Multimodal Data Classification using Signal Quality Indices and Empirical Similarity-Based Reasoning

Man Xu[1], Jiang Shen[2], Haiyan Yu[3]

[1] Nankai University, Tianjin, China
[2] Tianjin University, Tianjin, China
[3] Chongqing University of Posts & Telecommunications, Chongqing, China

Abstract

All bedside monitors are prone to heterogeneity and mis-labeled data, yet each multimodal sample data contains different sets of multi-dimensional attributes. To reduce the incidence of false alarms in the Intensive Care Unit (ICU), a new interactive classifier was proposed. In the algorithm, case was represented with signal quality Indices(SQIs) and RR interval features. With the function wabp, the annotations were obtained from the target signal after preprocessing. Five features were used as the inputs to a case-based reasoning classifier, retrieving the cases with empirical similarity. With the posted 750 records of the PhysioNet/CinC 2015 Challenge, the classifier was trained for answering the alarm types of the query segments. Compared with conventional threshold-based alarm algorithms, the performance of our proposed algthom reduces the maximum number of false alarms while avoiding the suppression of true alarms. Evaluated with the hidden test dataset, both real-time and retrospective, the results show that the overall TPR is 83% and 82% respectively; and TNR 44% and 43% respectively. This algorithm offers a new way of thinking about retrieving heterogeneity patients with multimodal data and classifying the alarm types in the context of mis-labeled cases.

1. Introduction

All bedside monitors are prone to heterogeneity and mis-labeled data, yet each multimodal sample data contains different sets of multi-dimensional attributes. For example, in the domain of multimodal data classification in medical databases[1], which often contain records with different channels, while most records were mis-annotated, i.e. false alarms in the intensive care unit(ICU).

Previous investigations into reducing error annotations in data recorded from critically ill patients were relatively few. For multimodal data from different signals, information fusion [2] has been investigated for robust hear beat detection and adapted for signal quality assessment of pulsatile signals. Most of these methods provided decisions with threshold-based alarm algorithms [3], in which the multiple labels were classified with the preset thresholds. Although they had improved the performance of decision making, they cannot deal with the perturbation brought by the heterogeneity and mis-labeled data. Studies of human decision-making and cognition provided the key inspiration for Case-Based Reasoning (CBR) approaches [4-5]. [6-7] leveraged the power of examples and hot features with CBR, explaining machine learning results.

To reduce the incidence of false alarms in the Intensive Care Unit (ICU), a new interactive algorithm was proposed. Our contributions lie in two folds. First, this method can take account of the domain knowledge and its heterogeneity which has not been taken account in the threshold-based alarm algorithms. Second, this method can deal with the mis-labeled instances with reusing the knowledge from the historical instances, which can improve the performance in the perspective of accuracy.

2. Multimodal data classification

2.1. Physiological feature extraction and case representation

The cases in the case base are represented with the features extracted from SQIs and RR interval signals.

$$\text{Case}_i=(x_{i1}, x_{i2}, \ldots, x_{im}, C_i)$$

where x_{ij} (j=1,..,m)is one feature from the four selected SQIs and the heart rates, and C_i is the alarm type of the Case$_i$.

Although there are totally 5r feautres where r is the number of the channels in each instance. $m=(1+4)\times r$, and these five features are usful for classifing the instance. Thus, the instance in the casebase can also be denoted as

$$\text{Case}_i=((x_{i,r,1}, x_{i,r,2}, x_{i,r,3}, x_{i,r,4}, x_{i,r,5}), C_i)$$

where $x_{i,r,1}$, $x_{i,r,2}$, $x_{i,r,3}$, $x_{i,r,4}$ are the features from heart rates and $x_{i,r,5}$ is one from the SQIs, and C_i is the alarm type of Case$_i$.

The feature $x_{i,r,5}$ is abstracted with the SQI method[4]. The algorithm run on each of the channels separately, producing one of the 4 features for each channel: iSQI, kSQI, ppgSQI and sSQI.

We normalized the kSQI and sSQI to the range [0 1] by subtracting the median value and dividing by the standard deviation.

For obtaining the heart rate feature ($x_{i,r,1}$, $x_{i,r,2}$, $x_{i,r,3}$, $x_{i,r,4}$), we obtain:

(i) the max heart rate($x_{i,r,1}$) for Asystole (C^1).

(ii) the low heart rate($x_{i,r,2}$) of 5 consecutive beats for Bradycardia (C^2).

(iii) the high heart rate($x_{i,r,3}$) of 17 consecutive beats for Tachycardia (C^3),

(iv)the max heart rate ($x_{i,r,4}$) for Ventricular_Flutter _Fib (C^4) and Ventricular_ Tachycardia (C^5).

2.2. Retrieve with empirical similarity for heterogeneity instances

The complexity of the feature data varies a lot across a pool of of patients. One could address this problem of heterogeneity instances by introducing a weight capturing how good each evidence instance is. For the query q, to learn the explicit knowledge from the historical case i, the similarity is

$$s_{w,q,i} = \exp(-\sqrt{\sum_{j=1}^{m} w_j (x_{ij} - x_{qj})^2})$$

where the weight w_j of the feature can be preset by the domain experts, i. e. w_j =1/m .

Therefore, the integrating beliefs of the consequence for the query q is

$$\beta_q = \frac{\sum_i s_{w,q,i} \beta_i}{\sum_i s_{w,q,i}}$$

where β_i is 1 when the label of the historical instance i is true alarm, and 0 when false alarm.

The intergrated belief is adapted for multimodal data classification.

2.3. Bernoulli sampling for inference

The categories of the multimodal sample data are inevitably mislabeled, especially when the data are classified with certain thresholds, i.e SQI_{th} =0.9. Thus, the instances are often labeled two different labels in the neighbors of the thresholds. Some of the true alarm data are over the bounds and located at the area of false alarm. These data are outlines of this class and often mis-labeled or error annotated as the other label.

For the discrete annotation noise, the probability $\Pr$ of the event follows a district distribution. For the case i (i = 1, . . . ,n) in the nosiy data base, the perturbed belief is

$$\bar{\beta}_i = \begin{cases} \beta_i, real\ labeled (\Pr = \varepsilon) \\ 1-\beta_i, mislabeled (\Pr = 1-\varepsilon) \end{cases}$$

where β_i is the belief vector of the decision maker with the instance annotated as C_i. ε is the parameter of nosie distribution.

According to the stochastic process , we denote $\beta_1 = \varepsilon_1$. We assume that the actual target response β_i is given by the deterministic probabilistic belief β_q with additive noise, that is,

$$\beta_q = \frac{\sum_i s_{w,q,i} \beta_i}{\sum_i s_{w,q,i}} + \varepsilon_q, q = 1,...,Q$$

where $s_{w,q,i}$ is the exponential similarity function, ε_q is a sequence of i.i.d random variables with zero mean and variance, and Q is the size of the query set.

We assume the variable β_i follows a Bernoulli distribution with parameter ε. When the historical data has a mis-labeled rate of ε, for the query, $\beta_q = E(\bar{\beta}_q) + \varepsilon_q$, satisfying

$$\varepsilon_q = \frac{\sum_i s_{w,q,i} \cdot \varepsilon(2\beta_i - 1)}{\sum_i s_{w,q,i}}$$

where $s_{w,q,i}$ is the similarity measure of historical instance i to the query q.

2.4. Connections to threshold-based alarm algorithms

In the conventional alarm algorithms[1,3], the alarm types of the signals were set according to the decision rules:

If the signal quality(mean value of SQI) is good enough(less than the thresholds) and certain RR interval is less or more than the preset threshold,

Then set the alarm as 'F'

Else set the alarm as { Asystole, Bradycardia, Tachycardia, Ventricular_Tachycardia, Ventricular _Flutter_Fib }

Dislike these methods, we adapted the CBR framework to infer the answer with the features extracted from the SQIs and the RR intervas. Since the neighbors of the queries are retrieved from the historical instances, our method improves its interpretability without sacrificing performance. Besides, this method can deal with the outline instances with reusing the knowledge from the historical instances.

3. Physiological alarms application

3.1. Dataset

As there is no annotated PPG database published, we trained and evaluated our algorithm using an annotated ECG waveforms excerpted from the posted 750 records of the PhysioNet/CinC 2015 Challenge. The dataset includes signal quality annotations of each channel including ECG, arterial blood pressure (ABP) and PPG from 104 independent adult critical care stays. Although the samples consist of subsets of these signal data, all signals have been resampled (using anti-alias filters) to 12 bit, 250 Hz and have had FIR band pass [0.05 to 40Hz] and mains notch filters applied to remove noise.

A team of expert annotators reviewed each alarm and labeled it either 'true', 'false', or 'impossible to tell' (ommited here). For machine learning, the samples labeled as 'true' includes five types of alarm, C^1, C^2, C^3, C^4 and C^5. To estimate the alarm type of query segments, data was split into separate training and testing groups. From the 750 recordings, 90% of them are used for training and 10% for testing. Summary of the annotations is demonstrated in Table 1.

Table 1. Summary of the annotations in the datasets.

Data Set	C^1	C^2	C^3	C^4	C^5	Total
True	22	46	131	6	89	294
False	100	43	9	52	252	456
Total	122	89	140	58	341	750

One sample data of a104s.mat is demonstrated with its three dimensional data, shown in Figure 1.

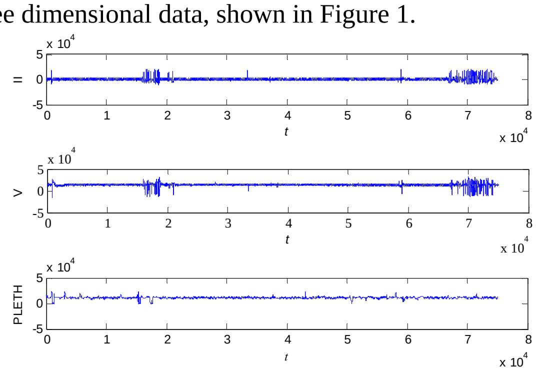

Figure 1.Data sample of a104s.mat.

The algorithm wabp was adapted to annotated onset times of the input waveforms, and an alarm was triggered 5 minutes from the beginning of each record. The exact time of the event that triggered the alarm varies somewhat from one record to another, but in order to meet the ANSI/AAMI EC13 Cardiac Monitor Standards, the onset of the event must be within 10 seconds of the alarm (i.e., between 4:50 and 5:00 of the record).

Then calculate the SQIs and heart rate of the samples, and construct the decision table. Ten sample cases with extracted features are shown as Table 2.

Table 2. Ten sample cases with extracted features.

Case	Samples	Features(Signal_r)		C	Labels
		SQI	rr	type	
X_1	a142s	$x_{1,r,5}$	$x_{1,r,1}, x_{1,r,2}, x_{1,r,3}, x_{1,r,4}$	C^1	1
X_2	a103l	$x_{2,r,5}$	$x_{2,r,1}, x_{2,r,2}, x_{2,r,3}, x_{2,r,4}$	C^1	0
X_3	b124s	$x_{3,r,5}$	$x_{3,r,1}, x_{3,r,2}, x_{3,r,3}, x_{3,r,4}$	C^2	1
X_4	b184s	$x_{4,r,5}$	$x_{4,r,1}, x_{4,r,2}, x_{4,r,3}, x_{4,r,4}$	C^2	0
X_5	t106s	$x_{5,r,5}$	$x_{5,r,1}, x_{5,r,2}, x_{5,r,3}, x_{5,r,4}$	C^3	1
X_6	t116s	$x_{6,r,5}$	$x_{6,r,1}, x_{6,r,2}, x_{6,r,3}, x_{6,r,4}$	C^3	0
X_7	f543l	$x_{7,r,5}$	$x_{7,r,1}, x_{7,r,2}, x_{7,r,3}, x_{7,r,4}$	C^4	1
X_8	f450s	$x_{8,r,5}$	$x_{8,r,1}, x_{8,r,2}, x_{8,r,3}, x_{8,r,4}$	C^4	0
X_9	v131l	$x_{9,r,5}$	$x_{9,r,1}, x_{9,r,2}, x_{9,r,3}, x_{9,r,4}$	C^5	1
X_{10}	v101l	$x_{10,r,5}$	$x_{10,r,1}, x_{10,r,2}, x_{10,r,3}, x_{10,r,4}$	C^5	0

Signal_r={ABP, PLETH, RESP, MCL, ECGII, ECGIII,ECGV,...}

For each record in the testing dataset, sensitivity and specificity are adopted to evaluate the performance of a classifier. TPR is its sensitivity (fraction of correctly predicted a particular alarm, i.e C^1 and Ture) and and TNR is its specificity (fraction of correctly predicted state, i.e., C^1 and False). A large TPR(TNR) indicates that the capacity of system classification for positive(negative) samples is strong. Thus, these measures are adopted to evaluate the performance for predicting all the types of the alarms.

3.2. Results

The similarity of the query X_1 is shown as Figure 2. The x-axis x-axis illustrates the retrieved cases in case bases, and y-axis the similarity. The similarity reaches its peaks(i.e, X_{623}), resulting in being labeled as C^1.

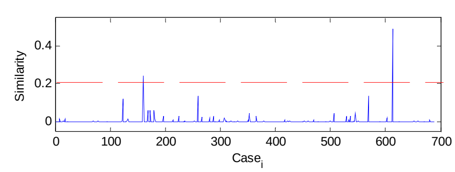

Figure 2. Similarity of the query X_1.

The classification of the sample data is shown as Figure 3. The x-axis illustrates the SQI feature, y-axis the RR interval feature, and z-axis the integrating belief(e.g for C^4 and C^5). The outline cases are classified with the reference to the similar historical instance. The results suppress false alarm using hr_{max} and SQI.

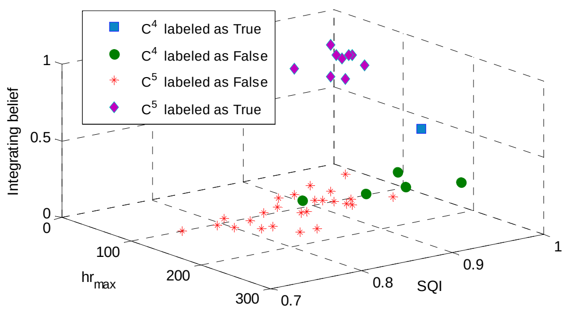

Figure 3. Classification of the samples with SQI and hr_{max} for C^4 and C^5.

Evaluated with the hidden test dataset, both real-time and retrospective, the results of this method show that the overall TPR is 83% and 82% respectively; and TNR 44% and 43% respectively. In particular, the TPR and TNR for the sample data are improved, as illustrated in Figure 4.

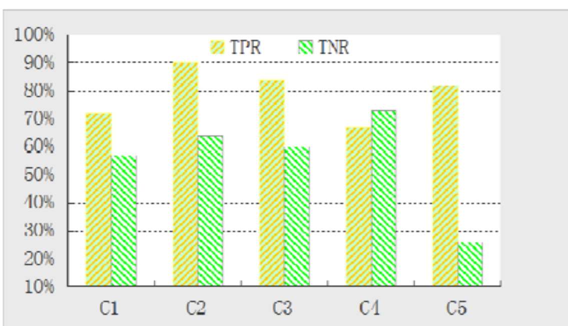

Figure 4. TPR and TNR for the sample dataset.

Compared with the threshold-based alarm algorithms, our algorithm offers a fresh perspective on reducing the maximum number of false alarms while avoiding the suppression of true alarms, and a new way of thinking on the intuition of classification in the context of mis-labeled cases.

4. Conclusions

To explore robust methods for heart beat detection using ECG and other physiological signals, we propose a multimodal machine learning framework that efficiently classifies the multimodal data using signal quality indices and empirical similarity-based reasoning. With the posted 750 records of the PhysioNet/CinC 2015 Challenge, our classifier was trained for answering the alarm types of the query segments. Evaluated with the hidden test dataset, both real-time and retrospective, the results of this method show that the true positive rate is 83% and true negative rate 43% for real-time data. In the futrue studies, to leverage of medical instances in physiological data base, more features will be abstracted and fused in both frequency and time domain.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant No. 71171143, 71571105, 71201087) and the Fundamental Research Funds for the Central Universities(Grant No. NKZXB1458), China.

References

[1] Ding Q, Bai Y, Erol YB, et al. Multimodal information fusion for robust heart beat detection. In: Computing in Cardiology Conference (CinC), 2014; 2014: IEEE; 2014. p. 261-4.

[2] Clifford G, Long W, Moody G, Szolovits P. Robust parameter extraction for decision support using multimodal intensive care data. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 2009;367:411-29.

[3] Li Q, Clifford G. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. Physiological measurement 2012;33:1491.

[4] Xu M, Yu H, Shen J. New approach to eliminate structural redundancy in case resource pools using alpha mutual information. Journal of Systems Engineering and Electronics 2013;4:768-77.

[5] Kim B, Rudin C, Shah JA. The Bayesian Case Model: A Generative Approach for Case-Based Reasoning and Prototype Classification. In: Advances in Neural Information Processing Systems; 2014; 2014. p. 1952-60.

[6] Xu M, Yu H, Shen J. New algorithm for CBR-RBR fusion with robust thresholds. Chinese Journal of Mechanical Engineering 2012;25:1255-63.

[7] Syed Z, Guttag J. Unsupervised Similarity-Based Risk Stratification for Cardiovascular Events Using Long-Term Time-Series Data. Journal of Machine Learning Research 2011;11:999-1024.

Address for correspondence.

Haiyan Yu.
Chongwen Road #2, Nan'an District, Chongqing, P.R.China, Zip code: 400065.