IMPROVEMENT OF EPSILON COMPLEXITY ESTIMATION AND AN APPLICATION TO SEIZURE PREDICTION

A thesis presented to the faculty of
San Francisco State University
In partial fulfilment of
The Requirements for
The Degree

Master of Arts
In
Mathematics

by

Nathanael Aff

San Francisco, California

August 2017

CERTIFICATION OF APPROVAL

I certify that I have read *IMPROVEMENT OF EPSILON COMPLEXITY ESTIMATION AND AN APPLICATION TO SEIZURE PREDICTION* by Nathanael Aff and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirements for the degree: Master of Arts in Mathematics at San Francisco State University.

Dr. Alexandra Piryatinska
Associate Professor of Mathematics

Dr. Tao He
Assistant Professor of Mathematics

Dr. Chun Kit Lai
Assistant Professor of Mathematics

IMPROVEMENT OF EPSILON COMPLEXITY ESTIMATION AND AN APPLICATION TO SEIZURE PREDICTION

Nathanael Aff
San Francisco State University
2017

The ε-complexity of a continuous function is a measure of the amount of information needed to reconstruct a function with an absolute error not larger than ε. For Hölder class functions, ε-complexity is characterized by a pair of real numbers, the complexity coefficients. The complexity coefficients have been shown to be useful features for the segmentation and classification of time series. In this work, we extend the set of approximation methods used in estimating the complexity coefficients. The performance of these approximation methods is tested on a number of simulated time series. In addition, we test the conjecture that, for a given generating mechanism, the mean of the complexity coefficients is constant. For our set of simulations, we find that the mean of the estimated complexity coefficients is constant on a constant Hölder class of functions. Finally, we apply the ε-complexity coefficients to the prediction of seizures in epileptic mice. We use this technique to identify which EEG signal preceded a seizure with over 80% accuracy.

I certify that the Abstract is a correct representation of the content of this thesis.

Chair, Thesis Committee Date

ACKNOWLEDGMENTS

TABLE OF CONTENTS

1 Introduction 1

Bibliography 5

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

Introduction

Many natural phenomena generate time series that exhibit complex functional and statistical characteristics. For example, an electroencephalogram(EEG) records the electrical potential generated by the synchronized firing of neurons. EEG signals are marked by both transient variations in waveforms and regime changes, distinct breaks in the character and statistical distribution of the EEG. While the exact generating mechanisms of variations in EEG may not be known, characteristics of the observed signal can be used to identify changes in the underlying dynamics.

The theory of $\varepsilon-$complexity as developed by Darkhovsky and Piryatinska provides a method of identifying these regime changes in complex signals such as EEG. The definition of $\varepsilon-$complexity is related to Kolmogorov's definition of the complexity of a sequence. Roughly, Kolmogorov identified the complexity of a discrete sequence as the size of the program, or the amount of information, needed to produce that sequence. The $\varepsilon-$complexity of a continuous function, on the other hand, is the amount of information — in this case, the proportion of the function, needed to reconstruct that function with an absolute error not greater than ε. Darkhovsky and Piryatinska have shown that the $\varepsilon-$complexity of a continuous function can be identified by two real numbers[1]. These are the $\varepsilon-$complexity coefficients or simply the complexity coefficients of the function. In practical applications, a signal like EEG is given by a some finite set of samples. The theory of $\varepsilon-$complexity is adapted to a discrete setting by considering discrete sequences, such as time series, as restrictions of continuous functions to a uniform grid.

The procedure for estimating the $\varepsilon-$complexity of a function entails the iterative approximation of that function on a smaller set of samples at each iteration. The initial implementation of the procedure for estimating $\varepsilon-$complexity coefficients used piecewise polynomials to reconstruct to approximate the original function. We implement the estimation procedure with an enlarged set of approximation methods — basis splines, cubic splines, and an interpolating subdivision method termed the lifting scheme. We test the estimation of $\varepsilon-$complexity on simulated data using each approximation method. First, we compare the reconstruction accuracy of of each method. Then we test the performance of the complexity coefficients estimated with each approximation method when used to classify groups of related time series. We also compare the computational efficiency of each approximation method. Although the basis spline method has the lowest approximation error across all simulations, we find that low approximation error does not correspond to improvements in classification accuracy. Each approximation method performs similarly in the classification task. However, the cubic spline method is the most computationally efficient, and is used to estimate the complexity coefficients in the applications discussed in Chapters 4 and 5.

Darkhovsky and Piryatinska have shown that for a Hölder class of functions the ε-complexity coefficients capture a linear relationship between the log of the approximation error and the log of the proportion of the function used in the approximation. They have conjectured that a constant generating mechanism corresponds to constant mean complexity coefficients. In Chapter 4 we use a number of simulations of deterministic and stochastic processes to test whether, on average, the complexity coefficients are constant for a constant Hölder class of functions. For these processes, a single parameter determines both the Hölder class and the fractal dimension of the time series generated by the processes. For these simulations, the slope complexity coefficient B and an estimator of fractal dimension behave simlarly as the parameters of the processes are varied. For three of the four processes tested, we find that the complexity coefficient B is on average, constant for a constant Hölder class of functions.

In the final chapter, the complexity coefficients are applied to the prediction of seizures in epileptic mice. Features from a 4 minute window of EEG are used to predict whether a stimulus applied to epilepsy-prone mice will result in a seizure. A time series of length n can be considered as a vector in an n-dimensional space. For example, 4 minutes of EEG from a single channel is a represented by 2 million observations. A common method of handling such high dimensional data is to map the data to a lower dimensional set of features. For time series like EEG, features are often calculated on a uniform partition or on a sliding window taken over the length of the signal. For an inhomogeneous time series like EEG, this may lead to features being calculated windows with widely varying characteristics. The $\varepsilon-$complexity was designed to identify such variations in a signal. In order to calculate our features on more homogeneous time periods, we use change points in the complexity coefficients to segment the EEG. The average value of features on these segments are then used as predictors of a seizure response.

Models using this procedure to dynamically segment the EEG signal are compared to several models which use features calculated on uniform partitions of the signal. For the best performing models, we are able to accurately classify seizure outcomes with over 80% accuracy. A model with features calculated on uniform partitions performs as well or better than the models using changes in the complexity coefficients to segment the signal. Due to the small number of trials and some inherent limitations of the data, additional tests would be needed to know if how well the model performance reported here generalizes to other contexts.

Finally, we have created an `R` language package, `ecomplex`, that implements the ε-complexity algorithm. The default settings of this package are based on the simulation experiments described in Chapter 3. Methods used for generating simulations, computing EEG features, and the classification algorithm used in Chapter 5 have been also been included in `R` packages to enable further study or replication of the experiments described in this work.

Bibliography

[1] Boris Darkhovsky and Alexandra Piryatinska, *Epsilon-complexity of continuous functions*, (2013).