Probability ridges and distortion flows: Visualizing multivariate time series using a variational Bayesian manifold learning method

Alessandra Tosi[1], Iván Olier[2], and Alfredo Vellido[1] *

[1]Dept. de Llenguatges i Sistemes Informàtics - Universitat Politècnica de Catalunya
Barcelona, 08034, Spain
[2]Manchester Institute of Biotechnology - The University of Manchester, M1 7DN, UK

Abstract. Time-dependent natural phenomena and artificial processes can often be quantitatively expressed as multivariate time series (MTS). As in any other process of knowledge extraction from data, the analyst can benefit from the exploration of the characteristics of MTS through data visualization. This visualization often becomes difficult to interpret when MTS are modelled using nonlinear techniques. Despite their flexibility, nonlinear models can be rendered useless if such interpretability is lacking. In this brief paper, we model MTS using Variational Bayesian Generative Topographic Mapping Through Time (VB-GTM-TT), a variational Bayesian variant of a constrained hidden Markov model of the manifold learning family defined for MTS visualization. We aim to increase its interpretability by taking advantage of two results of the probabilistic definition of the model: the explicit estimation of probabilities of transition between states described in the visualization space and the quantification of the nonlinear mapping distortion.

Keywords: Multivariate time series, Nonlinear dimensionality reduction, Mapping distortion, Magnification Factors, Visualization, Generative Topographic Mapping, Variational Bayesian methods.

1 Introduction

Most applied analysis of MTS involves, in one way or another, problems with specific targets such as prediction, forecasting, or anomaly detection. A less explored avenue of research is the exploratory analysis of MTS using machine learning and computational intelligence methods [1].

Data exploration may be a key stage in knowledge extraction from MTS using complex nonlinear methods, as it opens the door to their interpretability [2]. As in any other process of knowledge extraction from data, the analyst could benefit from the exploration of the characteristics of MTS consisting of a high number of individual series through their visualization [3]. The direct visualization of such high-dimensional data, though, can easily be beyond the

---

interpretation capabilities of human experts. Therefore, the exploration of MTS can be assisted by dimensionality reduction (DR) methods. In particular, the visualization of MTS using nonlinear DR (NLDR) methods [4] can provide the expert with inductive reasoning tools as a means to hypothesis generation [3, 5]. Visualization can thus facilitate interpretation, which is paramount given that NLDR methods can be rendered useless in practice if interpretability is lacking.

In this brief paper, we merge two strands of previous research on data visualization. The first one involves the visualization of MTS using Statistical Machine Learning (SML) NLDR methods [6]. The second tackles one of the main interpretability bottlenecks of NLDR techniques: the difficulty of expressing the nonlinear mapping distortion they introduce in the data visualization space in an intuitive manner. Specifically, we attempt to increase the interpretability of the Variational Bayesian Generative Topographic Mapping Through Time (VB-GTM-TT), a variational Bayesian variant of a constrained hidden Markov model (HMM) [7] of the manifold learning family, defined for MTS visualisation [8]. For this, we use two results of the probabilistic definition of the model: the explicit estimation of probabilities of transition between states described in the visualization space and the quantification of the distortion introduced by the nonlinear mapping of the MTS in the form of Magnification Factors (MF).

Note that this paper does not address the assessment of the quality of the mapping as such. In fact, the proposed visualization strategies are meant to be independent from it. Although VB-GTM-TT is used here for illustration (as a method that, even if prone to limitations such as local minima, has been shown to perform robustly in the presence of noise), the proposed approach could be extended to alternative MTS DR models for which distortion and probability of state transition (or some approximations to them) were quantifiable.

2   Methods

2.1   Variational Bayesian GTM Through Time

The Generative Topographic Mapping (GTM: [9]) is a NLDR latent variable model of the manifold learning family. It can be seen as a mixture of distributions whose centres are constrained to lay on an intrinsically low-dimensional space. Given that the generative model specifies a mapping from latent space to observed data space, such latent space can be used for data visualization when its dimensionality is 1 or 2. Unless regularization is included, the GTM is prone to overfitting. Adaptive regularization for GTM was proposed in [10].

The GTM was redefined as a constrained HMM in [6]. The resulting GTM Through Time (GTM-TT) can be considered as a GTM model in which the latent states are linked by transition probabilities, in a similar fashion to HMM. This model, even if useful for MTS clustering and visualization, did not implement any regularization process.

Recently, the GTM was reformulated within a variational full Bayesian framework in [11], which was extended to the analysis of MTS in [8]. The result was

the VB-GTM-TT: a model that integrates regularization explicitly and provides adaptive optimization of most of the model parameters involved. Assuming a sequence of N hidden states $\mathbf{Z} = \{z_1, z_2, \ldots, z_n, \ldots, z_N\}$ and the observed MTS $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \ldots, \mathbf{x}_N\}$, the complete-data likelihood for VB-GTM-TT is given by:

$$p\left(\mathbf{Z}, \mathbf{X} | \boldsymbol{\Theta}\right) = p\left(z_1\right) \prod_{n=2}^{N} p\left(z_n | z_{n-1}\right) \prod_{n=1}^{N} p\left(\mathbf{x}_n | z_n\right). \tag{1}$$

The model parameters are $\boldsymbol{\Theta} = (\pi, \mathbf{A}, \mathbf{Y}, \beta)$, where $\pi = \{\pi_j\} : \pi_j = p\left(z_1 = j\right)$ are the initial state probabilities; $\mathbf{A} = \{a_{ij}\} : a_{ij} = p\left(z_n = j | z_{n-1} = i\right)$ are the transition state probabilities; and

$$\{\mathbf{Y}, \beta\} : p\left(\mathbf{x}_n | z_n = j\right) = \left(\frac{\beta}{2\pi}\right)^{D/2} \exp\left(-\frac{\beta}{2} \|\mathbf{x}_n - \mathbf{y}_j\|^2\right)$$

are the emission probabilities, which are controlled by spherical Gaussian distributions with common inverse variance β and a matrix $\mathbf{Y}$ of K centroids $\mathbf{y}_j, 1 \leq j \leq K$. They can be considered as hidden variables and integrated out to describe the marginal likelihood as:

$$p\left(\mathbf{Z}, \mathbf{X}\right) = \int p\left(\boldsymbol{\Theta}\right) p\left(\mathbf{Z}, \mathbf{X} | \boldsymbol{\Theta}\right) d\boldsymbol{\Theta},$$
$$\text{where } \boldsymbol{\Theta} = (\boldsymbol{\pi}, \mathbf{A}, \mathbf{Y}, \beta). \tag{2}$$

VB-GTM-TT assumes its parameters to be independent, so that $p(\boldsymbol{\Theta}) = p(\boldsymbol{\pi})p(\mathbf{A})p(\mathbf{Y})p(\beta)$, where the set of prior distributions $p\left(\boldsymbol{\Theta}\right)$ are defined as:

$$p\left(\pi\right) = \texttt{Dir}\left(\{\pi_1, \ldots, \pi_K\} | \nu\right)$$
$$p\left(\mathbf{A}\right) = \prod_{j=1}^{K} \texttt{Dir}\left(\{a_{j1}, \ldots, a_{jK}\} | \lambda\right)$$
$$p\left(\mathbf{Y}\right) = \left[(2\pi)^K |\mathbf{C}|\right]^{-D/2} \prod_{d=1}^{D} \exp\left(-\frac{1}{2}\mathbf{y}_{(d)}^T \mathbf{C}^{-1} \mathbf{y}_{(d)}\right)$$
$$p\left(\beta\right) = \Gamma\left(\beta | d_\beta, s_\beta\right).$$

Here, $\texttt{Dir}\left(\cdot\right)$ represents the Dirichlet distribution and $\Gamma\left(\cdot\right)$ is the Gamma distribution. The vector ν, the matrix λ and the scalars d_β and s_β correspond to the hyperparameters of the model which are fixed *a priori*. The prior over the parameter $\mathbf{Y}$ defines the mapping from the hidden states to the data space as a Gaussian Process (GP), where $\mathbf{y}_{(d)}$ is each of the row vectors (centroids) of the matrix $\mathbf{Y}$ and $\mathbf{C}$ is a matrix where each element is defined by the covariance function as:

$$C_{i,j} = c(\mathbf{u}_i, \mathbf{u}_j) = \exp\left(-\frac{\|\mathbf{u}_i - \mathbf{u}_j\|^2}{2\alpha^2}\right), \quad i, j = 1 \ldots K. \tag{3}$$

The α parameter controls the flexibility of the mapping from the latent space to the data space. The vector $\mathbf{u}_j$, $j = 1 \ldots K$ corresponds to the state j in a latent space of usually lower dimension than that of the data space (for MTS visualization purposes). Thus, a topography over the states is defined by the GP as in the standard GTM. The VB-GTM-TT is optimized using variational approximation techniques. A more detailed description of the VB-GTM-TT and its formulation is provided in [8, 12].

2.2   Distortion measures and local metrics

NLDR techniques usually attempt to minimize the unavoidable distortion they introduce in the projection of the high-dimensional data from the observed space onto lower-dimensional spaces. For a more faithful interpretation of models, many distortion measures have been proposed and adapted to visualization techniques for different methods. While reducing dimensionality, different levels of local mapping distortion are generated, leading to a loss of information that we aim to recover, to some extent, to improve the interpretability of the model.

An interesting approach, proposed in [13] for GTM (and extended to Self Organizing Maps), is the calculation of the MFs. The concept of magnification has been applied to manifold learning methods in order to quantify the distortion due to the embedding of a manifold in a high-dimensional space. Importantly, the distortion caused by the mapping can be explicitly computed in a continuous way over the low-dimensional latent space of visual representation.

From the theory of differential geometry, we can describe the local geometry of a q-dimensional differential manifold through the mapping between two sets of coordinates systems $\xi^j \mapsto \zeta(\xi)$ defined in the high dimensional space. The Jacobian of this transformation can be written as:

$$J = \left(\frac{\partial \zeta^i}{\partial \xi^j}\right)_{i,j}. \tag{4}$$

Every point on the manifold has local geometrical properties which are given by its metric tensor $\mathbf{g}_{i,j}$, which is defined by:

$$\mathbf{g}_{i,j} = \delta_{i,j}\frac{\partial \zeta}{\partial \xi}\frac{\partial \zeta}{\partial \xi}, \tag{5}$$

where δ is the Kronecker delta. From (4) and (5), it follows that $\mid J \mid = \mid \mathbf{g} \mid^{\frac{1}{2}}$.

2.3   Magnification Factors for VB-GTM-TT

As stated in [13], the MF can be explicitly computed for the batch-SOM and GTM. In this paper, we provide the calculation of the MF for the VB-GTM-TT model. For this, we first consider the jointly Gaussian random variables:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_* \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} C & C_{(*,\cdot)} \\ C_{(\cdot,*)} & C_{(*,*)} \end{bmatrix} \right), \tag{6}$$

where y_* is a test point and $C_{(\cdot,\cdot)}$ is the covariance matrix defined according to (3). Due to the properties of Gaussian distributions, we can explicitly write the posterior probability as follows:

$$\mathbf{y}_* | \, \mathbf{u}_*, \mathbf{U}, \mathbf{Y}, \theta \sim \mathcal{N} \left(\mathbf{C}_{(*,\cdot)} \mathbf{C}^{-1} \mathbf{Y}, \mathbf{C}_{(*,*)} - \mathbf{C}_{(*,\cdot)} \mathbf{C}^{-1} \mathbf{C}_{(\cdot,*)} \right). \tag{7}$$

The Jacobian J of this mapping can be obtained computing the derivatives of $\langle (\mathbf{y}_* | \, \mathbf{u}_*, \mathbf{U}, \mathbf{Y}, \theta) \rangle$ with respect to u, using:

$$\frac{\partial c_{(*,j)}}{\partial \mathbf{u}_*^{\mathbf{l}}} = \frac{1}{\alpha^2} (u_*^l - u_j^l) \exp \left(-\frac{\|\mathbf{u}_* - \mathbf{u}_j\|^2}{2\alpha^2} \right), \quad l = 1 \ldots q, j = 1 \ldots K, \tag{8}$$

being q the dimension of the latent space. As a result, the MF is calculated as:

$$\mu_* = det^{-\frac{1}{2}} \left(J J^T \right) \tag{9}$$

The MF does not only provide us with a quantification of the local mapping distortion that separates areas of the visual map which have undergone much compression or stretching from those which have not; it also tells us about data sparsity: the model distorts the most in areas which are mostly empty of data and the least in densely populated areas. For this reason, the MF has been used as an indicator of the existence of data clusters and the boundaries between those clusters [14]. For MTS, we would expect the time series to flow over time through areas of low MF mostly when the MTS evolve slowly, whereas fast transitions between MTS regimes might require crossing areas of higher distortion.

2.4 Cumulative state transition probabilities

Another metric that might help improving the interpretability of the mapping is the likelihood for a state to be transited by any of the potential trajectories through states. Again, this can explicitly be quantified, for each state j defined by VB-GTM-TT, as the estimated cumulative state transition probability (CSTP) defined as the sum of the probabilities of transition from all states to it:

$$\mathcal{CSTP}_j = \sum_{i=1}^{K} a_{ij}. \tag{10}$$

We would expect the MTS trajectory to happen through areas of high $CSTP$, because these should be areas of highly likely transition. As such, the $CSTP$ plays the opposite role to MF, because the areas of large manifold stretching (high MF) should mostly be areas that the MTS is unlikely to cross (low $CSTP$).

3   Experiments and discussion

3.1   Materials and experimental setup

We illustrate the proposed MTS visualization using two different datasets. The first is an *artificial* 3-variate time series, with $1,000$ time points. The second set is the *Shuttle-data* from Space Shuttle mission STS-57[1]: a time series consisting of 1,000 points described by 6 features. This data set has previously been used for cluster detection in [15].

3.2   MTS Visualization

The considered MTS are particularly suitable for the illustration of the proposed visualization techniques due to the nature of their regimes and transitions periods. The *artificial* dataset (displayed in Fig.(1), top-row, left) is characterized by two intervals with regular regimes, divided by a sudden transition at point 700. The VB-GTM-TT model was trained over a 8×8, $2 - D$ grid of hidden states and each of the MTS points was mapped by VB-GTM-TT to a particular state in the grid. The result of this mapping assignment is shown in Fig.(1) (top-row, right). Before point 700, the periodicity of the data is well-captured by the roughly circular structure of populated states. The sudden transition to a higher-amplitude periodic interval is also neatly visualized.

On the other hand, *Shuttle_Data* presents four periods of little variability *A-C-D-E* and one period of high (quasi-periodic) variability *B*, which are separated by sudden transitions, as evidenced by their display in Fig.(2) (top row, left). The VB-GTM-TT model was trained over a 13×13 grid of hidden states and each of the MTS points was mapped by VB-GTM-TT to a particular state in the grid, as shown in Fig.(2) (top row, right). There is a clear interpretation for this state membership mapping, as the *Shuttle-data* trajectory is confined to a limited number of its states (a common characteristic of VB-GTM-TT mappings, in which over-complexity is penalized). Only a few of them are relatively big: these are mostly stationary states with little MTS change in intervals C, D and E. The quasi-periodic interval B evolves slowly through a cloud of states on the top-left and center of the map.

The MFs were computed for *artificial* and *Shuttle-data* and represented in Figs.1 and 2 (bottom, right) through color maps over the grid of hidden states. For both datasets, it might seem at first sight that the MTS cross through areas of high MF (high distortion), a behaviour that would refute the hypothesis that the densely data populated areas correspond to low mapping distortion. In fact, this is not the case: the MTS mostly flows over time through *channels* of low distortion surrounded by borders of high distortion. These borders seem to act as barriers that compel the MTS to follow a given trajectory. In fact, these barriers are only breached (with the MTS moving briskly towards higher MF) in sudden transitions between regimes. These can clearly be seen for *Shuttle-data* if we plot

---

[1] Which can be requested from www.cs.ucr.edu/∼eamonn.
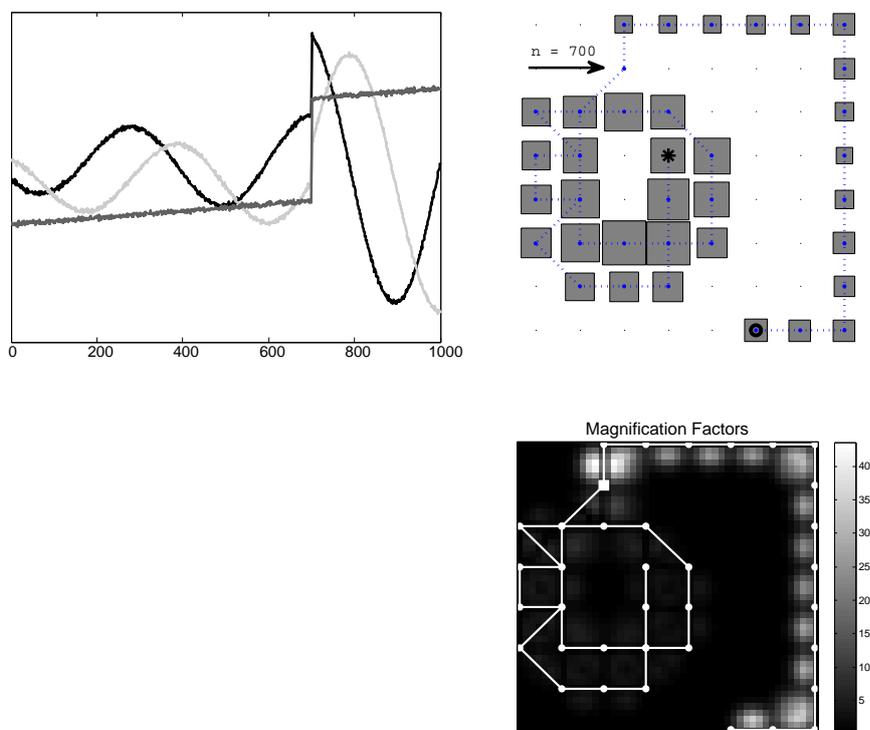
Magnification Factors

Fig. 1. Top row, left: Artificial dataset: a 3-variate time series, characterized by a sudden transition at $n = 700$. Top row, right: State-membership map of VB-GTM-TT, with a 8×8 grid of hidden states represented as squares, whose relative size is proportional to the time data points assigned to them; the starting point of the MTS is represented as a star and the ending point is represented as a circle. The sudden transition point is signaled by an arrow. Bottom row, right: MF gray-shade color map, represented in the VB-GTM-TT latent space visualization grid. The trajectory of the MTS over the map is displayed as a white solid line.

the value of MF over time, as in Fig.(2) (bottom row, left): MF narrow spikes of varying magnitude (particularly strong in the transition from B to C) appear in the transitions between time intervals. These spikes take values well over the mean MF of the map. This result suggests that the evolution of the MF over time could directly be used to detect sudden regime transitions in MTS.
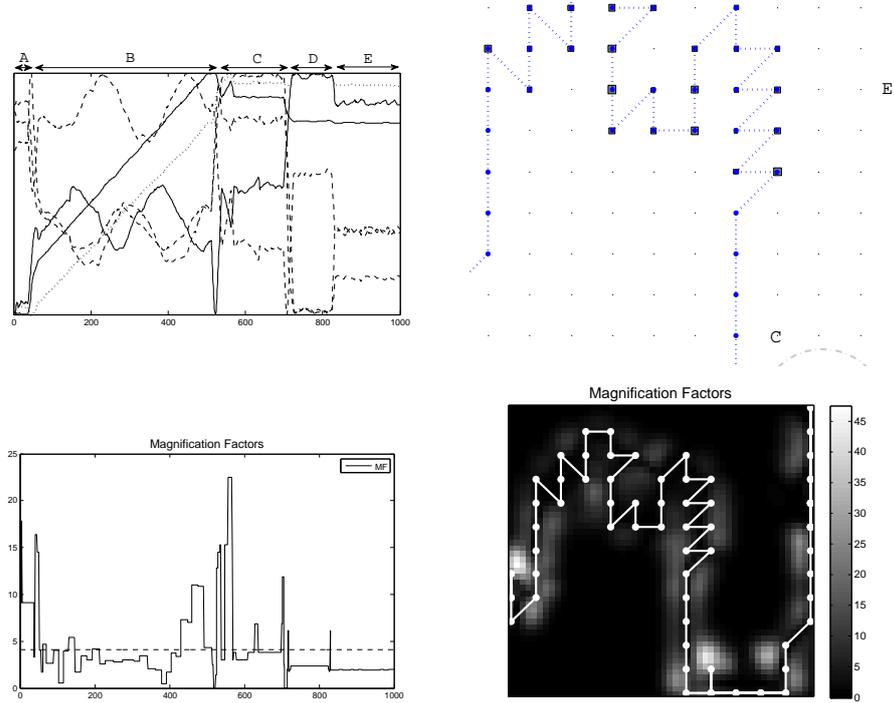
Fig. 2. Top left: Plot of the *Shuttle-data*; the five intervals or regimes separated by sudden transitions are identified as A, B, C, D and E. Top right: State-membership map generated by VB-GTM-TT, with a 13×13 grid of hidden states represented as squares; the relative size of these squares is again proportional to the ratio of time data points assigned to them; the starting point of the MTS is represented as a star, while the ending point is represented as a circle. Bottom Left: The Magnification Factors as a function of time, including the mean MF over all states (represented as a dashed line); narrow peaks of distortion are detected precisely in the areas of sudden transitions. Bottom Right: MF gray-shade color map, represented in the VB-GTM-TT latent space visualization grid; white areas correspond to high distortion.

The $CSTP$ maps in Fig.3 are very consistent with their MF counterparts, and complement them. Alternatively displayed as $3 - D$ maps over the grid of hidden states, they provide an intuitive illustration of the previously described behaviour. Following a geographical representation visual metaphor, the MTS can be seen to flow across cumulative state transition probability *ridges*, where rapid transitions between regimes see the MTS moving through relatively lower-valued *depressions* in those *ridges*. An opposite graphical metaphor could be used for the MF distortion, with the MTS flowing through its *valleys*, that is, across areas of the map characterized by low MF values.
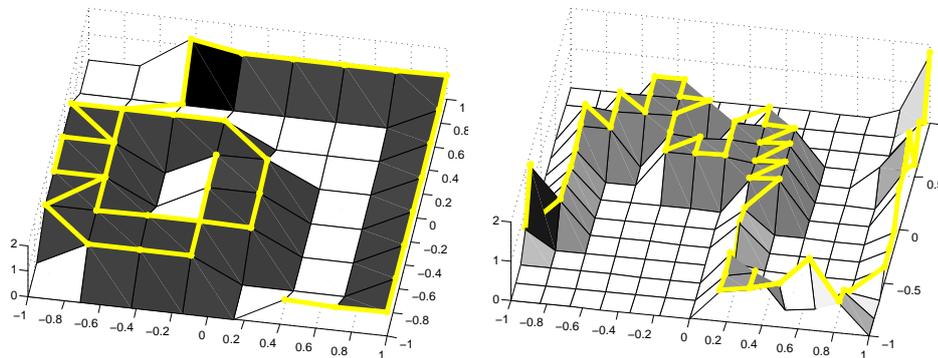
Fig. 3. A $3-D$ representation for the CSTP plots. The values in the vertical axis correspond to the CSTP values over the latent space. Left: *artificial* data; right: *Shuttle-data*.

4   Conclusions

Data visualization can be of great assistance in knowledge extraction processes. High dimensionality is always a barrier for visualization. In the case of MTS, this is compounded by their i.i.d. nature, because the search for patterns over time is often relevant in their study. Dimensionality reduction can make visualization operative for high-dimensional MTS. The use of NLDR methods to this purpose poses a challenge of model interpretability due to the existence of locally-varying distortion.

In this study, we have proposed two methods to improve interpretability for VB-GTM-TT, a manifold learning NLDR method. The model mapping distortion has been explicitly quantified in the latent space continuum and the probabilistic nature of the method has allowed us to define a cumulative probability of state transition. The reported preliminary experiments have shown that both metrics can provide interesting insights that enhance the low-dimensional visualization of the MTS provided by the model.

This exploration approach is quite flexible and could be extended to other dimensionality reduction models for MTS analysis, provided their local distortion can be quantified. Examples of this may include Gaussian process latent variable models (GP-LVM, [16]) and dynamical models (GPDM, [17]) or temporal Laplacian eigenmaps ([18]). It could also be extended to alternative display methods, such as the recently proposed cartograms[2] [19],[20] and topographic maps [21]. Following the line of this paper, we also aim to investigate the VB-GTM-TT model and its properties in more depth, focusing on its generalization capabilities and its use for prediction.

---

[2] www.worldmapper.org/

References

1. Fu, T.C.: A Review on Time Series Data Mining. Engineering Applications of Artificial Intelligence, 24(1), 164–181 (2011)
2. Vellido, A., Martín-Guerrero, J.D., Lisboa, P.J.G.: Making Machine Learning Models Interpretable. In: ESANN 2012, pp. 163–172. d-Side Pub. (2012)
3. Vellido, A., Martín, J.D., Rossi, F., Lisboa, P.J.G.: Seeing is Believing: The Importance of Visualization in Real-World Machine Learning Applications. In: ESANN 2011, pp.219–226, d-Side Pub. (2011)
4. Lee, J.A., Verleysen, M.: Nonlinear Dimensionality Reduction. Springer (2007)
5. Van Belle, V., Lisboa. P.: Research Directions in Interpretable Machine Learning Models. In: ESANN 2013, pp. 533–541. i6doc.com Pub. (2013)
6. Bishop, C.M., Hinton, G.E., Strachan, I.G.D.: GTM Through Time. Fifth International Conference on Artificial Neural Networks, pp. 111–116 (1997)
7. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257–286 (1989)
8. Olier, I., Vellido, A Variational Formulation for GTM Through Time. In: International Joint Conference on Neural Networks (IJCNN 2008), pp. 517-522 (2008)
9. Bishop, C.M., Svensén, M., Williams, C.K.I.: GTM: The Generative Topographic Mapping. Neural Computation, 10, 215–234 (1998)
10. Bishop, C.M., Svensén, M., Williams, C.K.I.: Developments of the Generative Topographic Mapping. Neurocomputing, 21(1), 203–224 (1998)
11. Olier, I., Vellido, A.: Variational Bayesian Generative Topographic Mapping. Journal of Mathematical Modelling and Algorithms, 7(4), 371–387 (2008).
12. Olier, I., Amengual, J., Vellido, A.: A Variational Bayesian Approach for the Robust Estimation of Cortical Silent Periods from EMG Time Series of Brain Stroke Patients. Neurocomputing, 74(9): 1301–1314 (2011)
13. Bishop, C.M., Svensén, M., Williams, C.K.I.: Magnification Factors for the SOM and GTM Algorithms. In: Proceedings of the 1997 Workshop on Self-Organizing Maps (WSOM), pp. 333-338 (1997)
14. Tosi, A., Vellido, A.: Robust Cartogram Visualization of Outliers in Manifold Learning, In: ESANN 2013, pp.555-560. i6doc.com Pub. (2013)
15. Lin, J., Vlachos, M., Keogh, E., Gunopulos, D.: Iterative Incremental Clustering of Time Series. In Advances in Database Technology-EDBT 2004, pp. 106–122. Springer, Berlin, Heidelberg (2004)
16. Lawrence, N.: Probabilistic Non-Linear Principal Component Analysis with Gaussian Process Latent Variable Models. The Journal of Machine Learning Research, 6, 1783–1816 (2005)
17. Wang, J.M., Fleet, D.J., Hertzmann, A.: Gaussian Process Dynamical Models for Human Motion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2): 283–298 (2008)
18. Lewandowski, M., Martínez-del-Rincón, J., Makris, D., Nebel, J.C.: Temporal Extension of Laplacian Eigenmaps for Unsupervised Dimensionality Reduction of Time Series. In: 20th International Conference on Pattern Recognition (ICPR), pp.161-164. IEEE (2013)
19. Tosi, A., Vellido, A.: Cartogram Representation of the Batch-SOM Magnification Factor. In: ESANN 2012, pp. 203–208 (2012)
20. Vellido, A., García, D., Nebot, À. Cartogram Visualization for Nonlinear Manifold Learning Models. Data Mining and Knowledge Discovery, 27(1): 22–54 (2013)
21. Gianniotis, N. Interpretable magnification factors for topographic maps of high dimensional and structured data. IEEE CIDM, pp 238–245, (2013)