A New Supervised Learning Approach: Statistical Adaptive Fourier Decomposition (SAFD)

Chunyu Tan[1(✉)], Liming Zhang[1], and Tao Qian[2]

[1] Faculty of Science and Technology, University of Macau, Macao, China
yb57416@connect.um.edu.mo, lmzhang@um.edu.mo
[2] Macau University of Science and Technology, Macao, China
tqian@must.edu.mo

Abstract. This paper proposes a new type of supervised learning approach - statistical adaptive Fourier decomposition (SAFD). SAFD uses the orthogonal rational systems, or Takenaka-Malmquist (TM) systems, to build up a learning model for the training set, based on which predictions of unknown data can be made. The approach focuses on the classification of signals or time series. AFD is a newly developed signal analysis method, which can adaptively decompose different signals into different TM systems that introduces the Fourier type but non-linear and non-negative time-frequency representation. SAFD fully integrates the learning process with the adaptability character of AFD, in which a small number of learned atoms are adequate to capture structures and features of the signals for classification. There are three advantages in SAFD. First, the features are automatically detected and extracted in the learning process. Secondly, all parameters are selected automatically by the algorithm. Finally, the learned features are mathematically represented and the characteristics can be further studied based on the induced instantaneous frequencies. The efficiency of the proposed method is verified by electrocardiography (ECG) signal classification. The experiments show promising results over other feature based learning approaches.

Keywords: Statistical adaptive Fourier decomposition · Heart beat classification · Time-frequency representation

1 Introduction

Supervised learning means that the learner observes some labeled example input–output pairs as the training set and learns a general hypothesis that maps from input to output, then makes predictions for all unseen instances using the learned hypothesis [1]. There are a number of popular supervised learning techniques in the literature, which can be divided into five categories [2]. They are Logic based algorithms, such as decision trees [3]; Perceptron-based techniques, such as neural networks [4]; Statistical learning algorithms, such as Bayesian networks [5];

T. Gedeon et al. (Eds.): ICONIP 2019, CCIS 1143, pp. 397–404, 2019.
https://doi.org/10.1007/978-3-030-36802-9_42

Instance-based learning, such as lazy-learning and nearest neighbor algorithms [6]; and Support Vector Machines (SVM). Each method has its own strengths and limitations. Supervised learning tasks can be further grouped into the regression and classification applications. There are a large number of classification applications related to various types of signals, including medical data, radar data, and financial data, etc. Signal processing offers effective techniques in analyzing various data, however, it has not been widely used in the supervised learning area and is not included in the above five categories.

This paper proposes a new learning technique based on adaptive Fourier decomposition (AFD), which is a newly developed signal processing technique [7]. Unlike ordinary transforms based on a pre-selected basis, AFD is based on a particular redundant dictionary leading to mono-components as composition units (atoms) of signals. AFD decomposes any given signal by suitably choosing dictionary atoms, according to some optimization principle, to form an atomic system specially for the signal to be expanded. Due to its adaptivity, different signals are represented by different systems. Combining the statistical results, signal with similar features can be represented by a common system. Based on this principle, the proposed approach can be used to classify signals. There are different variants of AFD, we choose n-best AFD in this paper that uses n-Blaschke-forms. The best n-Blaschke approximation (n-best AFD) is an alternative version of best approximation to Hardy space functions by rational functions of degree not exceeding n. The n- best approximation, due to its optimization nature, is more effective and more stable to approximate analytic signals.

The contributions of the paper are summarized as follows:

- This paper proposes a new type of supervised learning approach SAFD, in which the features are represented mathematically by the adaptively obtained well-defined orthogonal rational systems. Unlike the other transform based learning with pre-selected fixed basis, SAFD adaptively selects a parameter-determined system representing the features, and yet with fast convergence.
- Detecting and extracting the most relevant features are among very challenging tasks that often introduce manual intervention in other feature based learning models. SAFD provides a fully automatic feature detection and extraction learning process that is a superb nature in comparison with the manual feature selections. The feature detection and extraction make it to be of similarity with neural network (NN). It is, however, unlike NN, due to its automatic parameter finding and explicit mathematical representation.
- Parameter selection itself is another challenging task in learning process. SAFD offers automatic parameter selection. No parameter pre-selection is needed in SAFD.
- The commonly used feature based classification methods usually consist of two steps, including feature extraction and feature classification. The extracted features need to be put into a classifier to implement the classification. SAFD can achieve the classification by directly projecting the signal to the obtained model systems and comparing the residues without applying a separate classifier.

- The mathematical representation of the frequency features is provided in the paper, which lays foundation for feature analysis in the future.

The rest of the paper is organized as follows. The mathematical principle of the proposed SAFD is elaborated in Sect. 2. Section 3 presents the SAFD based signal classification approach in detail. In Sect. 4, the effectiveness of the method is evaluated by ECG classification. Conclusions are drawn in Sect. 5.

2 The Principle of the Proposed Statistical Adaptive Fourier Decomposition (SAFD)

The n-best Blaschke form approximation is based on the n-orthogonal rational function system, or the n-Takenaka-Malmquist system (the n-TM system in brief) [8].

Let $y(t)$ be a real-valued signal. The associated analytic signal of $y(t)$ is defined as [8]

$$y^{+}(t) = \frac{1}{2}\big(y(t) + iHy(t) + c_0\big), \tag{1}$$

where c_0 is the 0-th Fourier coefficient, and H is the Hilbert transformation. In the following part, we also denote y^+ as y for convenience.

For an analytic signal y, the n-best Blaschke form approximation is a function of the form

$$\tilde{y}(z) = \sum_{k=1}^{n} c_k B_k, \tag{2}$$

which best approximates y under a selection of the n-tuple of the parameters $a_1, a_2, \ldots, a_n$. Where $c_k = \langle y^+, B_k\rangle = \langle y^+, B_{a_1,\ldots,a_k}\rangle$ is the k-th coefficient of B_k.

The main learning process of SAFD is embedded in the parameter selection. An optimal selection of the parameters $a_1, a_2, \ldots, a_n$ is based on minimizing the square-distance between y and $\tilde{y}$, that is

$$||y - \sum_{k=1}^{n} \langle y, B_{a_1,\ldots,a_k}\rangle B_{a_1,\ldots,a_k}|| = \min_{\{b_1,\ldots,b_k\}\subset\mathbb{D}} ||y - \sum_{k=1}^{n} \langle y, B_{b_1,\ldots,b_k}\rangle B_{b_1,\ldots,b_k}||. \tag{3}$$

Cyclic AFD provided in [8] is an effective algorithm for n-best Blaschke-form approximation to solve the above optimization problems (3). It adaptively selects one more optimized parameter for each cycle.

3 SAFD Based Signal Classification

In this section, we first present the SAFD based signal classification approach, which consists of three steps, including pre-processing, learning process, and classification. Then we provide the mathematical representation of the learned features.

3.1 Pre-processing

First of all, the selected signals to be learned need to be suitably normalized as pre-processing of the training set. Then based on the statistical principle, signal averaging technique is applied, which can increase the signal-to-noise ratio. Assume signals can be divided into M classes in the training set. The i-th class is denoted as $C_i = (s_{i,1}, \dots, s_{i,N_i})$, where N_i is the number of signals in the i-th class, $i = 1, \dots, M$. For each class, we randomly select a certain number of signals, $(s_{i,1}, \dots, s_{i,i_p})$, as the training set of C_i. Then the training signals of the i-th class are averaged by $\bar{s_i} = \frac{\sum_{j=1}^{i_p} s_{i,j}}{i_p}$.

3.2 The Learning Process

The learning process described in Sect. 2 is applied to $\bar{s_i}$ to select the optimal parameters $\{a_1^i, \dots, a_{n_i}^i\}$, n_i is the number of the parameters of the i-th class. Then the corresponding weighted Blaschke products [8] are

$$B_k^i(z) = B_{a_1^i,\dots,a_k^i} = \frac{\sqrt{1-|a_k^i|^2}}{1-\bar{a_k^i}z} \prod_{j=1}^{k-1} \frac{z-a_j^i}{1-\bar{a_j^i}z}, \tag{4}$$

$\{B_k^i\}_{k=1}^{n_i}$ is the trained n-TM system of C_i.

3.3 Classification

To an unknown test signal s, it is represented as a linear combination of each trained $\{B_k^i\}_{k=1}^{n_i}$ from the i-th class, $i = 1, \dots, M$, that is, $\tilde{s}^i = \sum_{k=1}^{n_i} c_k^i B_k^i$, where $c_k^i = \langle s, B_k^i \rangle$. The residuals of the orthogonal projections on the trained n-TM systems $\{B_k^i\}_{k=1}^{n_i}$ are

$$||s - \sum_{k=1}^{n_i} c_k^i B_k^i||, i = 1, \dots, M. \tag{5}$$

Then s can be determined to the class l based on the following equation:

$$R^l(i) = \min_{i \in \{1,\dots,M\}} R^i(s) = \min_{i \in \{1,\dots,M\}} ||s - \tilde{s}^i||. \tag{6}$$

3.4 Mathematical Representation of the Frequency Features

The n-best Blaschke form approximation $\tilde{y}$ of the signal y in (2) gives a non-negative time-frequency representation of y. Furthermore, the instantaneous frequency (IF) feature can be extracted by [9]

$$\theta_n'(t) = \frac{|a_n|\cos(t-\theta_{a_n}) - |a_n|^2}{1 - 2|a_n|\cos(t-\theta_{a_n}) + |a_n|^2} + \sum_{l=1}^{n-1} \frac{1-|a_l|^2}{1-|a_l|\cos(t-\theta_{a_l}) + |a_l|^2}, \tag{7}$$

where $\theta_n = |a_n|e^{i\theta_{a_n}}$.

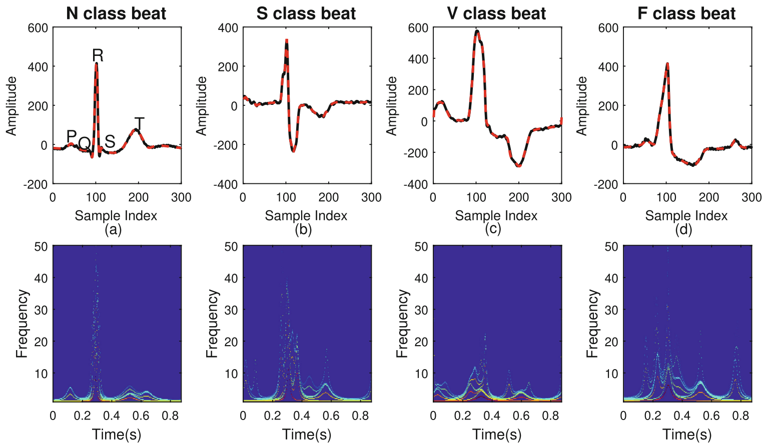

Fig. 1. Four heartbeat examples. Black lines represent the training signals and red lines represent the reconstructed signals by n-best AFD with $n = 12$. The row below is the associated time-frequency representations of the four heartbeat examples. (Color figure online)

4 Experiments

The effectiveness of the proposed learning technique is evaluated by electrocardiography (ECG) signals for heart beat classification. We evaluate our approach on well known MIT-BIH arrhythmia database[1]. The experiments are conducted on a computer with 16 GB RAM and 2.71 GHz Inter Core i5 processor and the code is implemented in MATLAB 2016a.

The annotations provided by the database are used as the labeled references for training and also used for testing the classified results. Following the Association for Advancement of Medical Instrumentation (AAMI) standard [10], the different heart beat types in MIT-BIH database are grouped into five classes: N, S, V, F and Q class. The Q class is commonly discarded according to the recommended practice and is not considered in the following heart beat classification task. The examples of the four heart beat classes are illustrated in Fig. 1.

4.1 SAFD Based ECG Classification

There are three steps in the SAFD based ECG classification approach. They are ECG signal segmentation, learning process, and classification process. In the first step, the raw ECG signals are divided into heartbeat segments following the R detection, which locates R-peak points using the provided beat locations. The training data normalization in ECG classification is as follow. We chose 300 sampling points as the segment length, 100 sampling points before the beat location and 200 sampling points after it. In this way, each segment contains

[1] http://www.physionet.org/physiobank/database/mitdbl.

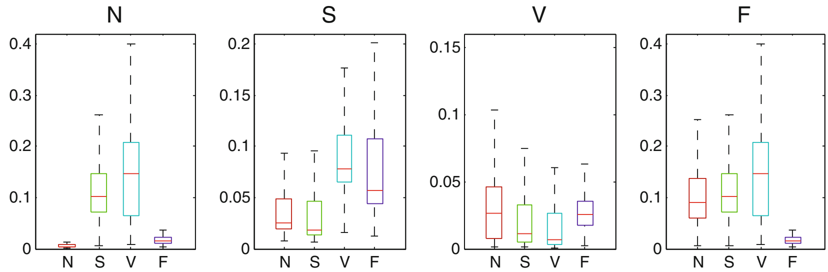

Fig. 2. The boxplots of the residual distribution for each class are projected on trained N, S, V, and F class, respectively.

a whole heart cycle, including P wave, QRS complex wave, and T wave, as illustrated in Fig. 1(a).

The proposed learning technique is trained to extract features that can represent N, S, V, and F classes, respectively. 13640 heartbeats are sampled randomly in MIT-BIH database, in which 834 heartbeats are F class, 8330 heartbeats are N class, 2072 heartbeats are S class, and 2404 heartbeats are V class. For the sake of extracting features that can represent each class and reflecting some statistical characteristics, 100 heartbeats are randomly selected in each class for training. The remaining heartbeats are used for classification validation.

In learning process, we first get the average signal in N, S, V, and F classes, respectively. In our experiment, $N = 12$ is selected for all four classes. Next, the SAFD learning process is applied to each of the labeled average N, S, V, and F classes, respectively. Four examples selected randomly from each of N, S, V, and F classes and are reconstructed by using the respective n-TM systems as illustrated in Fig. 1. It can be seen that the automatically selected features can perfectly represent the heartbeat classes they belong to. Furthermore, those respective n-TM systems possess positive IF features and there is no intersection among all IFs. The positive IFs effectively reflect the time-varying characteristics of signals, such as the morphology of heartbeats. The respective time-frequency representations of the four heartbeat examples are shown in Fig. 1.

In the classification process, tested signals are projected onto the obtained four n-TM systems, and the residuals are worked out to see which n-TM system gives rise to the minimum residual energy. The residual distributions of N, S, V, and F classes are graphically represented in Fig. 2. As shown in Fig. 2, residuals of heartbeats reach to a minimum when the heartbeats are consistent with their class labels, which suggest that the residuals have a good discriminative representation for classifying different beat types. The performance of the proposed learning technique has a 81.44% overall accuracy and other detailed results are shown in Table 1.

Table 1. Comparison the proposed method and the previous works. (# features: the number of features are extracted.)

Method	# features	N		S		V		F		Tot.
		Se	+P	Se	+P	Se	+P	Se	+P	Acc
De Chazal et al. [11]	52	86.9	99.2	75.9	38.5	77.7	81.9	89.43	0.08	81.9
Llamedo et al. [12]	39	77.55	99.47	76.46	41.34	82.94	87.97	95.36	4.23	78.0
Zhang et al. [13]	46	88.9	99.0	79.1	36.0	85.5	92.8	93.8	13.7	86.7
Herry et al. [14]	6	83.13	98.93	81.14	31.93	77.50	79.05	83.25	6.91	82.70
Proposed	–	83.58	95.45	82.25	72.48	68.53	97.23	95.78	32.38	81.44

4.2 Performance Comparisons

The experiment results of the proposed method are compared with some selected state-of-the-art feature based ECG classification methods. They are all tested and validated on the MIT-BIH arrhythmia database and follow the AAMI standard. The comparison results are illustrated in Table 1. The compared methods use a variety of features to represent the ECG signals and different types of classifiers for classification.

The performance is evaluated in terms of the sensitivity (Se) and positive predictivity ($+P$) [11]. Though the overall accuracy of the proposed method is not the best, the results are comparable. The most important is that the first three methods rely on very complicated and high dimension feature sets, which lead to three disadvantages, including manual feature selection, high computational cost, and high requirements for classifiers next step. The fourth method significantly reduces the feature number, however, the feature selection approach is still very complicated and manual selection is needed. The proposed approach classifies the ECG signal based on the selected parameters while no classifier is used. It is to be improved along with further studies on the adaptive parameter selection method.

4.3 Running Time Analysis

It takes approximately 0.373 s to complete one learning process and total $0.373 * 4 \approx 1.49$ s for the whole system training by the proposed SAFD with $N = 12$. Once the training of ECG signals is completed, the classification of ECG heartbeat is readily done. The time required for projecting a tested signal to the n-TM system space is only $0.006s$. Note that this is a very short time moment, and much shorter than the time needed to finish one heart beat that the length of each heartbeat segmentation is $300/360 \approx 0.83s$. Thus, the proposed algorithm has a great potential for the real-time monitoring system. We will leave this real-time implementation as future work. Since none of other methods shown in Table 1 provides the training time, we cannot make a comparison.

5 Conclusion

This paper presents a new type of supervised learning approach SAFD, which provides fully automatic feature selection and extraction with well defined mathematical representation. It offers a new fully explainable automatic learning process for signals. The effectiveness of the proposed learning technique is demonstrated by ECG classification. This study lays foundation to further analysis of the time-frequency characteristics of the learned features.

Acknowledgment. This study is supported by the research grants: The Science and Technology Development Fund of Macao SAR FDCT 079/2016/A2, 0123/2018/A3, and MYRG 2017-00218-FST, 2018-00111-FST.

References

1. Norvig, P., Russell, S.J.: Artificial Intelligence: A Modern Approach, 3rd edn. Pearson Education Inc., New Jersey (2010)
2. Kotsiantis, S.: Supervised machine learning: a review of classification techniques. Informatica **31**, 249–268 (2007)
3. Kotsiantis, S.: Decision trees: a recent overview. Artif. Intell. Rev. **39**(4), 261–283 (2013)
4. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)
5. Cheng, J., et al.: Learning Bayesian networks from data: an information-theory based approach. Artif. Intell. **137**(1–2), 43–90 (2002)
6. Zhang, M., Zho, Z.: ML-KNN: a lazy learning approach to multi-label learning. Pattern Recognit. **40**(7), 2038–2048 (2007)
7. Qian, T., Zhang, L., Li, Z.: Algorithm of adaptive Fourier decomposition. IEEE Trans. Signal Process. **59**(12), 5899–5906 (2011)
8. Qian, T.: Cyclic AFD algorithm for the best rational approximation. Math. Methods Appl. Sci. **37**(6), 846–859 (2014)
9. Dang, P., et al.: Transient time-frequency distribution based on mono-component decompositions. Int. J. Wavelets, Multiresolution Inf. Process. **11**(3), 1350022 (2013)
10. AAMI: Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms. ANSI/AAMI EC38 (1998)
11. De Chazal, P., et al.: Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Trans. Biomed. Eng. **51**(7), 1196–1206 (2004)
12. Mariano, L., Martinez, J.P.: Heartbeat classification using feature selection driven by database generalization criteria. IEEE Trans. Biomed. Eng. **58**(3), 616–625 (2011)
13. Zhang, Z., et al.: Heartbeat classification using disease specific feature selection. Comput. Biol. Med. **46**, 79–89 (2014)
14. Herry, C.L., et al.: Heart beat classification from single lead ECG using the synchrosqueezing transform. Physiol. Meas. **38**(2), 171 (2017)