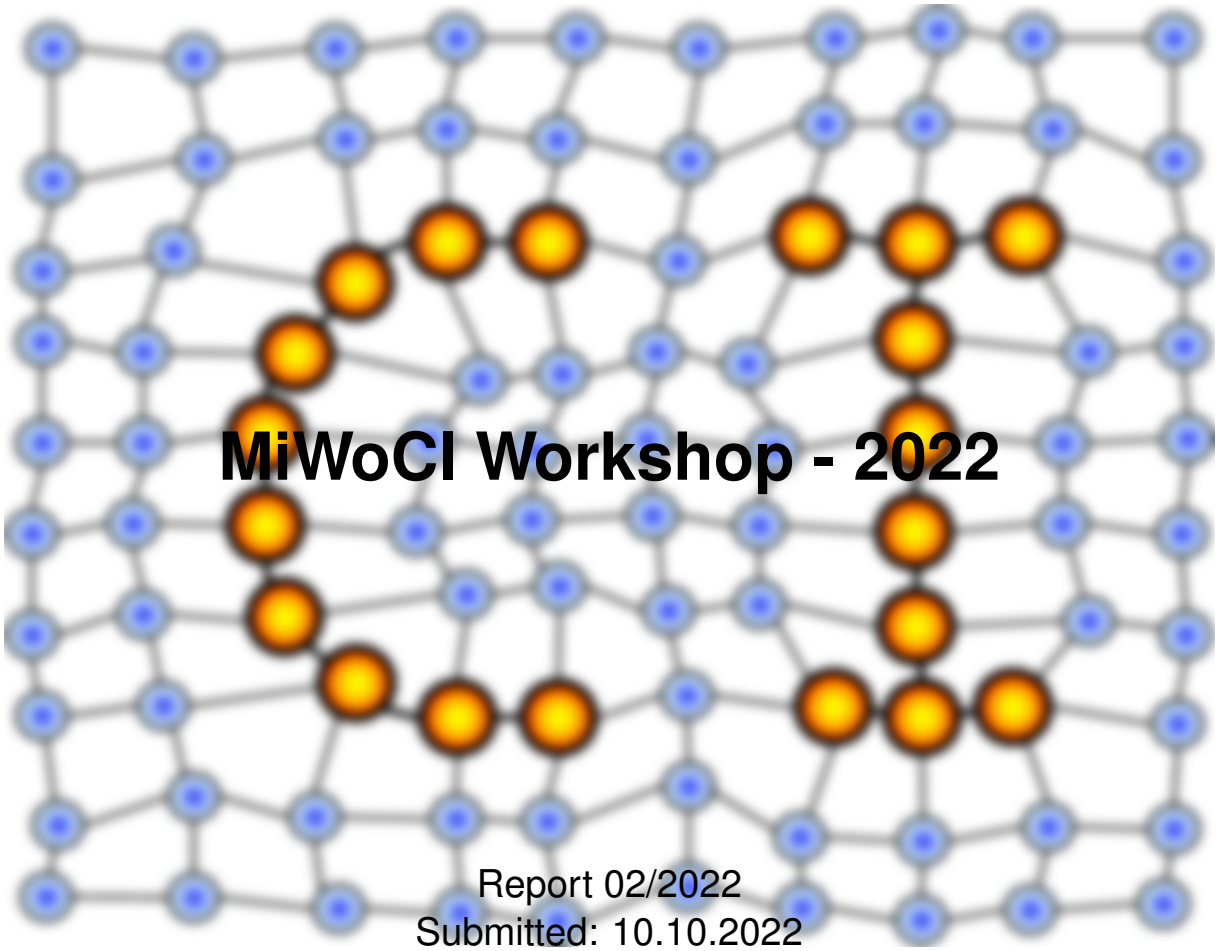
MACHINE LEARNING REPORTS



MiWoCI Workshop - 2022

Report 02/2022
Submitted: 10.10.2022
Published: 24.10.2022

Frank-Michael Schleif[1,2*,3*], Sven Hellbach[4], Marika Kaden[2], Thomas Villmann[2] (Eds.)
(1) University of Applied Sciences Wuerzburg-Schweinfurt, Sanderheinrichsleitenweg 20, 97074 Wuerzburg, Germany (2) University of Applied Sciences Mittweida, Technikumplatz 17, 09648 Mittweida, Germany (3) University of Birmingham, School of Computer Science, Edgbaston, B15 2TT Birmingham, UK
(4) University of Applied Sciences Zwickau, Kornmarkt 1, 08012 Zwickau, Germany

Impressum

Publisher:	University of Applied Sciences Mittweida Technikumplatz 17, 09648 Mittweida, Germany
Editor:	Prof. Dr. Thomas Villmann Prof. Dr. Frank-Michael Schleif
Technical-Editor:	Prof. Dr. Frank-Michael Schleif
Contact:	fschleif@techfak.uni-bielefeld.de
URL:	http://techfak.uni-bielefeld.de/∼ fschleif/mlr/mlr.html
ISSN:	1865-3960

Abstracts of the 14^{th} Mittweida Workshop on Computational Intelligence - MiWoCI 2022 -

Frank-Michael Schleif, Sven Hellbach, Marika Kaden, and Thomas Villmann

Machine Learning Report 02/2022

Preface

The 14 th international *Mittweida Workshop on Computational Intelligence* (MiWoCI) gathering together more than 50 scientists from different universities including Bielefeld, Groningen, UAS Mittweida, UAS Würzburg-Schweinfurt, UAS Zwickau, Dr. Ing. h.c. F. Porsche AG in Weissach and NEC Laboratories Europe, Heidelberg. This year we could again gathering together in Mittweida, Germany. For all who could not attend in person the workshop was hybrid. Thus, from 22.8 - 24.8.2022 the tradition of scientific presentations, vivid discussions, and exchange of novel ideas at the cutting edge of research was continued. They were connected to diverse topics in computer science, automotive industry, and machine learning.

This report is a collection of abstracts and short contributions about the given presentations and discussions, which cover theoretical aspects, applications, as well as strategic developments in the fields.

Contents

1 Neural Architecture Search for fine tuning Transformer transfer learning
(presenter: Philip Kennweg) 4

2 Soft mode in the dynamics of over-realizable on-line learning for SCM
(presenter:Frederieke Richert) 5

3 Sparse Factor Autoencoders for Education
(presenter: Benjamin Paassen) 6

4 The Sensoric Response Principle: Towards Relevance Learning with Biochemical Structured Data
(presenter: Katrin Bohnsack) 7

5 Adaptive Subspace Kernel Fusion - the genuine benefits of MiWoCI boat trips
(presenter: Maximilan Münch) 9

6 Leveraging Identifiability Analysis for Learning in the Model Space
(presenter: Janis Norden) 10

7 Pressure prediction and anomaly detection in Water Distribution Networks using Graph Neural Networks *(presenter: Muhammad Inaam Ashraf)* 11

8 Federated learning with concept drift : a theoretical approach
(presenter: Fabian Hinder) 13

9 Federated learning vector quantization for dealing with drift between nod
(presenter: Valerie Vaquet) 14

10 Handling Feature Drift Using Relevance Learning
(presenter: Moritz Heusinger) 16

11 Adversarial Robustness of GLVQ-models
(presenter: Sascha Saralajew) 17

12 Prototype-based One-Class-Classification Learning using Local Representations
(presenter: Daniel Staps) 18

13 Informed Posterior Construction for Partially Observed Dynamical Systems
(presenter: Elisa Oostwal) 19

14 Regression Neural Gas - Prototypes with Label Function
(presenter: Ronny Schubert) 21

15 Two or three things that have been observed in the context of Deep Learning *(presenter: Michael Biehl)* **23**

16 Insights into confidences and rejects for prototype-based classifiers *(presenter: Johannes Brinkrolf)* **24**

17 Real-time quality control for steel-based mass production *(presenter: Michiel Straat)* **25**

18 Quantum Computing and its Challenges for Prototype-Based Models *(presenter: Alexander Engelsberger)* **26**

Neural Architecture Search for Sentence Classification with BERT

Philip Kenneweg, Sarah Schroeder and Barbara Hammer

Bielefeld University - Faculty of Technology
Inspiration 1, 33615 Bielefeld - Germany

Abstract

Pre training of language models on large text corpora is common practice in Natural Language Processing. Following, fine tuning of these models is performed to achieve the best results on a variety of tasks. In this paper we question the common practice of only adding a single output layer as a classification head on top of the network. We perform an AutoML search to find architectures that outperform the current single layer at only a small compute cost. We validate our classification architecture on a variety of NLP benchmarks from the GLUE dataset. The source code is open-source and free (MIT licensed) software, available at
`https://github.com/TheMody/NASforSentenceEmbeddingHeads`.

Soft mode in the dynamics of over-realizable online learning for soft committee machines

Frederieke Richert

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

Abstract

Overparametrized deep neural networks trained by stochastic gradient descent are successful in performing many tasks of practical relevance. However, the theoretical understanding of the influence of the overparametrization is still insufficient. One possibility to address this problem is in the context of statistical mechanics of learning, where tools from statistical mechanics are employed to describe the average behavior of a model network. A sense of overparametrization can be introduced by the investigation of an over-realizable student-teacher scenario, where the student network has a larger number of hidden units than the teacher network. For online learning of a two-layer soft committee machine in the over-realizable case, evidence is presented that the approach to perfect learning occurs in a power-law fashion. This is different from the realizable case, where an exponential decay was observed in the asymptotic learning regime. The crux in the presented model is appropriately chosen hidden-to-output weights which significantly alter the learning behavior of the student network [1]. For a student hidden layer size which is an integer multiple of the teacher hidden layer size, learning of all student nodes is observed, while they replicate one of the teacher nodes each. Similar cooperative behavior was also shown in the context of using the ReLU activation function in [2] and stands in contrast to earlier investigations [3], where overparametrized student networks reduced their size to the teacher size by truncating the additional hidden units during the training process.

References

[1] Frederieke Richert, Roman Worschech, and Bernd Rosenow. "Soft mode in the dynamics of over-realizable online learning for soft committee machines". In: *Physical Review E*, Vol. 105, No. 5 (May 2022). DOI: 10.1103/PhysRevE.105.L052302.

[2] Michiel Straat, and Michael Biehl. "On-line learning dynamics of ReLU neural networks using statistical physics techniques". 2019. arXiv: 1903.07378.

[3] David Saad, and Sara A. Solla. "On-line learning in soft committee machines". In: *Physical Review E*, Vol. 52, No. 4 (Oct. 1995). DOI: 10.1103/PhysRevE.52.4225.

Sparse Factor Autoencoders for Education

Benjamin Paaßen

German Research Center for Artificial Intelligence (DFKI)

Abstract

When grading assignments, teachers are confronted with at least two questions: How much ability is contained in the students' answers? And, conversely, which ability level would best explain their answers? We formalize these questions as the encoding and decoding steps of an autoencoder. Encoding corresponds to grading the assignment, decoding corresponds to mapping the grades back to answers, which should match the actually observed answers. Finally, we encourage sparsity in the mapping between tasks and dimensions of ability. This yields our proposed model, the sparse factor autoencoder, a fully interpretable model grounded in educational theory. We evaluate the model on synthetic as well as real data, demonstrating its speed in training and prediction, while maintaining high accuracy and plausible parameters. The implementation can be found at `https://github.com/bpaassen/sparfae`.

References

[1] Lan, A., Waters, A., Studer, C., and Baraniuk, R. (2014). Sparse Factor Analysis for Learning and Content Analytics. Journal of Machine Learning Research 15(57): 1959–2008. `http://jmlr.org/papers/v15/lan14a.html`

[2] Paaßen, B., Dywel, M., Fleckenstein, M., and Pinkwart, N. (2022). Sparse Factor Autoencoders for Item Response Theory. Proceedings of the EDM. accepted.

[3] Wu, M., Davis, R., Domingue, B., Piech, C., and Goodman, N. (2020). Variational Item Response Theory: Fast, Accurate, and Expressive. Proceedings of the EDM: 257–268. `https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_22.pdf`

The sensor response principle: Towards relevance learning with biochemical structured data

Katrin Sophie Bohnsack

Saxon Institute of Computational Intelligence and Machine Learning, Hochschule Mittweida, Germany

Abstract

Applications of learning procedures in the biochemical domain frequently require comparison of structured data in the form of graphs. Direct comparison by means of graph kernels has emerged as a promising alternative to traditional structure-derived topological descriptors or fingerprints. However, the choice for one or another kernel narrows the model's view to certain graph characteristics and requires sound domain or task specific knowledge in advance. This shortcoming is accompanied by high computational complexity, making graph kernels unfeasible for huge amounts of data. Inspired by [2], we approach this computational bottleneck by characterizing each graph by it's proximities to a predefined (relatively small set of) reference(s). We interpret the new representation as the graph's sensor response, which may serve as input to learning algorithms, no longer restricted to median or relational variants. Relevance learning on these response vectors by means of Generalized Matrix Learning Vector Quantization [3] constitutes an alternative form of adaptive proximity mixing and is closely related to multiple kernel learning. The presentation will summarize results from [1] and include recent findings regarding the influence of the choice and number of references from the original data space.

References

[1] Katrin Sophie Bohnsack, Marika Kaden, Julius Voigt, and Thomas Villmann. Efficient classification learning of biochemical structured data by means of relevance weighting for sensoric response features. In *ESANN 2022 Proceedings (Accepted)*, 2022.

[2] Elzbieta Pekalska and Robert P W Duin. *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*, volume 64 of *Series in Machine Perception and Artificial Intelligence*. WORLD SCIENTIFIC, November 2005.

[3] Petra Schneider, Michael Biehl, and Barbara Hammer. Adaptive Relevance Matrices in Learning Vector Quantization. *Neural Computation*, 21(12):3532–3561, December 2009.

Sensor Response Principle: Applied to Sequences in Molecular Biology

Julius Voigt

Saxon Institute of Computational Intelligence and Machine Learning, Hochschule Mittweida, Germany

Abstract

A standard procedure for the inclusion of sequence data from molecular biology into machine learning algorithms does not yet exist. A new approach, based on the the sensor response principle (SRP) [1], which Katrin Sophie Bohnsacks presents in her talk. In my talk, I apply the SRP to protein sequences. Thereby the sequence similarities measured by sequence alignment with different substitution matrices. The measurements are then used as input for learning with GMLVQ [2].
The presentation will attempt to describe the method and discuss the relevance of the chosen reference sequences.

References

[1] Katrin Sophie Bohnsack, Marika Kaden, Julius Voigt, and Thomas Villmann. Efficient classification learning of biochemical structured data by means of relevance weighting for sensoric response features. In *ESANN 2022 Proceedings (Accepted)*, 2022.

[2] Petra Schneider, Michael Biehl, and Barbara Hammer. Adaptive Relevance Matrices in Learning Vector Quantization. *Neural Computation, 21(12):3532–3561*, December 2009.

Adaptive Subspace Kernel Fusion
& the genuine benefits of MiWoCI boat trips

Maximilian Münch

University of Applied Sciences Würzburg-Schweinfurt
University of Groningen

Abstract

In current times, data and information analysis have become increasingly challenging. On the one side, expectations on the model's capabilities and on the other side, the complexity of input data are higher than ever. Data is nowadays not only given in one single vectorial format and ready to be processed by high-sophisticated machine learning models but rather in multiple modalities at the same time. Additionally, the data is (often) provided as structured data, such as time series, textual descriptions, or graphs.

For both challenges, kernel methods [1] and similarity-based learning [2] have proven to be highly efficient and robust for structured data analysis. As an extension of classic kernel methods to multiple modalities, so-called *Multiple Kernel Learning* (MKL) has become very popular [3] due to its great performance when it comes to heterogeneous data. In contrast to kernel methods, many approaches from the deep learning domain are highly useful for embedding structured data into a vector space but still struggle with heterogeneous data. In general, the goal of MKL is the derivation of one strong kernel as a combination of multiple weak base kernels in order to further process it by an arbitrary kernel method. Nevertheless, these models are still limited by mathematical constraints of the intended machine learning model such as positive definiteness.

In this talk, we propose a technique that exploits the spectral properties of multiple kernels in order to learn a new representation of the data over multiple modalities. Within this learning procedure, our method limits its results to the mathematical properties of the intended model and hence can learn only a valid representation.

References

[1] John Shawe-Taylor and Nello Cristianini. Kernel Methods for Pattern Analysis. Cambridge University Press, 2004.

[2] Yihua Chen, Eric K. Garcia, Maya R. Gupta, Ali Rahimi, and Luca Cazzanti. Similarity-based classification: Concepts and algorithms. Journal of Machine Learning Research 10, 747-776, 2009.

[3] Fabio Aiolli and Michele Donini. EasyMKL: a scalable multiple kernel learning algorithm. Neurocomputing, 169, 215-224, 2015.

Leveraging Identifiability Analysis for Learning in the Model Space

Janis Norden

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

Abstract

At the core of the talk lies the idea of doing Machine Learning (ML) on time series data. Our work falls into the theory-guided domain of ML (see [1] for an overview). The theory-guided approach aims to combine elements of classical, model-based analysis with more recent techniques of modern ML. Motivated by examples from the biomedical domain, we will consider a scenario in which the available data is sparse and irregularly sampled. Measurement noise is assumed to be present as well. Given the sensitive nature of the biomedical domain, interpretability of any ML outcomes is highly desirable. To this end, we will review a framework, introduced in [2], aiming to accommodate all of the above: Making use of domain expert knowledge in the form of parameterized Ordinary Differential Equations (ODEs), individual time series are represented as particular instances of ODE models. To account for the various ways in which uncertainty about the model parameters enters the picture, a Bayesian approach is adopted and each time series is treated as a posterior distribution over model parameters. These posterior distributions inherit degeneracy from the underlying ODE model whenever the model is not structurally identifiable (see [3] for a recent review on the topic). Incorporating knowledge about structural identifiability can be used to pre-process the training data and make learning more efficient.

References

[1] A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar (2017). Theory-guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Transactions on knowledge and data engineering*, 29(10):2318–2331.

[2] Y. Shen, P. Tino, & K. Tsaneva-Atanasova (2016). A Classification Framework for Partially Observed Dynamical Systems. *Physical Review E*, 95. 10.1103/PhysRevE.95.043303.

[3] F. Anstett-Collin, L. Denis-Vidal, G. Millérioux (2020). A priori identifiability: An overview on definitions and approaches. *Annual Reviews in Control*, Volume 50, Pages 139-149.

Pressure prediction and anomaly detection in Water Distribution Networks using Graph Neural Networks

Inaam, Ashraf

Bielefeld University, CITEC, Germany

Abstract

Water is a vital resource and thus optimizing its usage and distribution is pivotal to the welfare of people in specific and to the environment in general. European Research Council's Water Futures Project is aiming to achieve this objective [2]. As part of the project, we employ Machine Learning methods to build smart water systems. Specifically, Deep Learning tools like Graph Convolutional Neural Networks (GCN) [1] present a promising proposition for modeling the Water Distribution Networks (WDN). Given the limited availability of pressure sensors (and thus water levels) in a WDN, GCNs can be used to predict pressure values throughout the network. To achieve that, a WDN is represented as a graph with highly sparse connections and even more sparse node features. We present encouraging results using a GCN based model for a reasonably sized simulated WDN. The model uses a very small number of sensor values (4.2%) to predict pressures at every node in the network. These predictions are further utilized for the downstream task of anomaly detection. Two major types of anomalies in WDNs are sensor faults and leakages. We apply our GCN based model to detect anomalies with a high success rate. We also demonstrate that the neighborhood aggregation approach of GCNs helps the model in localizing the anomalies to a considerable extent.

References

[1] Guohao Li et al. *DeeperGCN: All You Need to Train Deeper GCNs*. 2020. DOI: 10.48550/ARXIV.2006.07739. URL: https://arxiv.org/abs/2006.07739.

[2] *Smart Water Futures Designing the next generation of urban drinking water systems*. 2022. URL: https://waterfutures.eu/.

Federated Learning with Conecpt Drift: A Theretical Approach

Fabian Hinder, Valerie Vaquet,
Johannes Brinkrolf, and Barbara Hammer

Bielefeld University - Cognitive Interaction Technology (CITEC)
Bielefeld - Germany

Abstract

When data is collected from multiple locations (or nodes), problems with the amount of data to be transferred or with privacy can render classic machine learning algorithms useless. In this case federated learning approaches where models are trained on each node separately, i.e. on the locally collected datasets, and then are fused to obtain a global model offer a solution. One challenge in federated learning, in contrast to distributed learning, is that one usually does not assume that the locally collected observations follow the same distribution. Typical problems are the cases where some classes or features are not available at all nodes. One problem that is common in the real world but not considered in the classical spectrum of federated learning is concept drift, i.e., the conditional distribution of the label is not the same for all nodes. This makes fusing the models in a meaningful way significantly more difficult. In this work we focus on the problem of federated learning with concept drift from a theoretical point of view: We provide a formal framework that generalizes the concepts of classical learning theory to be applicable to the problem of federated learning. Based thereon we derive a general solution for federated learning with concept drift as generalized learning problem. We conclude our considerations by applying our ideas to several classical learning models.

Federated learning vector quantization for dealing with drift between sensor nodes

Valerie Vaquet, Fabian Hinder,
Johannes Brinkrolf, and Barbara Hammer

Bielefeld University, CITEC, Germany

Abstract

The increasing availability of sensor technologies enables the use of machine learning techniques in many applications. Frequently, the systems outperform humans as the sensors are more accurate or can perceive patterns humans cannot. However, when designing machine learning based systems which are learning from sensor data one needs to account for two challenges. First, sensors frequently have their own characteristics due to the properties of the physical components. This induces shifts between the measurements collected by similar sensors which hinder the transfer of a model trained on the data collected by one sensor to that measured by another similar one.

A second issue is that one cannot assume that the sensors are positioned in the same physical location and thus might rely on different data sets for the training procedure. Considering for example different production facilities, sensors might get different samples with ground truth measurements across the locations. The challenges introduced by this setup are usually addressed by federated learning. Hinder et al. [2] propose a framework for federated learning that additionally account for sensor shifts.

In this contribution [3], we propose an adaption of federated LVQ [1] following the framework by Hinder et al. [2]. We evaluate the method on hyper-spectral measurements which are obtained by different sensors with sensor shift [4].

References

[1] Johannes Brinkrolf and Barbara Hammer. “Federated Learning Vector Quantization”. In: *29th European Symposium on Artificial Neural*

Networks, Computational Intelligence and Machine Learning, ESANN 2021, Online event (Bruges, Belgium), October 6-8, 2021. 2021. DOI: 10.14428/esann/2021.ES2021-141. URL: https://doi.org/10.14428/esann/2021.ES2021-141.

[2] Fabian Hinder et al. "Federated Learning with Conecpt Drift: A Theretical Approach". In: *Machine Learning Reports 2022*.

[3] Valerie Vaquet et al. "Federated learning vector quantization for dealing with drift between nodes". In: Bruges, 2022. URL: https://pub.uni-bielefeld.de/record/2964534.

[4] Valerie Vaquet et al. "Investigating Intensity and Transversal Drift in Hyperspectral Imaging Data". In: *29th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2021, Online event (Bruges, Belgium), October 6-8, 2021*. 2021. DOI: 10.14428/esann/2021.ES2021-64. URL: https://doi.org/10.14428/esann/2021.ES2021-64.

Handling Feature Drift Using Relevance Learning

Moritz Heusinger

University of Applied Sciences Würzburg - Schweinfurt

Abstract

Online learning is a challenging task for predictive models where the problem of feature drift is an important but still widely unaddressed issue. Feature drift occurs when subsets of features become relevant or irrelevant during the continuous learning task. We investigate how the concept of Relevance Learning can be used to tackle the problems in non-stationary environments. Specifically, we focus on Concept Drift and Feature Drift, which both lead to changes in the data distribution over time.

Adversarial Robustness of Nearest Prototype Classifiers

Sascha Saralajew

NEC Laboratories Europe GmbH, Heidelberg, Germany

Abstract

In the last years, the performance of neural networks has constantly improved, leading to models that show superhuman abilities. However, the task of training adversarially robust machine learning models is not entirely solved. Recently, we showed that when training a seminorm-based Nearest Prototype Classifier (NPC) with the well-known generalized learning vector quantization loss function, the method becomes adversarially robust with respect to the used seminorm. In comparison to robustified neural networks, the trained NPCs showed comparable robustness with the advantage of less training and evaluation complexity. In this talk, we summarize results about the adversarial robustness of NPCs and discuss recently published improvements that show that NPCs are state of the art in terms of adversarial robustness, even outperforming robustified neural networks.

Prototype-based One-Class-Classification Learning using Local Representations

Daniel Staps

Saxon Institute of Computational Intelligence and Machine Learning,
Hochschule Mittweida, Germany

Abstract

In this talk a prototype based model for one-class-classification problems, remaining an important challenge in machine learning, is presented. The cost function includes representation learning aspects as well as the evaluate of the one-class-classification appropriately. The prototype-based model ensures a local representation of the target class. Following this guidance, an interpretable one-class-classifier model is obtained.

Informed Posterior Construction for Partially Observed Dynamical Systems

Elisa Oostwal

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

Abstract

An important task that can be performed with machine learning algorithms is classification of input data, e.g. of time-series. These may be irregularly or sparsely sampled and may contain noise. These limitations in the amount and quality of data make traditional statistical classification difficult. However, if the process that produces the time-series data can be explained by a dynamical system, then a mechanistic model can be introduced to incorporate domain knowledge, effectively guiding the classification process [4]. In this context, time-series data can be considered as partial observations of the underlying dynamical system and the machine learning task becomes *classification of partially observed dynamical systems.*

A framework which is particularly suited for this job is *Learning in the Model Space* (LiMS) [2]. Instead of performing classification on the time-series directly, a *posterior* is constructed for every time-series data, which essentially quantifies the level of belief for every possible parameterization (realization) of the given mechanistic system [6]. The obtained classes can in turn be used to distinguish different clusters and, by making use of the interpretable mechanistic model, explain the differences between groups. LiMS also effectively takes into account the variation in model parameters and uncertainty due to observational noise. Examples of applications of this approach can be found in [3], [1], [5].

The main difficulty of this Bayesian approach is obtaining the posterior distribution, which is typically not analytically tractable. However, several approximation methods exist such as a finite-grid approximation and sampling. The goal of our project is to establish an efficient sampling algorithm by making use of the knowledge of the underlying system which produced the time-series data. The idea would be to exploit the unidentifiabilities inherent to the dynamical system. In what ways we will be able to incorporate and use this information remains one of the open questions.

References

[1] Kerstin Bunte, David Smith, Michael Chappell, Zaki Hassan-Smith, Jeremy Tomlinson, Wiebke Arlt, and Peter Tino. “Learning pharmacokinetic models for in vivo glucocorticoid activation.” In: *Journal of Theoretical Biology*, Vol. 455, No. 7 (Jul. 2018). DOI: 10.1016/j.jtbi.2018.07.02.

[2] Huanhuan Chen, Peter Tino, Ali Rodan, and Xin Yao. "Learning in the Model Space for Fault Diagnosis". In: *IEEE Transactions on Neural Networks* (Mar. 2013). DOI: 10.1109/TNNLS.2013.2256797.

[3] Mark Girolami. "Bayesian inference for differential equations." In: *Theoretical Computer Science*, Vol. 408, No. 1 (Nov. 2008), p. 4-16. DOI: 10.1016/j.tcs.2008.07.005.

[4] Anuj Karpatne, Gowtham Atluri, James H. Faghmous, Michael S. Steinbach,Arindam Banerjee, Auroop R. Ganguly, Shashi Shekhar, Nagiza F. Samatova,and Vipin Kumar. "Theory-guided Data Science: A New Paradigm for Scientific Discovery". In: IEEE Transactions On Knowledge And Data Engineering, Vol. 29, No. 10 (Oct. 2017). DOI: 10.1109/TKDE.2017.2720168.

[5] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. "Machine learning of linear differential equations using gaussian processes." In: *Journal of Computational Physics*, Vol. 348 (Nov. 2017), p. 683-693. DOI: 10.1016/j.jcp.2017.07.050.

[6] Yuan Shen, Peter Tino, and Krasimira Tsaneva-Atanasova. "A Classification Framework for Partially Observed Dynamical Systems". In: *Phys. Rev. E* Vol. 95, No. 4 (Apr. 2017). DOI: 10.1103/PhysRevE.95.043303.

Regression Neural Gas - Prototypes with adaptable Label Functions

Ronny Schubert

Saxon Institute of Computational Intelligence and Machine Learning (SICIM), Hochschule Mittweida, Germany

August 14, 2022

Abstract

This work shall serve as a short survey on interpretable prototype-based models with adaptable label functions. Such models are considered to be a generalization to the case of discrete class functions for supervised machine learning tasks in the prototypical framework. The considered models are among others the *RBF-Network* [1] and its generalized version [2], the *Neural Gas* Algorithm for time-series prediction [3] and variants like [4, 5], as well as the *Regression Learning Vector Quantization* [6]. Furthermore, we shall briefly discuss obstacles which need to be taken into account when developing the architecture of a corresponding model. Conclusively, we provide an extension of the *Fuzzy-Labeled Neural Gas* Algorithm [4] for approximation, which we call *Regression Neural Gas*.

References

[1] John Moody, Christian J. Darken; Fast Learning in Networks of Locally-Tuned Processing Units. Neural Comput 1989; 1 (2): 281-294. doi: https://doi.org/10.1162/neco.1989.1.2.281.

[2] T. Poggio and F. Girosi, A Theory of Networks for approximation and learning. Report Number AI-1140. MIT Artificial Intelligence Laboratory, Cambridge, MA., July 1989.

[3] T. M. Martinetz, S. G. Berkovich and K. J. Schulten, "'Neural-gas' network for vector quantization and its application to time-series prediction," in IEEE Transactions on Neural Networks, vol. 4, no. 4, pp. 558-569, July 1993, doi: 10.1109/72.238311.

[4] Th. Villmann, B. Hammer, F. Schleif, T. Geweniger, W. Herrmann, Fuzzy classification by fuzzy labeled neural gas, Neural Networks, Volume 19, Issues 6-7, 2006, Pages 772-779, ISSN 0893-6080, https://doi.org/10.1016/j.neunet.2006.05.026.

[5] Marika Kaden. Integration of Auxiliary Data Knowledge in Prototype Based Vector Quantization and Classification Models. PhD thesis, 06 2016.

[6] M. Grbovic and S. Vucetic, "Regression Learning Vector Quantization," 2009 Ninth IEEE International Conference on Data Mining, 2009, pp. 788-793, doi: 10.1109/ICDM.2009.145.

Two or three things that have been observed in the context of Deep Learning

Michael Biehl

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

Abstract

Several interesting phenomena have been discovered or re-discovered rrecently in the context of Deep Learning. In this presentation I will briefly introduce and discuss learning curves characterized by the observation of *generalization beyond over-fitting*. The focus will be on learning curves which show a so-called *double descent* behavior [1] or display delayed generalization in a phenomenon that has been termed *grokking* [2].

References

[1] M. Looga,b, T. Vieringa, A. Meya, J.H. Krijthea, D.M. J. Tax. *A brief prehistory of double descent.* PNAS, Vol. 117, no. 20 , 10625-10626 (2020)

[2] A. Power, Y. Burda, H. Edwards, I. Babuschkin. *GROKKING: generalization beyond overfitting on small algorithmic datasets.* arXiv:2201.02177v1 [cs.LG] (2022)

Confidences and Rejects for Learning Vector Quantization

Johannes Brinkrolf

Bielefeld University, CITEC, Germany

Abstract

Learning Vector Quantization (LVQ) enjoys a great popularity as efficient and intuitive classification scheme, accompanied by a strong mathematical substantiation of its learning dynamics and generalization ability [1, 2]. However, popular deterministic LVQ variants do not allow an immediate probabilistic interpretation of its output and an according reject option in case of insecure classifications. This talk will treat two main topics.

First, we will introduce reject options and give deeper insights of the well-known one for LVQ proposed by Fisher et al. [3] dubbed *relative similarity*. The focus is on the strict boundaries and the classifier's margin. Secondly, we investigate how to extend and integrate pairwise LVQ schemes to an overall probabilistic output, and we compare the benefits and drawbacks of this proposal to a recent heuristic surrogate measure for the security of the classification, which is directly based on the LVQ classification scheme. Experimental results indicate that an explicit probabilistic treatment often yields superior results as compared to a standard deterministic LVQ method, but metric learning is able to annul this difference.

References

[1] Michael Biehl, Anarta Ghosh, and Barbara Hammer. "Dynamics and Generalization Ability of LVQ Algorithms". In: *Journal of Machine Learning Research* 8 (2007), pp. 323–360. URL: http://dl.acm.org/citation.cfm?id=1314511.

[2] Michael Biehl et al. "Stationarity of Matrix Relevance LVQ". In: *2015 International Joint Conference on Neural Networks, IJCNN 2015, Killarney, Ireland, July 12-17, 2015*. 2015, pp. 1–8. DOI: 10.1109/IJCNN.2015.7280441. URL: https://doi.org/10.1109/IJCNN.2015.7280441.

[3] Lydia Fischer, Barbara Hammer, and Heiko Wersing. "Efficient rejection strategies for prototype-based classification". In: *Neurocomputing* 169 (2015), pp. 334–342. DOI: 10.1016/j.neucom.2014.10.092. URL: http://dx.doi.org/10.1016/j.neucom.2014.10.092.

Real-time quality control for steel-based mass production using Machine Learning on non-invasive sensor data

Michiel Straat

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

Abstract

Insufficient steel quality in mass production can cause extremely costly damage to tooling, production downtimes and low quality products. Automatic, fast and cheap strategies to estimate essential material properties for quality control, risk mitigation and the prediction of faults are highly desirable. In this work we analyse a high throughput production line of steel-based products. Currently, the material quality is checked using manual destructive testing, which is slow, wasteful and covers only a tiny fraction of the material. To achieve complete testing coverage our industrial collaborator developed a contactless, non-invasive, electromagnetic sensor to measure all material during production in real-time. Our contribution is three-fold: 1) We show in a controlled experiment that the sensor can distinguish steel with deliberately altered properties. 2) During several months of production 48 steel coils were fully measured non-invasively and additional destructive tests were conducted on samples taken from them to serve as ground truth. A linear model is fitted to predict from the non-invasive measurements two key material properties (yield strength and tensile strength) that normally have to be obtained by destructive tests. The performance is evaluated in leave-one-coil-out cross-validation. 3) The resulting model is used to analyse the material properties and the relationship with reported product faults on real production data of approximately 108 km of processed material measured with the non-invasive sensor. The model achieves an excellent performance (F3-score of 0.95) predicting material running out of specifications for the tensile strength. The combination of model predictions and logged product faults shows that if a significant percentage of estimated yield stress values is out of specification, the risk of product faults is high. Our analysis demonstrates promising directions for real-time quality control, risk monitoring and fault detection.

Towards full quantum prototype learning

Alexander Engelsberger

Saxon Institute of Computational Intelligence and Machine Learning,
Hochschule Mittweida, Germany

Abstract

Prototype learning algorithms like Generalized Learning Vector Quantization (GLVQ) are put together by steps where benefits from quantum computing could be expected. Dense coding [6] and fidelity estimation routines [2] allows the calculation of high dimensional dissimilarities with logrithmically growing number of qubits. Amplitude amplification principles allow the design of efficient search algorithms, like Grover's search [3], for winner determination.

We solved a missing problem, that is distinctive for prototype learning, by introducing a quantum schema for the attraction-repulsion-scheme [1]. Adding the idea of quantum sorting introduced in [5] all heuristic prototype learning steps are available in the same quantum formalism, the so-called amplitude encoding. Making the transfer of heuristic schemes, like LVQ 1 [4] possible.

If intermediate measurements can be avoided, a complete quantum circuit solving the prototype learning algorithm can be constructed, paving the way to experimental setups on cloud quantum computer systems.

References

[1] A. Engelsberger, R. Schubert, and T. Villmann. Steps Forward to Quantum Learning Vector Quantization for Classification Learning on a Theoretical Quantum Computer. In *Advances in Self-Organizing Maps, Learning Vector Quantization, Clustering and Data Visualization*, Prague, 2022. Springer Nature Switzerland AG.

[2] X. Gitiaux, I. Morris, M. Emelianenko, and M. Tian. SWAP Test for an Arbitrary Number of Quantum States. 2021.

[3] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing - STOC '96*, pages 212–219, Philadelphia, Pennsylvania, United States, 1996. ACM Press.

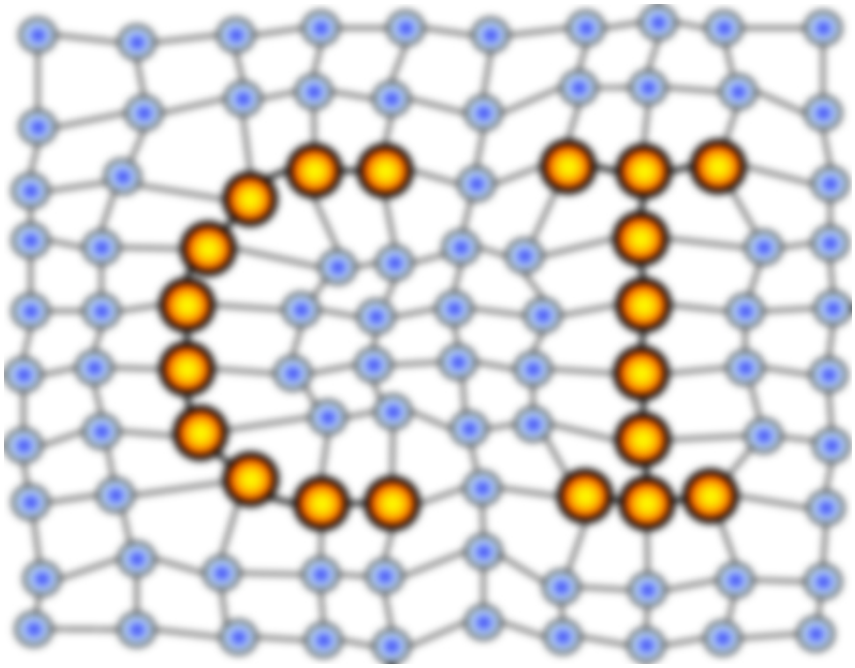
[4] T. Kohonen. Learning vector quantization. *Neural Networks*, 1:303, Jan. 1988.

[5] L. F. Quezada, G.-H. Sun, and S.-H. Dong. Quantum Version of the k-NN Classifier Based on a Quantum Sorting Algorithm. *Annalen der Physik*, 534(5):2100449, May 2022.

[6] M. Weigold, J. Barzen, F. Leymann, and M. Salm. Encoding patterns for quantum algorithms. *IET Quantum Communication*, 2(4):141–152, Dec. 2021.

MACHINE LEARNING REPORTS

Report 02/2022



Impressum

Machine Learning Reports ISSN: 1865-3960

▽ Publisher/Editors

Prof. Dr. rer. nat. Thomas Villmann
University of Applied Sciences Mittweida
Technikumplatz 17, 09648 Mittweida, Germany
• http://www.mni.hs-mittweida.de/

Prof. Dr. rer. nat. Frank-Michael Schleif
University of Bielefeld
Universitätsstrasse 21-23, 33615 Bielefeld, Germany
• http://www.cit-ec.de/tcs/about

▽ Copyright & Licence

Copyright of the articles remains to the authors.

▽ Acknowledgments

We would like to thank the reviewers for their time and patience.