Advances in artificial neural networks, machine learning and computational intelligence

Oneto, Luca; Bunte, Kerstin; Sperduti, Alessandro

Published in:
Neurocomputing

DOI:
10.1016/j.neucom.2020.03.059

Neurocomputing 416 (2020) 172–176

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 27th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2019)

This special issue of Neurocomputing presents 18 original articles that are extended versions of selected papers from the 27th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2019), 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 centers in Europe. It is jointly organized by UCL (Université Catholique de Louvain – Louvain-la-Neuve) and KU Leuven (Katholiek Universiteit – Leuven), and it is steered by Prof. Michel Verleysen from UCL. In addition to regular sessions, the conference regularly welcomes special sessions organized by renowned scientists in their respective fields. These sessions focus on specific topics, for instance deep learning, reinforcement learning, kernel methods, randomized learning approaches, biomedical applications, robotics, physics, and environmental data analysis, clustering, data privacy, data visualization, and big data analytics.

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. Around 200 researchers from 20 countries attended the 27th ESANN in April 2019. Around 100 oral and poster contributions have been presented. Based on the reviewers' and special session organizers' 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 18 articles presented in this volume were accepted for publication. They can be grouped as follows:

1. **Learning in structured domains**
   (a) Davide Bacciu, Alessio Micheli, and Marco Podda
   Edge-based sequential graph generation with recurrent neural networks
   Graph generation with Machine Learning is an open problem with applications in various research fields. In this work, authors propose to cast the generative process of a graph into a sequential one, relying on a node ordering procedure. Authors use this sequential process to design a novel generative model composed of two recurrent neural networks that learn to predict the presence of edges in a graph: the first network generates one endpoint of each edge, while the second network generates the other endpoint conditioned on the state of the first. Authors test their approach extensively on five different datasets, comparing with two well-known baselines coming from graph literature, and two recurrent approaches, one of which holds state-of-the-art performances. Evaluation is conducted considering quantitative and qualitative characteristics of the generated samples. Results show that their approach is able to yield novel, and unique graphs originating from very different distributions, while retaining structural properties very similar to those in the training dataset.
   (b) Nicoló Navarin, Dinh Van Tran, and Alessandro Sperduti
   A framework for the definition of complex structured feature spaces
   In this paper, authors propose a general framework that, starting from the feature space of an existing base graph kernel, allows to define more expressive kernels which can learn more complex concepts, meanwhile generalizing different proposals in literature. Experimental results on eight real-world graph datasets from different domains show that the proposed framework instances are able to get a statistically significant performance improvement over both the considered base kernels and framework instances previously defined in literature, obtaining state-of-the-art results on all the considered datasets.
2. **Adversarial learning**
   (a) Jonathan Peck, Bart Goossens, and Yvan Saeys
   Detecting adversarial manipulation using inductive Venn-ABERS predictors
   Inductive Venn-ABERS predictors (IVAPs) are a type of probabilistic predictors with the theoretical guarantee that their predictions are perfectly calibrated. In this paper, authors propose to exploit this calibration property for the detection of adversarial examples in binary classification tasks. By rejecting predictions if the uncertainty of the IVAP is too high, authors obtain an algorithm that is both accurate on the original test set and resistant to adversarial examples. This robustness is observed on adversarials for the underlying model as well as adversarials that were generated by taking the IVAP into account. The method appears to offer competitive robustness compared to the state-of-the-art

https://doi.org/10.1016/j.neucom.2020.03.059

in adversarial defense, yet it is computationally much more tractable.

(b) Cyprien Ruffino, Romain Hérault, Eric Laloy, and Gilles Gasso
Pixel-wise conditioned generative adversarial networks for image synthesis and completion
Generative Adversarial Networks (GANs) have proven successful for unsupervised image generation. Several works have extended GANs to image inpainting by conditioning the generation with parts of the image to be reconstructed. Despite their success, these methods have limitations in settings where only a small subset of the image pixels is known beforehand. In this paper, authors investigate the effectiveness of conditioning GANs when very few pixel values are provided. Authors propose a modelling framework which results in adding an explicit cost term to the GAN objective function to enforce pixel-wise conditioning. Authors investigate the influence of this regularization term on the quality of the generated images and the fulfillment of the given pixel constraints. Using the recent PacGAN technique, authors ensure that they keep diversity in the generated samples. Conducted experiments on FashionMNIST show that the regularization term effectively controls the trade-off between quality of the generated images and the conditioning. Experimental evaluation on the CIFAR-10 and CelebA datasets evidences that authors method achieves accurate results both visually and quantitatively in term of Fréechet Inception Distance, while still enforcing the pixel conditioning. Authors also evaluate their method on a texture image generation task using fully-convolutional networks. As a final contribution, authors apply the method to a classical geological simulation application.

3. **Privacy and fairness in learning from data**

(a) Luca Oneto, Michele Donini, Massimiliano Pontil, and John Shawe-Taylor
Randomized learning and generalization of fair and private classifiers: from PAC-Bayes to stability and differential privacy
Authors address the problem of randomized learning and generalization of fair and private classifiers. From one side authors want to ensure that sensitive information does not unfairly influence the outcome of a classifier. From the other side authors have to learn from data while preserving the privacy of individual observations. Authors initially face this issue in the PAC-Bayes framework presenting an approach which trades off and bounds the risk and the fairness of the randomized (Gibbs) classifier. Authors approach is able to handle several different state-of-the-art fairness measures. For this purpose, authors further develop the idea that the PAC-Bayes prior can be defined based on the data-generating distribution without actually knowing it. In particular, authors define a prior and a posterior which give more weight to functions with good generalization and fairness properties. Furthermore, authors show that this randomized classifier possesses interesting stability properties using the algorithmic distribution stability theory. Finally, authors show that the new posterior can be exploited to define a randomized accurate and fair algorithm. Differential privacy theory will allow us to derive that the latter algorithm has interesting privacy preserving properties ensuring authors threefold goal of good generalization, fairness, and privacy of the final model.

(b) Andrew Yale, Saloni Dash, Ritik Dutta, Isabelle Guyon, Adrien Pavao, and Kristin Bennett
Generation and evaluation of privacy preserving synthetic health data
Authors develop metrics for measuring the quality of synthetic health data for both education and research. Authors use novel and existing metrics to capture a synthetic dataset's resemblance, privacy, utility and footprint. Using these metrics, authors develop an end-to-end workflow based on authors generative adversarial network (GAN) method, HealthGAN, that creates privacy preserving synthetic health data. Authors workflow meets privacy specifications of their data partner: (1) the HealthGAN is trained inside a secure environment; (2) the HealthGAN model is used outside of the secure environment by external users to generate synthetic data. This second step facilitates data handling for external users by avoiding de-identification, which may require special user training, be costly, or cause loss of data fidelity. This workflow is compared against five other baseline methods. While maintaining resemblance and utility comparable to other methods, HealthGAN provides the best privacy and footprint. Authors present two case studies in which their methodology was put to work in the classroom and research settings. Authors evaluate utility in the classroom through a data analysis challenge given to students and in research by replicating three different medical papers with synthetic data. Data, code, and the challenge that authors organized for educational purposes are available.

4. **Dimensionality reduction and feature selection**

(a) Cécile Hautecoeur, François Glineur
Nonnegative matrix factorization over continuous signals using parametrizable functions
Nonnegative matrix factorization is a popular data analysis tool able to extract significant features from nonnegative data. Authors consider an extension of this problem to handle functional data, using parametrizable nonnegative functions such as polynomials or splines. Factorizing continuous signals using these parametrizable functions improves both the accuracy of the factorization and its smoothness. Authors introduce a new approach based on a generalization of the Hierarchical Alternating Least Squares algorithm. Proposed method obtains solutions whose accuracy is similar to that of existing approaches using polynomials or splines, while its computational cost increases moderately with the size of the input, making it attractive for large-scale datasets.

(b) Lukas Pfannschmidt, Jonathan Jakob, Fabian Hinder, Michael Biehl, Peter Tino, and Barbara Hammer
Feature relevance determination for ordinal regression in the context of feature redundancies and privileged information
Advances in machine learning technologies have led to increasingly powerful models in particular in the context of big data. Yet, many application scenarios demand for robustly interpretable models rather than optimum model accuracy; as an example, this is the case if potential biomarkers or causal factors should be discovered based on a set of given measurements. In this contribution, authors focus on feature selection paradigms, which enable us to uncover relevant factors of a given regularity based on a sparse model. Authors focus on the important specific setting of linear ordinal regression, i.e., data have to be ranked into one of a finite number of ordered categories by a linear projection. Unlike previous work, authors consider the case that features are potentially redundant, such that no unique minimum set of relevant features exists. Authors aim for an identification of all strongly and all weakly relevant features as well as their type of relevance (strong or weak); authors achieve this goal by determining feature relevance bounds, which correspond to the minimum and maximum feature relevance, respectively, if searched over all equivalent models. In addition to this, authors discuss how this

setting enables us to substitute some of the features e.g. due to their semantics, and how to extend the framework of feature relevance intervals to the setting of privileged information, i.e. potentially relevant information is available for training purposes only, but cannot be used for the prediction itself.

5. **Weightless neural networks**
   (a) Leopoldo A. D. Lusquino Filho, Luiz F. R. Oliveira, Aluizio Lima Filho, Gabriel P. Guarisa, Lucca M. Felix, Priscila M. V. Lima, and Felipe M. G. França
   Extending the weightless WiSARD classifier for regression
   This paper explores two new weightless neural network models, Regression WiSARD and ClusRegression WiSARD, in the challenging task of predicting the total palm oil production of a set of 28 differently located sites under different climate and soil profiles. Both models were derived from Kolcz and Allinson's n-Tuple Regression weightless neural model and obtained improved mean absolute error (MAE) rates, whilst being four orders of magnitude faster during the training phase. Additionally the models have been tested on three classic regression datasets, obtaining competitive performances with respect to other models often used in this type of task.
   (b) Leandro Santiago, Letícia Verona, Fábio Rangel, Fabrício Firmino, Daniel S. Menasche, Wouter Caarls, Mauricio Breternitz Jr, Sandip Kundu, Priscila M. V. Lima, and Felipe M. G. França
   Weightless neural networks as memory segmented bloom filters
   Weightless Neural Networks (WNNs) are Artificial Neural Networks based on RAM memory broadly explored as solution for pattern recognition applications. Memory-oriented solutions for pattern recognition are typically very simple, and can be easily implemented in hardware and software. Nonetheless, the straightforward implementation of a WNN requires a large amount of memory resources making its adoption impracticable on memory constrained systems. In this paper, authors establish a foundational relationship between WNN and Bloom filters, presenting a novel unified framework which encompasses the two. In particular, authors indicate that a WNN can be framed as a memory segmented Bloom filter. Leveraging such finding, authors propose a new model of WNNs which utilizes Bloom filters to implement RAM nodes. Bloom filters reduce memory requirements, and allow false positives when determining if a given pattern was already seen in data. Authors experimentally found that for pattern recognition purposes such false positives can build robustness into the system. The experimental results show that their model using Bloom filters achieves competitive accuracy, training time and testing time, consuming up to 6 orders of magnitude less memory resources when compared against the standard Weightless Neural Network model.

6. **Algorithms based on neural networks**
   (a) Iryna Korshunovaa, Yarin Galb, Arthur Grettonc, and Joni Dambre
   Conditional BRUNO: a neural process for exchangeable labelled data
   Authors present a neural process which models exchangeable sequences of high-dimensional complex observations conditionally on a set of labels or tags. Their model combines the expressiveness of deep neural networks with the data-efficiency of Gaussian processes, resulting in a probabilistic model for which the posterior distribution is easy to evaluate and sample from. In addition to that, the computational complexity scales linearly with the number of observations. The advantages of the proposed architecture are demonstrated on a challenging few-shot view reconstruction task which requires generalization from short sequences of viewpoints, and a contextual bandits problem.
   (b) Lukas Enderich, Fabian Timm, and Wolfram Burgard
   SYMOG: learning symmetric mixture of Gaussian modes for improved fixed-point quantization
   Deep neural networks (DNNs) have been proven to outperform classical methods on several machine learning benchmarks. However, they have high computational complexity and require powerful processing units. Especially when deployed on embedded systems, model size and inference time must be significantly reduced. Authors propose SYMOG (symmetric mixture of Gaussian modes), which significantly decreases the complexity of DNNs through low-bit fixed-point quantization. SYMOG is a novel soft quantization method such that the learning task and the quantization are solved simultaneously. During training the weight distribution changes from an unimodal Gaussian distribution to a symmetric mixture of Gaussians, where each mean value belongs to a particular fixed-point mode. Authors evaluate their approach with different architectures (LeNet5, VGG7, VGG11, DenseNet) on common benchmark data sets (MNIST, CIFAR-10, CIFAR-100) and compare it with state-of-the-art quantization approaches. Authors achieve excellent results and outperform 2-bit state-of-the-art performance.
   (c) Balthazar Donon, Benjamin Donnot, Isabelle Guyon, Zhengying Liu, Antoine Marot, Patrick Panciatici, and Marc Schoenauer
   LEAP nets for system identification and application to power systems
   Using neural network modeling, authors address the problem of system identification for continuous multivariate systems, whose structures vary around an operating point. Structural changes in the system are of combinatorial nature, and some of them may be very rare; they may be actionable for the purpose of controlling the system. Although authors ultimate goal is both system identification and control, authors only address the problem of identification in this paper. Authors propose and study a novel neural network architecture called LEAP net, for Latent Encoding of Atypical Perturbation. Authors method maps system structure changes to neural net structure changes, using structural actionable variables. Authors demonstrate empirically that LEAP nets can be trained with a natural observational distribution, very concentrated around a "reference" operating point of the system, and yet generalize to rare (or unseen) structural changes. Authors validate the generalization properties of LEAP nets theoretically in particular cases. Authors apply their technique to power transmission grids, in which high voltage lines are disconnected and re-connected with one-another from time to time, either accidentally or willfully. Authors discuss extensions of their approach to actionable variables, which are continuous (instead of discrete, in the case of their application) and make connections between their problem setting, transfer learning, causal inference, and reinforcement learning.
   (d) [***]Jensun Ravichandran, Marika Kaden, Sascha Saralajew, and Thomas Villmann
   Variants of DropConnect in learning vector quantization networks for evaluation of classification stability

[***] This paper was erroneously published in a regular volume of Neurocomputing instead of this special issue, but should be considered to be part of this special issue. The full publication details of this papers are as follows: Neurocomputing, Volume 403C, 25 August 2020, Pages 121–132.

Dropout and DropConnect are useful methods to prevent multilayer neural networks from overfitting. In addition, it turns out that these tools can also be used to estimate the stability of networks regarding disturbances. Prototype based networks gain more and more attraction in current research because of their inherent interpretability and robust behavior. Popular prototype-based classifiers are support vector machines and the heuristically motivated Learning Vector Quantizer (LVQ). The Generalized Matrix LVQ (GMLVQ) is an extension of LVQ which can be interpreted as a special multilayer network containing a projection and a prototype layer. First in this paper, authors extend the linear projection layer of GMLVQ to a non-linear mapping by employing different non-linear activations functions. Second, authors compare the classification decision stabilities of the linear and the non-linear GMLVQ regarding DropConnect while taking the neural network perspective. Thus authors can adopt DropConnect ideas known from multilayer perceptron learning to investigate stability and robustness of GMLVQ. To this end, the evaluation of the stability is done in terms of a information theoretic stability measure based on the Shannon-Entropy. Authors demonstrate the approach for three real world data sets from Raman spectroscopy, multispectral remote sensing and the well-known MNIST data set.

7. **Other topics in learning from data**
   (a) Jiajun Pan and Hoel Le Capitaine
   Metric learning with submodular functions
   Most of the metric learning approaches mainly focus on using single feature weights with L_p norms, or pair of features with Mahalanobis distances to learn the similarities between the samples, while ignoring the potential value of higher-order interactions in the feature space. In this paper, authors investigate the possibility of learning weights associated to coalitions of features whose cardinality can be greater than two, with the help of set-functions. Exploiting the property of submodular set-function, authors propose to define a metric for continuous features based on Lovasz extension of submodular functions, and then present a dedicated metric learning approach. In order to cope with the increase in computational complexity introduced by submodular constraints, authors exploit the ξ-additive fuzzy measure to reduce the order of interactions that are taken into account, consequently decreasing the computational complexity. Experiments on various datasets show the effectiveness of the proposed approach.
   (b) Christoph Raaba, Moritz Heusinger, and Frank-Michael Schleif
   Reactive soft prototype computing for concept drift streams
   The amount of real-time communication between agents in an information system has increased rapidly since the beginning of the decade. This is because the use of these systems, e.g. social media, has become commonplace in today's society. This requires analytical algorithms to learn and predict this stream of information in real-time. The nature of these systems is non-static and can be explained, among other things, by the fast pace of trends. This creates an environment in which algorithms must recognize changes and adapt. Recent work shows vital research in the field, but mainly lack stable performance during model adaptation. In this work, a concept drift detection strategy followed by a prototype-based adaptation strategy is proposed. Validated through experimental results on a variety of typical non-static data, authors solution provides stable and quick adjustments in times of change.
   (c) Markus Kaisera, Clemens Ottea, Thomas A. Runklera, Carl Henrik Ek
   Bayesian decomposition of multi-modal dynamical systems for reinforcement learning
   In this paper, authors present a model-based reinforcement learning system where the transition model is treated in a Bayesian manner. The approach naturally lends itself to exploit expert knowledge by introducing priors to impose structure on the underlying learning task. The additional information introduced to the system means that authors can learn from small amounts of data, recover an interpretable model and, importantly, provide predictions with an associated uncertainty. To show the benefits of the approach, authors use a challenging data set where the dynamics of the underlying system exhibit both operational phase shifts and heteroscedastic noise. Comparing their model to NFQ and BNN+LV, authors show how their approach yields human-interpretable insight about the underlying dynamics while also increasing data-efficiency.
   (d) Rodrigo S. Sousa, Priscila G. M. dos Santos, Tiago M. L. Veras, Wilson R. de Oliveira, Adenilton J. da Silva
   Parametric Probabilistic Quantum Memory
   Probabilistic Quantum Memory (PQM) is a data structure that computes the distance from a binary input to all binary patterns stored in superposition in the memory. This data structure allows the development of heuristics to speed up artificial neural networks architecture selection. In this work, authors propose an improved parametric version of the PQM to perform pattern classification, and authors also present a PQM quantum circuit suitable for Noisy Intermediate Scale Quantum (NISQ) computers. Authors present a classical evaluation of a parametric PQM network classifier on public benchmark datasets. Authors also perform experiments to verify the viability of PQM on a 5-qubit quantum computer.

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 in less than a year after the conference, timely before the ESANN meeting of 2020. 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

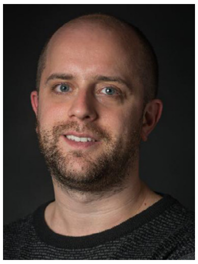
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Luca Oneto*
DIBRIS, Università di Genova, Italy

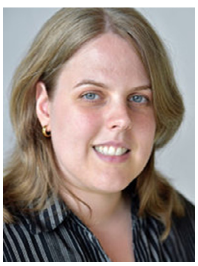
Kerstin Bunte
University of Groningen, Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, the Netherlands

Alessandro Sperduti
Dipartimento di Matematica "Tullio Levi-Civita", Università di Padova, Italy

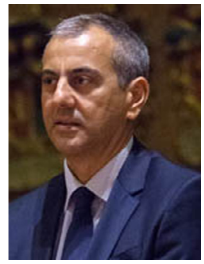
*Corresponding author.
E-mail address: luca.oneto@zenabyte.com (L. Oneto)



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.



Kerstin Bunte graduated at the Faculty of Technology at the University of Bielefeld, Germany. Since July 2016 she is assistant professor and Rosalind Franklin Fellow at the University of Groningen. Dr. Bunte received her Ph.D. in Computer Science in 2011 from the University of Groningen. She had research visits at the University of Rochester (USA) and postdoctoral positions have taken her to Bielefeld University (Germany), Aalto University (Finland), and Université catholique de Louvain (Belgium). In 2015 she got a European Marie Skłodowska-Curie Fellowship (Project ID: 659104) at the University of Birmingham and is currently partner in the European ITN: SUNDIAL, Project ID: 721463 and the Regions of Smart Factories project. Her recent work has focused on the development of interpretable machine learning techniques (explainable AI) for interdisciplinary data analysis, dimensionality reduction and visualization. Further information can be obtained from http://www.cs.rug.nl/~kbunte/.



Alessandro Sperduti is Professor in Computer Science at the Department of Mathematics "Tullio Levi-Civita" of the University of Padua, Italy. He holds a Master degree and a PhD in Computer Science from the University of Pisa, where he has been an assistant professor and associate professor at the Department of Computer Science. His main research interests are in Neural Networks, and Kernel Methods, with a focus on learning in structured domains. Currently he is also working on the integration of multimedia data with background knowledge encoded in ontologies. He is also active in the Process Mining research area, that he helped to promote with the creation of the IEEE CIS Task Force, currently running an international conference series (ICPM) with the sponsor of world-wide industrial organizations and consultancy agencies (e.g., SAP and Deloitte). He has been chair of the Data Mining Technical Committee, and of the Neural Networks Technical Committee of the IEEE Computational Intelligence Society. He has delivered several tutorials in main Artificial Intelligence conferences (e.g., IJCAI and WCCI), and summer schools (e.g., International Summer School on Deep Learning 2017, International Summer School on Deep Learning On-chip 2017). He was the recipient of the 2000 AI*IA (Italian Association for Artificial Intelligence) "MARCO SOMALVICO" Young Researcher Award.