Thematic section

MSML
Mathematics and Statistics in Machine Learning

ORGANIZERS:

Enrique Dominguez (Universidad de Malaga)

Andrzej Nowakowski (Uniwersytet Łódzki)

SCHEDULE OF THE SECTION
Mathematics and Statistics in Machine Learning

- Monday – September 4th

  16:00–16:30 Krzysztof Ślot, *Hybrid neuromorphic architecture with neural representation extraction and hyperdimensional analyzer for image and video classification*

  16:30–17:00 Krzysztof Podlaski, *Automatic method for transcription of handwritten archival documents*

  coffee break

  17:30–18:00 Domingo Lopez Rodriguez, *Explainable Machine Learning using Formal Concept Analysis*

  18:00–18:30 Arkadiusz Tomczyk, *Applications and explainability of graph neural networks*

  18:30–19:00 Kamil Kołodziejski, *Introduction to vector autoregression model*

- Tuesday – September 5th

  14:30–15:00 Bartosz Zieliński, *Learning logical rules using neural networks*

  15:00–15:30 Radosław Matusik, *Fixed-time anti-synchronization for reaction-diffusion neural networks*

  15:30–16:00 Janusz Gajda, *Neural networks boosted by fractional operators*

  coffee break

  16:30–17:00 Marta Lipnicka, *Learning of Neural Network with Optimal Control Tools*

Neural networks boosted by fractional operators

Janusz Gajda

University of Warsaw
Faculty of Economic Sciences
email: jgajda@wne.uw.edu.pl

Abstract

We will discuss the appearance of fractional operators in the context of data analysis with the use of machine learning methods. We discuss the concept of fractional (non-integer order) differentiation on real data of four datasets based on stock prices of main international stock indexes: WIG 20, S&P 500, DAX and Nikkei 225. For fractionally differenced series we use artificial neural networks (ANN) to build a predicitive model. Our work is based on the paper [1].

[1] Walasek R., Gajda J., *Fractional differentiation and its use in machine learning*, International Journal of Advances in Engineering Sciences and Applied Mathematics 13 (2021), no. 2-3, p. 270–277.

Introduction to Vector Autoregression model

Kamil Kołodziejski

Lodz University of Technology
Institute of Mathematics
email: kamil.kolodziejski44@gmail.com

Abstract

When we want to predict future values based on historical observations, we rely on time series models. If our dataset consists of multiple time series that influence each other, we can use Vector Autoregression.

This paper introduces the VAR model [1], presents the Yule-Walker method of parameters estimation and Levinson-Durbin algorithm which solves these equations [3]. To find the optimal order of the VAR model, the Akaike Information Criterion will be shown. Next, there will be discussion about the forecasting problem using the linear predictors [2]. The presentation will be enhanced with the VAR model training examples in R programming language.

[1] Brockwell P.J., Davis R.A., *Introduction to Time Series and Forecasting*, Third edition, 2016.

[2] Brockwell P.J., Davis R.A., *Time Series: Theory and Methods*, Second Edition, 1991.

[3] Musicus B.R., *Levinson and Fast Choleski Algorithms for Toeplitz and Almost Toeplitz Matrices* (1988).

Learning of neural network with optimal control tools

Marta Lipnicka

University of Łódź
Faculty of Math and Computer Sciences
email: marta.lipnicka@wmii.uni.lodz.pl

Abstract

Any learned artificial neural network on a given set of observations represents a function of several variables with vector values or real values. However, in general it is unknown except very simply cases and we have trouble to tell anything about its properties behind very general results received from learning data. In applications, such as medicine, it needs to say not only that on training data we get some error, we have to know that an error is not greater than some ε for all data for which we consider the system.

It is well known that the learned neural network define a function. However, we check correctness of it only on finite number of observed data. We develop an optimal control approach allowing to find approximation of an unknown function realizing the given observable data, parametrized by a set of controls and defined as ordinary differential equations. Moreover, to measure discrepancy, of the output of the network we define a functional into which we include probability distribution function estimating distribution of the data.

We develop a dual dynamic programming ideas to formulate a new optimization problem. We apply it to derive and to prove sufficient approximate optimality conditions for approximate neural network which should work correctly for given ε with respect to built functional, on a data different than the set of observations.

[1] Kosmatka K., Nowakowski A., *Estimating supervisor set using machine learning and optimal control*, Advances in Computational Intelligence 116 (2019), 1089–1096.

[2] Lipnicka M., Nowakowski A., *Optimal control in learning neural network*, In: Computation and Optimization in Information Systems and Management Sciences. MCO 2021. In: Lecture Notes in Networks and Systems 33, 2022.

[3] Lipnicka M., Nowakowski A., *Optimal control using to approximate probability distribution of observation set*, Mathematical Methods in the Applied Sciences 1-16 (2022), doi: 10.1002/mma.8391.

[4] Nowakowski A., *The dual dynamic programming*, Proceedings of the American Mathematical Society 116 (1992), 1089–1096.

[5] Vapnik A., *The Nature of Statistical Learning Theory*, Springer, New York, 2000.

Fixed-time anti-synchronization for reaction-diffusion neural networks

Radosław Matusik

University of Lodz
Faculty of Mathematics and Computer Science
email: radoslaw.matusik@wmii.uni.lodz.pl

joint work with Anna Michalak
University of Lodz, Faculty of Economics and Sociology
and Andrzej Nowakowski

University of LodzFaculty of Mathematics and Computer Science

Abstract

We consider the reaction-diffusion neural network for which coefficients and neural function depend on time and spatial variable. We study fixed-time anti-synchronization (FTAS) problem. We develop a dual dynamic programming theory to derive verification theorem allowing to find and verify the best fixed-time for anti-synchronization of the system.

[1] Chen S., Lim C.-C., Shi P., Lu Z., *Synchronization control for reaction-diffusion FitzHugh-Nagumo systems with spatial sampled data*, Automatica 93 (2018), no. 1, 352–361.

[2] Chen W.-H., Luo S., Zheng W. X., *Impulsive synchronization of reaction-diffusion neural networks with mixed delays and its application to image en- cryption*, IEEE Transactions on Neural Networks and Learning Systems 27 (2016), no. 12, 2696–2710.

[3] Chen Y., Yu W. W., Tan S. L., Zhu H. H., *Synchronizing nonlinear com- plex networks via switching disconnected topology*, Automatica 70 (2016), 189–194.

[4] Ji G., Hu C., Yu J., Jiang H., *Finite-time and fixed-time synchronization of discontinuous complex networks: A Unified control framework design*, Journal of the Franklin Institute 355 (2018), no. 11, 4665–4685.

[5] Lu L., Li C., Wang Z. et al., *Anti-synchronization transmission of the laser signal using uncertain neural network*, Optik 126 (2015), no. 22, 3385–3389.

[6] Nowakowski A., *The dual dynamic programming*, Proceedings of the American Mathematical Society 116 (1992), no. 4, 1089–1096.

[7] Rakkiyappan R., Dharani S., Zhu Q. X., *Synchronization of reaction-diffusion neural networks with time-varying delays via stochastic sampled-data controller*, Nonlinear Dynamics 79 (2015), no. 1, 485–500.

[8] Wang J.-L., Wu H.-N., Guo L., *Novel adaptive strategies for synchronization of linearly coupled neural networks with reaction-diffusion terms*, IEEE Transactions on Neural Networks and Learning Systsems 25 (2014), no. 2, 429–440.

[9] Wang Z., Cao J., Cai Z., Rutkowski L., *Anti-Synchronization in Fixed Time for Discontinuous Reaction-Diffusion Neural Networks With Time-Varying Coefficients and Time Delay*, IEEE Transactions on Cybernetics 50 (2020), no. 6, 2758–2769.

[10] Wu Y., Liu L., Hu J., Feng G., *Adaptive antisynchronization of multilayer reaction-diffusion neural networks*, IEEE Transactions on Neural Networks and Learning. Systsems 29 (2018), no. 4, 807–818.

Automatic method for transcription of handwritten archival documents

Krzysztof Podlaski

University of Lodz
Faculty of Physics and Applied Informatics
email: krzysztof.podlaski@uni.lodz.pl

Abstract

Digitalizing archival documents is an ongoing and important operation for preserving human heritage. Most of the existing archives store documents as well as their scanned copies. Unfortunately, the images of the documents are not suitable for easy use. These documents must be transcribed to be effectively used by people or automatic systems. In this talk, we present the usage of a Convolutional Recurrent Neural Network (CRNN) for the transcription of archives from The State Lodz Archives in Poland. The documents are handwritten in the thirties of the twentieth century. We describe the preprocessing of the images and the CRNN implementation suited for this task. We discuss the effectiveness of the transcription. It proves that the CRNN network can be used for the task, as most of the words were properly transcribed.

[1] Arthur Flor de Sousa Neto, Byron Leite Dantas Bezerra, Alejandro Hector Toselli, and Estanislau Baptista Lima. 2020. HTR-Flor: A Deep Learning System for Offline Handwritten Text Recognition. In *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE.

[2] Honggang Zhang, Jun Guo, Guang Chen, and Chunguang Li. 2009. HCL2000 - A Large-scale Handwritten Chinese Character Database for Handwritten Character Recognition. In *2009 10th International Conference on Document Analysis and Recognition*. IEEE.

Applications and explainability of graph neural networks

Arkadiusz Tomczyk

Lodz University of Technology
Institute of Information Technology
email: arkadiusz.tomczyk@p.lodz.pl

Abstract

Graph data structures are to be found everywhere. Examples are social networks (consisting of people and / or things) and chemical compounds (composed of atoms). A special case of graphs are images where single points (pixels) are organized in a regular grid. Usually both graph nodes and edges are described by means of feature vectors characterizing them. As a result of which, it is possible to predict properties of a graph as a whole or properties of its separate nodes and / or edges (structural prediction). Tools, which are currently used for that purpose, are modern architectures of neural networks (convolutional and recurrent networks, transformers). Such prediction has many crucial applications, e.g.: drug discovery, recommender systems (including security systems) and vision systems. During the presentation a short introduction to neural networks operating on graphs will be given. It will focus on the problem of convolution generalization, attention mechanism as well as review of basic graph convolutional operators: Graph Convolutional Network, Graph Attention Network, Graph Transformer and Mixture Model Network. Next, two practical application areas of graph neural networks will be shown. The first one will be prediction of chemical compounds' properties. Here, not only classification and regression tasks will be considered, but the problem of metric learning will be examined as well. The second one will be devoted to computer vision. It will be argued that changing the representation of image content and its description using graph structures can be beneficial for that domain. In both cases the aspect of explainable artificial intelligence will be emphasized, where the goal is not only to have a well-generalizing model, but also to understand the working principles of the obtained networks. The discussed explanation experiments will base both on attention coefficients and on classic attribution techniques.

Learning logical rules using neural networks

Bartosz Zieliński

University of Łódź
Department of Informatics
email: bartosz.zielinski@fis.uni.lodz.pl

Abstract

Inductive logic programming (ILP) [1] is one of the traditional, rule-based approaches to machine learning. The purpose of numerous variations of ILP is to construct logic programs from examples which explain those examples. ILP approach has many advantages, especially when compared with deep learning: it is easy to incorporate background knowleadge, the resulting models are usually human-understandable (unless they are very large), and it can train on small number of examples (unlike deep learning models which usually require very large training sets). On the other hand, it does not scale very well (compared to neural networks) and it does not work very well with noisy data [2].

Recent years saw development of various neuro-symbolic techniques which combine rule based approaches with deep learning. For example, in differentiable ILP the neural model is interpretable as a description of a logic program (or rather a probability distribution over rules). Usually, the resulting "logic programs" use fuzzy logic (see e.g., [2], [3]).

In the talk I will introduce the basics of differentiable ILP and present its applications, especially to recognition of patterns in text based on limited number of examples.

[1] Cropper A., Dumančić S., *Inductive logic programming at 30: a new introduction*, Journal of Artificial Intelligence Research 74 (2022), 765–850.

[2] Evans R., Grefenstette E., *Learning explanatory rules from noisy data*, Journal of Artificial Intelligence Research 61 (2018), 1–64.

[3] Sen P. et al., *Neuro-symbolic inductive logic programming with logical neural networks*, Proceedings of the AAAI Conference on Artificial Intelligence 36 (2022), no. 8.

[4] Susskind Z. et al., *Neuro-symbolic AI: An emerging class of AI workloads and their characterization*, arXiv:2109.06133 (2021).