Neurocomputing 571 (2024) 127098

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Editorial

Advances in artificial neural networks, machine learning and computational intelligence

Selected papers from the 30th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2022)

This special issue of Neurocomputing presents 16 original articles (one of them being a tutorial selected from a special session) that are extended versions of selected papers from the 30th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2022), a major event for researchers in the field of artificial neural networks and related topics. This single-track conference is held annually in Bruges, Belgium, a UNESCO World Heritage Site and one of the most beautiful medieval city centres in Europe. It is jointly organized by UCLouvain (Louvain-la-Neuve) and KULeuven (Leuven), and it is steered by Prof. Michel Verleysen from UCLouvain. In addition to regular sessions, the conference regularly welcomes special sessions organized by renowned scientists in their respective fields. These sessions focus on specific methodological aspects (e.g., continual learning, anomaly and change point detection, deep semantic segmentation models in computer vision, deep learning for graphs) and applications (e.g., machine learning and information-theoretic methods for molecular biology and medicine).

The contributions in this special issue show that ESANN covers a broad range of topics in neural computing and neuroscience, from theoretical aspects to state-of-the-art applications. The 30th ESANN was held in October 2022 in a hybrid format (due to the COVID-19 pandemic), with 106 in-person participants and 18 remote participants from 19 countries. Around 100 oral and poster contributions were presented. Based on the reviewers' and special session organisers' recommendations, as well as on the quality of the oral presentations at the conference, a number of authors were invited to submit an extended version of their conference paper for this special issue of Neurocomputing. All extended manuscripts went through an additional review process by at least two independent experts, and the 16 articles presented in this volume were accepted for publication:

- Luca Oneto, Sandro Ridella and Davide Anguita
  Do we really need a new theory to understand over-parametrization?
  This century saw an unprecedented increase of public and private investments in Artificial Intelligence (AI) and especially in (Deep) Machine Learning (ML). This led to breakthroughs in their practical ability to solve complex real-world problems impacting research and society at large. Instead, our ability to understand the fundamental mechanism behind these breakthroughs has slowed down because of their increased complexity, while in the past breakthroughs often emerged from foundational research. This questioned researchers about the necessity for a new theoretical framework able to help researchers catch up on this lag. One of the still not well-understood mechanisms is the so-called over-parametrization, namely the ability of certain models to increase their generalization performance (reduce test error) when the number of parameters is above the interpolating threshold (zero training error). In this paper, the authors show that this phenomenon can be better understood using known theories (surveying them in the process) and provide empirical evidence for both shallow and deep learning algorithms.
- Guillermo Castillo-García, Laura Morán-Fernández, Veronica Bolón-Canedo
  Feature selection for domain adaptation using complexity measures and swarm intelligence
  The authors propose to apply the Sticky Binary Particle Swarm Optimization algorithm to perform feature selection for domain adaptation, a specific type of transfer learning in which the source and the target domain have a common feature space, a common task, but different distributions. Differently to other works in literature, the authors adopt complexity metrics instead of error functions, obtaining a fast approach yielding a sparse feature representation.
- Alessio Micheli, Domenico Tortorella
  Addressing heterophily in node classification with graph echo state networks
  Graph Neural Networks have been extensively adopted for node classification tasks. While the most adopted models tend to work well on graphs with a high ratio of intra-class edges, in the opposite case the smoothed representations based on close neighbors computed by convolutional graph neural models are no longer effective. Several architectural variations in message-passing models to reduce excessive smoothing or rewiring the input graph to improve longer-range message passing have been proposed in the literature. In this paper, the authors address the challenges of heterophilic graphs with Graph Echo State Network (GESN), a reservoir computing model for graphs, where node embeddings are recursively computed by an untrained message-passing function.
- Katrin Sophie Bohnsack, Julius Voigt, Marika Kaden, Florian Heinke, Thomas Villmann
  Multi-proximity based embedding scheme for learning vector quantization-based classification of biochemical structured data

https://doi.org/10.1016/j.neucom.2023.127098
17 November 2023
Available online 12 December 2023

The authors propose a scalable data embedding technique for structured data that allows for the direct application of standard vector-based machine learning models without the need for explicit feature extraction. The proposed approach relies on multiple notions of data proximity, making it suitable for handling mixed data types or incorporating domain knowledge. Experimental results on graph and sequence datasets from the biochemical domain show that the method can efficiently be combined with interpretable machine learning approaches like relevance-based learning vector quantization for sophisticated classification learning.

- Madalina Olteanu, Fabrice Rossi, Florian Yger
  Meta-survey on outlier and anomaly detection
  In the field of outlier and anomaly detection, numerous reviews, surveys, and textbooks have sought to summarize the existing literature, encompassing a wide range of methods from both the statistical and data mining communities. The authors present the first systematic meta-survey of general surveys and reviews in the field. Employing a classical systematic survey approach, the study collects nearly 500 papers using two specialized scientific search engines. From this comprehensive collection, a subset of 56 papers that claim to be general surveys on outlier detection is selected using a snowball search technique to enhance field coverage. A meticulous quality assessment phase further refines the selection to a subset of 25 high-quality general surveys. Using this curated collection, the paper investigates the evolution of the outlier detection field over a 20-year period, revealing emerging themes and methods.
- Maximilian Münch, Manuel Röder, Simon Heilig, Christoph Raab, Frank-Michael Schleif
  Static and adaptive subspace information fusion for indefinite heterogeneous proximity data
  The paper faces the problem of machine learning applications where the available data is heterogeneous. Authors focus on kernel methods, proposing two approaches for such heterogeneous data analysis: static and adaptive subspace kernel fusion. The first approach is a kernel-based method extracting the essential parts of the subspace of each input modality and creating one single fused representation of the data. The second approach utilizes an adaptation step by integrating the weighting of spectral properties into the fusion process in order to improve the data's representation with respect to a given classification task. The results show an advantage in fusing the underlying subspace information of heterogeneous data for achieving superior performance in machine learning tasks.
- Quinten Van Baelen, Peter Karsmakers
  Constraint guided gradient descent: Training with inequality constraints with applications in regression and semantic segmentation
  Typically, the training of deep neural network exploits solely data in the form of input–output pairs ignoring available domain knowledge. Authors propose in this paper the Constraint Guided Gradient Descent (CGGD) framework, that enables the injection of domain knowledge into the training procedure. The domain knowledge is assumed to be described as a conjunction of inequality constraints which appears to be a natural choice for several applications. Compared to other neuro-symbolic approaches, the proposed method converges to a point that makes an arbitrarily small error with respect to any inequality constraint on the training data and does not require to first transform the constraints into some ad-hoc term that is added to the learning objective. Authors empirically show that the proposed CGGD makes training less dependent on the initialization of the network, improves the constraint satisfiability on all data, improves the generalization of the model to unseen data, and relaxes the need for annotated data.
- Luis A.Q. Villon, Zachary Susskind, Alan T.L. Bacellar, Igor D.S. Miranda, Leandro S. de Araújo, Priscila M.V. Lima, Mauricio Breternitz Jr., Lizy K. John, Felipe M.G. França, Diego L.C. Dutra
  A conditional branch predictor based on weightless neural networks
  Conditional branch prediction allows the speculative fetching and execution of instructions before knowing the direction of conditional statements. Machine learning techniques are a promising approach to building branch predictors. However, those traditional solutions demand large input sizes, which impose a considerable overhead. Moreover, they should be efficiently implemented in hardware. Authors propose a conditional branch predictor based on the WiSARD (Wilkie, Stoneham, and Aleksander's Recognition Device) weightless neural network model, a kind of neural network. The WiSARD-based predictor implements one-shot online training designed to address branch prediction as a binary classification problem. This work aims to explore the potential gain in accuracy and estimated hardware costs of using such a model for branch prediction.
- Valerio De Caro, Claudio Gallicchio, Davide Bacciu
  Continual adaptation of federated reservoirs in pervasive environments
  When performing learning tasks in pervasive environments, the main challenge arises from the need of combining federated and continual settings. The former comes from the massive distribution of devices with privacy-regulated data. The latter is required by the low resources of the participating devices, which may retain data for short periods of time. In this paper, the authors propose a setup for learning with Echo State Networks (ESNs) in pervasive environments. The proposal focuses on the use of Intrinsic Plasticity (IP), a gradient-based method for adapting the reservoir's non-linearity. First, the objective function of IP is extended to include the uncertainty arising from the distribution of the data over space and time. Then, authors propose Federated Intrinsic Plasticity (FedIP), which is intended for client–server federated topologies with stationary data, and adapts the learning scheme provided by Federated Averaging (FedAvg) to include the learning rule of IP. Finally, this algorithm is extended to equip clients with CL strategies for dealing with continuous data streams. Authors evaluate the proposed algorithms on an incremental setup built upon real-world datasets from human monitoring, showing that both of them improve the representation capabilities and the performance of the ESN, while being robust to catastrophic forgetting.
- Arun Pandey, Hannes De Meulemeester, Bart De Moor, Johan A.K. Suykens
  Multi-view kernel PCA for time series forecasting
  In this paper, the authors propose a kernel principal component analysis model for multi-variate time series forecasting, where the training and prediction schemes are derived from the multi-view formulation of Restricted Kernel Machines. The training problem is simply an eigenvalue decomposition of the summation of two kernel matrices corresponding to the views of the input and output data. When a linear kernel is used for the output view, it is shown that the forecasting equation takes the form of kernel ridge regression. When that kernel is non-linear, a pre-image problem has to be solved to forecast a point in the input space. The model is compared with closely related models on several standard time series datasets.
- Fabian Hinder, Valerie Vaquet, Johannes Brinkrolf, Barbara Hammer
  Model-based explanations of concept drift
  Concept drift refers to the phenomenon that the distribution generating the observed data changes over time. If drift is present, machine learning models can become inaccurate and need adjustment. While there do exist methods to detect concept drift or to adjust models in the presence of observed drift, the question of explaining drift, i.e., describing the potentially complex and high

dimensional change of distributions in a human-understandable fashion, has hardly been considered so far. This problem is of importance since it enables an inspection of the most prominent characteristics of how and where drift manifests. Hence, it allows human understanding of the change and it increases acceptance of life-long learning models. In this paper, the authors present a novel technology characterizing concept drift in terms of the characteristic change of spatial features based on various explanation techniques. To do so, they propose a methodology to reduce the explanation of concept drift to an explanation of models that are trained in a suitable way to extract relevant information from the drift. This way, a large variety of explanation schemes is available, and a suitable method can be selected for the problem at hand. Authors outline the potential of this approach and demonstrate its usefulness in several examples.

- André Artelt, Roel Visser, Barbara Hammer
  "I do not know! but why?" — Local model-agnostic example-based explanations of reject
  Machine learning based decision making systems in safety critical areas place high demands on the accuracy and generalization ability of the underlying model. A common strategy to deal with uncertainties and possible mistakes is offered by learning with reject option, i.e. a model can refrain from prediction in ambiguous cases and leave the decision to a human expert. Yet, as for the models themselves, human decision-making is hampered by the fact that reject options are often implemented as black-box rules: Experts cannot readily understand the reasons for rejection. The authors propose a model-agnostic framework that enriches classification with reject option by explanation mechanisms. They combine conformal prediction as a popular mathematically based technology of certainty estimation with local surrogates derived for the region of interest. In this way, they can provide local explanations in terms of example-based explanation methods, including counterfactual, semi-factual, and factual methods. The performance of the proposed technology is assessed through a series of benchmarks using 6 different data sets; the associated code is open source.
- Dingge Liang, Marco Corneli, Charles Bouveyron, Pierre Latouche
  The graph embedded topic model
  Most of existing graph neural networks (GNNs) developed for the prevalent text-rich networks typically treat texts as node attributes. This kind of approach unavoidably results in the loss of important semantic structures and restricts the representational power of GNNs. In this work, authors introduce a document similarity-based graph convolutional network (DS-GCN) encoder to combine graph topologies and document semantics for text-rich network representation. Then, a graph decoder, based on the latent position model, is used to reconstruct the graph while preserving the network topology. The document matrix is rebuilt by a document decoder, based on the embedded topic model, which considers both topic and word embeddings. By including a cluster membership variable for each node in the network, authors develop an end-to-end clustering technique relying on a new deep probabilistic model called the graph embedded topic model (GETM). Numerical experiments on three simulated scenarios emphasize the ability of GETM in fusing the graph topology structure and the document embeddings, and highlight its node clustering performance. Moreover, an application on the Cora-enrich citation network is conducted to demonstrate the effectiveness and interest of GETM in practice.
- Paolo Andreini, Simone Bonechi, Giovanna Maria Dimitri
  Enhancing glomeruli segmentation through cross-species pre-training
  The importance of kidney biopsy, a medical procedure in which a small tissue sample is extracted from the kidney for examination, is increasing due to the rising incidence of kidney disorders. This procedure helps diagnosing several kidney diseases which are cause of kidney function changes, as well as guiding treatment decisions, and evaluating the suitability of potential donor kidneys for transplantation. In this work, a deep learning system for the automatic segmentation of glomeruli in biopsy kidney images is presented. A novel cross-species transfer learning approach, in which a semantic segmentation network is trained on mouse kidney tissue images and then fine-tuned on human data, is proposed to boost the segmentation performance. The experiments conducted using two deep semantic segmentation networks, MobileNet and SegNeXt, demonstrated the effectiveness of the cross-species pre-training approach leading to an increased generalization ability of both models.
- Danilo Franco, Vincenzo Stefano D'Amato, Luca Pasa, Nicolò Navarin, Luca Oneto
  Fair graph representation learning: Empowering NIFTY via Biased Edge Dropout and Fair Attribute Preprocessing
  The increasing complexity and amount of data available in modern applications strongly demand Trustworthy Learning algorithms that can be fed directly with complex and large graphs data. In fact, on one hand, machine learning models must meet high technical standards (e.g., high accuracy with limited computational requirements), but, at the same time, they must be sure not to discriminate against subgroups of the population (e.g., based on gender or ethnicity). Graph Neural Networks (GNNs) are currently the most effective solution to meet the technical requirements, even if it has been demonstrated that they inherit and amplify the biases contained in the data as a reflection of societal inequities. In fact, when dealing with graph data, these biases can be hidden not only in the node attributes but also in the connections between entities. Several Fair GNNs have been proposed in the literature, with uNIfying Fairness and stabiliTY (NIFTY) (Agarwal et al. 2021) being one of the most effective. In this paper, we will empower NIFTY's fairness with two new strategies. The first one is a Biased Edge Dropout, namely, we drop graph edges to balance homophilous and heterophilous sensitive connections, mitigating the bias induced by subgroup node cardinality. The second one is Attributes Preprocessing, which is the process of learning a fair transformation of the original node attributes. The effectiveness of our proposal will be tested on a series of datasets with increasingly challenging scenarios. These scenarios will deal with different levels of knowledge about the entire graph, i.e., how many portions of the graph are known and which sub-portion is labeled at the training and forward phases.
- Pierre Poitier, Jérôme Fink, Benoît Frénay
  Towards better transition modeling in recurrent neural networks: The case of sign language tokenization
  Recurrent neural networks (RNNs) are a popular family of models widely used when facing sequential data such as videos. However, RNNs make assumptions about state transitions that could be damageable. This contribution presents two theoretical limitations of RNNs along with popular extensions proposed to mitigate those issues. The effectiveness of these methods is assessed in practice on the specific task of sign language (SL) video tokenization, as it remains challenging. Evaluated strategies enhance transition modeling with RNNs functioning as state machines. However, this performance gain diminishes in more complex architectures, indicating there is still room for improvement. Such improvement would help to build powerful SL tokenizers usable in future pipelines in natural language processing.

The guest editors would like to thank all authors for their interesting contributions and all reviewers for their excellent work. Authors and reviewers were asked to respect a very tight schedule, which allowed this issue to be published roughly a year after the conference. We would also like to thank the Neurocomputing editorial board for giving us the opportunity to publish this issue, as well as Elsevier's people for the

very efficient and seamless management of the publication procedure. Finally, our most sincere gratitude goes to Prof. Michel Verleysen for his excellent conference organization, and his strong support to the realization of this special issue.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Nicolò Navarin is associate professor in Computer Science at the Department of Mathematics "Tullio Levi-Civita", University of Padua, Italy. He got his Ph.D. in computer science from the University of Bologna, Italy, in 2014. He has been a visiting Researcher at the University of Freiburg, Germany, at the Università della Svizzera Italiana, Lugano, Switzerland, and at 3IA Côte d'Azur - Interdisciplinary Institute for Artificial Intelligence, Sophia Antipolis, France. He has been a research fellow at the University of Nottingham, UK and at the University of Padua. His research interests lie in the field of machine learning, including kernel methods and neural networks for structured data, online and continual learning, trustworthy ML, and applications to bioinformatics, business process mining, computer vision and computational psychology. At the University of Padua, he is affiliated to the Human Inspired Technology (HIT) research centre. He is a doctoral committee member of the Ph.D. Program in "Brain, Mind and Computer Science". Dr. Navarin has been serving as PC member in major machine learning conferences (ICML, IJCAI, NeurIPS, ECML, AAAI, ICLR), and he has been actively involved in the organization of several special sessions (ESANN, WCCI, IJCNN) and conferences (INNS BDDL 2019, ICPM 2020, IEEESSCI 2021, IEEE WCCI 2022). He is an associate editor for the journals Evolving Systems (Springer) and Neurocomputing (Elsevier).

Dounia Mulders is a postdoctoral fellow in the research groups of Michel Verleysen and André Mouraux at UCLouvain (Belgium). Prior to that, she was a BAEF postdoctoral fellow in the Fiete Lab at MIT (USA) from 2020 to 2021, a research affiliate in the Fiete Lab from 2021 to 2023 and a research affiliate in the Nox Lab at the University of Cambridge (UK) from 2022 to 2023. She received the M.Sc. degree in Mathematical Engineering in 2016 and the Ph.D. degree in Engineering Science and Technology in 2020, both from UCLouvain, Belgium. Her research is primarily focused on understanding the central mechanisms of pain perception and learning in humans, using computational tools to analyze behavioral data and neural recordings.

Luca Oneto was born in Rapallo, Italy in 1986. He received his B.Sc. and M.Sc. in Electronic Engineering at the University of Genoa, Italy respectively in 2008 and 2010. In 2014 he received his Ph.D. from the same university in the School of Sciences and Technologies for Knowledge and Information Retrieval with the thesis "Learning Based On Empirical Data". In 2017 he obtained the Italian National Scientific Qualification for the role of Associate Professor in Computer Engineering and in 2018 he obtained the one in Computer Science. He worked as Assistant Professor in Computer Engineering at University of Genoa from 2016 to 2019. In 2018 he was co-funder of the spin-off ZenaByte s.r.l. In 2019 he obtained the Italian National Scientific Qualification for the role of Full Professor in Computer Science and Computer Engineering. In 2019 he became Associate Professor in Computer Science at University of Pisa and currently is Associate Professor in Computer Engineering at University of Genoa. He has been involved in several H2020 projects (S2RJU, ICT, DS) and he has been awarded with the Amazon AWS Machine Learning and Somalvico (best Italian young AI researcher) Awards. His first main topic of research is the Statistical Learning Theory with particular focus on the theoretical aspects of the problems of (Semi) Supervised Model Selection and Error Estimation. His second main topic of research is Data Science with particular reference to the problem of Trustworthy AI and the solution of real world problems by exploiting and improving the most recent Learning Algorithms and Theoretical Results in the fields of Machine Learning and Data Mining.

Guest editors
Nicolò Navarin *
Department of Mathematics, University of Padua, Italy
E-mail address: nnavarin@math.unipd.it.

Dounia Mulders
IONS and ICTEAM institutes, UCLouvain, Belgium

Luca Oneto
DIBRIS, Università di Genova, Italy

* Corresponding editor.