Demo: Time Series Online Measurement for Python (TSOMpy)

Michael Menth and Frederik Hauser
Department of Computer Science
University of Tuebingen, Germany
{menth, frederik.hauser}@uni-tuebingen.de

ABSTRACT

TSOMpy is a Python library for online measurement of time series, i.e., it provides functions to calculate moving averages, moving histograms, and time-dependent rates. The demo illustrates various methods for these concepts and points out their differences. The tool can be used to apply online measurement to time series randomly generated according to specified stochastic processes or to own data sets. The library furthermore allows the reproduction of the tables and figures presented in [1].

1. INTRODUCTION

Adaptive systems monitor their states and use these measurement values for control purposes. Mostly, neither the last observation value nor long-term averages are of interest, but averages of observations over the recent past are needed. We address this issue by online measurement that includes moving averages (MA), moving histograms (MH), and time-dependent rate measurement (TDRM). With online measurement only past samples can be used since future samples are not yet available. In [1], we proposed a framework for the definition of MAs, MHs, and TDRMs. We pointed out shortcomings of existing methods and suggested improvements.

We developed a library for time series online measurement in Python (TSOMpy) implementing all investigated methods of [1] including test and plot functions. The proposed demo illustrates the use of TSOMpy, reproduces the experiments and their visualizations in [1], and allows to apply the various MA, MH, and TDRM methods to configurable stochastic processes and own sample traces.

We implemented TSOMpy in Python because this programming language is most suitable for rapid prototyping and it is one of the most widely used programming language in statistics and data science today. It allows programmers

to produce readable code through plain syntactical concepts, and profits from a wealth of available external community-based libraries that can be easily used. Python is an interpreted programming language supporting, e.g., the object-oriented, the procedural, and the functional programming paradigms. We use Python in version 3.5.0 and limited the dependency on external libraries to matplotlib [2] for data visualization.

The source code of TSOMpy will be published under a GPLv3 license on Github [2].

2. RELATED WORK

General averages like CumSum or (biased) exponential MAs (EMA) are simple and quickly programmed or available as functions in many programming languages or external libraries. One example is the qcc package [3] for the R programming language. The Python community also provides numerous external libraries for basic online measurement: Numpy [4] supports a method for a customizable, weighted average. Pandas [5] includes implementations for CumSum and the (biased) EMA. Other libraries provide CumSum and (biased) EMA for analysis of stock market data [6] or implement the window-based moving average (WMA) [7]. However, we could not find packages implementing advanced online measurement methods like unbiased EMA, time-dependent MAs, time-dependent MHs, novel TDRM methods, and allow their evaluation and comparison. Therefore, we developed the TSOMpy library.

3. CONCEPT

TSOMpy allows the generation of time series according to stochastic processes or alternatively read time series from CSV files. It implements the online measurement methods presented in [1] and allows their application to the time series with a graphical user interface. Additional functions reproduce the tables and figures shown in [1]. In the following, we explain the time series generation offered by TSOMpy. We give an overview of the implemented online measurement methods including class structures and challenges for their illustration.

3.1 Generation and Import of Time Series

Evenly spaced time series can be represented by a vector of values which are associated with evenly spaced time instants whose inter-sample time is a basic time unit Δt. In contrast, unevenly spaced time series consist of both a vector of values and a vector of time instants with possibly different, non-zero inter-sample times. TSOMpy supports generation of constant, equally, exponentially, normally, Gamma-, and

Table 1: Class names of online measurement methods and their full names with reference to their descriptions in [1].

Class name	Full method's name
MAs (for evenly-spaced time series)	
CumMean	Cumulative Mean (3.1.2)
WMA	Window MA (3.1.3)
DWMA	Disjoint Windows MA (3.1.4)
UEMA	Unbiased Exponential MA (3.1.5)
EMA	Exponential MA (3.1.6)
Time-dependent MAs (for unevenly-spaced time series)	
TWMA	Time Window MA (3.2.2)
DTWMA	Disjoint Time Windows MA (3.2.3)
UTEMA	Unbiased Time-Exponential MA (3.2.4)
TEMA	Time-Exponential MA (3.2.5)
MHs (for evenly-spaced time series)	
MH_CumMean	MH with CumMean (5.1)
MH_WMA	MH with WMA (not considered in [1])
MH_DWMA	MH with DWMA (not considered in [1])
MH_UEMA	MH with UEMA (5.1)
MH_EMA	MH with EMA (not considered in [1])
Time-dependent MHs (for unevenly-spaced time series)	
TDMH_TWMA	TDMH with TWMA (not considered in [1])
TDMH_DTWMA	TDMH with DTWMA (not considered in [1])
TDMH_UTEMA	TDMH with UTEMA (5.2)
TDMH_TEMA	TDMH with TEMA (not considered in [1])
Time-dependent RMs (for unevenly-spaced time series)	
TDRM_TWMA	TDRM with Time Window MA (6.1.2)
TDRM_DTWMA	TDRM with DTWMA (6.1.3)
TDRM_DTWMA_UEMA	TDRM with DTWMA and UEMA (6.1.4)
TDRM_UTEMA	TDRM with UTEMA (6.1.5)
TDRM_UTEMA_CPA	TDRM with UTEMA and Continuous Packet Arrivals (6.1.6)

Pareto-distributed values and also values created according to an autoregressive process. It offers constant, equally, exponentially, Gamma-, and Pareto-distributed inter-sample times for the time vectors. The distributions can be configured with appropriate parameters. Both time and value vectors can be imported from CSV files. It is possible to combine imported time or value vectors with generated counterparts.

3.2 Online Measurement Methods

Table 1 lists all implemented online measurement methods. Normal MAs and MHs are applicable only to evenly spaced time series while time-dependent MAs (TDMA) and MHs (TDMH) are extended for application to unevenly spaced time series. TDRM methods are applied only to unevenly spaced time series. They can also be applied to mere time vectors with a constant value vector or to a value vector combined with an evenly spaced time vector.

We used an object-oriented class hierarchy for the implementation of all online measurement methods. Normal MA methods inherit from "MovingAverage", TDMA methods inherit from "TimeDependentMovingAverage", MH methods inherit from "MovingHistogram", and TDMH methods in-

herit from "TimeDependentMovingHistogram". "TimeDepenentRateMeasurement" inherits from "TimeDependentMovingAverage" and TDRM methods inherit from both "TimeDependentMovingRate" and possibly from some specific TDMA method.

Some TDRM methods produce step curves with jumps at sample instants and drops some constant time afterwards. Others lead to curves with jumps at sample instants and continuously decreasing phases in between. Illustrations are provided in [1]. TSOMpy produces such curves and offers functions to calculate variance, differences, and accumulates.

4. DEMONSTRATION

The demonstration explains various online measurement methods and points out their differences using visualization and quantitative metrics.

The module *TsomDemo* includes classes demonstrating the usage of TSOMpy, e.g., the reproduction of the tables and figures in [1] with the same or different parameters. It also provides a user interface to apply the online measurement methods to imported data or to specified random data. In addition to visualization, calculated averages and rates can be exported as CSV for further investigation, e.g., in a statistical software tool like R. TSOMpy includes functions that may be used by other processes to calculate the proposed metrics and to visualize them.

5. CONCLUSION

TSOMpy is a Python library for time series online measurement. The associated tool and the demo illustrate various algorithms for the moving averages (MA), moving histograms (MH), and time-dependent rate measurement (TDRM) presented in [1]. Moreover, it reproduces the results presented in the tables and figures of that paper and allows the community to evaluate the proposed online measurement methods with own experiments and data. On the one hand, the tool may serve educational purposes and be used for the processing of research data. On the other hand, the functions provided by the library may directly be integrated in control algorithms of adaptive systems.

6. REFERENCES

[1] M. Menth and F. Hauser, "On Moving Averages, Histograms and Time-Dependent Rates for Online Measurement," in *ACM/SPEC ICPE*, 2017.

[2] ——, "TSOMpy – Time Series Online Measurement in Python," https://www.github.com/uni-tue-kn/TSOMpy, 2017.

[3] L. Scrucca. (2014, Oct.) Cran: Quality control charts (qcc). [Online]. Available: https://cran.r-project.org/web/packages/qcc/

[4] N. Developers. (2016, May) Numpy. [Online]. Available: http://www.numpy.org/

[5] T. P. D. Team. (2016, Dec.) Python data analysis library. [Online]. Available: http://pandas.pydata.org/

[6] C. Zhuang. (2016, Nov.) Python package index: Stockstats 0.2.0. [Online]. Available: https://pypi.python.org/pypi/stockstats

[7] S. Reifschneider. (2011, Nov.) Github repository: Python-movingaverage. [Online]. Available: https://github.com/linsomniac/python-movingaverage