A Complexity-Invariant Measure Based on Fractal Dimension for Time Series Classification

Ronaldo C. Prati, Centro de Matemática, Computação e Cognição, Universidade Federal do ABC, Santo André, SP, Brazil
Gustavo E. A. P. A. Batista, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP, Brazil

ABSTRACT

Classification is an important task in time series mining. It is often reported in the literature that nearest neighbor classifiers perform quite well in time series classification, especially if the distance measure properly deals with invariances required by the domain. Complexity invariance was recently introduced, aiming to compensate from a bias towards classes with simple time series representatives in nearest neighbor classification. To this end, a complexity correcting factor based on the ratio of the more complex to the simpler series was proposed. The original formulation uses the length of the rectified time series to estimate its complexity. In this paper the authors investigate an alternative complexity estimate, based on fractal dimension. Results show that this alternative is very competitive with the original proposal, and has a broader application as it does neither depend on the number of points in the series nor on a previous normalization. Furthermore, these results also verify, using a different formulation, the validity of complexity invariance in time series classification.

Keywords: Complexity Invariance, Dynamic Time Warping, Euclidean Distance, Fractal Dimensions, Time Series Classification

1. INTRODUCTION

Time series data are being generated by numerous applications, with domains ranging from medical and biological experimental observations (e.g. electro encephalograms and cardiograms), financial and trade transactions recordings (e.g. stock market daily fluctuations, periodic variations of mercantile prices of goods and services), data from dynamic processes in scientific experiments (e.g. data generated by the Large Hadron Collider and signals recorded by radio telescopes) and readings of sensor networks (e.g. recordings from high precision agriculture sensors and advanced video game movement control tracking), to name but a few. This ubiquity has propelled the interest in time series data mining in recent years.

In the case of time series classification, numerous research papers report empirical evidence that a simple nearest neighbor classification approach is very competitive, and

DOI: 10.4018/jncr.2012070104
often superior to many state of the art classification methods (Ding et al., 2008). This is especially true if the distance measure properly captures the “invariances” required by the domain at hand (Batista et al., 2011). For instance, some domains require invariance to “warping” (local non-linear accelerations and decelerations). This is the case of most motion capture applications, where the technique used for measuring similarity can compensate from (sub) sequences which may vary in time or speed; similarities in movement patterns would be detected even if in one video the subject was moving slowly and in another it is moving quicker, or even if there were accelerations and decelerations during the course of one observation. The Dynamic Time Warping (DTW) technique (Berndt & Clifford, 1994) is a well known and widely used method which allows the time series to be “stretched” or “compressed” to provide a better match with another time series. Other invariances include uniform scaling, offset, amplitude scaling, phase, occlusion, uncertainty and wandering baseline. Several methods have been proposed to handle them - see (Batista et al., 2011) for a thoughtful discussion about invariances and methods to handle invariances in time series distance calculation.

Batista et al. (2011) introduced new kind of invariance. This invariance was called complexity invariance by the authors. The idea is rather simple and appealing: the “complexity” of a time series may introduce a bias in nearest neighbor classification because complex time series tend to be “closer” to simple time series rather than to other complex time series if the distance measure is not compensated for complexity. Therefore, if a class has simpler time series representations than others, this class will be preferred in nearest neighbor classification and introduce a bias towards the class with simpler representatives. The authors intuitively defined the complexity of a series as “having many peaks, valleys and features”. To capture the notion of complexity difference between two time series, they propose a complexity correction factor based on the ratio of the lengths of the longest (the more complex) to the shortest (simpler) “rectified” series. This correction factor is then multiplied by the distance between the two series in order to make it “complexity-invariant”.

Experimental evidence reported in their paper shows the suitability of the proposed complex invariance correction in numerous domains. However, as the authors report in their paper, the correcting factor used in the experiments is an ad-hoc formulation, and is one of possible complexity measures empirically chosen.

In this paper, we evaluate the use of an alternative way to estimate complexity of time series based on fractal dimension. The fractal (or Hausdorff) dimension is a measure of roughness (or smoothness) for time series and spatial data. As the fractal dimension is an intrinsic measure of the series, it does not depend on the number of the observations of the time series, nor a previous normalization step of amplitude and offset, as the original formulation does.

This paper aims to evaluate the use of this alternative approach based on fractal dimension to estimate the complexity of time series. This complexity estimate is then used to compound the complexity correction factor in the complexity invariance distance. Results of an empirical study using the same experimental setup as in Batista et al. (2011) show that the use of the fractal dimension is very competitive to the original formulation. Therefore, these results corroborate to the hypothesis that complexity correction is an important invariance to be considered in time series classification using a different complexity estimation procedure. Furthermore, the fractal dimension is a more flexible way to compute the complexity of a time series, as it is an intrinsic measure that can be calculated directly from the time series.

The remaining of this paper is structured as follows: Section 2 presents the concept of complexity invariance as originally proposed. Section 3 reviews methods for calculating fractal dimension from time series. Section
13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:
www.igi-global.com/article/complexity-invariant-measure-based-fractal/76377?camid=4v1

www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

Application of Natured-Inspired Technique to Odia Handwritten Numeral Recognition
www.igi-global.com/chapter/application-of-natured-inspired-technique-to-odia-handwritten-numeral-recognition/187696?camid=4v1a

A Structural Model to Investigate Factors Affect Patient Satisfaction and Revisit Intention in Jordanian Hospitals
www.igi-global.com/article/structural-model-investigate-factors-affect/62072?camid=4v1a
Computational Approaches for Modeling Intrinsic Noise and Delays in Genetic Regulatory Networks
www.igi-global.com/chapter/computational-approaches-modeling-intrinsic-noise/38235?camid=4v1a

Optimisation of Engineering Systems With the Bees Algorithm
www.igi-global.com/article/optimisation-of-engineering-systems-with-the-bees-algorithm/211932?camid=4v1a