

STATISTICAL ANALYSIS AND DATA MINING

Research Article

Significant motifs in time series

Nuno C. Castro , Paulo J. Azevedo

First published: 09 February 2012

<https://doi.org/10.1002/sam.11134>

Cited by: 7



About



Access

JLS



ire

Abstract

Time series motif discovery is the task of extracting previously unknown recurrent patterns from time series data. It is an important problem within applications that range from finance to health. Many algorithms have been proposed for the task of efficiently finding motifs. Surprisingly, most of these proposals do not focus on how to evaluate the discovered motifs. They are typically evaluated by human experts. This is unfeasible even for moderately sized datasets, since the number of discovered motifs tends to be prohibitively large. Statistical significance tests are widely used in the data mining communities to evaluate extracted patterns. In this work we present an approach to calculate time series motifs statistical significance. Our proposal leverages work from the bioinformatics community by using a symbolic definition of time series motifs to derive each motif's p -value. We estimate the expected frequency of a motif by using Markov Chain models. The p -value is then assessed by comparing the actual frequency to the estimated one using statistical hypothesis tests. Our contribution gives means to the application of a powerful technique—statistical tests—to a time series setting. This provides researchers and practitioners with an important tool to evaluate automatically the degree of relevance of each extracted motif. © 2012 Wiley Periodicals, Inc. *Statistical Analysis and Data Mining* 5: 35–53, 2012

Citing Literature



About Wiley Online Library



Help & Support



Opportunities



Connect with Wiley



Copyright © 1999-2018 John Wiley & Sons, Inc. All rights reserved

WILEY