DOI 10.34185/1562-9945-2-151-2024-08
UDC 519.85:004.42
E. Koshel

NEURAL NETWORK-ASSISTED CONTINUOUS EMBEDDING OF UNIVARIATE DATA STREAMS FOR TIME SERIES ANALYSIS

Abstract. Univariate time series analysis is a universal problem that arises in various science and engineering fields and the approaches and methods developed around this problem are diverse and numerous. These methods, however, often require the univariate data stream to be transformed into a sequence of higher dimensional vectors (embeddings). In this article, we explore the existing embedding methods, examine their capabilities to perform in real time, and propose a new approach that couples the classical methods with the neural network based ones to yield results that are better in both accuracy and computational performance. Specifically, the Broomhead King inspired embedding algorithm implemented in a form of an autoencoder neural network is employed to produce unique and smooth representation of the input data fragments in the latent space.
Keywords: time series analysis, time series embedding, dimensionality reduction

Introduction. Time series analysis is the process of extracting information from the series of data points in an attempt to obtain useful information about the system that produced it or to predict the next values in the series. The simplest form of the time series is a univariate series or just a sequence of one-dimensional measurements. If the system that the data was obtained from is sufficiently complex, the time series may exhibit periodic, quasi-periodic, chaotic behavior, drift, etc. It is easy for a human to see whether these behaviors are present or not simply by looking at the plotted sequence of data points, but this can hardly qualify as proper analysis. Luckily, there are methods developed in different fields that can be used for identification and quantification of different properties of the time series data. Most of these methods, however, have a hard requirements for the dimensionality of the problem and can yield nonsensical results if the dimension of the problem they are applied to is too low or too high. This issue can be solved by embedding the available time series data into a higher-dimensional space where the analysis methods can be applied.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Time series embedding is a mapping between the segments of the time series data and some vector space $\mathbb{R}^n$. This mapping is expected to be an injective function that produces a smooth trajectory of embedding vectors for adjacent time series data segments. In the following sections, we examine some of the methods for producing time series embeddings, see how well they are able to uphold the requirements listed earlier, compare their ability to preserve the qualitative characteristics of their source system, their resistance to noise, and the overall computational cost of each method.

Literature Review. The broad definition of time series embedding provides a lot of flexibility in selecting the mapping functions. Below is the review of some of the classical transformation methods.

One of the most well-known ways of embedding the time series data is the short-time Fourier transform (STFT) and related wavelet transform-based methods (WTs) which are widely used for feature extraction from a variety of signals [7, 10, 17] by transitioning from temporal to frequency domain representation of signal frames. The STFT algorithm is defined as follows:

1. Break down the signal into short overlapping frames.
2. Apply a window function to each frame to suppress the high frequency artifacts at the edges.
3. Pad the windowed frame with zeros to achieve the required frequency resolution and/or achieve the length required for fast Fourier transform (FFT) to have maximum efficiency.
4. Compute the discrete Fourier transform of the padded windowed frame.
5. Compute the squares of the first half of the resulting vector to obtain the relative magnitude of the frequencies that make up the original signal frame.

The STFT's result is a sequence of n-dimensional vectors where n is the number of "frequency bins" that depends on the chosen length of the frame. Selecting longer frames will result in more available frequency bins and thus higher resolution in the frequency domain representation of the signal. But this will not only introduce latency but also lower the temporal resolution of the STFT because the frequency information will be spread out over the longer stretches of the signal.

Another well-known class of embedding methods is time-delay based and is intended for phase-space reconstruction of dynamical systems but also used for feature extraction [3, 5] and data mining applications [14]. These methods operate under the assumption that there exists a twice-differentiable *observation function* $\alpha: \mathbb{R}^m \to \mathbb{R}$ that maps all points on the m-dimensional attractor of the smooth map

$f: \mathbb{R}^m \to \mathbb{R}^m$ to real numbers so that an embedding function can be constructed in the following form:

$$\varphi(X) = \left(\alpha(X), \alpha\big(f(X)\big), \dots, \alpha\big(f^{k-1}(X)\big)\right),$$
$$\alpha\left(f^{i}(X)\right) = x_{t-\tau_i}, \quad (1)$$

where X is a point on the attractor in $\mathbb{R}^m$ space, x_t is the time-series observation at time t, and τ_i are lag values that specify the distance in time between the current point and its neighbors.

When performing a time-delay embedding, the researchers must solve two problems. First one is selecting a proper embedding dimension k which, according to the Takens's theorem [16], would ensure that the dynamics of the embedded trajectory will be completely equivalent to the dynamics of the original system. The second problem is selecting the proper lag values τ_i which is especially difficult considering that the Takens's theorem doesn't provide any guidance regarding this and assumes that the time series is not noisy and only considers the minimal lag time of 1.

To solve these problems, either separated or unified approaches can be taken. Separated approaches start by selecting the time delay value τ and use it to estimate the optimal embedding dimension employing one of the False Nearest Neighbors algorithm variations [2, 8, 9, 13]. The optimal embedding dimension is usually selected based on values of some statistic. Unified methods recognize the interdependence between the delay values and the embedding dimensions and employ methods that allow them to select the τ_i values dynamically as they increase the dimension of the embedding during the "embedding cycles", essentially optimizing the delay times and the dimensions in parallel. One prominent example of the unified approaches is the PECUZAL algorithm [15] which uses two statistics to objectively evaluate the quality of the embedding and decide whether or not to continue increasing the number of dimensions and select appropriate delay values for each dimension.

There is also an alternative to the delay embedding which applies more complex transformations to the time series to obtain trajectories in a higher-dimensional space. One such alternative is the Broomhead-King method for extracting qualitative dynamics from experimental data [1]. It applies the Takens's theorem directly to the time series with the smallest possible delay and an

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

embedding dimension d. Then singular value decomposition (SVD) is performed on the resulting embedding:

$$X = \frac{1}{\sqrt{N}} \begin{pmatrix} x_1 & x_2 & \cdots & x_d \\ x_2 & x_3 & \cdots & x_{d+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N-d+1} & x_{N-d+2} & \cdots & x_N \end{pmatrix} = U \cdot S \cdot V^{tr}, \quad (2)$$

where N is the length of the time series, x_i is the value from time series adjusted to have the mean of 0, U is the new embedding space, S is a diagonal matrix that holds the "importance values" of each column of U. Since by the definition of SVD the matrix U is a unitary one, its columns form an orthonormal basis which means that the components of the high-dimensional trajectory it describes are guaranteed to not contain correlations and redundant information. The properties of the matrix S also allow the researchers to judge the level of importance of each of the columns of U and reduce its dimensionality by dropping columns with importance lower than some specified threshold.

Research methodology and results. The methods discussed above provide a reliable set of tools for time series embedding, dimensionality estimation and reduction. To make the further discussion easier, let us introduce some definitions and generalizations.

First, observe how the definitions of all the methods from the previous section can be rewritten in terms of matrix multiplication. The STFT is based on the discrete Fourier transform which can be expressed as a square complex-valued transformation matrix that the signal is multiplied by:

$$W = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega & \cdots & \omega^{N-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \omega^{N-1} & \cdots & \omega^{(N-1)(N-1)} \end{pmatrix}, \quad (3)$$

where $\omega = e^{i\pi/N}$.

Similarly, the delay embedding methods can be reduced to a problem of finding a rectangular matrix that yields an embedding vector given a signal segment:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ a_{\tau_{max}1} & a_{\tau_{max}2} & \cdots & a_{\tau_{max}d} \end{pmatrix}, \quad (4)$$

where columns of A are unitary vectors and $\forall j \exists i: a_{ij} = 1$. The number of columns d is the embedding dimension and the number of rows is given by the largest selected delay value. As for the Broomhead-King approach, the transformation matrix can be trivially derived from its definition in (2):

$$A = (V^{tr})^{-1} \cdot S^{-1} = V \cdot S^{-1}. \qquad (5)$$

Notably, the matrices (4) and (5) both describe which values of the time series segment contribute to the embedding vector by assigning coefficients to them. The difference is that the matrix (4) is very rigid and can only "select" one value from the time series per its column. The matrix (5), however, is a lot more flexible and can incorporate multiple time series values into a single embedding vector component, which is very similar to how the discrete Fourier transformation matrix operates.

Second, let's examine the equation (1) more closely. It introduces the assumption that there exists some unknown observation function α that associates real numbers to high-dimensional points on the attractor. Its definition does not constrain it to any particular structure and only requires it to be twice-differentiable. In practice, however, it is almost always assumed to be a linear transformation that selects a single coordinate from each point on the attractor. This assumption also informs the $\alpha\left(f^{i}(X)\right) = x_{t-\tau_i}$ equation which in the real world applications almost never the case due to the presence of noise and the fact that neither the map f nor the function α are known. Considering (3) and (5), the second part of the equation (1) can be rewritten as follows:

$$\alpha\left(f^{i}(X)\right) = \beta(x_t, x_{t-1}, \ldots, x_{t-l}) \qquad (6)$$

where $\beta: \mathbb{R}^l \to \mathbb{R}$ is the *feature extraction function* that maps time series segments to the codomain of the unknown observation function α.

All of the above implies that if an optimal embedding can be achieved using matrices of type (4) and the properties of the SVD transformation show that an equally valid alternative embedding can be achieved with matrices of type (5), then an optimization problem can be formulated that optimizes the equation (6) with respect to a loss function designed to enforce the objective statistics used to evaluate the quality of the embedding in unified embedding methods.

Since both α and β functions are unknown, a data-driven approach for their modeling can be employed. More specifically, an autoencoder neural network architecture can be used to find a model that accurately maps the time series data to the codomains of α and β and back. The main idea of an autoencoder is to couple two functions – encoder E and decoder D – such that $x \approx D\big(E(x)\big)$ with the goal of

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

either increasing or decreasing the dimensionality of x in some latent space. Applying this idea with the context of equations (1) and (4), we get the following:

$$E([x_t, x_{t-1}, \dots, x_{t-l}]) = \beta(x_t, x_{t-1}, \dots, x_{t-l}) = Y,$$
$$D(Y) = \alpha\left(f^{k-1}(X)\right) = x_t. \quad (7)$$

The advantage of using this neural network-based approach is that the structure of functions E and D is not limited to linear transformations like all the methods discussed in the previous section and can contain any number of nonlinearities that can be tailored to a specific time series being analyzed.

Let's now examine the performance of each of the discussed embedding methods using some synthetic data. It's best to chose a well-known system to make the benchmarking of it easier to reproduce in different computing environments, which is why the chaotic regime of the Lorenz system will be examined in the examples below. However, to acknowledge the previous remarks about the observation function α , it will be altered and instead of picking a specific coordinate of the points on the attractor, it will compute the norm of each point. To evaluate the preservation of the qualitative characteristics of the original attractor in the embedding space, their Lyapunov spectra will be compared.

All computations and simulations will be done using the Julia programming language and its various packages.

The system and the observation function that is used to generate the synthetic time series data are as follows:

$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = x(\rho - z) - y,$$
$$\dot{z} = xy - \beta z$$
$$\sigma = 10, \rho = 28, \beta = 8/3, \quad (8)$$
$$\alpha(X) = \|X\|_2$$

The shape of the time series generated by (7) and modified to have a mean of 0.

To obtain the STFT embedding of the test time series, we set the matrix (3) size to 10 and apply it to windows of the time series with same length which have 9 overlapping samples. This transformation is different from all the others because it yields complex numbers in its output which makes it difficult to visualize. For the purposes of demonstration, we will avoid plotting the norms of the embeddings and instead simply omit their imaginary parts. This will help us to keep the output smooth and comparable to other embedding methods.

Next, the PECUZAL algorithm is applied to the time series and is used to select the embedding dimension and the delay values. In this example, the Julia implementation of the algorithm from the DelayEmbeddings.jl [4] package was used. No extra parameters were supplied to the algorithm aside from the time series itself.

The Broomhead-King algorithm was supplied two parameters for constructing the embedding – the time series data itself and the target embedding dimension of 20. The matrix (5) was constructed using the algorithm's output and all columns of the resulting matrix were discarded except the first 3. In this example, the Julia implementation of the algorithm from the ChaosTools.jl [4] package was used.

Finally, for constructing the autoencoder model, two simple linear models were selected and trained:

$$\begin{aligned} E(x) &= A_e x + b_e, \\ D(y) &= A_d y + b_d, \end{aligned} \tag{9}$$

where A_e is a N by m matrix that maps N points of the trajectory into the m dimensional latent space, A_d is a m by 1 matrix that maps the points in the latent space to the last point of the converted segment of the time series as shown in (7). The b_e and b_d are the bias vectors.

To compare the quality of each embedding, let's calculate their respective Lyapunov spectra to see how well they preserve the dynamics of the original system. Since the embedded trajectories do not have analytical representations, numerical approach to calculating the spectrum will be used. The algorithm that is used is a QR decomposition-based one described in [6] but instead of using the analytical representation of the original map, the local Jacobian matrix is computed numerically for each point based on its nearest neighbors using the algorithm from [11]. The Lyapunov spectra values are given in Table 1.

Another embedding quality measure that we can use is the L statistic that the PECUZAL algorithm uses to evaluate the quality of its own embeddings, the higher the value of the statistic, the less correlation there is between the components of the embedding vector The values of the L statistic for the original system trajectory and each of the embeddings are presented in the Table 2.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Table 1

Lyapunov spectrum values for the actual attractor compared to the values calculated for the trajectories in the embedding spaces

	True values	STFT	PECUZAL	Broomhead-King	Autoencoder
λ_1	0.91	3.9	2.18	0.04	1.9
λ_2	0	1.08	0.5	0	0.22
λ_3	-14.57	-28.02	-13	-0.19	-12.07

As evident from both the Table 1 and Table 2, most embedding methods were able to accurately reconstruct the trajectory in a way that preserved the original dynamics of the system. The only outlier is the Broomhead-King approach that turned out to not have the amount of information in its first few columns required to create reliable embeddings.

Table 2

L statistic values for the actual attractor compared to the values calculated for the trajectories in the embedding spaces

	True values	STFT	PECUZAL	Broomhead-King	Autoencoder
	-0.09	1.41	0.43	-0.48	0.4

Another thing to point out is that the autoencoder can sometimes produce intersecting trajectories if the loss function is not constructed properly or its non-linearities do not allow for the valid behavior to be achieved [12].

Conclusions. In this article, we demonstrated and evaluated the performance of different embedding methods and provided a basic framework for constructing new embedding algorithms based on the autoencoder neural network approach.

REFERENCES

1. Broomhead D.S. and King G. P. (1986). Extracting qualitative dynamics from experimental data. Physica D: Nonlinear Phenomena, Volume 20 (2-3), Pages 217-236. https://doi.org/10.1016/0167-2789(86)90031-X.

2. Cao, L. (1997). Practical method for determining the minimum embedding dimension of a scalar time series. Physica D: Nonlinear Phenomena, Volume 110 (1-2), 43-50, https://doi.org/10.1016/S0167-2789(97)00118-8.

3. Chakraborty, B. (2014).A Proposal for Classification of Multisensor Time Series Data based on Time Delay Embedding. International Journal on Smart Sensing and Intelligent Systems,7(5) 1-5. https://doi.org/10.21307/ijssis-2019-120
4. Datseris, G. (2018). DynamicalSystems.jl: A Julia software library for chaos and nonlinear dynamics. Journal of Open Source Software, 3(23), Pages 598, https://doi.org/10.21105/joss.00598
5. Frank, J., Mannor, S., & Precup, D. (2010). Activity and Gait Recognition with Time-Delay Embeddings. Proceedings of the AAAI Conference on Artificial Intelligence, 24(1), Pages 1581-1586. https://doi.org/10.1609/aaai.v24i1.7724
6. Geist K., Parlitz U., and Lauterborn W., (1990). Comparison of Different Methods for Computing Lyapunov Exponents, Progress of Theoretical Physics, Volume 83 (5), Pages 875–893, https://doi.org/10.1143/PTP.83.875
7. Gupta, V. and Mittal, M. (2019). QRS Complex Detection Using STFT, Chaos Analysis, and PCA in Standard and Real-Time ECG Databases. J. Inst. Eng. India Ser. B 100, 489–497. https://doi.org/10.1007/s40031-019-00398-9
8. Hegger R. and Kantz H. (1999). Improved false nearest neighbor method to detect determinism in time series data. Phys. Rev. E, Volume 60, Pages 4970-4973, https://doi.org/10.1103/PhysRevE.60.4970
9. Kennel M. B., Brown R., and Abarbanel H. D. I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys. Rev. A, Volume 45, Pages 3403–3411, https://doi.org/10.1103/PhysRevA.45.3403.
10. Kıymık M. K., Güler İ., Dizibüyük A., and Akın M. (2005). Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application, Computers in Biology and Medicine, Volume 35, Issue 7, 603-616, ISSN 0010-4825,
https://doi.org/10.1016/j.compbiomed.2004.05.001.
11. Koshel, E. (2020). Local jacobian estimation for delay embedded time series data. In XVIII International scientific and practical conference, Dnipro (pp. 152-153).
12. Koshel Y. V. and Belozyorov V. Y. (2023). Univariate Time Series Analysis with Hyper Neural ODE. Journal of Optimization, Differential Equations and Their Applications (JODEA), 31(2), 50-66.
13. Krakovská A., Mezeiová K., and Budáčová H. (2015). Use of False Nearest Neighbours for Selecting Variables and Embedding Parameters for State Space Reconstruction. Journal of Complex Systems, Volume 2015,
https://doi.org/10.1155/2015/932750.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

14. Nalmpantis, C., Vrakas, D. (2019). Signal2Vec: Time Series Embedding Representation, Engineering Applications of Neural Networks. EANN 2019. Communications in Computer and Information Science, vol 1000. https://doi.org/10.1007/978-3-030-20257-6_7
15. Pecora L. M., Moniz L., Nichols J., and Carroll T. L. (2007). A unified approach to attractor reconstruction. Chaos, Volume 17 (1). https://doi.org/10.1063/1.2430294
16. Takens, F. (1981). Detecting strange attractors in turbulence. In: Rand, D., Young, LS. (eds) Dynamical Systems and Turbulence, Warwick 1980. Lecture Notes in Mathematics, vol 898. Springer, Berlin, Heidelberg. https://doi.org/10.1007/BFb0091924
17. Tüske, Z., Golik, P., Schlüter, R., & Drepper , F. R. (2011). Non-stationary feature extraction for automatic speech recognition, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 5204-5207, doi: 10.1109/ICASSP.2011.5947530.

Received 12.03.2024.
Accepted 15.03.2024.

Нейронно-мережевий підхід до неперервного вкладення одновимірних потоків даних для аналізу часових рядів в реальному часі

Задача вкладення одновимірних часових рядів у багатовимірні простори дуже розповсюджена і зустрічається у багатьох галузях досліджень. Методи, які розв'язують цю задачу, зазвичай покладаються на ту чи іншу форму вкладання з затримкою, яке реконструює фазовий простір невідомої системи шляхом асоціювання значень часового ряду з історичними даними. Ми пропонуємо більш гнучкий метод, який використовує правила для комбінування значень часового ряду, щоб створити вкладення, які є більш репрезентативними щодо взаємозв'язку даних часового ряду одне з одним.

Кошель Євген - аспірант кафедри комп'ютерних технологій Дніпровського національного університету імені Олеся Гончара.

Koshel Eugene - postgraduate student, department of computer technologies, Oles Honchar Dnipro National University.