Probabilistic Classification Approaches for Cardiac Arrest Rhythm Interpretation during Resuscitation

Ali Bahrami Rad[1], Trygve Eftestøl[1], Jan Terje Kvaløy[1],
Unai Ayala[2], Jo Kramer-Johansen[3], Kjersti Engan[1]

[1] University of Stavanger, Stavanger, Norway
[2] University of the Basque Country (UPV/EHU), Bilbao, Spain
[3] Oslo University Hospital, Oslo, Norway

Abstract

Our ultimate objective is to develop methodology for resuscitation data analysis that involves monitoring of the patients response, the quality of therapy, and to understand the interplay between therapy and response. To this end, methods to reliably detect the rhythm types during a resuscitation episode are needed. The objective of this study was to develop machine learning algorithms to recognize the rhythms appearing during a resuscitation episode.

In this study, we used a probabilistic framework to classify different cardiac arrest rhythms. We propose two different classifiers; naïve Bayes and logistic regression classifier.

1. Introduction

During advanced life support professional rescuers provide chest compressions, ventilation, drugs and electrical shocks to the cardiac arrest victim. The recorded changing cardiac rhythms give evidence of the patient's response to the provided therapy. CPR delivery can be monitored through impedance and accelerometer signals providing information about the chest compressions and ventilation [1, 2]. Survival rates within various emergency medical services (EMS) vary in the range 5-50%. Our ultimate objective is to develop methodology for data analysis that involves monitoring the quality of therapy and the patient's response, and to understand the interplay between therapy and response. This will enable resuscitation data analysis, so that the factors explaining the reasons for success and failure can be discovered [3]. Thus, best practice can be transferred from service to service to increase the survival rates. To be able to monitor the patient response to therapy, methods to reliably detect the rhythms during a resuscitation episode are needed. The objective of this study was to develop machine learning algorithms to recognize the rhythms typically appearing during a resuscitation episode.

2. Materials and methods

The probabilistic classification approaches can provide an objective framework for dealing with uncertainty which is in the heart of almost every pattern recognition problems with real-world data. In this study we used two different probabilistic classification approaches; the first one is a naïve Bayes classifier which is a type of generative classifier, and the second one is a logistic regression classifier which is a discriminative classifier.

The structure of this section is as follows. In section 2.1 we describe ECG database. Section 2.2 explains the feature generation stage. And, in sections 2.3 and 2.4 we discuss naïve Bayes and logistic regression classifiers, respectively.

2.1. ECG database

The database was extracted from a large out-of-hospital cardiac arrest patients study. The original study was conceived to measure the CPR quality in three geographic locations between March 2002 and September 2004. A modified version of Laerdal's Heartstart 4000 defibrillator was used to record surface ECG with a sampling rate of 500 Hz and a resolution of 1.031 μV per least significant bit, and several other reference signals. Episodes were annotated by expert reviewers using five rhythm types: ventricular fibrillation (VF) and fast ventricular tachycardia (VT) in the shockable category and asystole (AS), pulseless electrical activity (PE) and pulse generating rhythm (PR) in the non-shockable category.

For this study, artifact-free segments with a single rhythm type annotation and a duration of ten seconds (10s) were extracted. After reviewing the annotations a total of 1121 segments were included in the database, which is composed of 269 VF, 25 VT, 262 AS, 411 PE, and 154

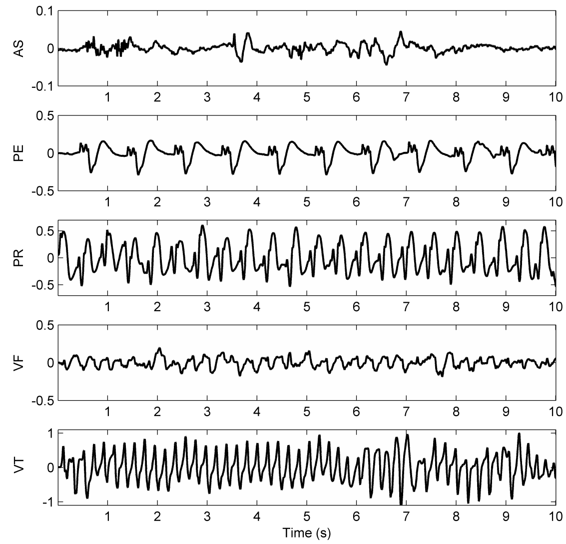

Figure 1. Representation of the 10s segments of five different rhythm types (AS, PE, PR, VF and VT) during cardiac arrest. The y axes are the amplitudes of the ECG signals in mV which are shown in the different ranges for better representation.

PR. Fig. 1 shows an example of these different rhythms during cardiac arrest. Records were resampled to 250 Hz.

2.2. Feature generation

By means of wavelet transform, the multiresolution analysis of the ECG signal is possible. This gives us the possibility to view the ECG signal as a summation of smooth and detailed parts, which makes it a powerful way to extract informative features for classification of different rhythms during cardiac arrest.

In this work, 16 different features were extracted in the wavelet domain (discrete wavelet transform). We have used Daubechies 4 (D4) wavelets in 4 level decomposition. All features are based on the approximation coefficients in level 4 and detail coefficients from level 1 to 4.

We have used two different statistical descriptors of these 5 types of coefficients; variance and the first quartile (10 features). Moreover, the parameters of the fourth-order autoregressive (AR) model of the approximation coefficients in level 4 has been used (5 features including the estimated variance of the white noise input). We used Burg's [4] method for estimation of AR parameters. And, finally the number of peaks of the autocorrelation of the approximation coefficients in level 4 is the last feature.

2.3. Naïve Bayes classifier

Bayes classifiers are generative classifiers where the joint probability density $p(\mathbf{x}, y) = p(y)p(\mathbf{x}|y)$ of input vector $\mathbf{x}$ and label y is learned over the training set, then by using Bayes rule the posterior probability $p(y|\mathbf{x})$ is calculated. Naïve Bayes (NB) classifier is based on Bayes classifier with the addition that the features are assumed to be conditionally independent given class label. In other words, we can write the class-conditional probability density function $p(\mathbf{x}|y)$ in the following way

$$p(\mathbf{x}|y_i) = \prod_{j=1}^{D} p(x_j|y_i), \quad i = 1, 2, \ldots, M \tag{1}$$

where $\mathbf{x}$ is a D-dimensional feature vector, and y_i is a class label. In this study we have 5 different cardiac arrest rhythms (AS, PE, PR, VT, VF) and 16-dimensional feature vectors, thus $M = 5$ and $D = 16$.

From Eq. 1 and Bayes classification rule, the NB classifier can be summarized as

$$y_{NB} = \arg\max_{y_i} \; p(y_i) \prod_{j=1}^{D} p(x_j|y_i). \tag{2}$$

In this study we used a uniform class prior $p(y_i)$, and for estimation of class-conditional densities $p(x_j|y_i)$, we used kernel density estimation method. We used Gaussian kernel with zero mean and unit variance. In [5] this method is named as "flexible naïve Bayes". The only difference is that contrary to [5], in our study we used Silverman's rule of thumb [6] for calculation of bandwidth (smoothing parameter), which is optimal for normal distributions.

Final point here is that although in naïve Bayes classifiers features are assumed to be conditionally independent given class label, it turns out [7] that the performance of these classifiers can be surprisingly high under the violation of the independence assumption.

2.4. Logistic regression classifier

A logistic regression (LR) classifier is a discriminative classifier where the probability $p(y|\mathbf{x})$ is learned directly. In LR classifier we assume a specific form for the probability in the following way

$$p(y_k|\mathbf{x}, \mathbf{w}_1, \ldots, \mathbf{w}_M) = \frac{\exp(\mathbf{w}_k^T \mathbf{x})}{\sum_{i=1}^{M} \exp(\mathbf{w}_i^T \mathbf{x})} \tag{3}$$

where $\mathbf{w}_k$ is the weight vector corresponding to class k, and $k = 1, 2, \ldots, M$. Again, $\mathbf{x}$ is a D-dimensional feature vector, and y_k is a class label. Since the sum of these probabilities must be equal to 1, the weight vector for one of the classes do not need to be estimated. Thus, without loss of generality we can assume $\mathbf{w}_M = 0$.

Table 1. The confusion matrices of NB and LR classifiers (for 3s segments).

	NB Classifier					LR Classifier				
	AS	PE	PR	VF	VT	AS	PE	PR	VF	VT
AS	42	5	0	1	0	43	5	0	0	0
PE	11	63	17	9	0	11	74	9	6	0
PR	0	8	22	5	4	1	22	10	5	1
VF	0	4	2	69	18	2	22	1	68	0
VT	0	0	0	4	6	0	0	1	9	0

Estimation of parameters $\mathbf{w}_k$ can be done using maximum likelihood (ML) method. To do this we need to maximize the log-likelihood function

$$\ell(\mathbf{w}) = \sum_{n=1}^{N} \sum_{k=1}^{M} t_{nk} \ln p(y_k|\mathbf{x}_n, \mathbf{w}_1, \ldots, \mathbf{w}_M) \qquad (4)$$

where N is the number of training samples, and we utilize the 1-of-M coding technique in which $t_{nk} = 1$ when $\mathbf{x}_n$ belong to class y_k, and $t_{nk} = 0$ otherwise. Then, the ML solution is

$$\mathbf{w}_{ML} = \arg \max_{\mathbf{w}} \ \ell(\mathbf{w}). \qquad (5)$$

This maximization can be performed with a variety of optimization algorithms. In this work, we used maximum entropy (MaxEnt) Matlab implementation which is available online [8] in which a limited memory quasi-Newton (L-BGFS) method is used to solve the optimization problem. After finding these parameters, the test data can be classified as follows

$$y_m = \arg \max_{y_i} p(y_i|\mathbf{x}, \mathbf{w}_{ML}). \qquad (6)$$

3. Experiments and results

We have conducted experiments to compare NB and LR classifiers both for classifying all 5 rhythms, and for classifying shockable (VF+VT), and non-shockable (AS+PE+PR) rhythms. In all experiments the same feature vector with 16 features, as described in section 2.2 is used. For training the classifiers, approximately 75% of the ECG data is used. The training data consists of 176 VF, 15 VT, 214 AS, 311 PE, and 115 PR. The remaining 25% of the ECG data is used as the test data set for evaluation of the performance of the classifiers. The test data consists of 93 VF, 10 VT, 48 AS, 100 PE, and 39 PR. As it was stated in sections 2.3 for NB classifier we used a uniform class prior and for class-conditional densities we used kernel density estimation method on the training data. In addition, for LR classifier we learn the parameters from the training data with ML method.

All reported results in this section refer to the test data set. For evaluation of the performance of different classifiers, the confusion matrices are shown in Table 1. In this table the classification is done for 3s segments of the ECG signals. The detection accuracy for the five different rhythms (AS, PE, PR, VF, VT) for 3s segments are 70% and 67% for NB and LR classifiers respectively, while the sensitivity/specificity for shockable/non-shockable rhythms are 94/90 and 75/94(%).

Table 2. The confusion matrix of LR classifiers in which in the training stage VT samples are counted three times as much as the samples from other classes.

	LR Classifier				
	AS	PE	PR	VF	VT
AS	43	5	0	0	0
PE	11	75	7	7	0
PR	1	22	9	6	1
VF	2	21	1	69	0
VT	0	0	1	4	5

Table 2 shows the confusion matrix for LR classifier when each sample of VT is counted three times as much as the samples in other classes. We did this due to the class imbalance problem in which the number of VT samples is much less than the other classes.

Finally, Table 3 shows the same results as Table 1 with a difference that in this case the 16-dimension feature vector is extracted from 10s segments of the ECG data (in Table 1 we used 3s segments). Here, the detection accuracy for the five different rhythms are 70% and 77% for NB and LR classifiers respectively, while the sensitivity/specificity for shockable/non-shockable rhythms are 93/88 and 85/96(%).

4. Discussion and conclusions

The results show that most of the time the misclassification occurs within shockable/non-shockable classes. Thus, in classification of the five different rhythms even if there are misclassifications, usually shockable/non-shockable rhythms are detected correctly (except for VF/PE in LR classifier).

Another point is that in both classifiers (NB and LR)

Table 3. The confusion matrices of NB and LR classifiers (for 10s segments).

	NB Classifier					LR Classifier				
	AS	PE	PR	VF	VT	AS	PE	PR	VF	VT
AS	42	4	1	1	0	45	3	0	0	0
PE	14	55	17	11	3	8	78	8	6	0
PR	0	10	21	2	6	0	23	15	0	1
VF	0	1	5	76	11	1	8	3	81	0
VT	0	1	0	2	7	0	0	3	3	4

there is a high number of misclassification for PE/PR. Thus, in this case we need to extract more informative features or/and design finer classifiers.

In LR classifier the performance of the classification of VT was not good due to the class imbalance problem. But, the results of Table 2 suggest that we can partly address that problem by oversampling the VT class (we counted each sample of VT three times as much as the other classes).

According to [9], the discriminative classifiers like LR are most of the times preferred due to the lower asymptotic error. However, the generative classifiers like NB approach their asymptotic error much faster. Ng [9] assumed that the class conditional density $p(\mathbf{x}|y)$ is Gaussian with a diagonal covariance matrix. But in this study, we used kernel density estimation method to model each $p(x_j|y_i)$, and this can be a reason for the NB to perform better than LR in our experiments.

The last point which we would like to discuss here is the effect of ECG segment length on the classification task. These results suggest that for classification of short segments (3s segments) the NB classifier is preferred. On the other hand, for long segments (10s segments) the initial experiments suggest that the LR classifier shows good potential. It probably has to be used with a correction for the class imbalance problem. In the case of NB classifier, we somehow do not have the class imbalance problem since in the training part each class-conditional density is estimated independent of others (still we should be careful that very few samples can result the unreliable estimation for the class-conditional density, but this is different from the class imbalance problem).

Finally, this study is our initial work to address automatic cardiac arrest rhythm analysis. The results are promising, and the performance might be improved by adding sub-algorithms like QRS detection. Different features might give additional information, and we will explore this in future work.

References

[1] Wik L, Kramer-Johansen J, Myklebust H, et al. Quality of cardiopulmonary resuscitation during out-of-hospital cardiac arrest. JAMA 2005;293(3):299–304.

[2] Stecher FS, Olsen J, Stickney RE, Wik L. Transthoracic impedance used to evaluate performance of cardiopulmonary resuscitation during out-of-hospital cardiac arrest. Resuscitation 2008;79(3):432–437.

[3] Eftestøl T, Thorsen KA, Tøssebro E, Rong C, Steen PA. Representing resuscitation data-considerations on efficient analysis of quality of cardiopulmonary resuscitation. Resuscitation 2009;80:311–317.

[4] Burg JP. A new analysis technique for time series data. In NATO Advanced Study Institute of Signal Processing with emphasis on Underwater Acoustics. New York: IEEE Press; 1968.

[5] John GH, Langley P. Estimating continuous distributions in Bayesian classifiers. In 11th Conference on Uncertainty in Artificial Intelligence. San Mateo: Morgan Kaufmann Publishers, 1995; 338–345.

[6] Silverman BW. Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC, 1998.

[7] Domingos P, Pazzani M. Beyond independence: Conditions for the optimality of the simple Bayesian classifier. Machine Learning 1997;29:103–130.

[8] Weinman J, 2013. URL `http://www.cs.grinnell.edu/~weinman/code/`.

[9] Ng AY, Jordan MI. On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In Advances in Neural Information Processing Systems 14 (NIPS'01). 2002; 841–848.

Address for correspondence:

Ali Bahrami Rad
Department of Electrical Engineering and Computer Science, Faculty of Science and Technology, University of Stavanger, 4036 Stavanger, Norway
ali.bahrami.rad@uis.no