University of Groningen

**Advances in artificial neural networks, machine learning and computational intelligence
Selected papers from the 26th European Symposium on Artificial Neural Networks,
Computational Intelligence and Machine Learning (ESANN 2018)**

Oneto, Luca; Bunte, Kerstin; Schleif, Frank-Michael

Published in:
Neurocomputing

DOI:
[10.1016/j.neucom.2019.01.081](https://doi.org/10.1016/j.neucom.2019.01.081)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Oneto, L., Bunte, K., & Schleif, F-M. (2019). Advances in artificial neural networks, machine learning and computational intelligence Selected papers from the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018). *Neurocomputing*, 342, 1-5. <https://doi.org/10.1016/j.neucom.2019.01.081>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Editorial

Advances in artificial neural networks, machine learning and computational intelligence

Selected papers from the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018)

This special issue of Neurocomputing presents 17 original articles that are extended versions of selected papers from the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018), 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 by one of the most beautiful medieval city centers in Europe. It is organized in collaboration by UCL (Université Catholique de Louvain – Louvain-la-Neuve) and KULeuven (Katholiek Universiteit – Leuven) and 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 particular topics, for instance deep learning, reinforcement learning, kernel methods, randomized learning approaches, biomedical, 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. More than 100 researchers from 20 countries participated in the 26th ESANN in April 2018. Around 100 oral and poster communications have been presented this year. 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 17 articles presented in this volume were accepted for publication. They can be grouped as follows:

1. Statistical Learning Theory

- (a) Khadija Musayeva, Fabien Lauer, and Yann Guermeur
Rademacher complexity and generalization performance of multi-category margin classifiers

One of the main open problems in the theory of multi-category margin classification is the form of the optimal dependency of a guaranteed risk on the number C of

categories, the sample size m and the margin parameter γ . From a practical point of view, the theoretical analysis of generalization performance contributes to the development of new learning algorithms. In this paper, authors focus only on the theoretical aspect of the question posed. More precisely, under minimal learnability assumptions, authors derive a new risk bound for multi-category margin classifiers. Authors improve the dependency on C over the state of the art when the margin loss function considered satisfies the Lipschitz condition. Authors start with the basic supremum inequality that involves a Rademacher complexity as a capacity measure. This capacity measure is then linked to the metric entropy through the chaining method. In this context, authors improvement is based on the introduction of a new combinatorial metric entropy bound.

- (b) Rynkiewicz Joseph
Asymptotic statistics for multilayer perceptron with ReLU hidden units

In numerous tasks, deep networks are state of the art. However, they are still not well understood from a statistical point of view. In this article, author tries to contribute to filling this gap, and considers regression models involving deep multilayer perceptrons (MLP) with rectified linear (ReLU) functions for activation units. It is a difficult task to study the statistical properties of such models. The main reason is that in practice these models may be heavily overparameterized. For the sake of simplicity, author focuses here on the sum of square errors (SSE) cost function which is the standard cost function for regression purposes. In this framework, author studies the asymptotic behavior of the difference between the SSE of estimated models and the SSE of the theoretical best model. This behavior gives us information on the overfitting properties of such models. In this paper new methodology is introduced to deal with models with a loss of identifiability, i.e. in the case that the true parameter cannot be identified uniquely. Hence, author does not have to assume that a unique parameter vector realizes

the best regression function, which seems to be a too strong assumption for heavily overparameterized models. Results shed new light on the overfitting behavior of MLP models.

2. New learning algorithms

- (a) Luca Oneto, Sandro Ridella, and Davide Anguita

Local Rademacher Complexity Machine

Support Vector Machines (SVMs) are a state-of-the-art and powerful learning algorithm that can effectively solve many real world problems. SVMs are the transposition of the Vapnik–Chervonenkis (VC) theory into a learning algorithm. This paper presents the Local Rademacher Complexity Machine (LRCM), a transposition of the Local Rademacher Complexity (LRC) theory, the state-of-the-art evolution of the VC theory, into a learning algorithm. Analogously to what has been done for the SVMs, authors present first the theoretical ideas behind the LRC theory, then show how these ideas can be translated into a learning algorithm, the LRCM, and then how the LRCM can be made efficient and kernelizable. By exploiting a series of real world datasets, authors show the effectiveness of the LRCM against the SVMs.

- (b) Tommi Kaärkkäinen

Extreme minimal learning machine: ridge regression with distance-based basis

The extreme learning machine (ELM) and the minimal learning machine (MLM) are nonlinear and scalable machine learning techniques with a randomly generated basis. Both techniques start with a step in which a matrix of weights for the linear combination of the basis is recovered. In the MLM, the feature mapping in this step corresponds to distance calculations between the training data and a set of reference points, whereas in the ELM, a transformation using a radial or sigmoidal activation function is commonly used. Computation of the model output, for prediction or classification purposes, is straightforward with the ELM after the first step. In the original MLM, one needs to solve an additional multilateration problem for the estimation of the distance-regression based output. A natural combination of these two techniques is proposed and experimented in this paper: to use the distance-based basis characteristic in the MLM in the learning framework of the regularized ELM. In other words, author conducts ridge regression using a distance-based basis. The experimental results characterize the basic features of the proposed technique and surprisingly, indicate that overlearning with the distance-based basis is in practice avoided in classification problems. This makes the model selection for the proposed method trivial, at the expense of computational costs.

- (c) Davide Bacciu and Daniele Castellana

Bayesian mixtures of hidden tree Markov models for structured data clustering

The paper deals with the problem of unsupervised learning with structured data, proposing a mixture model approach to cluster tree samples. First, authors discuss how to use the switching-parent hidden tree Markov model, a compositional model for learning tree distributions, to define a finite mixture model where the number of components is fixed by a hyperparameter. Then, authors show how to relax such an assumption by introducing a Bayesian non-parametric mixture model where the number of necessary hidden tree components is learned from data. Experimental validation on synthetic and real datasets show the benefit of mixture models over simple hidden tree models in clustering applications. Further, authors provide a characterization of the behavior of the two mixture models for different choices of their hyperparameters.

3. Algorithms based on neural networks

- (a) Josef Feigl and Martin Bogdan

Neural networks for personalized item rankings

Most users typically interact with products only through implicit feedback such as clicks or purchases rather than explicit user-provided information like product ratings. Learning to rank products according to individual preferences using only implicit feedback can be helpful to make useful recommendations. In this paper, a neural network architecture to solve collaborative filtering problems for personalized rankings on implicit feedback datasets is presented. It is shown how a layer of constant weights forces the network to learn pairwise item rankings. Additionally, similarities between the proposed neural network and a matrix factorization model trained with the Bayesian personalized ranking optimization criterion are proven. The experiments indicate state-of-the-art performance for the task of personalized ranking.

- (b) Jan M. Wülfing, Sreedhar S. Kumar, Joschka Boedecker, Martin Riedmiller, and Ulrich Egert

Adaptive long-term control of biological neural networks with deep reinforcement learning

Driving activity patterns in the brain to desired levels is an important therapeutic strategy for many neurological disorders. The highly dynamic nature of neuronal networks and changes with disease progression create an urgent need for closed-loop control. Without adequate mathematical models of such complex networks, it remains unclear how tractable control problems can be formulated and solved for neurobiological systems. Reinforcement Learning (RL) is a promising tool to address such challenges, but has rarely been used for the long-term control of live, plastic neural networks. This is a difficult problem since it requires the controller to adapt to poorly characterized non-stationary background processes that alter stimulus-response relations over time. Authors captured these challenges in a novel control task, defined as clamping response strengths to predefined levels over long durations in a living model system, namely generic BNNs in vitro grown on microelectrode arrays. Authors show that by defining appropriate state-action spaces and employing powerful non-linear RL methods such as Deep RL, adaptivity to non-stationary background processes can be achieved in a series of experiments: in 27/29 networks, the learned controllers were able to improve performance compared to a random and a (linear) LSPI controller by a large margin.

- (c) Tayfun Alpay, Fares Abawi, and Stefan Wermter

Preserving activations in recurrent neural networks based on surprisal

Learning hierarchical abstractions from sequences is a challenging and open problem for Recurrent Neural Networks (RNNs). This is mainly due to the difficulty of detecting features that span over long time distances with also different frequencies. In this paper, authors address this challenge by introducing surprisal-based activation, a novel method to preserve activations and skip updates depending on encoding-based information content. The preserved activations can be considered as temporal shortcuts with perfect memory. Authors present a preliminary analysis by evaluating surprisal-based activation on language modeling with the Penn Treebank corpus and find that it can improve performance when compared to baseline RNNs and Long Short-Term Memory (LSTM) networks.

4. Dimensionality reduction and feature selection

- (a) Rebecca Marion, Adrien Bibalb, and Benoît Frénay

BIR: a method for selecting the best interpretable multidimensional scaling rotation using external variables

Interpreting nonlinear dimensionality reduction models using external features (or external variables) is crucial in many fields, such as psychology and ecology. Multidimensional scaling (MDS) is one of the most frequently used dimensionality reduction techniques in these fields. However, the rotation invariance of the MDS objective function may make interpretation of the resulting embedding difficult. This paper analyzes how the rotation of MDS embeddings affects sparse regression models used to interpret them and proposes a method, called the Best Interpretable Rotation (BIR) method, which selects the best MDS rotation for interpreting embeddings using external information.

- (b) Borja Seijo-Pardo, Amparo Alonso-Betanzos, Kristin P. Bennett, Verónica Bolón-Canedo, Julie Josse, Mehreen Saeed, Isabelle Guyon

Biases in feature selection with missing data

Feature selection is of great importance for two possible scenarios: (1) prediction, i.e. improving (or minimally degrading) the predictions of a target variable while discarding redundant or uninformative features and (2) discovery, i.e. identifying features that are truly dependent on the target and may be genuine causes to be determined in experimental verifications (for example for the task of drug target discovery in genomics). In both cases, if variables have a large number of missing values, imputing them may lead to false positives; features that are not associated with the target become dependent as a result of imputation. In the first scenario, this may not harm prediction, but in the second one, it will erroneously select irrelevant features. In this paper, authors study the risk/benefit trade-off of missing value imputation in the context of feature selection, using causal graphs to characterize when structural bias arises. Authors aim is also to investigate situations in which imputing missing values may be beneficial to reduce false negatives, a situation that might arise when there is a dependency between feature and target, but the dependency is below the significance level when only complete cases are considered. However, the benefits of reducing false negatives must be balanced against the increased number of false positives. In the case of binary target variable and continuous features, the *t*-test is often used for univariate feature selection. In this paper, authors also introduce a de-biased version of the *t*-test allowing us to reap the benefits of imputation, while not incurring the penalty of increasing the number of false positives.

5. Machine learning challenges

- (a) Mirko Polato and Fabio Aielli

Boolean kernels for rule based interpretation of Support Vector Machines

Machine learning started as an academic-oriented domain, but nowadays it is becoming more and more widespread across diverse domains, such as retail, healthcare, finance, and many more. This non-academic face of machine learning creates a new set of challenges. The usage of such complex methods by non-expert users has increased the need for interpretable models. To this end, in this paper authors propose an approach for extracting explanation rules from Support Vector Machines. The core idea is based on using kernels with feature spaces composed by logical propositions. On top of that, a searching algorithm tries to retrieve the most relevant features/rules that can be used to explain the trained model. Experiments on both categorical and real-valued datasets show the effectiveness of the proposed approach.

- (b) Johannes Brinkrolf, Christina Göpfert, and Barbara Hammer

Differential privacy for learning vector quantization

Prototype-based machine learning methods such as learning vector quantization (LVQ) offer flexible classification tools, which represent a classification in terms of typical prototypes. This representation leads to a particularly intuitive classification scheme, since prototypes can be inspected by a human partner in the same way as data points. Yet, it bears the risk of revealing private information included in the training data, since individual information of a single training data point can significantly influence the location of a prototype. In this contribution, authors investigate the question how to algorithmically extend LVQ such that it provably obeys privacy constraints as offered by the notion of so-called differential privacy. More precisely, authors demonstrate the sensitivity of LVQ to single data points and hence the need of its extension to private variants in case of possibly sensitive training data. Authors investigate three technologies which have been proposed in the context of differential privacy, and authors extend these technologies to LVQ schemes. Authors investigate the effectiveness and efficiency of these schemes for various data sets, and authors evaluate their scalability and robustness as regards the choice of meta-parameters and characteristics of training sets. Interestingly, one algorithm, which has been proposed in the literature due to its beneficial mathematical properties, does not scale well with data dimensionality, while two alternative techniques, which are based on simpler principles, display good results in practical settings.

6. Applications

- (a) Stefan Heinrich, Peer Springstübe, Tobias Knöppler, Matthias Kerzel, and Stefan Wermter

Continuous convolutional object tracking in developmental robot scenarios

Tracking arbitrary objects in natural environments is a challenging task in visual computing. A central problem is the need to adapt to changing appearances under strong transformation and occlusion. Authors propose a tracking framework that utilizes the strength of Convolutional Neural Networks to create a robust and adaptive model of the object from training data produced during tracking. An incremental update mechanism provides increased performance and reduces the computational costs for training during tracking, allowing for robust real-time tracking with state-of-the-art performance. Together with optimizations for deploying the framework on humanoid robots and distributed devices, this shows its viability for research in developmental robotics on questions around infant cognition or active exploration.

- (b) Miguel Angrick, Christian Herff, Garrett Johnson, Jerry Shih, Dean Krusienski, and Tanja Schultz

Interpretation of Convolutional Neural Networks for speech spectrogram regression from intracranial recordings

The direct synthesis of continuously-spoken speech from neural activity could provide a fast and natural way of communication for users suffering from speech disorders. Mapping the complex dynamics of neural activity to spectral representations of speech is a demanding task for regression models. Convolutional Neural Networks have recently shown promise for finding patterns in neural signals and might be a good candidate for this particular regression task. However, the intrinsic agency of the resulting networks is challenging to interpret and thus provides little opportunity to gain insights on neural processes underlying speech. While activation maximization can be used to get a glimpse

into what a network has learned for a classification task, it usually does not benefit regression problems.

In this paper, authors show that Convolutional Neural Networks can be used to reconstruct an audible waveform from invasively-measured brain activity. By adapting activation maximization, authors present a method that can provide insights from neural networks targeting regression problems. Based on experimental data, authors achieve statistically-significant correlations between spectrograms of synthesized and original speech. Authors interpretation approach shows that trained models reveal that electrodes placed in cortical regions associated with speech production tasks have a large impact on the reconstruction of speech segments.

- (c) Niall Twomey, Haoyan Chen, Tom Diethe, and Peter Flach
An application of Hierarchical Gaussian Processes to the detection of anomalies in star light curves

This study is concerned with astronomical time-series called light-curves that represent the brightness of celestial objects over a period of time. Authors consider the task of finding anomalous light-curves of periodic variable stars. Authors employ a Hierarchical Gaussian Process to create a general and stable model of time-series for anomaly detection, and apply this approach to the light-curve problem. Hierarchical Gaussian Processes require only a few additional parameters compared to conventional Gaussian Processes and incur negligible additional computational complexity. Moreover, since the additional parameters are objectively optimized in a principled probabilistic framework one does not need to resort to grid searches for parameter selection. Experimentally, authors demonstrate that authors approach outperforms several baselines on both synthetic and light-curve data. Of particular interest is that the proposed method generalizes very well from small subsets of the data, achieving near perfect precision of outlier detection even with as few as seven instances.

- (d) Mohammad Mohammadi, Nicolai Petkov, Kerstin Bunte, Reynier F. Peletier, and Frank-Michael Schleif
Globular cluster detection in the GAIA survey

Many techniques for the detection of interesting stellar structures in astronomical data sets require full phase-space or color information, which is not always available. The first data release of the GAIA satellite, for example, provided highly accurate positions and magnitudes for more than one billion sources. Therefore, the question arises if such structures can also be automatically found without waiting for more detailed information in future data releases. In this contribution authors propose and compare two conceptually different strategies to find globular clusters in the GAIA DR1 survey. The first approach is a nearest neighbor retrieval and the second an anomaly detection. Both techniques are able to find most of the known globular clusters within the observation frames consistently, as well as potential candidates for further investigation. Furthermore, authors address approximation approaches to scale the strategy to larger data.

- (e) Aleke Nolte, Lingyu Wang, Maciej Bilicki, Benne Holwerda, Michael Biehl
Galaxy classification: a machine learning analysis of GAMA catalogue data

Authors present a prototype-based machine learning analysis of five different labeled galaxy catalogues containing parameters from the Galaxy and Mass Assembly (GAMA) survey. Extending the work previously presented at the ESANN 2018 conference – an analysis that is based on Generalized Relevance Matrix Learning Vector Quantization, authors find

that neither the data from the single catalogues individually nor a combined dataset based on all 5 catalogues does fully support the popular visual-inspection-based galaxy classification scheme employed to categorize the galaxies. In particular, only one class, the Little Blue Spheroids, is consistently separable from the other classes. For each of the catalogues, authors present the galaxy parameters that are most discriminative for the achieved class distinctions.

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 2019. 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 strong support of this special issue and his excellent conference organization.

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 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 the one in Computer Science. He worked as Assistant Professor in Computer Engineering at University of Genoa from 2016 to 2019 and he is currently Associate Professor in Computer Science at University of Pisa with particular interests in Statistical Learning Theory and Data Science.

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. Her recent work has focused on the development of interpretable machine learning techniques for interdisciplinary data analysis, dimensionality reduction and visualization. Further information can be obtained from <http://www.cs.rug.nl/~kbunte/>

Frank-Michael Schleif (Dipl.-Inf., University of Leipzig, Ph.D., TU-Clausthal, Germany) was a Marie Curie Senior Research Fellow at the University of Birmingham, Birmingham, UK and a Post-Doctoral Fellow in the group of Theoretical Computer Science (TCS) at the University of Bielefeld, Bielefeld, Germany, where he also received a *venia legendi* in applied computer science in 2013. He was also a software developer and consultant for the Bruker Corp. Since 2016 he is with the University of Applied Sciences, Wuerzburg, Germany, where he is a Professor for Database Management and Business Intelligence. His current research interests include data management, computational intelligence techniques and machine learning for non-metric models and large scale problems. Several research stays have taken him to UK, the Netherlands, Japan and the USA. He is a member of the German chapter of the European Neural Network Society (GNNS), the GI and the IEEE-CIS. He is editor of the Machine Learning Reports and member of the editorial board of the Neural Processing Letters.

Guest Editors

Frank-Michael Schleif

*Department of Computer Science, University of Applied Sciences
Wuerzburg-Schweinfurt, Germany*

Luca Oneto*

Dipartimento di Informatica, Università di Pisa, Italy

Kerstin Bunte

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

*Corresponding author.

E-mail address: luca.oneto@unige.it (L. Oneto)

Received 21 January 2019

Accepted 23 January 2019