Classifying Ischemic Events Using a Bayesian Inference Multilayer Perceptron and Input Variable Evaluation Using Automatic Relevance Determination

MG Smyrnakis, DJ Evans

Aston University, Birmingham United Kingdom

Abstract

In this paper we present a Bayesian inference Multilayer Perceptron (MLP) which was used to classify the events of the Long Term ST Database (LTSTDB) as ischaemic or non-ischaemic episodes with an accuracy of 89.1%, sensitivity of 82.3% and specificity of 91.2% when the accuracy of the winning paper was 90.7%. The Automatic Relevance Determination (ARD) method was used to identify which of the extracted features that were used as input in the Bayesian inference MLP were the most important with respect to the models performance. ARD indicated that ΔT, a combination of the ST deviation and the duration of the episode, inspired from Langley et al [1], was the most important feature for determining Ischaemic episodes, given the data. A simple MLP which had as input variable of only ΔT was trained to verify the results of the ARD method. The classification accuracy was 85.8% on the test set. We can conclude from the results that the most important extracted feature was ΔT.

1. Introduction

Myocardial ischaemia is one of the most common fatal diseases of the western industrial world. It is a heart problem which is caused by the lack of oxygen and nutrients to the contractile cells (muscles) and leads to dangerous arrhythmias and myocardial infractions. The methods which are employed to detect myocardial ischaemia are based on the measurement of blood flow and oxygen supply of the heart. Two of these methods are coronary angiography and exercise test which are either very expensive or very exhaustive for the patients . These are the reasons why these methods are applied only to high risk patients.

Myocardial ischaemia can also be detected from the abnormalities that are depicted in the ST segment of the electrocardiogram (ECG) despite the fact it does not contain any information about the blood flow and the oxygen supply of the heart [2]. This method is cheaper than coronary angiography and demands less effort from the patient than the exercise test. Nevertheless, we should bare in mind that apart from myocardial ischaemia the abnormalities that are observed in the ST segment of an electrocardiogram can be also the result of many other factors such as changes in the heart rate, the position of the subject, noise in ECG.

The development of a classifier that will be able to identify whether the changes in the ST segment are caused from ishaemia or from other reasons, was the challenge of Physionet and Computers in Cardiology of 2003.

The Long term ST Database (LTSTDB) was used for the purposes of this challenge. LTSTDB contains 86 records of 19-26 hour ECG, 43 of these records are available from Physionet as a training set for the algorithms which were developed to detect ischaemia.

In each one of these 43 records the significant ST episodes were specified. The procedure which was used in order to determine these episodes is [3]:

a) an episode begins when the ST deviation exceeds the 50μV

b) the ST deviations must reach a threshold value V_{min} for a period of time T_{min}

c) the episode ends when the ST deviation becomes less than 50μV continuously in the following 30 seconds

Three different protocols were used to define whether these events were ischaemic or non-ischaemic according to different combinations of the values V_{min} and T_{min}. The protocol which was used for this paper was the protocol b. The values of V_{min} and T_{min} for that protocol were 100 μV and 30 seconds respectively. Using these values of V_{min} and T_{min} we located 1772 episodes in the LTSTDB, 1369 of which were non-ischaemic and the rest 403 were ischaemic.

In this work we also used the Physionet annotation files in order to be determined the J point of the beginning of each episode, the position of the R peak and the values of the ST deviation for each episode.

The remainder of this paper is set out as follows: The next section contains a brief description of the methods which were used in order to obtain the results. The third section contains the results. In the last chapter there is a discussion over the methods which were used, a comparison between the results obtained in the test set of the Bayesian inference MLP and the results of the validation set of each entry in the 2003 and 2005 Physionet challenges and in the end the conclusion.

2. Methods

For each available episode of LTSTDB the signal from the J point of the first beat to 80 ms before the next R peak was extracted. When the duration of that segment was less than 70 ms then the remaining signal was padded with zeros. In the case of the duration being more than 70ms then the signal of the first 70 ms was extracted. Our initial data set was 1772 segments of the first beat of each episode. The first four principal components which represented at least the 90% of the variance of this dataset were extracted.

Two more features were used. The value of the ST deviation (ΔST) at the beginning of each episode and a combination of ΔST and the duration of each episode (ΔT), inspired from Langley's algorithm [1]. ΔT is the difference between T_e and T_s, where T_s is the time at the beginning of each episode, and T_e can be computed as follows.

If the value of ΔST is greater than the value V_{thres}=50 mV then the value of T_e is the time of the starting point of a time interval where the value of ΔST will be less than V_{thres} for at least 40 seconds. If ΔST is smaller than the threshold value then $T_e = T_s$ and ΔT is equal to zero. Figure 1 depicts a schematic representation of the procedure that is used for the determination of ΔT.

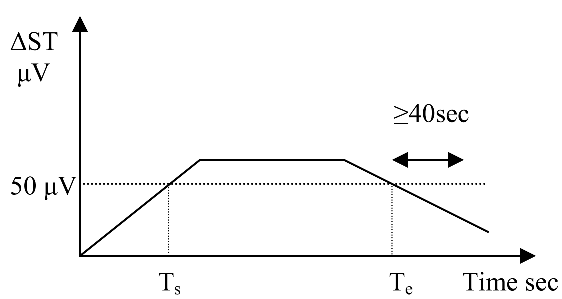

Figure 1-ΔT extraction procedure

So we concluded with six features the four principal components of the initial dataset, ΔST and ΔT. These constitute the 1772 observations of the dataset, one for each episode of the LTSTDB.

The MLP is a feed-forward neural network. For N number of input units, H hidden units and K outputs the formula of the MLP is the following:

$$y_k = f\left(\sum_{j=0}^{H} w_{kj}^{(2)} g\left(\sum_{i=0}^{N} w_{ij}^{(1)} x_i\right)\right)$$

where y_k is the k^th output of the MLP, f is the output activation function, $w_{kj}^{(2)}$ are the weights of the j^th hidden unit of the k^th output, g is the hidden layer activation function, $w_{ji}^{(1)}$ are the weights of the i^th input of the j^th hidden unit and x_i is the i^th input [4].

A Bayesian inference MLP is a classifier that combines the theory of MLP with Bayes theorem using probabilities. Adopting Bishop's notation [4] for the weights of an MLP we have that:

$$p(w \mid D) = \frac{p(D \mid w)p(w)}{p(D)}$$

where p(D/w) is the probability of the data given the weights, p(w) is the prior distribution of the weights and p(D) is a normalization factor. Usually a Gaussian prior is used for the weights distribution. The form of that prior is

$$p(w) = \frac{1}{Z_w(a)} \exp(-aE_w)$$

where $Z_w(a)$ is a normalization factor of the form: $\int p(D \mid w)p(w)dw$ and E_w is a regularization factor.

For classification problems cross-entropy error function is used. The error function log-likelihood becomes p(D|w)=exp(-G(D|w)) where G is the cross-entropy function. The function of the weights become: $p(w \mid D) = \frac{1}{Z_s} \exp(-G - aE_w)$ where Z_s is a normalization constant.

Since we have defined the distribution of the weights and the error function we can determine the form of the output distribution. The output will have the following form: $p(C_1 \mid x, D) = \int g(a)p(a \mid x, D)da$ where g is the logistic activation function. An approximation of that integral proposed by Mackay is the following: $p(C_1 \mid x, D) = g(k(s)a_{MP})$ where $k(s) = (1 + \frac{\pi s^2}{8})^{-1/2}$ α_{MP} is the hyperparameter α which maximize the posterior distribution of the weights, and s is the standard deviation of the hyperparameters distribution. To determine α_{MP} we could integrate over the hyperparameters or use the evidence procedure [5] which is an iterative method equivalent to type II maximum likelihood.

Automatic Relevance Determination (ARD) is a method that uses Bayesian inference to identify the variables of the model which are more important than the others [6]. That can be achieved by setting a different hyperparameter *α* to each variable. Since the hyperparameter *α* is equal to the inverse of the variance, low values of α indicate a greater variance of the distribution of weights. That is essential because wider distributions mean that the range of the weights for the specific variable will be wider. Whenever the neural network allows the weights to have a high value, that indicates the importance of these weights for the final result, as they will dominate the output of the classifier in contrast to the other variables that have smaller values. So we can determine which of the features are more important by comparing the values of the corresponding hyperparameter [7]. There is no impartial approach to deciding which hyperparameter represents an important variable and which corresponds to a non significant feature, especially in the case of variables having different mean and variance. That is the reason why in ARD method the variables which are used are preferred to have zero mean and unit variance.

From the 1772 observations the 886 were used as training set and the rest were used as test set, since the regularization factor is included in the way of the estimation of the Bayesian inference and there is no need for a validation set. For the MLP with input ΔT the test set was split into two equal parts into validation and test set. Early stopping technique was employed for regularization and that is the reason why the existence of the validation set was essential.

A Bayesian inference MLP with 6 inputs, 8 hidden units and 1 output was trained for 1400 iterations, using evidence procedure to assess the most probable hyperparameter α_{MP}. In addition the hyperparameter α was split into six parts a_1-a_6 to implement the ARD method.

The MLP which was used to verify the results of the ARD method had 1 input, 4 hidden units, one output and the training stopped after 700 iterations. The scaled conjugate gradient algorithm was used. for the optimization of the weights [4,7].

3. Results

This section contains all the results, of the test set, of the Bayesian inference MLP, of the ARD method which was used to identify which of the extracted features were more important and also the results of the validation and test set of an MLP with input only ΔT.

As it is depicted in Table 1 for the test set of the Bayesian inference MLP the accuracy was 89.1%, the sensitivity 82.3% and specificity 91.2 %. From these results we can deduce that this algorithm classifies more accurately the non-ischaemic episodes than the ischaemic since the sensitivity is approximately 9% smaller than specificity.

Table 1 – Results of the test set of the Bayesian inference MLP

Accuracy	Sensitivity	Specificity
89.1 %	82.3 %	91.2 %

The ARD method was used to evaluate the extracted features that used in the Bayesian inference MLP. The results are depicted in table 2. The variables α_1-α_4 represents the corresponding hyperparameters of the principal components 1-4 respectively. The variable α_5 is the hyperparameter of the variable ΔST, and α_6 is the hyperparameter that corresponds to the variable ΔT. The feature with the smaller hyperparameter is α_6 with value α=0,015. We can conclude from the above that the most important feature for the determination of the output of the classifier with respect to the data is the variable ΔT.

After the results of the ARD method an MLP with input variable only ΔT was trained to verify the results of ARD. The overall results in the validation set were worse than these of the test set.

Table 2 – Results of the ARD method for the hyperparameters of the input variables.

α_1	α_2	α_3	α_4	α_5	α_6
0.252	0.634	0.101	0.091	0.231	0,015

The greater difference between the results of the two data sets was observed in their ability to identify the ischaemic episodes. As it depicts in Table 3 in the test set sensitivity was 12% greater than validation set, since the sensitivity for the validation set was 66.3% and for the test set was 78.3%. The difference in accuracy and specificity of the validation and test set were smaller. For the validation set the accuracy was 79.9% and the specificity 84.0%. The accuracy of the test set is 85.8 and specificity was 87.6%.

Table 3 – Results of validation and test set of the MLP with input ΔT

	Accuracy	Sensitivity	Specificity
Validation set	79.9%	66.3 %	84.0 %
Test set	85.8%	78.3 %	87.6 %

We observe that the results of the test set were similar to these of the test of the Bayesian inference MLP which had as input the six extracted features. This is additional verification of the results of the ARD, using the specific dataset that, that the combination of the ST deviation with the duration of an episode is a very important feature for the development of a classifier which distinguishes ischaemia based on the ST segments of ECG.

4. Discussion and conclusions

The accuracy of the Bayesian inference MLP was 89.1%. As it depicts in table 4 this was 1.6% smaller than the accuracy of the challenge's test set achieved from Langley et al [1] who was the winner of the challenge of 2003. Comparing the results with the other two entries we notice that the results of Bayesian inference MLP were better (see Table 4 for comparison). The results of sensitivity and specificity are available for the entries of Zimmerman et al [8] and Povinelli [9]. The results were better than these of Zimmerman et al [7] in accuracy 10% higher, in sensitivity 80.6% higher and in specificity 12.3% higher. Two classifiers were proposed from Povinelli [9] which are based on different extracted features. The first one used as input the Reconstructed Phase Space (RPS) of the ST segment and T-wave and the second had as input the five first Principal Components. The classifier with input the RPS had worse results in all the measures (accuracy, sensitivity and specificity). The second classifier had better results in specificity 98.5%. Despite the very good results in specificity, the sensitivity was only 2%. That made the overall accuracy to be 50.3% which is not significantly different from chance [9], so the algorithm had worse overall results when it was compared with the other entries and the results of the Bayesian inference MLP.

Table 4 – Comparison of the results between the entries of the Physionet and Computers in Cardiology challenges of 2003 and 2005 and Bayesian MLP

Methods	Accuracy	Sensitivity	Specificity
Bayesian inference MLP	89.1%	82.3%	91.2%
Rule based [1]	90.7 %	–	–
Reconstructed Phase Space [8]	79.1 %	80.6 %	78.9 %
GMM & Principal Components [9]	50.3 %	2.0 %	98.5 %
Reconstructed Phase Space [9]	54.0 %	74.6 %	33.5 %

Concluding, we have proposed an algorithm based on Bayesian inference MLP which has very similar results to the rule based algorithm. of the winner of the challenge. The results of the ARD method showed that the specific combination of the ST deviation and the duration of the episode is a very important feature for an automated detector of ischaemia. The training of a simple MLP with only one input ΔT verified that ΔT was the most important extracted feature for the algorithm, since the accuracy of that classifier was 85.8%.

Acknowledgements

Michail Smyrnakis was supported for his MSc studies from the public welfare institution Bakala.

References

[1] Langley P, Bowers E, J. Wild MD, et al. An algorithm to distinguish ischaemic and non ischaemic ST changes in the Holter ECG, Computers in Cardiology, Thessaloniki, Greece, 2003;30;239-242

[2] Moody G. B, Jager F. Distinguishing ischemic from non-ischemic ST changes: The Physionet/Computers in Cardiology challenge. Computers in Cardiology 2003;41;235-238

[3] Jager F, Taddei A, Moody G, et al. Long-term ST database: A reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia. Medical and Biological Engineering and Computing. 2003;41;172-182

[4] Bishop M. C. Neural networks for pattern recognition. Oxford University press,1995

[5] Mackay D. The evidence framework applied to classification networks. Neural Computation. 1992;4;720-736

[6] Bayesian Non-linear modeling for the 1993 energy prediction competition. Maximum entropy and Bayesian Methods. 1993

[7] Nabney T.I. Netlab algorithms for pattern recognition. Springer 2001

[8] Zimmerman MW, Povinelli P, Johnson MT, et al. A reconstructed phase space approach for distinguishing ischemic from non-ischemic ST changes using Holter ECG data, Computers in cardiology, 2003;41;243-246

[9] Povinely R. J. Towards the prediction of transient ST changes. Computers in Cardiology 2005;32;663-666

Address for correspondence

Michail Smyrnakis
Room 1.19, School of Mathematics
University of Bristol, University Walk,
Bristol BS8 1TW
UK
m_smirnakis@yahoo.gr

Dr David Evans
Aston University, Room 271
Aston Triangle
Birmingham B4 7ET
UK
d.j.evans@aston.ac.uk