ENCASE: an ENsemble ClASsifiEr for ECG Classification Using Expert Features and Deep Neural Networks

Shenda Hong[1,2], Meng Wu[1,2], Yuxi Zhou[1,2],
Qingyun Wang[1,2], Junyuan Shang[1,2], Hongyan Li[1,2], Junqing Xie[3,4]

[1] School of EECS, Peking University, Beijing, China
[2] Key Laboratory of Machine Perception (Peking University), Ministry of Education, Beijing, China
[3] Medical Informatics Center, Peking University, Beijing, China
[4] School of Public Health, Peking University, Beijing, China

Abstract

We propose ENCASE to combine expert features and DNNs (Deep Neural Networks) together for ECG classification. We first explore and implement expert features from statistical area, signal processing area and medical area. Then, we build DNNs to automatically extract deep features. Besides, we propose a new algorithm to find the most representative wave (called centerwave) among long ECG record, and extract features from centerwave. Finally, we combine these features together and put them into ensemble classifiers. Experiment on 4-class ECG data classification reports 0.84 F_1 score, which is much better than any of the single model.

1. Introduction

ECG is a common non-invasive measurement that can reflect the physiology activities of heart. A typical 12-lead 300 Hz ECG monitor can produce hundreds of millions of points of each patient. Analyzing large scale ECG data can help physicians to detect many heart diseases like atrial fibrillation, myocardial infarction, acute hypotensive and so on.

There are many existing researches that propose various kinds of features and achieves high detection accuracy [1] [2] [3] [4]. These features are highly related to domain knowledge so that we call them expert features. Recently, DNNs (Deep Neural Networks) have achieved state-of-the-art results in many areas like image classification, audio recognition and natural language procession. Some researches that classify ECG data with DNNs also have achieve good results [5] [6]. However, there lack of work that combine them together for ECG data classification.

In this paper, We propose ENCASE, an ensemble classifier for ECG classification using combinations of various kinds of features. We first explore and implement expert features from large amount of formal literature. These features can be roughly divided into three groups: statistical features, signal processing features and medical features. Then, we build DNNs to automatically extract deep features. We pre-train DNNs on the training data, feed the testing data and extract the last hidden layer as deep features. Besides, we propose a new algorithm to find centerwave – the most representative wave among ECG waves of one patient. And extract features from centerwave. Finally, we combine expert features, deep features and centerwave features together, train several gradient decision boosting tree classifiers, and ensemble these classifiers together to output predictions. Experiment on 4-class ECG data classification reports 0.84 F_1 score, which is much better than any of the single model. We also report information gain with the help of XGBoost [7], which reveal the importance and interpretation of these features.

ENCASE is a general and flexible framework. It can add features incrementally, equip any DNNs, and ensemble any classifier. It can also detect more classes of heart disease if providing more data. Thus, ENCASE can be used in real world applications.

2. Methods

2.1. Data Description and Preprocessing

The dataset containing 8528 records of short 1-lead 300 Hz ECG recordings, varying length from 2700 points to 18300 points. These records are labeled with 4 classes: normal sinus rhythm (N, 5154 records), atrial fibrillation (A, 771 records), alternative rhythm (O, 2557 records) and noise (P, 46 records). Details can be found in [8].

Before feature extraction, we preprocess raw data to get the following five kind of data:

- **Long data**: We use tool from sample code to read them

 ISSN: 2325-887X DOI:10.22489/CinC.2017.178-245

into numeric time series, and save raw values by records directly.

- **Short data**: We split the long data into short waves using QRS detector by Joachim Behar from sample code. Further improvement of QRS detector is introduced in Section 2.4.
- **QRS data**: We calculate the length of consecutive QRS interval of each long data.
- **Centerwave**: The most representative wave among short data of each long data. Details are introduced in Section 2.2.2.
- **Expanded data**: Expanding to get more data using slide window and stride. In this paper, we choose window size 6000, strides are chosen dynamically for balanced dataset, detail are introduced in Section 2.4.

2.2. Features Extraction

In this section, we will introduce three kind of feature extractors in detail: expert features, centerwave features and deep features.

2.2.1. Expert Features

There are large amount of researches about extracting features for ECG data analyzing. Roughly speaking, these features can be divide into three groups.

- **Statistical features**: These features use statistic to summarize a sequence of ECG data, and give values that describe some characteristic of the data. Typical statistic including count, mean, maximum, minimum, range, variance, skewness, kurtosis, percentile and so on.
- **Signal procession features**: These features first transform ECG data from time domain into frequency domain, then extract frequency related features. For example, one may first implement FFT (Fast Fourier Transform), or DWT (Discrete Wavelet Transform). And then compute power, frequency band power, Shannon entropy, SNR (Signal Noise Ratio) and so on.
- **Medical features**: These features are base on medical domain knowledge. One group of features compute the variation based on QRS data. For example, [1] compute the sample entropy (SampleEn), [2] compute the coefficient of variation and density histograms (CDF), [3] compute the thresholding on the median absolute deviation (MAD), [4] compute the heart rate variability (Variability). Another group of features compute statistic based on P, Q, R, S and T waves. We follow the wave detection method described in [9], then extract statistic features from short data and centerwave like interval, duration, amplitude, location, slope and area [10]. Moreover, we also come up with some effective features like zig-zag (number of turns in data), zero-crossing (number of cross y axis), auto-correlation (with lags from 1 to 12).

2.2.2. Centerwave Features

In this part, we introduce centerwave – the most representative wave among short data of one patient, and extract features from it. The reasons comes from two aspects. On the one hand, we observe that some of misclassified samples are contaminated by noises. Since the ECG data is not much so long, noise filters can not handle it well. On the other hand, some classes are controlled by the morphology of the wave, not the overall rhythm. Find and analyze the representative wave directly would give better performance.

We find centerwave in three steps:

- Step1: For each patient, calculate pairwise distance matrix of his/her short data using DTW distance [11]. Since each short data are of unequal length. To solve this, we use to compute distance of unequal length time series.
- Step2: Group short waves into several clusters based on distance matrix using spectral clustering [12], where waves are similar intra clusters, not much similar inter clusters.
- Step3: Find the center of the biggest cluster. This can be done by treating distance matrix of the cluster as a graph, where nodes represent short waves, edges represent distance between short waves. Then we find the graph center – the node that is closest to other nodes, which means that this short data is the most representative wave among all short data.

Then we extract both statistical features and medical features on the centerwave. Details can be found in Section 2.2.1. Besides, we also treat raw values of centerwave as features. However, these centerwaves don't have the same length. To solve this, we resample them by linear interpolation. In our experiment, we resample all centerwaves to 200 points.

2.2.3. Deep Features

Recent works [5] [6] have demonstrate the effectiveness of DNNs on ECG classification. These DNNs have 1-D CNN (1-Dimension Convolutional Neural Networks) layers that can naturally integrate and extract hierarchy features automatically [13]. However, these DNNs models are end-to-end and hard to enhanced with extra expert features. For the purpose of benefits from both features from DNNs and expert. We transform DNNs models to deep feature extractors. In detail, we remove the output layer, extract values of last hidden layer as features. This transformation is an general process that can be used in any DNNs. So we can focus on the architecture of deep feature extractors.

The first deep feature extractor is based on a deep residual convolution networks [6] and is trained using expanded data (see Section 2.1). We want to extract multi-view and more accurate features by stacking more layers [14]. We improve [6] by adding a Bi-directional LSTM layer and vote for result. The architecture is shown below:

Input: Expanded data
Conv, BN, ReLU
Conv, BN, ReLU, Dropout, Conv
BN, ReLU, Dropout, Conv, BN, ReLU, Dropout, Conv
BN, ReLU, Dropout, Conv, BN, ReLU, Dropout, Conv
BN, ReLU, Dropout, Conv, BN, ReLU, Dropout, Conv
BN, ReLU, Dropout, Conv, BN, ReLU, Dropout, Conv
BN, ReLU, Bi-directional LSTM
Dropout, Dense
Output: Deep Features (32 dim)

Table 1. The architecture of our improved deep residual convolution feature extractor

The second deep feature extractor is based on RNNs (Recurrent Neural Networks) and is trained using centerwave (see Section 2.1). Specifically, we use LSTM (Long Short-Term Memory) cell as basic RNN cell, Aside of CNN models that can extract hierarchy features, we also want to extract time related features, this can be done by RNNs. And we take centerwave (less than 600 points) as inputs, because long data is too long (more than 6000 points) for RNN. The architecture is shown below:

Input: Centerwave
Dynamic LSTM, Dense
Output: Deep Features (32 dim)

Table 2. The architecture of LSTM deep feature extractor

2.3. Ensemble Classifies with Feature Combination

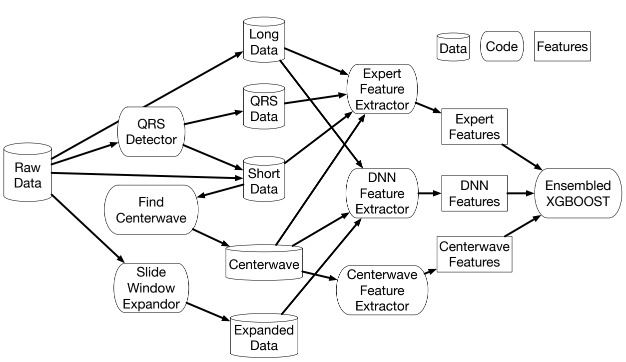

Figure 1. Framework of ENCASE

Now we have expert features, centerwave features and deep features. We concatenate three part of feature vectors into one combined feature vector, train several individual classifiers, and ensemble them by average the predicted probabilities.

It has been shown that ensemble classifiers are often much more accurate than the individual classifiers that make them up [15]. We also choose XGBoost [7] (eXtreme Gradient Boosting of decision trees) as individual classifier. XGBoost has been verified in many data mining contests that achieve better and more stable performance.

2.4. Improvement

There are some important processes that further improve ENCASE performance.

Recursive QRS detector. There are two key parameters in original QRS detector. THRES (energy threshold of the detector) and REF_PERIOD (refractory period in sec between two R-peaks). The original QRS detector sometimes output long sequence composed of several undivided QRS. To handle this, we multiply THRES and REF_PERIOD by factor of 0.68, each time the algorithm output segment longer than 600 (twice of sample frequency), and apply new QRS detector on the undivided long sequence.

Dynamic oversample. The given raw data is unbalanced that label N is much more than other labels. We propose dynamic oversample to handle this. Specifically, when generating expanded data for deep feature extractor, we use small stride for label O and big stride for label N. So that the modified expanded data are much more balanced than raw data. The modified expanded data would train a better deep feature extractor.

Model evaluation and selection. The typical offline evaluation schema is to split the data into training data and testing data, build the model on training data and evaluate on testing data. For more accurate evaluation of ENCASE, we use k-fold cross validation to split the data iteratively, and run cross validation multiple times. The final average F1 score would be more credible. We build multiple ENCASE models with different setups to choose the best model. The final ENCASE ensemble five XGBoost, and each XGBoost has 3000 trees with max_depth = 9, min_child_weight = 3.

3. Results

In this section, we first compare the effectiveness of different features, then we evaluate the performance of different classifiers on the 4 classes ECG classification task.

3.1. Features Importance

We evaluate features importance by the average information gain used in each decision trees. Results are shown in Table 3.

Feature Type	Total	Top-300	Top-150	Top-50	Top-20
Expert Features: Statistical	106	74	49	10	1
Expert Features: Signal procession	9	2	1	0	0
Expert Features: Medical	185	101	45	14	2
Centerwave Features: Raw	200	58	16	0	0
Centerwave Features: Expert	58	16	4	0	0
Deep Features: CNN	32	31	31	25	17
Deep Features: RNN	32	18	4	1	0

Table 3. Rank of features importance

We can see that most statistical features and medical features by experts are effective. It is worth nothing that deep features from CNN and RNN are both significantly effective, deep CNN features even has 17 in Top-20.

3.2. Performance

We evaluate each method by run cross validation multiple times. The measurements are F1 score related Details can be found in [8]. Results are shown in Table 4. In Features column, E stands for expert features, C stands for centerwave features, D stands for deep features.

Method	Features/Data	F_{1N}	F_{1A}	F_{1O}	F_{1P}	F_1	F_{1_NAO}
[1]	SampleEn	0.7502	0.3874	0.0030	0.0003	0.2852	0.3802
[2]	CDF	0.7961	0.2893	0.3093	0.0433	0.3595	0.4649
[3]	MAD	0.7822	0.0257	0.1812	0.0003	0.2474	0.3297
[4]	Variability	0.8321	0.5231	0.3681	0.0093	0.4332	0.5714
ResNet [6]	Expanded Data	0.7846	0.8643	0.6887	0.7317	0.7673	0.7792
ResNet Improved	-	0.9138	0.8060	0.7728	0.5000	0.7481	0.8309
LR	E	0.8843	0.7312	0.6852	0.5441	0.7112	0.7669
	E + C	0.8854	0.7353	0.6885	0.5700	0.7198	0.7697
	E + C + D	0.9095	0.8408	0.7714	0.6652	0.7967	0.8406
XGBoost	E	0.9031	0.7756	0.7445	0.6428	0.7665	0.8077
	E + C	0.9053	0.7770	0.7538	0.6354	0.7679	0.8120
	E + C + D	0.9207	0.8667	0.8074	0.8061	0.8502	0.8649
ENCASE	E	0.9059	0.7908	0.7543	0.6574	0.7771	0.8170
	E + C	0.9086	0.7899	0.7622	0.6603	0.7803	0.8202
	E + C + D	0.9204	0.8692	0.8068	0.8156	0.8530	0.8655
ENCASE Online	E + C	0.92	0.84	0.74	-	-	0.83
	E + C + D	0.92	0.85	0.74	-	0.83	0.84

Table 4. Results of different methods

We can see that ENCASE performs better than other methods. Notice that label P is so unbalanced that would lead to unstable result. More reasonable measurement should be F_{1_NAO} which not consider F_{1P}.

4. Conclusion

In this paper, we propose an ensemble classifier ENCASE to combine expert features, centerwave features and deep features together for ECG classification. ENCASE is a flexible framework that supports incremental features extraction and classifier updating. Experiments shows that ENCASE performs better than other methods.

References

[1] Alcaraz R, Abásolo DE, Hornero R, Rieta JJ. Optimal parameters study for sample entropy-based atrial fibrillation organization analysis. Computer Methods and Programs in Biomedicine 2010;99(1):124–132.

[2] Tateno K, Glass L. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of rr and deltarr intervals. Med Biol Eng Comput 2001; 39(6):664–671.

[3] Linker D. Accurate, automated detection of atrial fibrillation in ambulatory recordings. Cardiovasc Eng Technol 2016;.

[4] Carrara M, Carozzi L, Moss T, Pasquale MD, Cerutti S, Ferrario M, Lake D, Moorman J. Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy. Physiol Meas 2015; 36(9):18731888.

[5] Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ecg classification by 1-d convolutional neural networks. IEEE Transactions on Biomedical Engineering March 2016;63(3):664–675.

[6] Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY. Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. ArXiv e prints July 2017;.

[7] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In KDD. 2016; 785–794.

[8] Clifford G, Liu C, Moody B, Silva I, Li Q, Johnson A, Mark. R. Af classification from a short single lead ecg recording: the physionet computing in cardiology challenge 2017. Computing in Cardiology Rennes IEEE 2017;44.

[9] Umer M, Bhatti B, Tariq M, Zia-ul Hassan M, Khan M, Zaidi T. Electrocardiogram feature extraction and pattern recognition using a novel windowing algorithm. Advances in Bioscience and Biotechnology 2014;5:886–894.

[10] Karpagachelvi S, Arthanari M, Sivakumar M. ECG feature extraction techniques - A survey approach. CoRR 2010; abs/1005.0957.

[11] Berndt DJ, Clifford J. Using dynamic time warping to find patterns in time series. In Knowledge Discovery in Databases: AAAI Workshop. 1994; 359–370.

[12] Shi J, Malik J. Normalized cuts and image segmentation. PAMI 2000;22(8):888–905.

[13] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In ECCV. 2014; 818–833.

[14] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In CVPR. 2016; 770–778.

[15] Dietterich TG. Ensemble methods in machine learning. In Multiple Classifier Systems. 2000; 1–15.

Address for correspondence:

Hongyan Li
School of EECS / Peking University
No.5 Yiheyuan Road, Haidian District / Beijing / China
tel.: +86-010-62754911
lihy@cis.pku.edu.cn