ_{ii} denoting the precision (inverse characteristic length-scale) of each feature of the input matrix.

Given some observations $\mathbf{y}$ of the function $\mathbf{f}$ at certain input values X, and given a noise model $p(\mathbf{y}|\mathbf{f}, \mathbf{X})$, one can use Bayes' theorem to obtain a posterior over the function values at the inputs

$$p(\mathbf{f}|\mathbf{y}, \mathbf{X}, \theta) = \frac{p(\mathbf{y}|\mathbf{f}, \mathbf{X}, \theta)\, p(\mathbf{f}|\mathbf{X}, \theta)}{p(\mathbf{y}|\mathbf{X}, \theta)} \quad (3)$$

where θ denotes the parameters of the GP prior (ARD parameters). One can then obtain a predictive distribution for the function value f^* at a new input point $\mathbf{x}^*$ by averaging the conditional distribution of $p(f^*|\mathbf{f})$ under the posterior (3)

$$p(f^*|\mathbf{y}, \mathbf{X}, \mathbf{x}^*, \theta) = \int p(\mathbf{f}^*|\mathbf{f}, \mathbf{X}, \mathbf{x}^*, \theta)\, p(\mathbf{f}|\mathbf{y}, \mathbf{X}, \theta)\, d\mathbf{f}.$$

If the noise model $p(\mathbf{y}|\mathbf{f})$ is Gaussian, then we are dealing with a regression problem and one can obtain an analytical expression for the posterior (3). In classification, the noise model is non-Gaussian; in this paper, we will take the noise model to be the logistic function

$$p(y = 1|f) = \frac{1}{1 + \exp(-f)}.$$

In this case, the denominator of equation (3) cannot be computed analytically and one must seek approximate solutions. In this paper, we use the Laplace approximation [7]. This computes a second order Taylor expansion of the un-normalised posterior $p(\mathbf{y}|\mathbf{f}, \mathbf{X}, \theta)\, p(\mathbf{f}|\mathbf{X}, \theta)$ about its mode and then approximates the true posterior distribution with a Gaussian centered at the true mode and with covariance given by the Hessian of the un-normalised posterior.

2.2. Experimental setup

In this study experimental data were taken from the MIT-BIH Arrhythmia database [8], for training and evaluation purposes of the proposed classifier. Specific recordings were selected according to the exhibited type of arrhythmia; each recording was sampled at $360Hz$, and had sufficient amount of Normal and premature ventricular contraction (PVC) beats, for training and evaluating the model. Annotation provided by the database was used to separate the beats before any preprocessing.

2.3. Data processing

Two different types of transforms were considered in the analysis of the ECG signal. The first one is based on the Fourier Transform while the second one on the Wavelet Transform.

2.3.1. Fast Fourier Transform (FFT)

Each beat segment, consisting of 360 data points (one minute), was transformed into the frequency domain using a Fast Fourier Transform with a Hanning window. The frequency based representation of each beat consisted of 180 frequencies, since it was sampled at $360Hz$.

2.3.2. Wavelet Transform (WT)

The second type of features were obtained by the Discrete Wavelet Transform. The Continuous Wavelet Transform (CWT) of a signal $x(t)$ is defined as:

$$W_a x(b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \psi\left(\frac{t-b}{a}\right) \mathrm{dt},\ \mathrm{a} > 0. \quad (4)$$

The discrete wavelet transform uses a dyadic scale factor $a = 2^k$ for $k \in \mathbf{Z}^+$. The wavelet that was used in this work was from the Daubechie family [9]. It is noted that the high frequency phenomena of a signal are captured at the smallest scales, namely 2^2 and 2^1,while most of the details of the signal are contained from the third to the fifth scale. Consequently, coefficients from these three scales were selected as input features for the classification, resulting in 900 features.

3. Results

3.1. Feature selection

To determine which frequencies are more relevant for classification, for the features acquired from FFT, ARD was used. Using ARD the five most relevant frequencies were identified and then the characteristic length scale of

Table 1. Classifier Performance in terms of Accuracy (%)(**HC**:High Confidence)

	FFT			WT		
Recording	**Accuracy Thresh. 0.5 (%)**	**Accuracy Thresh. 0.8 (%)**	**Data of HC(%)**	**Accuracy Thresh. 0.5 (%)**	**Accuracy Thresh. 0.8 (%)**	**Data of HC(%)**
106	93.2	95.8	89.23	98.61	99.53	95.58
119	100	100	100	99.84	99.89	99.49
200	98.32	99.08	97.61	97.61	98.53	95.97
203	87.9	93.37	80.79	96.9	98.53	95.55
221	96.16	96.7	97.18	96.16	96.92	95.69
223	88.43	96.8	71.43	90.67	96.54	81.59
228	99.8	99.8	99.8	99.11	99.65	98.43
233	97.96	98.78	96.98	96.19	99.35	90.94

each input feature was optimized for each recording. Thorough experimental research indicate that using two features (frequencies) achieves better performance, instead of using a larger number of features.

The features obtained from the Wavelet transform, were projected into a two-dimensional space, using Principal Component Analysis (PCA). After PCA, the characteristic length scale of each feature was estimated again using ARD.

3.2. Performance evaluation

The performance measure that was used for the evaluation of the classifier is simple misclassification error. In the Gaussian Process framework, the misclassification error is computed by setting a threshold over the posterior probabilities, since GPs produce a measure of uncertainty instead of giving hard assignments to a class. This concept can be further cultivated, by setting a threshold of 0.8 and retaining test samples that have been assigned with posterior probabilities higher than 0.8. In this way, a measure of high confidence is obtained to evaluate the classifier.

For example figure 1 and 2 show the decision boundaries along with the data points in the Euclidean space,that were created by GPs of the wavelet transformed input data of recording 223. It is clearly observable, that high confidence regions are produced in the input space where the density, of the training data of each class, is high.

Table 1 shows the performance of the classifier, with optimized hyperparameters, in terms of the accuracy the test data set achieves. The first column of each transform indicates the accuracy the classifier achieves with a threshold of 0.5. The other two columns show the accuracy of the classifier when a threshold of 0.8 is used and the proportion of data that have been assigned with probabilities higher than 0.8.

Each recording was trained and tested four times. The beats in each recording was separated into four disjoint subsets, preserving the initial prior probabilities of each class. In each run three data subsets were used for training and one for testing. The accuracy each recording achieves in table 1 is the mean accuracy of the four trainings.

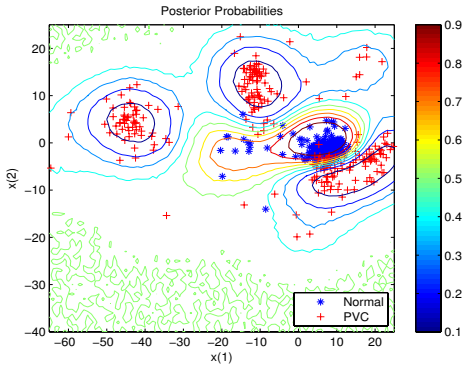

Figure 1. Decision Boundaries of recording 233(contour plot) - Scatter plot; where x(1), x(2) represent the features obtained from the wavelet transform after PCA

Figure 3 illustrates the effect the increase of the threshold has on the accuracy (solid line), and the proportion of the test data set, that has been assigned probabilities higher than the threshold (dashed line). It is noticed that the accuracy remains high but the proportion of the data that have been assigned with probabilities higher than the value of the threshold, decreases as the threshold increases.

4. Discussion and conclusions

In this paper we propose the use of Gaussian Processes for automatically classifying ECG signals into normal and ventricular beats. Furthermore, Automatic Relevance Determination was applied for the identification of the frequencies that were most relevant for classification, and then for the optimisation of the hyperparameters of the covariance function. A different methodology was followed for the features extracted by the wavelet transform of the raw ECG signal, where first PCA was applied to reduce the dimensionality of the inputs, and then ARD was used

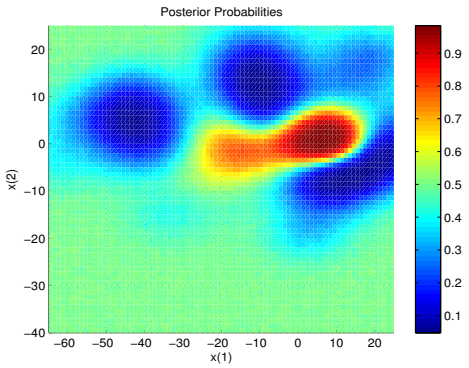

Figure 2. Decision Boundaries of recording 233

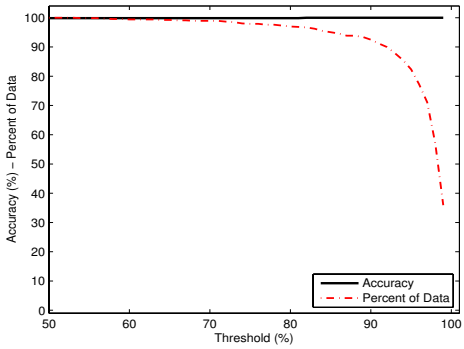

Figure 3. Model Accuracy as the threshold increases - Percent of test data set larger than the threshold

again to optimise the hyperparameters. On both types of features obtained, results indicate that Gaussian Processes are shown to perform with high precision, with an average accuracy above 90%. Moreover, a measure of performance exceeding 95% of accuracy is achieved, by considering only posterior probabilities of high confidence above a certain threshold.

Future work will emphasize on the use of different features for classifying beat hearts with Gaussian Processes, as well as the investigation of applying different approximation methods to the non-Gaussian likelihood. Moreover, future work will focus on extending the predictions to different subjects.

Acknowledgements

The authors would like to thank EPSRC for providing financial support for this project, under the grant R/115710-11-1.

References

[1] S. AFA, I. H. Combined wavelet transformation and radial basis neural network for classifying life-threatening cardiac arrhythmias. Med Bio Eng Comput 1999;37(1):566–573.

[2] Minami K, Nakajima H, Toyoshima T. Real-time discrimination of ventricular tachyarrhythmia with fourier-transform neural network. Biomedical Engineering IEEE Transactions on Feb 1999;46(2):179–185. ISSN 0018-9294.

[3] de Chazal P, Reilly R. Automatic classification of ecg beats using waveform shape and heart beat interval features. Acoustics Speech and Signal Processing 2003 Proceedings ICASSP 03 2003 IEEE International Conference on April 2003;2:II–269–72 vol.2. ISSN 1520-6149.

[4] Rasmussen CE, Williams CK. Gaussian Processes for Machine Learning. MIT press, 2005.

[5] Neal RM. Bayesian Learning for Neural Networks. Ph.D. thesis, University of Toronto, Canada, 1994.

[6] Neal RM. Bayesian Learning for Neural Networks. Springer, 1996. Lecture Notes in Statistics 118.

[7] Williams CKI, Barber D. Bayesian classification with gaussian processes. IEEE Trans Pattern Analysis Machine Intelligence 1998;20(12):1342–1351.

[8] Goldberger AL, Amaral LAN, 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 (June 13);101(23):e215–e220. Circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215.

[9] Daubechies I. Ten lectures on wavelets. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1992. ISBN 0-89871-274-2.

Address for correspondence:

Grigorios Skolidis
Dept. of Computer Science / University of Sheffield
Regent Court, 211 Portobello
Sheffield, S1 4DP, United Kingdom