Time series classification approaches through NeuralODE

Kirill Semkin[1*] and Vadim Strijov[2*]

[1*]MIPT Moscow Russia.
[2*]MIPT Moscow Russia.

*Corresponding author(s). E-mail(s): semkin.ki32@gmail.com;
strijov@forecsys.ru;

The task of time series classification is to assign class label to given time series. The number of classes is fixed. It is assumed that each class corresponds to some dynamical system. Each system generates phase trajectories within some probabilistic model. These trajectories are then mapped into empirical observations. The paper investigates how to restore dynamical systems out of experimental data using NeuralODE. We consider fixed and bayesian approaches to defining dynamical systems' parameters. Two classification principles are introduced: bayesian testing and transition to dynamical system's parameters space. Taken's theorem is used to approximate true phase trajectories. At last, time series from inertial measurement unit are classified using given techniques and compared with other machine learning models.

1

1 Introduction

The paper introduces and analyzes methods for time series classification. By *time series* we understand samples of some *stochastic process*. Consider $x(t)$ to be stochastic process defined on $t \in [0, T]$. Take increasing time stamps $0 \le t_1 < \ldots < t_n \le T$ and define time series as $x_i := x(t_i)$, $i \in 1 \ldots n$. So time series are stochastic variables by definition. *Classification* of time series on M classes is a process of assigning a *label* $y \in 1 \ldots M$ to a given time series x_i.

Time series classification finds a lot of applications. It is used in medicine to detect arrhythmias or other heart conditions by EEG [1]. Or to identify seizures and sleep stages from brainwave data [2]. In speech and audio processing it is used to recognize human emotions from speech signals [3] or bird species from audio recordings [4]. Other fields of application are human activity recognition from wearable sensors [5] and weather forecasting [6]. The latter two will be explored in the paper's experiment later.

There are many methods for time series classification in the literature. *Distance-based* approaches rely on measuring similarity between time series using distance metrics. The metric can be defined using Dynamic Time Warping [7] and the k-Nearest-Neighbors (k-NN) classifier can be used then [8]. *Feature-based* approaches extract finite set of statistical [9, 10] or spectral [11] features from time series for further classification. *Deep learning* approaches uses neural networks to automatically learn time series features from raw train data. Convolutional neural networks [1, 2], recurrent neural networks (RNN) [12] or their combination [3, 5, 6] are designed to process sequential data like time series. Classification with the RNN will be compared to our method in the experiment section.

The underlying mathematical model for time series in the paper is a *dynamical system*. By this we understand a time-invariant *ordinary differential equation (ODE)*

2

of the form

$$
\begin{cases}
\mathbf{z} = f(\mathbf{z}), \\
\mathbf{z}(0) = \mathbf{z}_0,
\end{cases}
\tag{1}
$$

where $\mathbf{z} \in \mathbb{R}^m$, $f(\mathbf{z}) : \mathbb{R}^m \to \mathbb{R}^m$ is a vector field, $\mathbf{z}_0$ is a starting point. *Phase trajectory* $\mathbf{z}(t)$ is a solution of (1) for $t \in [0, T]$. Note that the phase trajectory is not directly observed so we call the dynamical system *latent*. Next, introduce an *observation function* $\phi : \mathbb{R}^m \to \mathbb{R}$. It maps phase trajectory to intermediate deterministic representation $y(t) = \phi\big(\mathbf{z}(t)\big)$. Finally, we introduce a stochastic field $h : \mathbb{R} \times [0, T] \to \Omega$. It maps intermediate representation to observable stochastic process $x(t) = h(y(t), t)$ introduced at the begging. The process takes random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The stochastic field h can be treated as adding noise to original observations $y(t)$.

The reason for introducing an intermediate process y is to be able to use Takens theorem [13]. It basically states that if can estimate $y(t)$ from $x(t)$ then we are able to reconstruct latent phase trajectory $z(t)$. It is important for classification because we *associate each class label with some dynamical system* $f_y(\cdot)$, $y \in 1 \ldots M$. If we know all $f_y(\cdot)$ and $z(t)$ is given, we can inference the dynamical system that generated this phase trajectory. Thus we can inference class label and perform classification.

When we don't know vector fields $f_y(\cdot)$ we approximate them from labeled trajectories. The algorithm for approximation we will use is Neural Ordinary Differential Equations (NODE) [14]. It allows to use gradient optimization when an optimized objective depends on ODE trajectories. There are plenty of papers using NODE for regression tasks. For example, NODE are used in the context of physics-informed ODE [15, 16] and control learning [17, 18]. However very few papers analyze NODE in the context of time series classification. Authors of [19] reinterpreted convolutional neural networks in terms of latent ODE dynamics and used it for hyperspectral image classification. Paper [20] interestingly formulated and solved general classification

problem in terms of control theory and NODE. Our paper contributes to NODE classification in terms of latent dynamical systems for time series.

Original NODE paper [14] introduced its own generative latent function time-series model in the section 5. It basically uses the same latent dynamical system model. Observations are generated from latent trajectories through general probability distribution. The authors then used variational autoencoder [21] to inference the distribution and latent vector field. We point out three drawbacks of this approach. First, it struggles to define the starting point $\mathbf{z}_0$ of the latent trajectories from the observations. Authors need to introduce additional probability distribution to account for that. Second, it is not clear how introduced probability distributions should look like for particular tasks. Third, this approach does give explicit way to reconstruct latent trajectories from the observations. Authors have to resort to variational principle [22] and introduce additional parametric models for reconstruction. On the contrary, our time series model does not have problems with starting point and trajectory reconstruction thanks to the Takens theorem. Our general stochastic component $h(y(t), t)$ of the model only affects transition from intermediate state y to observations x and can be easily interpreted for particular tasks as a noise. Applying Takens theorem ideas and NODE to restore latent dynamics can be find in the literature [23, 24]. But unfortunately it is poorly covered for machine learning applications. Our paper aims to close this gap.

References

[1] Rajpurkar, P., Hannun, A., Haghpanahi, M., Bourn, C., Ng, A.: Cardiologist-level arrhythmia detection with convolutional neural networks (2017) https://doi.org/10.48550/arXiv.1707.01836

[2] Lawhern, V., Solon, A., Waytowich, N., Gordon, S., Hung, C., Lance, B.: Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces.

Journal of Neural Engineering **15** (2018) https://doi.org/10.1088/1741-2552/aace8c

[3] Trigeorgis, G., Ringeval, F., Brueckner, R., Marchi, E., Nicolaou, M.A., Schuller, B., Zafeiriou, S.: Adieu features? end-to-end speech emotion recognition using a deep convolutional recurrent network. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5200–5204 (2016). https://doi.org/10.1109/ICASSP.2016.7472669

[4] Lasseck, M.: Bird species identification in soundscapes (2019)

[5] Hammerla, N., Halloran, S., Ploetz, T.: Deep, convolutional, and recurrent models for human activity recognition using wearables. (2016)

[6] Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.K., WOO, W.-c.: Convolutional lstm network: A machine learning approach for precipitation nowcasting (2015)

[7] Berndt, D.J., Clifford, J.: Using dynamic time warping to find patterns in time series. In: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining. AAAIWS'94, pp. 359–370. AAAI Press, ??? (1994)

[8] Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E.: Querying and mining of time series data: experimental comparison of representations and distance measures. Proc. VLDB Endow. **1**(2), 1542–1552 (2008) https://doi.org/10.14778/1454159.1454226

[9] Nanopoulos, A., Alcock, R., Manolopoulos, Y.: Feature-based classification of time-series data, pp. 49–61. Nova Science Publishers, Inc., USA (2001)

[10] Wang, X., Smith, K., Hyndman, R.: Characteristic-based clustering for time series

data. Data Min. Knowl. Discov. **13**(3), 335–364 (2006) https://doi.org/10.1007/s10618-005-0039-x

[11] Mörchen, F.: Time series feature extraction for data mining using dwt and dft (2003)

[12] Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., Shroff, G.: Lstm-based encoder-decoder for multi-sensor anomaly detection (2016) https://doi.org/10.48550/arXiv.1607.00148

[13] Takens, F.: Detecting strange attractors in turbulence. In: Dynamical Systems and Turbulence, Warwick 1980: Proceedings of a Symposium Held at the University of Warwick 1979/80, pp. 366–381 (2006). Springer

[14] Chen, R.T.Q., Rubanova, Y., Bettencourt, J., Duvenaud, D.: Neural ordinary differential equations. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. NIPS'18, pp. 6572–6583. Curran Associates Inc., Red Hook, NY, USA (2018)

[15] Lai, Z., Mylonas, C., Nagarajaiah, S., Chatzi, E.: Structural identification with physics-informed neural ordinary differential equations. Journal of Sound and Vibration **508**, 116196 (2021) https://doi.org/10.1016/j.jsv.2021.116196

[16] Lai, Z., Liu, W., Jian, X., Bacsa, K., Sun, L., Chatzi, E.: Neural Modal ODEs: Integrating Physics-based Modeling with Neural ODEs for Modeling High Dimensional Monitored Structures. https://doi.org/10.48550/arXiv.2207.07883

[17] Yildiz, C., Heinonen, M., Lahdesmaki, H.: In: Advances in Neural Information Processing Systems

[18] Du, J., Futoma, J., Doshi-Velez, F.: Model-based reinforcement learning for semi-markov decision processes with neural odes. NIPS '20. Curran Associates Inc., Red Hook, NY, USA (2020)

[19] Paoletti, M.E., Haut, J.M., Plaza, J., Plaza, A.: Neural ordinary differential equations for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing **58**(3), 1718–1734 (2020) https://doi.org/10.1109/TGRS.2019.2948031

[20] Ruiz-Balet, D., Zuazua, E.: Neural ode control for classification, approximation, and transport. SIAM Review **65**(3), 735–773 (2023) https://doi.org/10.1137/21M1411433 https://doi.org/10.1137/21M1411433

[21] Kingma, D.P., Welling, M.: Auto-encoding variational bayes. CoRR **abs/1312.6114** (2013)

[22] Blei, D.M., Kucukelbir, A., McAuliffe, J.D.: Variational inference: A review for statisticians. Journal of the American statistical Association **112**(518), 859–877 (2017)

[23] Ouala, S., Fablet, R., Drumetz, L., Chapron, B., Pascual, A., Collard, F., Gaultier, L.: Physically informed neural networks for the simulation and data-assimilation of geophysical dynamics, pp. 3490–3493 (2020). https://doi.org/10.1109/IGARSS39084.2020.9323081

[24] Ouala, S., Nguyen, D., Drumetz, L., Chapron, B., Pascual, A., Collard, F., Gaultier, L., Fablet, R.: Learning latent dynamics for partially observed chaotic systems. Chaos: An Interdisciplinary Journal of Nonlinear Science **30**, 103121 (2020) https://doi.org/10.1063/5.0019309