Boosting for Real-Time Multivariate Time Series Classification

Haishuai Wang,† Jun Wu‡
†Centre for Quantum Computing & Intelligent Systems, University of Technology Sydney
‡School of Computer and Information Technology, Beijing Jiaotong University
haishuai.wang@student.uts.edu.au; wuj@bjtu.edu.cn

Abstract

Multivariate time series (MTS) is useful for detecting abnormity cases in healthcare area. In this paper, we propose an ensemble boosting algorithm to classify abnormality surgery time series based on learning shapelet features. Specifically, we first learn shapelets by logistic regression from multivariate time series. Based on the learnt shapelets, we propose a MTS ensemble boosting approach when the time series arrives as stream fashion. Experimental results on a real-world medical dataset demonstrate the effectiveness of the proposed methods.

Introduction

There is a significant public health concern regarding major complications and death following surgery. Forty million Americans undergo surgery yearly. Approximately five percent die within a year of their operations, and roughly ten percent suffer major in-hospital morbidity (e.g., stroke, heart attack, pneumonia). Early recognition of risk and appropriate management could often prevent or modify these adverse outcomes. Modern intraoperative monitoring yields a wealth of data from thousands of operating rooms. Integration and real-time analysis of these data streams has the potential to revolutionize perioperative care.

The constantly being produced data from medical sensors are often taken as time series, since time series consists of ordered sequences of pairs (timestamps, data elements) (Anguera et al. 2016; Sow, Turaga, and Schmidt 2013). However, the time series are commonly as multivariate time series (MTS) since the collected data probably from various measure sensors (e.g., heart rate, blood pressure, etc.). In the process of operation, medical sensors real-time generate data which forms time series data stream. Thus, detecting abnormality time series timely is significant to increase survival rate. This work aims to real-time distinguish abnormalities from normal multivariate time series.

To date, dynamic time warping (DTW) and shapelets are commonly used for time series classification (Grabocka et al. 2014; Górecki and Łuczak 2015). Unfortunately, existing DTW and shapelets approaches all work on static time series datasets with regression or simple classifier (e.g., decision tree). No effective strategy exists to support classification for multivariate time series stream. Intuitively, a trivial solution is to partition time series into a number of chunks and carry out classifier learning in each individual subset. This simple solution, however, does not allow time series subsets to collaborate with each other to train robust models (Wang et al. 2016). To solve the challenges, we propose, in this paper, a time series ensemble boosting algorithm for MTS stream classification by learning shapelets from MTS.

Boosting Algorithm for Real-time Multivariate Time Series Classification

In this section, we first propose a method to learn shapelet features from multivariate time series data. Then, we present our boosting algorithm to classify abnormality time series based on the learnt shapelet features.

Shapelet Features Extraction

Assume a multivariate time series databset T is composed of N-dimensional, each dimension has I training series instances, and each series contains Q-many ordered values, the multivariate time series is denoted as $T^{(N\times I)\times Q}$. Consider a sliding window of length L. The window starts at time j on i-th series of n-th dimension $T_i^n \in T$ is defined as $(T_{i,j}^n, \cdots, T_{i,j+L-1}^n)$. When the window slides along a time series, there are totally $Q - L + 1$ ordered sub-sequences (segments) of a series. We use label 1 to denote abnormality time series and 0 for others. Thus, we have a label matrix $Y \in \{0,1\}^{(N\times I)}$.

Shapelets are defined as the most discriminative time series segments. The K-most informative shapelets of length L in n-th dimension are denoted as $S^n \in \mathbb{R}^{K\times L}$.

Based on the learning shapelets on univariate time series approach presented by (Grabocka et al. 2014), we revise their method to learn shapelets on multivariate time series.

In n-th dimension of multivariate time series, the distance between the i-th series T_i^n and the k-th shapelet S_k^n is defined as the minimum distance $M_{i,k}^n$ among the distance between the shapelet S_k^n and each segment j of T_i^n, as shown in Eq. 1.

$$M_{i,k}^n = \min_{j=1,\cdots,J} \frac{1}{L} \sum_{l=1}^{L} (T_{i,j+l-1}^n - S_{k,l}^n)^2 \qquad (1)$$

Then, a linear learning model can predict approximate $\widetilde{Y}^n \in \mathbb{R}^{I\times Q}$ by using the minimum distances predictor M and linear weights $W \in \mathbb{R}^{(N\times K)}$, as shown in Eq. 2.

$$\widetilde{Y}_i^n = \sum_{n=1}^{N} W_0^n + \sum_{k=1}^{K} M_{i,k}^n W_k^n, \quad \forall i \in \{1,\cdots,I\}, \forall n \in \{1,\cdots,N\} \tag{2}$$

where $W_0^n \in \mathbb{R}$ is bias.

Eq. 3 shows the logistic regression operates by minimizing the logistic loss between the estimated $\widetilde{Y}$ and true Y,

$$\mathcal{L}(Y,\widetilde{Y}) = -Y \ln \sigma(\widetilde{Y}) - (1-Y)\ln(1-\sigma(\widetilde{Y})) \tag{3}$$

To learn the optimal shapelets S, we have the following regularized objective function:

$$\arg\min_{S,W} \mathcal{F}(S,W) = \arg\min_{S,W} \sum_{n=1}^{N}\sum_{i=1}^{I} \mathcal{L}(Y_i^n, \widetilde{Y_i^n}) + \sum_{n=1}^{N} \lambda_{W_n} \parallel W_n \parallel^2 \tag{4}$$

To solve the problem in Eq. 4, a differentiable approximation to the minimum function called soft-minimum function is introduced in (Grabocka et al. 2014).

Boosting Algorithm for MTS Classification

Boosting algorithm learns many weak classifiers repeatedly, and after many rounds, the boosting algorithm combines these weak classifiers into a single prediction rule that will be much more accurate than any one of the weak classifiers.

We suppose the time series stream arrives in batch, denoted as D_t. Given a time series chunk $D_t = \{(T_1, y_1), \cdots, (T_m, y_m)\}$, which contains a number of shapelets, let $S_t = \{s_1, \cdots, s_g\}$ denotes the full set of shapelets in D_t. We can use S as features to represent each time series T_i into a vector space as $x_i = \{x_i^{s_1}, \cdots, x_i^{s_g}\}$, where $x_i^{s_i} = 1$ if $s_i \in T_i$, and 0 otherwise. To build weak classifiers for boosting, we can use each shapelet s_j as a decision stump classifier (s_j, π_j), as follows:

$$\hbar(T_i; s_j, \pi_j) = \begin{cases} \pi_j : & s_j \in T_i \\ -\pi_j : & s_j \notin T_i \end{cases}$$

where $\pi_j \in Y = \{0, 1\}$ is a parameter controlling the label of the classifier.

The prediction rule in a local chunk D_t for a series T_i is a linear combination of all weak classifiers, as shown in Eq. 5,

$$\mathcal{H}(T_i) = \sum_{(s_j,\pi_j)\in S\times Y} \alpha_j \hbar(T_i; s_j, \pi_j) \tag{5}$$

where α_j is the weight of weak classifier $\hbar(T_i; s_j, \pi_j)$. If $\mathcal{H}(T_i) \geq 0$, it is a abnormality time series (1), or negative (0) otherwise.

Experiments

We conduct experiments on a standard ECG dataset that is commonly used in the literature about multivariate time series [1]. The ECG dataset comprises a collection of time series where the sequence of measurements recorded by one electrode during one heartbeat, and contains 200 samples with two dimensions. There are 133 samples are normal cases and 67 samples are abnormities. The length of a MTS sample is between 39 and 153. All abnormal heartbeats are representative of a cardiac pathology known as supraventricular premature beat. We divide time series into chunks by using a fixed size (e.g., 100). The classification task is to distinguish abnormal heartbeats from normal records.

[1] http://www.cs.cmu.edu/∼bobski/

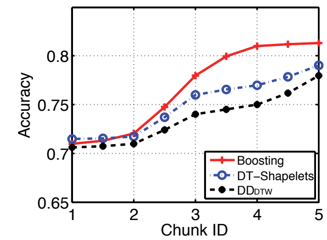

Figure 1: Accuracy under the chunk size 100.

For comparison purposes, we implement DDDTW (Górecki and Łuczak 2015) and DT-Shapelets as baseline approaches. DDDTW is a latest approach for MTS classification, using a parametric derivative dynamic time warping distance. DT-Shapelets is to use the learnt shapelets from MTS and decision tree as classifier instead of our boosting method. Fig. 1 shows the classification accuracy of all comparison algorithms in each chunk, in which the chunk size of time series stream is set to 100. Overall, the proposed ensemble boosting algorithm with shapelets learning approach performs better than other benchmarks. Besides, when more MTS are observed, the classification accuracy of boosting algorithm increases but tends to stable after four chunks arrived.

Acknowledgements

This work is partially supported by the NSFC NO.: 61301185, 61370070 and 61671048.

References

Anguera, A.; Barreiro, J.; Lara, J.; and Lizcano, D. 2016. Applying data mining techniques to medical time series: an empirical case study in electroencephalography and stabilometry. *Computational and Structural Biotechnology Journal* 14:185–199.

Górecki, T., and Łuczak, M. 2015. Multivariate time series classification with parametric derivative dynamic time warping. *Expert Systems with Applications* 42(5):2305–2312.

Grabocka, J.; Schilling, N.; Wistuba, M.; and Schmidt-Thieme, L. 2014. Learning time-series shapelets. In *SIGKDD*, 392–401. ACM.

Sow, D.; Turaga, D. S.; and Schmidt, M. 2013. Mining of sensor data in healthcare: a survey. In *Managing and Mining Sensor Data*. Springer. 459–504.

Wang, H.; Zhang, P.; Zhu, X.; and Tsang, I. 2016. Incremental subgraph feature selection for graph classification. *IEEE Transactions on Knowledge and Data Engineering*.