Uncertain Time Series Classification

Michael Franklin Mbouopda
Clermont Auvergne University, CNRS, ENSMSE, LIMOS, F-63000, Clermont-Ferrand, France
michael.mbouopda@uca.fr

Abstract

Time series analysis has gained a lot of interest during the last decade with diverse applications in a large range of domains such as medicine, physics, and industry. The field of time series classification has been particularly active recently with the development of more and more efficient methods. However, the existing methods assume that the input time series is free of uncertainty. However, there are applications in which uncertainty is so important that it cannot be neglected. This project aims to build efficient, robust, and interpretable classification methods for uncertain time series.

1 Motivation

Time series analysis has gained a lot of interest during the last decade. In fact, time series are used in many applications such as smoking and sleep detection [Adams and Marlin, 2018], outlier detection detection [Kieu et al., 2019] and astronomical objects classification [Allam Jr et al., 2018]. The field of time series classification has been particularly active recently with the development of more and more efficient methods [Ruiz et al., 2020; Bagnall et al., 2017]. However, the existing methods assume that the input data is free of uncertainty. This assumption does not hold in every application. For instance, the Plasticc dataset [Allam Jr et al., 2018] contains time series representing astronomical objects such as galaxies and stars. The measurement method produces time series that are not free of uncertainty. Fig. 1 shows a sample from the Plasticc dataset. The blue line is the best guess of the time series and the red region represent the uncertainty region. Any time series that lies in the red region could be the exact (unknown) time series.

More generally, imprecision in data can be caused by the environment, the collection method and technique, privacy constraints and other causes. This imprecision adds epistemic uncertainty in the collected data, and in some applications, it is important for machine learning models to take that into account in order to be efficient.

In order to apply the existing methods to uncertain data, one approach is to simply ignore uncertainty. This approach might work when the uncertainty is not important and can be neglected. The second approach is to remove uncertainty from the data using a preprocessing technique. This approach requires domain knowledge which is not always available. The last approach is to use a method that can automatically handle uncertainty in the data. We use the term end-to-end to qualify such a method. An end-to-end method can be used efficiently without any domain knowledge, since it learns by itself how to handle uncertainty. The field of uncertain time series analysis is still under-explored and the goal of this research project is to build end-to-end methods for uncertain time series analysis.

2 Project Description

Before giving the specific goals of this research project, let us define some core notions.

Definition 1 (Time series). A time series $T$ of length $m$ is an ordered sequence of $m$ observations $t_i$.

$$T = < t_1, t_2, ..., t_m >, t_i \in \mathbb{R}$$

Definition 2 (Uncertain time series). An uncertain time series (uTS) is a time series in which the observations are uncertain.

There are two models for representing an uncertain observation:

- the multi-set based model represents each uncertain observation by a finite set of real values. That is $t_i = \{t_{i1}, t_{i2}, ..., t_{ik}\}, k \in \mathbb{N}$
- the probability density function (PDF) based model represents each uncertain observation by a mean and a deviation. That is $t_i = t_i \pm \delta t_i$, meaning that the unknown exact value of $t_i$ belongs the interval $[t_i - \delta t_i; t_i + \delta t_i]$ and follows a particular probability distribution.
Definition 3 (uTS classification task). Given a dataset \( D = \{ < T_1, c_1 >, < T_2, c_2 >, \ldots, < T_n, c_n > \} \) of \( n \) uncertain time series \( T_i \), each with its class label \( c_i \), taken from a discrete finite set \( C \), the task of uTS classification consists of learning a function that maps the uTS to their class labels.

The goal of this research project is to build end-to-end methods that can resolve the uncertain time series classification task. The methods must have the following properties:

- **Efficiency**: The methods should produce accurate and reliable predictions
- **Robustness**: The methods should be robust to noise and to uncertainty variation in the data. This is very important since we know that the existing time series classifier are vulnerable to adversarial attacks [Fawaz et al., 2019; Karim et al., 2020]
- **Interpretability**: Finally, the methods should provide tools that help to explain or interpret predictions.

3 Contribution

We have proposed the Uncertain Shapelet Transform (UST) algorithm [Mbouopda and Nguifo, 2020], an adaptation of the shapelet transform algorithm [Hills et al., 2014] for uncertain time series. UST is to our knowledge the first and only end-to-end method for uncertain time series classification. UST is built by propagating uncertainty throughout the shapelet transform algorithm. This is achieved by the use of the Uncertain Euclidean Distance (UED) which is obtained by taking uncertainty into account in the Euclidean Distance. UST has two principal components: the uncertain similarity measure used to extract features and the classifier used after features extraction. By propagating uncertainty during features extraction, the classifier can take the propagated uncertainty into account in order to build better decision boundaries: The more the classifier is aware of uncertainty, the better are its predictions. The principal UST components can be configured regarding the dataset characteristics. We shown experimentally that UST is efficient and robust to uncertainty. UST predictions can be explained by visualizing the features (which are actually patterns appearing in the time series) that are more correlated with the predictions.

4 Perspectives

In our paper [Mbouopda and NguIFO, 2020], we applied UST on dataset downloaded from the UEA & UCR repository [Dau et al., 2019]. Since these datasets do not have uncertainty, we manually added synthetic uncertainty to the one we used. The first future work is to apply our method on some real uncertain datasets such as the Plasticc dataset [Allam Jr et al., 2018].

UST is based on a single type of features called shapelet. Simply said, a shapelet is a pattern that is common to time series that belong to the same class. It has been shown that there are many datasets that cannot be efficiently classified using only shapelets [Bagnall et al., 2017]. Therefore, another future work is to build methods that use different type of features.

Acknowledgments

This work is funded by the French Ministry of Higher Education, Research and Innovation and is supervised by Prof. Engelbert MEPHU NGUIFO (https://perso.isima.fr/~enmephun/)

References


