# A General Logical Approach to Learning from Time Series

**Guido Sciavicco**
Department of Mathematics and Computer Science, University of Ferrara, Italy

## Abstract

Machine learning from multivariate time series is a common task, and countless different approaches to typical learning problems have been proposed in recent years. In this talk, we review some basic ideas towards logic-based learning methods, and we sketch a general framework.

**2012 ACM Subject Classification** Theory of computation → Theory and algorithms for application domains

**Keywords and phrases** Machine learning, temporal logic, general approach

**Digital Object Identifier** 10.4230/LIPIcs.TIME.2024.1

**Category** Invited Talk

**Funding** We acknowledge the support of the FIRD project *Methodological Developments in Modal Symbolic Geometric Learning*, funded by the University of Ferrara, and the INDAM-GNCS project *Symbolic and Numerical Analysis of Cyberphysical Systems* (code CUP_E53C23001670001), funded by INDAM; Guido Sciavicco is a GNCS-INdAM member. Moreover, this research has also been funded by the Italian Ministry of University and Research through PNRR - M4C2 - Investimento 1.3 (Decreto Direttoriale MUR n. 341 del 15/03/2022), Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 8 "Pervasive AI", funded by the European Union under the "NextGeneration EU programme".

## 1 Extended Abstract

*Time series* are temporally ordered observations. Time series can be *univariate* or *multivariate*, depending on whether there is a single one or multiple measurements, and each measurement, known as *temporal variable* can be either numerical or categorical.

Time series are ubiquitous in computer science. In some areas, data have naturally the form of multivariate time series; this is the case, for example, of *predictive maintenance* [5] in industry, usually obtained via the recording of sensors' values (e.g., vibration sensors, gas exhaust sensors), of *(hospitalized) patients' monitoring* [2], during which the value of vital signs (e.g., oximetry, blood pressure, temperature) is recorded in order to quickly identify variations or deterioration of the condition. In other areas, time series arise from data whose temporal nature is often overlooked; examples range from *acoustic data* [3] (e.g., voice, cough samples, breath samples), in which case the data becomes temporal when audio is separated into its frequency component (for example via a Fourier transform) whose power changes over time – in the scale of the milliseconds, to *brain waves recording data* [4] (e.g from an electroencephalogram), in which the electrical power at different frequencies (again, after a Fourier transform) changes over time and across different electrodes, up to *textual data*: tokens (e.g., words) follow a linear order, and their different characterizations (e.g., syntactic type, semantic value) can be seen as temporal variables, implying that even text, in a sense, can be interpreted as a multivariate time series [1].

A collection of data instances, or *dataset*, is associated to several classic problems. Given a single multivariate time series, seen as its own dataset, the most natural machine learning problem is that of *forecasting*, defined as the problem of establishing the next value of a specific temporal variable given the past values of all variables. Otherwise, given a dataset of several multivariate time series, one can formulate a classification problem: if the dataset if labelled, then the problem is *supervised classification/regression*, and if it is not, it becomes *unsupervised classification*. The typical approaches to forecasting are based on examining the values of the time series in the last temporal points preceding the one for which the forecasting is required; by partitioning the time series into several chunks, each labeled with the value of the variable to be forecasted, then a single time series becomes a dataset itself, and forecasting can be reduced to regression). Moreover, both supervised classification/regression and unsupervised classification are pattern-searching problems; in the former case the search of patterns is guided by some measure of information on the class (e.g., entropy), and in the latter case by some measure of information on the pattern itself (e.g., frequency). In this sense, it can be said that with time series, at some abstract level there is only one machine learning paradigm of interest, that is, *classification*, or pattern extraction.

In this talk we focus on *logical* pattern extraction of datasets of multivariate time series. Patterns can be written in different logical languages, from propositional, to modal (temporal), to first-order, and beyond. One common idea to all symbolic methods and techniques is that a multivariate time series can be seen as a model of a logical framework, and pattern extraction is essentially a model-checking exercise.

Although the logical languages may vary, it is possible to give a general definition of propositional letter/atomic relation, that serves as a starting point. To this end, we consider a set of time points (e.g., from moment $t_1$ to moment $t_2$), and the value of all variables at those points (e.g.„ the level of vibration picked up by sensors $A$ and $B$); then we apply a function to the set all values (e.g., the average). Finally, we compare the result to a constant, ending up with an atomic sentence (e.g., the average vibration between sensors $A$ and $B$ is below $100Hz$). By varying the parameters that govern such generic atomic statements, one obtains a wide range of basic alphabets, which can be combined with different languages.

As we shall see, logical pattern extraction from datasets of time series has been and can be approached in ways that range from very simple (i.e., using propositional logic), to relatively complex (i.e., using point-based or interval-based temporal logics), to very complex (i.e., using higher order logics). Standard symbolic machine learning methods (e.g., decision trees and lists learning algorithms) are designed for propositional logic, but they are progressively adapting to more complex languages, and in this talk we give an overview of this landscape.

## References

1. D. Del Fante, F. Manzella, G. Sciavicco, and I. E. Stan. A post-modern approach to automatic metaphor identification. In *Proc. of the 9th Italian Conference on Computational Linguistics (CLIC)*, volume 3596 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2023. URL: https://ceur-ws.org/Vol-3596/short9.pdf.
2. R. Hegde, S. Ranjana, and C. D. Divya. Survey on development of smart healthcare monitoring system in iot environment. In *Proc. of the 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pages 395–399, 2021.
3. F. Manzella, G. Pagliarini, G. Sciavicco, and I. E. Stan. The voice of COVID-19: breath and cough recording classification with temporal decision trees and random forests. *Artif. Intell. Medicine*, 137:102486, 2023. `doi:10.1016/J.ARTMED.2022.102486`.
4. S. Mazzacane, M. Coccagna, F. Manzella, G. Pagliarini, V. A. Sironi, A. Gatti, E. Caselli, and G. Sciavicco. Towards an objective theory of subjective liking: A first step in understanding the sense of beauty. *PLOS ONE*, 18(6):1–20, June 2023.
5. Y. Ran, X. Zhou, P. Lin, Y. Wen, and R. Deng. A survey of predictive maintenance: Systems, purposes and approaches. *ArXiv*, abs/1912.07383, 2019. `arXiv:1912.07383`.