Chapter 2

Related work

Early classification of time series has been extensively studied for minimizing class prediction delay in time-sensitive applications such as medical diagnostic [26,30,41] and industrial process monitoring [15, 25, 43–45]. A primary task of an early classification approach is to classify a time series as soon as possible with limited number of data points. It is true that one can achieve better accuracy if one waits for more data points, but will miss the opportunities as well. Recent years have witnessed several approaches for solving the problem of early classification for Univariate Time Series (UTS) and MTS. This chapter discusses the related work to know the current status of the area.

The existing early classification approaches are divided into four categories based on their proposed solution strategies. The four categories include prefix based, shapelet based, model based, and miscellaneous approaches. In prefix based approaches [18, 19, 28, 46, 47], the strategy is to learn a minimum prefix length of time series using training instances and then classify a testing time series using its prefix of learned length. The shapelet based approaches such as [5, 20, 23, 26, 31, 33, 34, 41, 48–51] attempt to obtain a set of key shapelets from the training dataset and utilize them as class discriminatory features of time series during testing. In model based early classification approaches [15, 25, 52–57], the strategy is to develop a mathematical model by incorporating a cost based trigger function for making a reliable class prediction. Finally, the approaches [22,27,30, 32,44,58–60] that do not follow any of the above strategies, are included as miscellaneous approaches.

The rest of the chapter is organized as follows. Next section discusses the prefix based early classification approaches in detail. Sections 2.2 and 2.3 review the shapelet and model based approaches, respectively. In Section 2.4, we discuss the miscellaneous approaches. Finally, Section 2.5 concludes the chapter by summarizing the related work on the early classification of time series.

2.1 Prefix based early classification

This section presents the existing early classification approaches that have utilized prefix of time series for achieving earliness. The key idea is to learn a stable prefix length of each time series during training and then utilize them for classifying an incomplete time series during testing. During training, a separate classifier is first constructed for each prefix length of time series and then checked for the stability of relationship between the results of prefix space and full-length space. The classifier that achieves an adequate level of stability with minimum prefix length, is considered as early classifier and the corresponding prefix length is called as minimum prediction length [19, 28, 46] or MRL. This early classifier has the ability to classify an incomplete time series as soon as MRL is available. Table 2.1 summarizes the prefix based early classification approaches.

One of the first notable prefix based early classification approach is proposed in [18]. Two interesting methods, sequential rule classification and generalize sequential decision tree, has been introduced in [18] for early classification of symbolic sequences. For a given training dataset, these methods first extracts a large number of sequential rules from different length of prefix spaces and then selects top-k rules based on their support and prediction accuracy. The selected rules are later used for early classification of a testing time series.

Xing *et al.* [19] developed two different algorithms, named as Early 1-Nearest Neighbor (1-NN) and Early Classification of Time Series (ECTS), for UTS data. Early 1-NN algorithm computes the MRL of each time series of training dataset using 1-NN. These computed MRLs are first arranged in ascending order by their length and then used for early prediction of class label.

Early 1-NN has two major drawbacks: i) each time series can have different MRL and ii) computed MRLs are short and not robust enough due to overfitting problem of 1-NN. To overcome these drawbacks, ECTS algorithm [19] first clusters the time series based on their similarities in full-length space. It employed an agglomerative hierarchical clustering [61] with single linkage for clustering. The agglomerative clustering is parameterized by minimum support threshold to avoid the overfitting issue. Later, ECTS computes only one MRL for each cluster to have a more generalized set of MRLs for reliable early classification of an incomplete time series.

In [46], the authors presented an extension of ECTS, called as *Relaxed* ECTS, to find shorter MRLs. *Relaxed* ECTS relaxes the stability condition of Reverse Nearest Neighbor (RNN) while computing MRLs for the clusters. To compute MRL of any cluster, *Relaxed* ECTS requires only a subset of time series with stable RNN instead of all. It also speeds up the learning process.

Mori *et al.* [47] proposed an early classification framework based on discriminativeness and reliability of the classes over time. The framework employed Gaussian Process classifier to compute the class discriminative MPLs. It also assigned some thresholds to each class label to ensure the reliability of predictions. Such thresholds are computed from two highest posterior probabilities that are obtained by applying Gaussian Process classifier on training dataset. In addition, the authors in [47] also conducted a case study for early identification of bird species by using their chirping sounds.

The authors in [28] proposed an early classification approach based on piecewise aggregated approximation method [62]. The approach first applies a center sequence method [63] to transform each MTS instance of dataset into UTS and then reduces the length of the transformed UTS by using the approximation method.

Table 2.1: Summary of prefix based early classification approaches.

Paper ID	Year	Major Focus	Limitation	Type of time series
[18]	2008	Symbolic representation of time series	Time series has to be discretized properly	UTS
[19]	2009	MRLs computation using hierachical clustering	Clusters separation can not be guaranteed for small datasets	UTS
[46]	2012	Relaxation of RNN stability condition	Overfitting problem for small datasets	UTS
[47]	2017	Avoid unnecessary predictions before the availability of sufficient data	No check for stability while learning MRLs	UTS
[28]	2017	Transformation of MTS into UTS	Approximation of segments causes to lose identifiable information of the classes	MTS

2.2 Shapelet based early classification

This section presents a detailed review of the approaches that have used shapelets for early classification of time series. A shapelet is defined as a subsequence of time series with a certain length. It is also associated with a class label which is same as that of time series. The authors [64] have successfully implemented the idea of shapelets for time series classification, which became the motivation point for many researchers to utilize the shapelets for achieving the earliness in the classification. Moreover, the shapelets improve the interpretability of the classification results [20, 41, 48], which enhances the adaptability of the proposed approach for real-world applications such as medical diagnostic and industrial process monitoring.

In the existing work [23, 26, 41, 48, 49], the authors focused on to extract a set of perfect shapelets (called as key shapelets) from the given training dataset. Ideally, a perfect shapelet is powerful enough to distinguish all the time series of one class from the time series of other classes. However, it is impractical to find such perfect shapelets. The researchers therefore put their efforts towards developing a proper criterion that can provide a set of effective shapelets (if not perfect) for early classification [23,31,33,34,41]. A summary of shapelet based approaches is presented in Table 2.2.

The authors [20] are the first to address the early classification problem using shapelets. They developed an approach called Early Distinctive Shapelet Classification (EDSC) which utilizes the local distinctive subsequences as shapelets (or features) for early classification of time series. EDSC consists of two major steps: *feature extraction* and *feature selection*. In former step, it first finds all local distinctive subsequences from training dataset and then computes a distance threshold for each subsequence. EDSC employed two methods, kernel density estimation [65] and chebyshevs inequality [66], to compute the distance thresholds. Next, in feature selection step, the authors selected key shapelets based on their utility.

As EDSC does not provide any estimate of certainty while making the decision about class label of an incomplete time series, Ghalwash *et al.* [48] presented an extension of EDSC with an additional property of uncertainty estimate. The uncertainty estimate indicates the confidence level with which the prediction decision is made and if it is less than some user-defined confidence level then the decision may be delayed even after a shapelet is matched.

In [26], the authors utilized shapelets for early classification of gene expression data. A Multivariate Shapelets Detection (MSD) method is proposed to classify an incomplete MTS by extracting the key shapelets from training dataset. MSD finds several multivariate shapelets from all dimensions of MTS with same start and end points. It computes an information gain based distance threshold for each multivariate shapelet to facilitate the matching with incomplete MTS. In addition, the authors also formulated a weighted information gain based utility measure to select the key shapelets and to prune the needless shapelets in the process. In [41], the authors also proposed an approach, called as Interpretable Patterns for Early Diagnosis (IPED), for studying viral infection in humans using their gene expression data. Similar to MSD, IPED also extracts multivariate candidate shapelets from the training MTS but it allows to have a multivariate shapelet with different start and end points in different dimensions.

He *et al.* [23] attempted to solve an imbalanced class problem of ECG classification where training instances in abnormal class are much lesser than normal. They addressed this problem in the framework of early classification of MTS and proposed an early prediction approach for handling imbalanced distribution the MTS. The approach extracts all possible subsequences (candidate shapelets) of different length from each component of MTS separately. Unlike MSD [26], the extracted candidate shapelets are univariate and thus need not have same start and end points in all the dimensions.

Table 2.2: Summary of shapelet based early classification approaches.

Paper ID	Year	Major Focus	Limitation	Type of time series
[20]	2011	Utility based feature selection	No assurance of classification accuracy during training	UTS
[26]	2012	Separate distance threshold along each dimension of MTS	Extracted shapelets can not have variable length of dimensions	MTS
[23]	2013	Clustering based core shapelets selection	Limited to binary classification	MTS
[41]	2013	Formulation of convex optimization problem for key shapelets selection	Data points of time series must be obtained at regular interval	MTS
[48]	2014	Computation of shapelet ranks by incorporating accuracy and earliness	Generate huge number of candidate shapelets of varying length	UTS
[49]	2015	Evaluation strategy to check the quality of shapelets	Quality of shapelets heavily dependent on employed clustering method	MTS
[31]	2015	Pattern discovery using sequential and simultaneous combinations of shapelets	Computationally inefficient	MTS
[50]	2016	Automatic feature extraction using deep learning models	Required to set several parameters at each layer of the model	UTS
[51]	2019	Computation of distance between shapelet and time series in change-point space	Extracted shapelets tend to lose natural interpretability	UTS
[34]	2019	Key-point based shapelet extraction	Limited to binary classification	MTS

One of the major drawback of [26], [23], and [41], is that they do not incorporate the correlation among the shapelets of different components of MTS during classification. Such a correlation helps to improve the interpretability of the shapelets. To overcome this drawback, the authors in [5, 49] developed an approach, called as Mining Core Features for Early Classification (MCFEC), where core features are the key shapelets. MCFEC first obtains candidate shapelets from each component independently and then discovers the correlation among the shapelets of different components to enhance their interpretability.

The authors in [34] presented a confident early classification framework for MTS with interpretable rules where key shapelets are extracted by using a concept of local extremum and turning points. This framework first discovers interpretive rules from the sets of candidate shapelets and then estimates the confidence of each rule to select the key shapelets. The correlation among the components of MTS is also incorporated in the proposed framework.

2.3 Model based early classification

This section discusses the model based early classification approaches for time series data. Unlike prefix and shapelet based approaches, the model based approaches formulate a mathematical model to optimize the tradeoff between earliness and reliability (accuracy) of prediction. Most of these approaches aimed to design a decision or stopping rule by using the conditional probabilities. These conditional probabilities are either generated by generative classifiers or computed by fitting a discriminative classifier on training dataset. Table 2.3 presents a summary of model based approaches with their major focus and limitations.

The authors in [52, 67] formulated a decision rule to classify an incomplete test time series with some pre-defined reliability. They employed Gaussian Mixture Model estimation and joint Gaussian estimation for estimating the distribution of incomplete time series by modeling the complete time series of training dataset as random variables. Two generative classifiers, linear Support Vector Machines and Quadratic Discriminant Analysis [68], with the formulated decision rule were adopted to provide a desired level of accuracy in the early classification.

In [15], the authors developed an ensemble model based early classification approach to recognize a type of gas using an incomplete 8-dimensional time series generated by a sensors-based electronic nose. The ensemble model consists of a set of classifiers with a reject option which allows them to express their doubt about the reliability of the predicted class label. The set of classifiers are kept serially along the progress of time series to facilitate the prediction using small portion of data points. If sufficient data points are not arrived then prediction is carried out again by the next classifier when another portion of data points are arrived. This process is repeated until majority of the classifiers are confident enough about the predicted class label. Decision of choosing the reject option is also dependent on the cost of data collection time.

The work in [69] focused on minimizing response time to obtain the earliness in the classification. Here, an empirical risk function is developed to minimize the risk associated with early prediction and thus optimizes the response time and earliness with confidence.

Table 2.3: Summary of model based early classification approaches.

Paper ID	Year	Major Focus	Limitation	Type of time series
[52]	2012	Construction of optimal decision rules from incomplete time series	Several iterations are required for setting an appropriate value of reliability threshold	UTS
[15]	2013	Providing a reject option to expression doubt about reliability of prediction	Limited to probabilistic classifiers only	MTS
[53]	2015	Designing of cost based trigger function	Cost function heavily relies on clustering accuracy	UTS
[54]	2015	Likelihood based similarity between time series	High domain dependency	UTS
[69]	2016	Formulation of convex optimization problem for training	Computationally inefficient	UTS
[45]	2016	Designing of trigger function	Impose unnecessary computations due to non-myopic property	UTS
[25]	2017	Stopping rules using all posterior probabilities	Non-convex cost function	UTS
[55]	2019	fusion of multiple classifiers for early decision	Single confidence threshold is not sufficient for generalization	UTS
[57]	2019	Abstaining from premature predictions	No check for stability while discovering safeguard points	UTS
[56]	2019	Stopping rules using selected posterior probabilities	Non-convex cost function	UTS

Dachraoui *et al.* [53] proposed a non-myopic early classification approach where the term *non-myopic* means at each time step the classifier estimates an optimal time in the future when a reliable prediction can be made. The authors in [45] pointed out two weaknesses of [53]: i) assumption of low intra-cluster variability, which is impractical while obtaining membership probabilities using clustering and ii) clustering is carried out with complete time series, which may impact the estimation of optimal time. In [45], two different algorithms (*NoCluster* and *2Step*) are introduced to overcome these weaknesses while preserving adaptive and non-myopic properties.

The authors in [25] developed different stopping rules by using the class-wise posterior probabilities. These stopping rules also included some real-value parameters which are optimized by using Genetic algorithms. Besides that, the authors in [55, 57] computed a confidence threshold to indicate the data sufficiency for making early prediction.

2.4 Miscellaneous approaches

This section covers the early classification approaches that do not meet the inclusion criteria of other categories. Some of these approaches such as [22, 27, 29, 30, 32] did not attempt to optimize the tradeoff between accuracy and earliness. In other words, they performed early classification without ensuring the reliability of prediction. Table 2.4 summarizes the miscellaneous approaches for early classification of time series.

The authors in [27, 30] presented hybrid early classification models by combining a generative model with a discriminative model. At first, several generative models are trained over short segments of time series to learn the distribution of patterns in training data. These trained models generate an array of log likelihood values for the disjoint shapelets of a time series. Such array of likelihood is passed as features for training a discriminative model.

The work in [22,29] employed a stochastic process, called as Point Process model, to capture the *temporal dynamics* of different components of MTS. The *temporal dynamics* of each component is extracted independently and then *sequential cues* are computed to capture temporal order of events that have occurred over time among components. The authors also incorporated the correlation among components of MTS by using a variable order markov model.

In another [59], a reinforcement learning based early classification framework is introduced using a Deep Q-Network [70] agent. The framework uses a reward function to keep balance between accuracy and earliness. It also includes a suitable set of states and actions for the observations of the training time series. The agent learns an optimal decision making strategy during training which helps to pick a suitable action after receiving an observation in the incoming time series during testing.

Table 2.4: Summary of miscellaneous early classification approaches.

Paper ID	**Year**	**Major Focus**	**Limitation**	**Type of time series**
[27]	2015	Combining generative and discriminative models	Computationally inefficient due to large number of segments	MTS
[29]	2018	Utilization of correlation	No assurance of reliable prediction	MTS
[32]	2018	Combining deep learning techniques for early prediction	Unable to utilize correlation among the components of MTS	MTS
[59]	2018	Designing reward function to balance between accuracy and earliness	Difficult to understand the early classification problem as reinforcement learning	UTS
[58]	2019	Designing stopping rule using deep learning based posterior probabilities	Difficult to select appropriate values for parameters	UTS

Recently, a deep neural network based early classification framework is developed in [58] that focused on optimizing the tradeoff by estimating the stopping decision probabilities at all time stamps of time series. The authors formulated a new loss function to improve the accuracy and earliness of the classifier.

2.5 Conclusion

This chapter presented a categorization of the existing early classification approaches to get a quick understanding of the notable contributions that have been made over the years. The existing approaches are divided into four categories based on the strategies that have been followed to achieve earliness in the classification. The first category included the prefix based early classification approaches which primary focused on learning MRLs. These approaches are simple and easy to understand. Next category included the shapelet based approaches that have focused on early classification of gene expression data (*i.e.,* MTS) and thus suitable for medical applications. As the doctors may be reluctant to adapt an approach without interpretable results, the primary objective of these approaches was to obtain the key shapelets that can exclusively represent all the time series of one class. Such shapelets are easy to interpret by linking with the patient's disease. In another category, the model based approaches are reviewed where the researchers have focused on developing the stopping criteria or trigger function for early classification of time series. Finally, this chapter discussed miscellaneous approaches that have shown an interest in deep learning models for achieving earliness.

Table 2.5 presents a comparative summary of early classification approaches for MTS. After reviewing the existing literature, we observed that most of the researchers have focused on one or more of the following three parameters: interpretability, reliability, and correlation. The table shows that some of the existing approaches attempted to incorporate correlation in the early classification but they did not ensure the reliability or desired level of accuracy while predicting the class label of an incomplete MTS. Moreover, there exists no work that can handle the issues of MTS such as components of different length, faulty components, imbalance distribution of instances among classes, presence of unseen class, and noisy labels in the training dataset. Thus, further research is required to address these issues in the framework of early classification of MTS.

Table 2.5: Comparative summary of existing literature on early classification approaches for MTS.

Paper ID	Year	Interpretability	Reliability	Correlation
[26]	2012	√	—	—
[23]	2013	√	—	—
[41]	2013	√	—	—
[15]	2013	—	√	—
[22]	2014	—	—	√
[5]	2014	√	—	√
[49]	2015	√	—	√
[31]	2015	√	—	—
[27]	2015	√	√	—
[28]	2017	—	√	—
[29]	2018	—	—	√
[32]	2018	—	√	—
[34]	2019	√	√	—

Publication

- **Ashish Gupta**, H. P. Gupta, B. Biswas, and T. Dutta, "Approaches and Applications of Early Classification of Time Series: A Review," *IEEE Transactions on Artificial Intelligence.* vol. 1, no. 1, pp. 47-61, 2020.