# Symbolic Representation of Time Series: a Hierarchical Coclustering Formalization

Alexis Bondu, Marc Boullé, and Antoine Cornuéjols

EDF R&D, 1 avenue du Général de Gaulle 92140 Clamart France,
Orange Labs, 2 avenue Pierre Marzin 22300 Lannion France,
AgroParisTech, 16 rue Claude Bernard 75005 Paris France.
alexis.bondu@edf.fr,
marc.boulle@orange.com,
antoine.cornuejols@agroparistech.fr

**Abstract.** The choice of an appropriate representation remains crucial for mining time series, particularly to reach a good trade-off between the dimensionality reduction and the stored information. Symbolic representations constitute a simple way of reducing the dimensionality by turning time series into sequences of symbols. SAXO is a data-driven symbolic representation of time series which encodes typical distributions of data points. This approach was first introduced as a heuristic algorithm based on a regularized coclustering approach. The main contribution of this article is to formalize SAXO as a hierarchical coclustering approach. The search for the best symbolic representation given the data is turned into a model selection problem. Comparative experiments demonstrate the benefit of the new formalization, which results in representations that drastically improve the compression of data.

**Keywords:** Time series, symbolic representation, coclustering

## 1 Introduction

The choice of the representation of time series remains crucial since it impacts the quality of supervised and unsupervised analysis [1]. Time series are particularly difficult to deal with due to their inherently high dimensionality when they are represented in the time-domain [2] [3]. Virtually all data mining and machine learning algorithms scale poorly with the dimensionality. During the last two decades, numerous high level representations of time series have been proposed to overcome this difficulty. The most commonly used approaches are: the Discrete Fourier Transform [4], the Discrete Wavelet Transform [5] [6], the Discrete Cosine Transform [7], the Piecewise Aggregate Approximation (PAA) [8]. Each representation of time series encodes some information derived from the raw data[1]. According to [1], mining time series heavily relies on the choice of a representation and a similarity measure. Our objective is to find a **compact** and **informative** representation which is driven by the data. The symbolic representations constitute a simple way of reducing the dimensionality of the data by turning time series

[1] *"Raw data"* designates a time series represented in the time-domain by a vector of real values.

into sequences of symbols [9]. In such representations, each symbol corresponds to a time interval and encodes information which summarize the related sub-series. Without making hypothesis on the data, such a representation does not allow one to quantify the loss of information. This article focuses on a less prevalent symbolic representation which is called SAXO[2]. This approach optimally discretizes the time dimension and encodes typical distributions[3] of data points with the symbols [10]. SAXO offers interesting properties. Since this representation is based on a **regularized** Bayesian coclustering[4] approach called MODL[5] [11], a good trade-off is naturally reached between the dimensionality reduction and the information loss. SAXO is a parameter-free and data-driven representation of time series. In practice, this symbolic representation proves to be highly **informative** for training classifiers. In [10], SAXO was evaluated on public datasets and favorably compared with the SAX representation.

Originally, SAXO was defined as a heuristic algorithm. The two main contributions of this article are: i) the **formalization** of SAXO as a hierarchical coclustering approach; ii) the evaluation of its **compactness** in terms of coding length. This article is organized as follows. Section 2 briefly introduces the symbolic representations of time series and presents the original SAXO heuristic algorithm. Section 3 formalizes the SAXO approach resulting in a new evaluation criterion which is the main contribution of this article. Experiments are conducted in Section 4 on real datasets in order to compare the SAXO evaluation criterion with that of the MODL coclustering approach. Lastly, perspectives and future works are discussed in Section 5.

## 2 Related work

Numerous compact representations of time series deal with the curse of dimensionality by discretizing the time and by summarizing the sub-series within each time interval. For instance, the Piecewise Aggregate Approximation (PAA) encodes the mean values of data points within each time interval. The Piecewise Linear Approximation (PLA) [12] is an other example of compact representation which encodes the gradient and the y-intercept of a linear approximation of sub-series. In both cases, the representation consist of numerical values which describe each time interval. In contrast, the symbolic representations characterize the time intervals by categorical variables [9]. For instance, the Shape Definition Language (SDL) [13] encodes the shape of sub-series by symbols. The most commonly used symbolic representation is the SAX[6] approach [9]. In this case, the time dimension is discretized into regular intervals, the symbols encode the mean values per interval.

[2] SAXO *Symbolic Aggregate approXimation Optimized by data*.

[3] The SAXO approach produces clusters of time series within each time interval which correspond to the symbols.

[4] The coclustering problem consist in reordering rows and columns of a matrix in order to satisfy a homogeneity criterion.

[5] *Minimum Optimized Description Length*

[6] *Symbolic Aggregate approXimation.*

The symbolic representations appear to be really helpful for processing large datasets of time series owing to dimensionality reduction. However, these approaches suffer several limitations.

- Most of these representations are lossy compression approaches unable to quantify the loss of information without strong hypothesis on the data.
- The discretization of the time dimension into regular intervals is not data driven.
- The symbols have the same meaning over time irrespectively of their rank *(i.e. the ranks of the symbols may be used to improve the compression).*
- Most of these representations involve user parameters which affect the stored information *(ex: for the SAX representation, the number of time intervals and the size of the alphabet must be specified).*

The SAXO approach overcomes these limitations by optimizing the time discretization, and by encoding typical distributions of data points within each time interval [10]. SAXO was first defined as a heuristic which exploits the MODL coclustering approach.

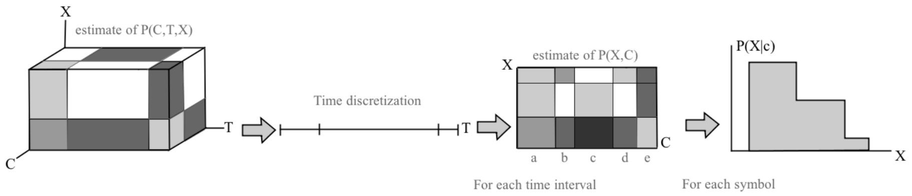
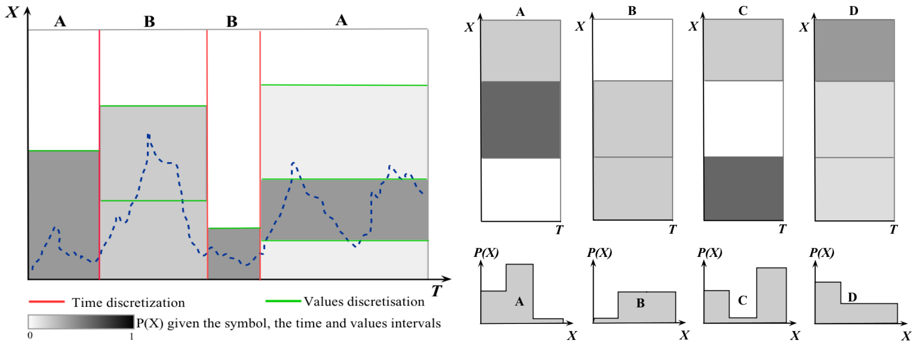


**Fig. 1.** Main steps of the SAXO learning algorithm.

Figure 1 provides an overview of this approach by illustrating the main steps of the learning algorithm. The joint distribution of the identifiers of the time series $C$, the values $X$, and the timestamp $T$ is estimated by a trivariate coclustering model. The time discretization resulting from the first step is retained, and the joint distribution of $X$ and $C$ is estimated within each time interval by using a bivariate coclustering model. The resulting clusters of time series are characterized by piecewise constant distributions of values and correspond to the symbols. A specific representation allows one to re-encode the time series as a sequence of symbols. Then, the typical distribution that best represents the data points of the time series is selected within each time interval. Figure 2(a) plots an example of recoded time series. The original time series *(represented by the blue curve)* is recoded by the **"abba"** SAXO word. The time is discretized into four intervals *(the vertical red lines)* corresponding to each symbol. Within time intervals, the values are discretized *(the horizontal green lines)*: the number of intervals of values and their locations are not necessary the same. The symbols correspond to typical distributions of values: conditional probabilities of $X$ are associated with each cell of the grid *(represented by the gray levels)*; Figure 2(b) gives an example of the alphabet associated with the second time interval. The four available symbols correspond to typical distributions which are both represented by gray levels and by histograms. By considering Figures 2(a) and 2(b), **b** appears to be the closest typical distribution of the second sub-series.



(a) (b)

**Fig. 2.** Example of a SAXO representation (a) and the alphabet of the second time interval (b).

As in any heuristic approach, the original algorithm finds a suboptimal solution for selecting the most suitable SAXO representation given the data. Solving this problem in an exact way appears to be intractable, since it is comparable to the coclustering problem which is NP-hard. The main contribution of this paper is to **formalize** the SAXO approach within the MODL framework. We claim this formalization is a first step to improving the quality of the SAXO representations learned from data. In this article, we define a new evaluation criterion denoted by $C_{saxo}$ *(see Section 3)*. The most probable SAXO representation given the data is defined by minimizing $C_{saxo}$. We expect to reach better representations by optimizing $C_{saxo}$, instead of exploiting the original heuristic algorithm.

## 3 Formalization of the SAXO approach

This section presents **the main contribution** of this article: the SAXO approach is formalized as a hierarchical coclustering approach. As illustrated in Figure 3, the originality of the SAXO approach is that the groups of identifiers *(variable $C$)* and the intervals of values *(variable $X$)* are allowed to change over time. By contrast, the MODL coclustering approach forces the discretization of $C$ and $X$ to be the same within time intervals. Our objective is to reach better models by removing this constraint.

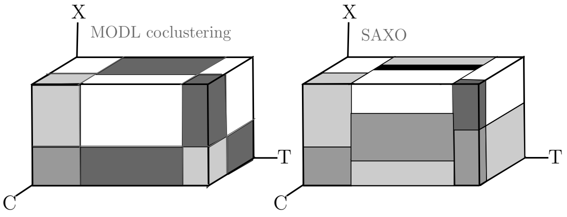


**Fig. 3.** Examples of a MODL coclustering model *(left part)* and a SAXO model *(right part)*.

A SAXO model is hierarchically instantiated by following two successive steps. First, the discretization of time is determined. The bivariate discretization $C \times X$ is then defined within each time interval. Additional notations are required to describe the sequence of bivariate data grids.

**Notations for time series:** *In this article, the input dataset $\mathcal{D}$ is considered to be a collection of $N$ time series denoted $S_i$ (with $i \in [1, N]$). Each time series consists of $m_i$ data points, which are couples of values $X$ and timestamps $T$. The total number of data points is denoted by $m = \sum_{i=1}^{N} m_i$.*

**Notations for the *t-th* time interval of a SAXO model:**

- $k_T$*: number of time intervals;*
- $k_C^t$*: number of clusters of time series;*
- $k_X^t$*: number of intervals of value;*
- $k_C(i,t)$*: index of the cluster that contains the sub-series of* $S_i$*;*
- $\{n_{i_C}^t\}$*: number of time series in each cluster* $i_C^t$*;*
- $m_t$*: number of data point;*
- $m_i^t$*: number of data points of each time series* $S_i$*;*
- $m_{i_C}^t$*: number of data points in each cluster* $i_C^t$*;*
- $\{m_{j_X}^t\}$*: number of data points in the intervals* $j_X$*;*
- $\{m_{i_C j_X}^t\}$*: number of data points belonging to each cell* $(i_C, j_X)$*.*

Eventually, a SAXO model $M'$ is first defined by a number of time intervals and the location of their bounds. The bivariate data grids $C \times X$ within each time interval are defined by: i) the partition of the time series into clusters; ii) the number of intervals of values; iii) the distribution of the data points on the cells of the data grid; iv) for each cluster, the distribution of the data points on the time series belonging to the same cluster. Section 3.1 presents the prior distribution of the SAXO models. The likelihood of a SAXO model given the data is described in Section 3.2. A new evaluation criterion which defines the most probable model given the data is proposed in Section 3.3.

### 3.1 Prior distribution of the SAXO models

The proposed prior distribution $P(M')$ exploits the hierarchy of the parameters of the SAXO models and is uniform at each level. The prior distribution of the number of time intervals $k_T$ is given by Equation 1. The parameter $k_T$ belongs to $[1, m]$, with $m$ representing the total number of data points. All possible values of $k_T$ are considered as equiprobable. By using combinatorics, the number of possible locations of the bounds can be enumerated given a fixed value of $k_T$. Once again, all possible locations are considered as equiprobable. Equation 2 represents the prior distribution of the parameter $\{m^t\}$ given $k_T$. Within each time interval $t$, the number of intervals of values $k_X^t$ is uniformly distributed *(see Equation 3)*. The value of $k_X^t$ belongs to $[1, m^t]$, with $m^t$ representing the number of data points within the *t-th* time interval. All possible values of $k_X^t$ are equiprobable. The same approach is applied to define the prior distribution of the number of clusters within each time interval *(see Equation 4)*. The value of $k_C^t$

belongs to $[1, N]$, with $N$ denoting the total number of time series. Once again, all possible values of $k_C^t$ are equiprobable. The possible ways of partitioning the $N$ time series into $k_C^t$ clusters can be enumerated, given a fixed number of clusters in the *t-th* time interval. The term $B(N, k_C^t)$ in Equation 5 represents the number of possible partitions of $N$ elements into $k_C^t$ possibly empty clusters[7]. Within each time interval, all distributions of the $m^t$ data points on the cells of the bivariate data grid $C \times X$ are considered as equiprobable. Equation 6 enumerates the possible ways of distributing $\{m^t\}$ data points on $k_X^t.k_C^t$ cells. Given a time interval $t$ and a cluster $i_C^t$, all distributions of the data points on the time series belonging to the same cluster are equiprobable. Equation 7 enumerates the possible ways of distributing $m_i^t$ data points on $n_{i_C}^t$ time series.

$$P(k_T) = \frac{1}{m} \tag{1}$$

$$P(\{m^t\}|k_T) = \frac{1}{\binom{m+k_T-1}{k_T-1}} \tag{2}$$

$$P(\{k_X^t\}|k_T, \{m^t\}) = \prod_{t=1}^{k_T} \frac{1}{m^t} \tag{3}$$

$$P(\{k_C^t\}|k_T) = \prod_{t=1}^{k_T} \frac{1}{N} \tag{4}$$

$$P\left(k_C(i,t)|k_T, \{k_C^t\}\right) = \prod_{t=1}^{k_T} \frac{1}{B(N, k_C^t)} \tag{5}$$

$$P\left(\{m_{j_C, j_X}^t\}|k_T, \{m^t\}, \{k_X^t\}, \{k_C^t\}\right) = \prod_{t=1}^{k_T} \frac{1}{\binom{m^t+k_C^t.k_X^t-1}{k_C^t.k_X^t-1}} \tag{6}$$

$$P\left(\{m_i^t\}|k_T, \{k_C^t\}, k_C(i,t), \{m_{j_C, j_X}^t\}\right) = \prod_{t=1}^{k_T}\prod_{i=1}^{k_C^t} \frac{1}{\binom{m_{i_C}^t+n_{i_C}^t-1}{n_{i_C}^t-1}} \tag{7}$$

In the end, the prior distribution of the SAXO models $M'$ is given by Equation 8.

$$\begin{aligned} P(M') = \frac{1}{m} \times \frac{1}{\binom{m+k_T-1}{k_T-1}} \times \prod_{t=1}^{k_T} \Bigg[ & \frac{1}{m^t} \times \frac{1}{N} \times \frac{1}{B(N, k_C^t)} \\ & \times \frac{1}{\binom{m^t+k_C^t.k_X^t-1}{k_C^t.k_X^t-1}} \times \prod_{i=1}^{k_C^t} \frac{1}{\binom{m_{i_C}^t+n_{i_C}^t-1}{n_{i_C}^t-1}} \Bigg] \end{aligned} \tag{8}$$

### 3.2 Likelihood of data given a SAXO model

A SAXO model matches with several possible datasets. Intuitively, the likelihood $P(D|M')$ enumerates all the datasets which are compatible with the parameters of the model $M'$. The first term of the likelihood represents the distribution of the ranks of the values of $T$. In other words, Equation 9 codes all the possible permutations of the data points within each time interval. The second term enumerates all the possible distributions of the $m$ data points on the $k_T$ time intervals, which are compatible with the parameter $\{m^t\}$ *(see Equation 10)*. In the same way, Equation 11 enumerates the distributions of

[7] The second kind of Stirling numbers $S\{{v \atop k}\}$ enumerates the possible partitions of $v$ elements into $k$ clusters and $B(N, k_C^t) = \sum_{i=1}^{k_C^t} S\{{N \atop i}\}$.

the $m^t$ data points on the $k^t_X.k^t_C$ cells of the bivariate data grids $C \times X$ within each time interval. The considered distributions are compatible with the parameter $\{m^t_{i_C,j_X}\}$. For each time interval and for each cluster, Equation 12 enumerates all the possible distributions of the data points on the time series belonging to the same cluster. Equation 13 enumerates all the possible permutations of the data points in the intervals of $X$, within each time interval. This information must also be coded over all the time intervals, which is equivalent to enumerating all the possible fusions of $k_T$ stored lists in order to constitute a global stored list *(see Equation 14)*. In the end, the likelihood of the data given a SAXO models $M'$ is characterized by Equation 15.

$$\frac{1}{\prod_{t=1}^{k_T} m^t!} \quad (9) \qquad \frac{1}{\frac{m!}{\prod_{t=1}^{k_T} m^t!}} \quad (10) \qquad \prod_{t=1}^{k_T} \frac{1}{\frac{m^t!}{\prod_{i_C=1}^{k^t_C} \prod_{j_X=1}^{k^t_X} m^t_{i_C,j_X}!}} \quad (11)$$

$$\prod_{t=1}^{k_T} \frac{1}{\frac{\prod_{i_C=1}^{k^t_C} m^t_{i_C}!}{\prod_{i=1}^{N} m^t_i!}} \quad (12) \qquad \prod_{t=1}^{k_T} \frac{1}{\prod_{j_X=1}^{k^t_X} m^t_{j_X}!} \quad (13) \qquad \frac{1}{\frac{m!}{\prod_{t=1}^{k_T} m^t!}} \quad (14)$$

$$P(D|M') = \frac{1}{m!^2} \times \prod_{t=1}^{k_T} \left[ \frac{\prod_{i_C=1}^{k^t_C} \prod_{j_X=1}^{k^t_X} m^t_{i_C,j_X}! \times \prod_{i=1}^{N} m^t_i!}{\prod_{j_X=1}^{k^t_X} m^t_{j_X}! \times \prod_{i_C=1}^{k^t_C} m^t_{i_C}!} \right] \quad (15)$$

### 3.3 Evaluation criterion

The SAXO evaluation criterion is the negative logarithm of $P(M') \times P(D|M')$ *(see Equation 16)*. The first three lines correspond to the prior term $-log(P(M'))$ and the last two lines represent the likelihood term $-log(P(M'|D))$. The most probable model given the data is found by minimizing $C_{saxo}(M')$ over the set of all possible SAXO models denoted by $\mathbb{M}'$.

$$\begin{aligned}
C_{saxo}(M') &= log(m) + log\binom{m + k_T - 1}{k_T - 1} + \sum_{t=1}^{k_T} log(m^t) \\
&+ k_T.log(N) + \sum_{t=1}^{k_T} log\left(B(N, k^t_C)\right) + \sum_{t=1}^{k_T} log\binom{m^t + k^t_C.k^t_X - 1}{k^t_C.k^t_X - 1} \\
&+ \sum_{t=1}^{k_T} \sum_{i_C=1}^{k^t_C} log\binom{m^t_{i_C} + n^t_{i_C} - 1}{n^t_{i_C} - 1} \\
&+ 2.log(m!) - \sum_{t=1}^{k_T} \sum_{i_C=1}^{k^t_C} \sum_{j_X=1}^{k^t_X} log(m^t_{i_C,j_X}!) \\
&+ \sum_{t=1}^{k_T} \left[ \sum_{i_C=1}^{k^t_C} log(m^t_{i_C}!) - \sum_{i=1}^{N} log(m^t_i!) + \sum_{j_X=1}^{k^t_X} log(m^t_{j_X}!) \right]
\end{aligned} \quad (16)$$

**Key ideas to retain:** Rather than having a heuristic decomposition of the SAXO approach in a two-step algorithm, we propose a single evaluation criterion based on the MODL framework. Once optimized, this criterion should yield better representations of time series. We compare the ability of both criterion to compress data. We aim at evaluating the interest of optimizing $C_{saxo}$ rather than the original trivariate coclustering criterion [14] *(denoted by $C_{modl}$).*

## 4 Comparative experiments on real datasets

According to the information theory and since both criteria are a negative logarithm of a probability, $C_{saxo}$ and $C_{modl}$ represent the coding length of the models. In this section, both approaches are compared in terms of coding length. The 20 processed datasets come from the *UCR Time Series Classification and Clustering repository* [15]. Some datasets are relatively small, we have selected the ones which include at least 800 learning examples. Originally, these datasets are divided into training and test sets which have been merged in our experiments. The objective of this section is to compare $C_{saxo}$ and $C_{modl}$ for each dataset. On the one hand, the criterion $C_{modl}$ is optimized by using the greedy heuristic and a neighborhood exploration mentioned described in [11]. The coding length of the most probable MODL model *(denoted by $MAP_{modl}$)* is then calculated by using $C_{modl}$. On the other hand, the criterion $C_{saxo}$ is optimized by exploiting the original heuristic algorithm illustrated in Figure 1 [10]. The coding length of best SAXO model *(denoted by $MAP_{saxo}$)* is given by the criterion $C_{saxo}$. Notice that both algorithms have a $\mathcal{O}(m\sqrt{m}\log m)$ time complexity. The order of magnitude of the coding length depends on the size of the data set and can not be easily compared over all datasets. We choose to exploit the compression gain [16] which consists in comparing the coding length of a model $M$ with the coding length of the simplest model $M_{sim}$. This key performance indicator varies in the interval $[0, 1]$. The compression gain is similarly defined for the MODL and the SAXO approaches such that:

$$\mathcal{G}ain_{modl}(M) = 1 - C_{modl}(M)/C_{modl}(M_{sim})$$
$$\mathcal{G}ain_{saxo}(M') = 1 - C_{saxo}(M')/C_{saxo}(M_{sim})$$

Our experiments evaluate the variation of the compression gain between the SAXO and the MODL approaches. This indicator is denoted by $\Delta_G$ and represents the relative improvement of the compression gain provided by SAXO. The value of $\Delta_G$ can be negative, which means that SAXO provides a worse compression gain than the MODL approach.

$$\Delta_G = \frac{\mathcal{G}ain_{saxo}(MAP_{saxo}) - \mathcal{G}ain_{modl}(MAP_{modl})}{\mathcal{G}ain_{modl}(MAP_{modl})}$$

Table 1 presents the results of our experiments and includes a particular case with a missing value for the dataset “TwoPatterns”. In this case, the first step of the heuristic

| Dataset | $\Delta_G$ | Dataset | $\Delta_G$ |
|---|---|---|---|
| Starlight curves | 63.86% | CBF | $-1.43\%$ |
| uWaveGestureX | 191.41% | AllFace | 383.24% |
| uWaveGestureY | 157.79% | Symbols | 23.16% |
| uWaveGestureZ | 185.13% | 50 Words | 400.68% |
| ECG Five Days | $-1.80\%$ | Wafer | 37.03% |
| MoteStrain | 627.84% | Yoga | 63.40% |
| CincEGCtorso | 32.93% | FacesUCR | $-18.39\%$ |
| MedicalImages | 191.32% | Cricket Z | 290.22% |
| WordSynonym | 264.93% | Cricket X | 285.87% |
| TwoPatterns | missing | Cricket Y | 296.40% |

**Table 1.** Coding length evaluation.

algorithm which optimizes $C_{saxo}$ *(see Figure 1)* leads to the simplest trivariate coclustering model that includes a single cell. This is a side effect due to the fact that the MODL approach is regularized. A possible explanation is that the temporal representation of time series is not informative for this dataset. Other representations such as the Fourier or the wavelet transforms could be tried. In most cases, $\Delta_G$ has a positive value which means SAXO provides a better compression than the MODL approach. This trend emerges clearly, the average compression improvement reaches $183\%$. We exploit the *Wilcoxon signed-ranks test* to reliably comparing both approaches over all datasets [17]. If the output value *(denoted by z)* is smaller than $-1.96$, the gap in performance is considered as significant. Our experiments give $z = -3.37$ which is highly significant. In the end, the compression of data provided by SAXO appears to be intrinsically better than the MODL approach. The prior term of $C_{saxo}$ induces an additional cost in terms of coding length. This additional cost is far outweighed by a better encoding of the likelihood.

## 5 Conclusion and perspectives

SAXO is a data-driven symbolic representation of time series which extends SAX in three ways: i) the discretization of time is optimized by a Bayesian approach rather than considering regular intervals; ii) the symbols within each time interval represents typical distributions of data points rather than average values; iii) the number of symbols may differ per time interval. The parameter settings is automatically optimized given the data. SAXO was first introduced as an heuristic algorithm. This article formalizes this approach within the MODL framework as a hierarchical coclustering approach *(see Section 3)*. A Bayesian approach is applied leading to an analytical evaluation criterion. This criterion must be minimized in order to define the most probable representation given the data. This new criterion is evaluated on real datasets in Section 4. Our experiments compare the SAXO representation with the original MODL coclustering approach. The SAXO representation appears to be significantly better in terms of data compression. In future work, we plan to use the SAXO criterion in order to define a similarity measure. Numerous learning algorithms, such as $K$-means and $K$-NN, could use such an improved similarity measure defined over time series. We plan to explore potential gains in areas such as: i) the detection of atypical time series; ii) the query of a database by similarity; iii) the clustering of time series.

## References

1. T. Liao, "Clustering of time series data: a survey," *Pattern Recognition*, vol. 38, pp. 1857–1874, 2005.
2. D. Bosq, *Linear Processes in Function Spaces: Theory and Applications (Lecture Notes in Statistics)*. Springer, 2000.
3. J. Ramsay and B. Silverman, *Functional Data Analysis*, ser. Springer Series in Statistics. Springer, 2005.
4. M. Frigo and S. Johnson, "The design and implementation of FFTW3," *Proceedings of the IEEE*, vol. 93, no. 2, pp. 216–231, 2005, special issue on "Program Generation, Optimization, and Platform Adaptation".
5. R. Polikar, *Physics and Modern Topics in Mechanical and Electrical Engineering*. World Scientific and Eng. Society Press, 1999, ch. The story of wavelets.
6. K. Chan and W. Fu, "Efficient Time Series Matching by Wavelets," in *ICDE '99: Proceedings of the 15th International Conference on Data Engineering*. IEEE Computer Society, 1999.
7. N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transfom," *IEEE Trans. Comput.*, vol. 23, no. 1, pp. 90–93, 1974.
8. C. Guo, H. Li, and D. Pan, "An improved piecewise aggregate approximation based on statistical features for time series mining," in *Proceedings of the 4th international conference on Knowledge science, engineering and management*, ser. KSEM'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 234–244.
9. J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A Symbolic Representation of Time Series, with Implications for Streaming Algorithms," in *8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, San Diego, 2003.
10. A. Bondu, M. Boullé, and B. Grossin, "SAXO : An Optimized Data-driven Symbolic Representation of Time Series," in *IJCNN (International Joint Conference on Neural Networks)*. IEEE, 2013.
11. M. Boullé, *Hands on pattern recognition*. Microtome, 2010, ch. Data grid models for preparation and modeling in supervised learning.
12. H. Shatkay and S. B. Zdonik, "Approximate Queries and Representations for Large Data Sequences," in *12th International Conference on Data Engineering (ICDE)*, 1996, pp. 536—-545.
13. R. Agrawal, G. Psaila, E. L. Wimmers, and M. Zait, "Querying Shapes of Histories," in *21th International Conference on Very Large Data Bases (VLDB '95)*, 1995, pp. 502—-514.
14. M. Boullé, "Functional data clustering via piecewise constant nonparametric density estimation," *Pattern Recognition*, vol. 45, no. 12, pp. 4389–4401, 2012.
15. E. Keogh, Q. Zhu, B. Hu, H. Y., X. Xi, L. Wei, and C. A. Ratanamahatana, "The UCR Time Series Classification/Clustering Homepage : www.cs.ucr.edu/∼eamonn/time_series_data/," 2011.
16. M. Boullé, "Optimum simultaneous discretization with data grid models in supervised classification: a Bayesian model selection approach," *Advances in Data Analysis and Classification*, vol. 3, no. 1, pp. 39–61, 2009.
17. J. Demšar, "Statistical Comparisons of Classifiers over Multiple Data Sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, Dec. 2006.