ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Domain Knowledge Integration in Machine Learning Systems – An Introduction

Marika Kaden[1], Sascha Saralajew[2], and Thomas Villmann[1] *

1- Saxon Institute for Computational Intelligence and Machine Learning (SICIM),
Mittweida University of Applied Sciences, Mittweida, Germany
[kaden1|villmann]@hs-mittweida.de

2- NEC Laboratories Europe GmbH, Heidelberg, Germany
Sascha.Saralajew@neclab.eu

Abstract. Knowledge integration into machine learning systems is a promising and successful strategy to achieve more plausible and consistent results. The plausibility is accompanied by better model interpretability due to the adjustment of the machine learning system to the domain specific requirements and restrictions. Further, informed machine learning can be seen as a particular task specific regularization of the model leading to better learning convergence and frequently also requiring a lower amount of training data. This short introduction paper addresses some recent aspects, how domain knowledge can be integrated into learning systems on different levels ranging from informed feature extraction to domain adjusted structure and model architecture.

1 Introduction

Machine learning currently is dominated by deep neural network architectures (DNN), which offer great flexibility and frequently yield superior performance [14]. This dominance has lead to an overwhelming number of successful applications in a broad variety of technical areas ranging from image and text processing and analysis, feature based data investigation and sequence analysis to the evaluation of structured data like graphs or general proximity relational data. The flexibility of DNN can be primarily attributed to the large model complexity [3]. Thus DNN are mainly used in unsupervised representation learning and coding as well as in supervised scenarios, i.e. regression and classification learning.

Yet, training of deep models usually requires huge training data sets and, hence, also need long training time. Further, due to the great model complexity the challenge to avoid local minima of the loss function is non-trivial [1, 5, 17]. To tackle this problem several regularization techniques are favored [3]. Further, as pointed out in [7], stable learning contributes to causal inference whereby stability maybe achieved augmenting the database by additional information.

Another possibility to deal with those difficulties is to integrate additional knowledge available about the data into the data handling by the machine

*M.K. is supported by the IAI-XPRESS and the DAIMLER project which both are parts of the interdisciplinarry project Artificial Intelligence Meets Space (AIMS, No. 50WK2270E) funded by the German Space Agency (DLR).

learning model. Further, the model structure can be adjusted to reflect prior knowledge regarding the application domain [44, 43]. Otherwise, integration of external domain knowledge also may support advanced model explanations after training and predictions in the application phase. Hence, knowledge integration helps to achieve interpretable machine learning systems as demanded more and more for successful artificial intelligence application in many areas [21], e.g. technical or medical AI-supported systems. Hence, informed machine learning can be seen as learning with hybrid information consisting of both the essential data and knowledge sources determining processing pipeline in contrast to pure data-driven machine learning. Generally, knowledge integration contributes to model acceptance and trustworthiness of the machine learning system and regarding explanations [10, 16, 32].

In this contribution we will highlight several strategies to incorporate domain knowledge into machine learning systems as they are used in several application areas like medical informed machine learning, user-centric explainability of machine learning approaches in healthcare as well as in physics, engineering and bioinformatics or other fascinating application areas [36, 37, 23, 25, 20, 15, 12].

2 How to integrate knowledge into machine learning systems

Traditional machine learning models frequently lack awareness of the intrinsic structure between data attributes, leading to decisions based on confounding variables, improper relationships, or latent variables without physical interpretation. The integration of domain knowledge into machine learning systems helps to avoid/reduce these effects and, therefore, supports scientific discoveries and data appropriate processing can be realized on various levels and from different perspectives. Yet, it is not always clear how and where domain knowledge can be employed adequately and to what extent this integration contributes to improved performance as well as leading to better explanations derived from it. A general framework is depicted in Fig. 1. In the following we will highlight some conceptual aspects and approaches for knowledge incorporation into machine learning models without any claim of completeness.

Feature extraction and data comparison The specific structure of the data samples to be handled by the machine learning system determines an appropriate processing. Depending on the task and additional knowledge about the given database an informed feature extraction may be applicable. For example, in image processing the application of *scale-invariant feature transforms* (SIFT, [22]) or *speeded up robust features* (SURF, [2]) are popular as well as Fourier and other integral transformations. Further, it turns out that specific proximity measures frequently are more suitable for particular data structures, e.g. structural similarity index for image comparison [4], graph-kernel based distances for graph-structure data [41], or functional norms like Sobolev-norms for time-

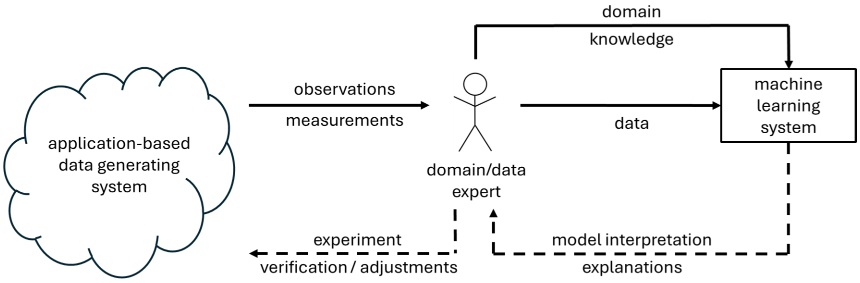

Fig. 1: Data analysis by an informed machine learning system taking a domain expert in the loop: Additional domain knowledge is provided to the model by the expert for adjustment and training. The informed system supports improved explanations and allow extended evaluations. Scheme adapted from [8].

series [40, 29, 19, 33] whereas divergences are favored to evaluate the dissimilarity between probabilistic vectors [39]. Kernels, as another important ingredient for various machine learning models also can benefit from domain specific information as explained in [24].

Thus, the selection of a suitable proximity measure or an adequate feature extraction has to be in accordance with the given data types and machine learning task. Thereby, attention has to be given to keep or better to enhance the task relevant information of the data.

Task dependent loss functions The choice of a loss function for the machine learning system matching the requirements of the model and, at the same time, the particularities of the data environment and the learning task is crucial for the performance of the trained model as well as for the learning process. Informed loss function explicitly favor or penalize task solutions which are in agreement with the prior domain knowledge. Hence, the domain knowledge plays the role of additional constraints for the general complex learning system contributing to a task specific regularization [38]. As a simple example, the choice of sum-squared-error loss or cross-entropy-loss based on divergences is triggered by the task type (regression or classification). In Support Vector Machines (SVM), prior information can be integrated into the constraints regarding the loss function [18, 35]. An example in medical application area is presented [26], where the survival time of tumor patients has to be predicted but only sparse database is available. The resulting regression model is obtained by convex optimization where the tumor grading and other correlations determine the constraints for the optimization to compensate the low data availability.

One of the most prominent examples for loss-constraint knowledge integration is machine learning application for dynamic systems in physics as proposed in [15]. Here, the differential equations of the dynamical system representing the physical laws act as limiting constraints for admissible solutions and, finally, a training data driven solution of the differential equation system is synthesized. Similar approaches for other physics or related engineering application areas are currently under consideration, for example [6, 34] in this ESANN 2024 proceedings. Those dynamical systems related knowledge integration also contributes to advanced machine learning systems for epidemiology research [30].

Thus, scientifically plausible and consistent results are enforced by informed determination of adequate loss functions. Doing so, frequently less training data are required compared to standard approaches with non-specific optimization [45]. Further, the generalization ability usually is improved by those informed loss-based regularization techniques due to the resulting overall variability restriction.

Domain specific model architectures Modern machine learning models are dominated by deep architectures, which usually offer best flexibility and superior performance. Yet, the internal decision paths inside the model architecture are often impossible to interpret. Further, those deep architectures usually require huge databases for training. Here, knowledge integration may help to deal with these insufficiences and can be understood as a kind of semantic-based regularization for learning and inference [11]. Usually, this semantic domain knowledge is give as provided relations between (parts of) data structures, ontologies or plausibility and probability constraints regarding the expected solutions. These information frequently are given as algebraic equations or relational data (graphs). Yet, other types of augmented information, e.g. forbidden solution areas due to ethical or other aspects.

Beside the above mentioned resulting loss constraints or modifications, the additional knowledge can be used to adjust the model architecture. Respective approaches are promising attempts in bio-medical areas as well as in engineering, if complex tasks with maybe complicate data structures have to be solved and/or limited data are available:

- *computational neuroscience*: The understanding of complex brain functions requires more realistic neuron models than perceptron [27], which are usually the basis of deep networks. Further, neuroanatomical properties including areal structure and local and long-range connectivity has to be reflected by a modeling artificial neural network for real cognitive brain function analysis.

- *medicine*: As pointed out in the beginning of this section, traditional approaches do not take into account the intrinsic structure between attributes and, hence, leading to decisions based on confounding variables or latent features without physical interpretation which could cause disastrous decisions in medicine. Therefore, known dependencies between data item have to be reflected by the design of the network structure, i.e. the connectivity graph within the neural network model has to mirror the known possible but not necessarily observable influences and relations [20, 36]. As an example we refer to cancer research an respective biologically-informed networks [46].

- *bioinformatics*: As in medicine, here reliable interpretability of biology-inspired deep neural networks have to adopt data available knowledge about hierarchies and dependencies to specify the neural network structure (number of layers and available trainable connections between them) [13]. In gene expression analysis, this knowledge could consist of discovered and verified pathways es well as observed regulatory processes between genes on different process levels [12]. Alternatively, those structural knowledge can be used to find an interpretable embedding scheme for low-dimensional classification learning by dependency graph matrix decomposition as explained in [42].

- *engineering*: Beside the physics-informed approaches, hybrid modeling in engineering combine a mechanistic simulation with a machine learning model to produce a more realistic behavior than considering both aspects independently [23]. In sensor fusion, several sensor can be combined based on a relation graph describing their technical dependencies [47]. Accordingly, the sensoric outputs established as possible node connections in a network for informed sensor data processing.

In general, domain specific adaptation of general neural network architectures lead to better interpretability of the model outcome. In particular, the trained model allows to analyze, which information contributes most to the model prediction and which other information is neglected.

3 Conclusion

Knowledge integration into machine learning systems is an increasingly important aspect to to adjust models to the given task domain. This integration can be on different levels including the general model architecture and used building blocks, task specific feature extraction and combination of sensoric information as well as problem specific mathematical modeling of the objective to be learned. This paradigm frequently contributes to achieve better perfomances and more plausible solutions in terms of the application domain. Further, informed machine learning leads to better output interpretability and evaluation and, hence, qualifies for general model explanations. Further, for post-hoc checks, where the scientific plausibility and consistency of the results is checked and possibly invalid results are removed, the domain knowledge can be used to evaluate those decisions enforcing knowledge consistent results.

Technically, integration of expert knowledge usually constitutes a regularization of the model and the solution space. Another advantage beside performance and interpretability of this strategy is that the domain specific regularization frequently yields better convergence during learning as well as usually the requirement regarding the amount of available training data is drastically reduced. Thereby, knowledge informed learning can be used if the model is learned from the scratch or a pre-trained model has to be adjusted for a specific application.

In consequence, more sparse machine learning systems are obtained with reduced data requirements and, hence, lower energy consumption during model training. In this sense, informed machine learning can contribute to achieve more sustainable AI systems [9, 28, 31].

References

[1] P. Baldi. *Deep Learning in Science.* Cambridge University Press, 2021.

[2] H. Bay, T. Tuytelaars, and L. van Goot. Speeded up robust features (SURF). *Computer Vision and Image Understanding*, 10(3):346–359, 2008.

[3] C. Bishop and H. Bishop. *Deep Learning – Foundations and Concepts.* Springer International Publishing, 2024.

[4] D. Brunet, E. Vrscay, and Z. Wang. On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing*, 21(4):1488–1499, 2012.

[5] M. Colbrook, V. Antun, and A. Hansen. The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem. *Proceedings of the national Academy of Science (PNAS)*, 119(12):1–10, 2022.

[6] N. Costa, F. Barros, J. Lima, and A. Restivo. Leveraging physics-informed neural networks as solar wind forecasting models. In M. Verleysen, editor, *Proceedings of the 32nd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN'2024), Bruges (Belgium)*, page in this volume, Louvain-La-Neuve, Belgium, 2024. i6doc.com.

[7] P. Cui and S. Athey. Stable learning establishes some common ground between causal inference and machine learning. *Nature Machine Intelligence*, 4:110–115, 2022.

[8] T. Dash, S. Chitlangia, A. Ahuja, and A. Srinivasan. A review of some techniques for inclusion of domain-knowledge into deep neural networks. *Nature Scientifi Reports*, 12(1040):1–15, 2022.

[9] P. Dhar. The carbon impact of artificial intelligence. *nature machine Intelligence*, pages 423–425, 2020.

[10] M. Dikmen and C. Burns. The effects of domain knowledge on trust in explainable AI and task performance: A case of peer-to-peer lending. *International Journal of Human - Computer Studies*, 162(102792):1–11, 2022.

[11] M. Diligenti, M. Gori, and C. Saccà. Semantic-based regularization for learning and inference. *Artificial Intelligence*, 244:143–165, 2017.

[12] H. Elmarakeby, J. Hwang, R. Arafeh, J. Crowdis, S. Gang, D. Liu, S. AlDubayan, K. Salari, S. Kregel, C. Richter, T. Arnoff, J. Park, W. Hahn, and E. Van Allen. Biologically informed deep neural network for prostate cancer discovery. *Nature*, 598:348–352, 2021.

[13] W. Esser-Skala and N. Fortelny. Reliable interpretability of biology-inspired deep neural networks. *NPJ Systems Biology and Applications*, 9(50):1–8, 2023.

[14] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning.* MIT Press, Cambridge, MA, 2016.

[15] G. Karniadakis, I. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. Physics-informed machine learning. *Nature Reviews Physics*, 2:422–440, 2021.

[16] A. Karpatne, G.Atluri, J. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, 29(10):2318–2331, 2017.

[17] K. Kawaguchi, J. Huang, and L. Kaelbling. Effect of depth and width on local minima in deep learning. *Neural Computation*, 31:1462–1498, 2019.

[18] F. Lauer and G. Bloch. Incorporating prior knowledge in support vector machines for classification: A review. *Neurocomputing*, 71(7-9):1578–1594, 2008.

[19] J. Lee and M. Verleysen. Generalization of the l_p norm for time series and its application

to self-organizing maps. In M. Cottrell, editor, *Proc. of Workshop on Self-Organizing Maps (WSOM) 2005*, pages 733–740, Paris, Sorbonne, 2005.

[20] F. Leiser, S. Rank, M. Schmidt-Kräpelin, S. Thiebes, and A. Synyaev. Medical informed machine learning: A scoping review and future research directions. *Artificial Intelligence in Medicine*, 145(102676):1–11, 2023.

[21] P. Lisboa, S. Saralajew, A. Vellido, R. Fernández-Domenech, and T. Villmann. The coming of age of interpretable and explainable machine learning models. *Neurocomputing*, 535:25–39, 2023.

[22] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[23] C. Mackay and D. Nowell. Informed machine learning methods for application in engineering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 237(24):5801–5818, 2023.

[24] O. Mangasarian, J. Shavlik, and E. Wild. Knowledge-based kernel approximation. *Journal of Machine Learning Research*, 5:1127–1141, 2004.

[25] L. Oberste and A. Heinzl. User-centric explainability in healthcare: A knowledge-level perspective of informed machine learning. *IEEE Transactions on Artificial Intelligence*, 4:840–857, 2023.

[26] B. Paaßen, N. Gaisa, M. Rose, and M.-S. Bösherz. Tumor grading via decorrelated sparse survival regression. In M. Verleysen, editor, *Proceedings of the 32nd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN'2024), Bruges (Belgium)*, page in this volume, Louvain-La-Neuve, Belgium, 2024. i6doc.com.

[27] F. Pulvermüller, R. T. andM.R. Henningsen-Schomers, and T. Wennekers. Biological constraints on neural network models of cognitive function. *Nature Reviews Neuroscience*, 22(8):488–502, 2021.

[28] K. Rafat, S. Islam, A. Al Mahfug, M. Ismail Hossain, F. Rahman, S. Momen, S. Rahman, and N. Mohammed. Mitigating carbon footprint for knowledge distillation based deep learning model compression. *Plos One*, 18(5: e0285668):1–22, 2023.

[29] J. Ramsay and B. Silverman. *Functional Data Analysis.* Springer Science+Media, New York, 2nd edition, 2006.

[30] A. Rodríguez, J. Cui, N. Ramakrishnan, B. Adhikari, and B. Aditya Prakash. EINNs: Epidemiologically-informed neural networks. In *Proceedings fo the 37th AAAI Conference on Artificial Intelligence (AAAI-23)*, pages 14453–14460. Association for the Advancement of ArtifcialIntelligence, 2023.

[31] F. Rohde, J. Wagner, A. Meyer, P. Reinhard, M. Voss, U. Petschow, and A. Mollen. Broadening the perspective for sustainable artificial intelligence: sustainability criteria and indicators for artificial intelligence systems. *Current Opinion in Environmental Sustainability*, 66(101411):1–12, 2024.

[32] R. Roscher, B. Bohn, M. Duarte, and J. Garcke. Explainable machine learning for scientific insights and discoveries. *IEEE Access*, 8:42200–42216, 2020.

[33] F. Rossi, N. Delannay, B. Conan-Gueza, and M. Verleysen. Representation of functional data in neural networks. *Neurocomputing*, 64:183–210, 2005.

[34] B. Schindler and T. Schmid. Physics-aware normalizing flows: Leveraging electric circuit models in adversarial learning. In M. Verleysen, editor, *Proceedings of the 32nd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN'2024), Bruges (Belgium)*, page in this volume, Louvain-La-Neuve, Belgium, 2024. i6doc.com.

[35] B. Schölkopf, P.Simard, and a. V. V. A. Smola. Priorknowledge in support vector kernels. In *Advances in Neural Information Processing Systems*, volume 10, pages 640–646. 1998.

[36] C. Sirocchi, A. Bogliolo, and S. Montagna. Medical-informed machine learning: integrating prior knowledge into medical decision systems. *BMC Medical Informatics and Decision Making*, 24(186):1–17, 2024.

[37] S. Taverniers, E. Hall, M. Katsoulakis, and D. Tartakovsky. Graph-informed neural networks. In *Proceedings of the AAAI 2021 Spring Symposium on Combining Artificial Intelligence and Machine Learning with Physical Sciences*, volume Vol-2964, pages 1–3. CEUR Workshop Proceedings (CEUR-WS.org), 2021.

[38] Y. Tian and Y. Zhang. A comprehensive survey on regularization strategies in machine learning. *Information Fusion*, 80:146–166, 2022.

[39] T. Villmann and S. Haase. Divergence based vector quantization. *Neural Computation*, 23(5):1343–1392, 2011.

[40] T. Villmann and F.-M. Schleif. Functional vector quantization by neural maps. In J. Chanussot, editor, *Proceedings of First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2009)*, pages 1–4. IEEE Press, 2009. ISBN 978-1-4244-4948-4.

[41] S. Vishwanathan, N. Schraudolph, R. Kondor, and K. Borgwardt. Graph kernels. *Journal of Machine Learning Research*, 11:1201–1242, 2010.

[42] J. Voigt, S. Saralajew, M. Kaden, L. Reuss, and T. Villmann. Biologically-informed shallow classification learning integrating pathway knowledge. In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC'2024) – Bioinformatics Workshop*, volume 1, pages 357–367. SCITEPRESS – Science and Technology Publications, Lda., 2024.

[43] L. von Rüden, J. Garcke, and C. Bauckhage. How does knowledge injection help in informed machine learning? In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2023.

[44] L. von Rüden, S. Mayer, K. B. B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, J. Pfrommer, A. Pick, R. Ramamurthy, M. Walczak, J. Garcke, C. Bauckhage, and J. Schuecker. Informed machine learning – A taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):614–633, 2023.

[45] Y. Wang, Q. Yao, J. Kwok, and L. Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys*, 53(3, Article No. 63):1–34, 2020.

[46] M. Wysocka, O. Wysocki, M. Zufferey, D. Landers, and A. Freitas. A systematic review of biologically-informed deep learning models for cancer: fundamental trends for encoding and interpreting oncology data. *BMC Bioinformatics*, 24(198):1–31, 2023.

[47] F. Zoghlami, M. Kaden, T. Villmann, G. Schneider, and H. Heinrich. AI-based multi sensor fusion for smart decision making: A bi-functional system for single sensor evaluation in a classification task. *Sensors*, 21(13):1–18, 2021.