Deconstructing the Final Frontier of Artificial Intelligence: Five Theses for a Constructivist Machine Learning

Thomas Schmid
Universität Leipzig
Fakultät für Mathematik und Informatik
Augustusplatz 10, D-04109 Leipzig, Germany
schmid@informatik.uni-leipzig.de

Abstract

Ambiguity and diversity in human cognition can be regarded a final frontier in developing equivalent systems of artificial intelligence. Despite astonishing accomplishments, modern machine learning algorithms are still hardly more than adaptive systems. Deep neural networks, for example, represent complexity through complex connectivity but are not able to allow for abstraction and differentiation of interpretable knowledge, i.e., for key mechanisms of human cognition. Like support vector machines, random forests and other statistically motivated algorithms, they do neither reflect nor yield structures and strategies of human thinking. Therefore, we suggest to realign the use of existing machine learning tools with respect to the philosophical paradigm of constructivism, which currently is the key concept in human learning and professional teaching. Based on the idea that learning units like classifiers can be considered models with limited validity, we formulate five principles to guide a constructivist machine learning. We describe how to define such models and model limitations, how to relate them and how relationships allow to abstract and differentiate models. To this end, we propose the use of meta data for classifiers and other models. Moreover, we argue that such meta data-based machine learning results in a knowledge base that is both created by the means of automation and interpretable for humans.

Over the last decade, it has become widely accepted to address computational systems intelligent. Not only journalists, but also scientists have adapted this habit in their publications. In fact, many classical engineering tasks like monitoring or regulating have profited from the employment of machine learning (Abellan-Nebot and Romero Subirón 2010; Mohanraj, Jayaraj, and Muraleedharan 2012). The same holds true for pattern recognition, most prominently in automated image and video analysis (Zafeiriou, Zhang, and Zhang 2015; Yang et al. 2011). And even though ultimate challenges like the infamous Turing test are left unsatisfied (You 2015), some exceptional results in specialized tasks like playing the game of go (Silver et al. 2016) make current learning machines look intelligent on a human level.

A final frontier for learning systems, however, is the variety of alternative cognitive functions observable in a diverse set of individuals or from ambiguous stimuli (Kornmeier and Bach 2012). While philosophy has acknowledged and embraced the subjectivity and limitations of human cognition during the last decades (Prawat and Floden 1994), current learning systems regard cognition a complex, yet technical task to be solved. In particular, established algorithms do neither provide convincing answers to the challenges provided by an ambiguous environment; nor do they offer concepts that explicitly allow for contradictory judgements comparable to differences in social perception.

The main reason for this shortcoming is that so far both algorithms and researchers have failed to incorporate a constructivist point of view. Constructivism implies not only cognition to be a highly individual phenomenon, but also humans to take an active role in their perception of the world – and that there is no such thing as a human-independent reality (Reich 2009). Yet algorithms and applications aiming to predict things other than laws of nature are implicitly founded on exactly this outdated asumption.

In the following, we introduce axioms that allow machine learning to follow constructivist principles. Key features of this approach are the use of modern tools from empirical sciences, model-oriented learning, the ability to handle ambiguity, the ability to integrate supervised and unsupervised learning into a unified framework, the ability to create an individual knowledge base and the ability to abstract, differentiate or discard learned knowledge automatically.

1. The key component of cognitive functionality is a model.

Since the introduction of artificial neural networks as a theoretical concept (McCulloch and Pitts 1943), many mathematicians and computer scientists have considered neurons the key component of learning systems. In education and psychology, however, cognitive functions are often seen as certain skills or abilities acquired and exposed by an individual human and described in terms like the concept of competence, which, e.g., is widely used in the modern European education system (Méhaut and Winch 2012).

Functionalistic psychology explains cognitive functions of humans by the concept of mental models (Rouse and Mor-

ris 1986). Initially, mental models have been used to understand motor control, e.g., of hand movements (Veldhuyzen and Stassen 1977). In a more general sense, however, mental models are described as "hypothetical constructs" (Wickens 2000) that can be ordered hierarchically (Rasmussen 1979) and allow a human to make predictions about his physical and social environment (Oatley 1985). It has also been postulated that such models cannot be of static nature but rather underlay continuous modifications (Oatley 1985).

Philosophers, too, consider models an important tool in human knowledge acquisition (Klaus 1967, p. 412) or even the only tool, respectively (Stachowiak 1973, p. 56). While varying and concurring theoretical definitions exist, most model concepts assume an image, an origin of the image and a relationship between them. This definition is, e.g., matched by the idea of mathematical modeling as proposed by Heinrich Hertz and others (Hertz 1894; Hamilton 1982). With the rise of robotics and artificial intelligence, engineers have adapted and extended this idea by postulating the concept of a cybernetic model, which involves a generalized subject and an object of the model (Rose 2009).

Cybernetics, however, did neither reflect time-related aspects nor issues involved with individual model subjects. This matter was adressed by Herbert Stachowiak, who was influenced by cybernetics when developing his *General Model Theory* (Hof 2018). He postulated any model to be limited to specific subjects, specific temporal ranges and specific purposes (Stachowiak 1973, p. 133). Limitations, to this end, are considered a matter of fact rather than a matter of definition. Thus, such models circumvent ambiguity by viewing an otherwise ambiguous model with unknown validity limits as a number of models of limited validity.

2. Learning constitutes from constructing, reconstructing or deconstructing models.

Modern education is dominated by the ideas of constructivism and constructivist learning (Fox 2001). At its heart, this approach is based on the assumption that humans acquire knowledge and competences actively and individually through processes called construction, reconstruction and deconstruction (Duffy and Jonassen 1992). Construction is associated with creation, innovation and production and implies searching for variations, combinations or transfers of knowledge (Reich 2004, p. 145). Analogously, reconstruction is associated with application, repetition or imitation and implies searching for order, patterns or models (Reich 2004, p. 145). Deconstruction is in the context of constructivism associated with reconsideration, doubt and modification and implies searching for omissions, additions and defective parts of acquired knowledge (Reich 2004, p. 145).

Learning algorithms have been used for half a century to transform sample data into models in a mathematical sense, that is: into generalized mathematical relationships between image and origin. The two major approaches or objectives, known as supervised and unsupervised learning, either do or do not require a given target parameter. Artificial neural networks and their relatives are among the most popular and prominent algorithms for learning with a given target parameter (Singh, Thakur, and Sharma 2016), but statistically motivated approaches like support vector machines (Cristianini and Shawe-Taylor 2000) or random forests (Breiman 2001) are also widely used for supervised learning; a specialized field of supervised learning is reinforcement learning, which is popular in robotics (Kober and Peters 2012) and adaptive control (Lewis, Vrabie, and Vamvoudakis 2012). For unsupervised learning, too, biologically inspired approaches like self-organizing maps (Kohonen 2001) as well as statistically motivated approaches like k-means (Jain 2010) are employed.

To some extent, machine learning parallels modern education concepts. A construction process in the constructivist sense may be matched by an unsupervised learning, i.e., identifying clusterings or dimensionality reduction, and can, e.g., be implemented with self-organizing maps, k-means, autoencoders or feature clustering (Schmid 2018). A reconstruction process in the constructivist sense may be matched by a supervised learning, i.e., classification or regression tasks, and can, e.g., be implemented with artificial neural networks or random forests (Schmid 2018). Few researchers, however, have discussed a constructivist approach to machine learning (Drescher 1989; Quartz 1993), and even less how to design a deconstruction process. While domain-specific applications with manual re-engineering options exist (Herbst and Karagiannis 2000), to the best of our knowledge, there is currently only one working implementation of an algorithmic deconstruction process (Schmid 2018).

3. Deconstructing models computationally requires model-based meta data.

In order to automate and implement a deconstruction process, successfully learned models must be held available for comparison or re-training. More over, possible matchings with novel models must be identifiable in an unambiguous manner by calculation or logical operations, respectively. For Stachowiak models, features regarding validity limitations exist for any model employed, which allows to discriminate models. Here, we outline how such meta data can be identified for models created by machine learning.

Machine learning implies learning from examples termed training vectors. For a supervised training vector $\mathcal{V}$ resulting from sensor data, for instance, this implies

$$\begin{aligned} \mathcal{V} &= (I, O) && (1)\\ &= (i_0, ..., i_{m-1}, o_0, ..., o_{n-1}) && (2) \end{aligned}$$

with a m dimensional input I and a n dimensional output O.

A Stachowiak-like training vector $\mathcal{V}^*$ will, in addition, possess three pragmatic features:

$$\mathcal{V}^* = (\mathcal{V}, T_\mathcal{V}, \Sigma_\mathcal{V}, Z_\mathcal{V}) \quad (3)$$

with

$$\begin{aligned} T_\mathcal{V} &= \tau && (4)\\ \Sigma_\mathcal{V} &\subset \Sigma && (5)\\ Z_\mathcal{V} &\subset Z && (6) \end{aligned}$$

where $T_\mathcal{V}$ is a point τ in time, and $\Sigma_\mathcal{V}$ and $Z_\mathcal{V}$ are subsets of the infinite sets of model subjects Σ and of model purposes

Z, respectively. When using sensor data as training data, T and Σ for each vector may be given by sensor meta data and Z by the application context of the data collection.

If a machine learning-based model $\mathcal{M}$ is considered an approximation of n training vectors $\mathcal{V}$ with

$$\mathcal{M} \sim \{\mathcal{V}_0, ..., \mathcal{V}_{n-1}\} \tag{7}$$

then meta data for a Stachowiak-like model $\mathcal{M}^*$ with

$$\mathcal{M}^* = (\mathcal{M}, T_\mathcal{M}, \Sigma_\mathcal{M}, Z_\mathcal{M}) \tag{8}$$

can be derived from the underlying n Stachowiak-like training vectors $\mathcal{V}^*$ (Schmid 2018) by:

$$T_\mathcal{M} = \left[min(T_{\mathcal{V}^*_0}, ..., T_{\mathcal{V}^*_{n-1}}), max(T_{\mathcal{V}^*_0}, ..., T_{\mathcal{V}^*_{n-1}})\right] \tag{9}$$

$$\Sigma_\mathcal{M} = \cup_{i=0}^{n-1} \Sigma_{\mathcal{V}^*_i} \tag{10}$$

$$Z_\mathcal{M} = \cup_{i=0}^{n-1} Z_{\mathcal{V}^*_i} \tag{11}$$

By extracting and administrating these meta data for every model a machine learning algorithm has learned, learned models become discriminable. Most importantly, overlaps or contradictions in model validity become thereby identifiable and may be resolved algorithmically.

4. Deconstructing models implies to either abstract, differentiate or discard them.

Using the pragmatic features T, Σ, Z of Stachowiak-like models as meta data, machine learning-generated models can be matched and discriminated automatically. In particular, this allows to implement learning through deconstruction of given models. With regard to the degree of meta data matching exposed by the respective models, four types of deconstruction operations will be distinguished here.

The degree of relationship between two given Stachowiak-like models $\mathcal{M}_a$ and $\mathcal{M}_b$ is termed

1. complete,
   if $T_{\mathcal{M}_a} = T_{\mathcal{M}_b}$, $\Sigma_{\mathcal{M}_a} = \Sigma_{\mathcal{M}_b}$, $Z_{\mathcal{M}_a} = Z_{\mathcal{M}_b}$.
2. subjective-intentional (ΣZ),
   if $T_{\mathcal{M}_a} \neq T_{\mathcal{M}_b}$, $\Sigma_{\mathcal{M}_a} = \Sigma_{\mathcal{M}_b}$, $Z_{\mathcal{M}_a} = Z_{\mathcal{M}_b}$;
3. temporal-intentional (TZ),
   if $T_{\mathcal{M}_a} = T_{\mathcal{M}_b}$, $\Sigma_{\mathcal{M}_a} \neq \Sigma_{\mathcal{M}_b}$, $Z_{\mathcal{M}_a} = Z_{\mathcal{M}_b}$;
4. temporal-subjective ($T\Sigma$),
   if $T_{\mathcal{M}_a} = T_{\mathcal{M}_b}$, $\Sigma_{\mathcal{M}_a} = \Sigma_{\mathcal{M}_b}$, $Z_{\mathcal{M}_a} \neq Z_{\mathcal{M}_b}$;

The deconstruction of two completely related models can basically either confirm the congruency and validity of both models or render both invalid, which leads to both being discarded. As a third option, this deconstruction process allows to test whether the combination of both models may be split in two submodels of more limited temporal validity.

Two ΣZ-related models $\mathcal{M}_a$ and $\mathcal{M}_b$ share a common set of model subjectives and model purposes, but differ in their temporal validity. ΣZ deconstruction therefore builds and evaluates a temporal union of $\mathcal{M}_a$ and $\mathcal{M}_b$; by this, the initial model is either replaced by a unified model with larger temporal validity – or left untouched.

Deconstruction of two TZ-related models $\mathcal{M}_a$ and $\mathcal{M}_b$, which share a congruent temporal validity and a common set of model purposes but differ regarding their subjective validity, either promotes $\mathcal{M}_a$ to a model $\mathcal{M}_c$ of intersubjective validity – or leaves $\mathcal{M}_a$ untouched.

$T\Sigma$ deconstruction leaves two $T\Sigma$-related models $\mathcal{M}_a$ and $\mathcal{M}_b$ untouched. Instead, it will construct novel models based on their outputs, yielding models of a higher level. This is possible only because $\mathcal{M}_a$ and $\mathcal{M}_b$ share a congruent temporal validity and a common set of model subjectives while differing in their model purpose. All together, a $T\Sigma$ deconstruction process can be regarded as a way of automated abstraction or generalization of knowledge.

5. A hierarchically ordered set of models constitutes an enriched knowledge base.

Any set of models created by machine learning algorithms represents information inherited in the underlying training data and can therefore be considered a knowledge base. In a constructivist approach, however, each model of such a set possesses explicit validity limitations, which contributes additional knowledge and complexity. Temporal gaps in the knowledge base, e.g., can thereby be identified explicitly. A hierarchical ordering also indicates hot spots of abstraction, i.e., models of higher hierarchy levels.

The degree of abstraction and differentiation within such a knowledge base can be quantified by assessing the number of models, the average temporal validity, etc. Alternatively, this can be achieved by building and visualizing a meta data-based ordering. Apart from their temporal validity, models of such a set can also be ordered according to their model purposes or level of abstraction, respectively. For uniform model subjects or identical learning algorithm, respectively, a three-dimensional plot may visualize both temporal extensions and the extend of successful abstraction.

Moreover, each individual model represents a supervised learning application, i.e., a classifier or regressor, and can be used as such after the knowledge base has been established. To this end, a hierarchically ordered set of models created by constructivist machine learning inherits not only structured, but also applicable knowledge. Models that match a given test sample – and are therein valid classifiers or regressors – can be identified by simply matching the meta data. Consequently, application of the knowledge base can be rejected in scenarios where no knowledge is available.

Conclusions

In the present work, we suggest principles for using existing machine learning algorithms with respect to constructivist theories of human learning. Based on five axioms introduced here, a constructivist approach of creating explainable knowledge can be implemented. This approch allows, in particular, to create an ambiguity-free knowledge base. While there is no restriction regarding potential applications for constructivist machine learning, it is likely that tasks where ambiguous knowledge and results need to be avoided will profit most from this learning paradigm. Future work on this approach will focus on parallelization strategies for constructivist machine learning and on developing task-oriented comparisons with human cognitive functionality.

References

Abellan-Nebot, J. V., and Romero Subirón, F. 2010. A review of machining monitoring systems based on artificial intelligence process models. *The International Journal of Advanced Manufacturing Technology* 47(1):237–257.

Breiman, L. 2001. Random forests. *Machine learning* 45(1):5–32.

Cristianini, N., and Shawe-Taylor, J. 2000. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press. chapter 6, 93–124.

Drescher, G. L. 1989. *Made-up minds : a constructivist approach to artificial intelligence*. Ph.D. Dissertation.

Duffy, T. M., and Jonassen, D. H. 1992. Constructivism: new implications for instructional technology. In Duffy, T. M., and Jonassen, D. H., eds., *Constructivism and the Technology of Instruction – A Conversation*. Hillsdale, NJ: Erlbaum. 1–16.

Fox, R. 2001. Constructivism examined. *Oxford Review of Education* 27(1):23–35.

Hamilton, A. G. 1982. *Numbers, Sets and Axioms: The Apparatus of Mathematics*. Cambridge University Press.

Herbst, J., and Karagiannis, D. 2000. Integrating machine learning and workflow management to support acquisition and adaptation of workflow models. *Intelligent Systems in Accounting, Finance and Management* 9(2):67–92.

Hertz, H. 1894. Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt. In Hertz, H., ed., *Gesammelte Werke*, volume 3. Leipzig: Barth.

Hof, B. E. 2018. The cybernetic "General Model Theory": Unifying science or epistemic change? *Perspectives on Science* 26(1):76–96.

Jain, A. K. 2010. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters* 31(8):651–666.

Klaus, G. 1967. *Wörterbuch der Kybernetik*. Berlin: Dietz.

Kober, J., and Peters, J. 2012. Reinforcement learning in robotics: a survey. In Wiering, M., and van Otterlo, M., eds., *Reinforcement Learning*. Springer. chapter 18, 579–610.

Kohonen, T. 2001. *Self-organizing maps*, volume 30 of *Springer Series in Information Sciences*. Springer, 3 edition.

Kornmeier, J., and Bach, M. 2012. Ambiguous figures – what happens in the brain when perception changes but not the stimulus. *Frontiers in Human Neuroscience* 6.

Lewis, F.; Vrabie, D.; and Vamvoudakis, K. 2012. Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers. *IEEE Control Systems* 32(6):76–105.

McCulloch, W. S., and Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology* 5(4):115–113.

Méhaut, P., and Winch, C. 2012. The European qualification framework: Skills, competences or knowledge? *European Educational Research Journal* 11(3):369–381.

Mohanraj, M.; Jayaraj, S.; and Muraleedharan, C. 2012. Applications of artificial neural networks for refrigeration, air-conditioning and heat pump systemsa review. *Renewable and Sustainable Energy Reviews* 16(2):1340–1358.

Oatley, K. 1985. Representations of the physical and social world. In Oatley, D. A., ed., *Brain and Mind*. London: Methuen. 32–58.

Prawat, R. S., and Floden, R. E. 1994. Philosophical perspectives on constructivist views of learning. *Educational Psychologist* 29(1):37–48.

Quartz, S. R. 1993. Neural networks, nativism, and the plausibility of constructivism. *Cognition* 48(3):223–242.

Rasmussen, J. 1979. On the structure of knowledge – a morphology of metal models in a man–machine system context. Technical Report Riso-M-2192, Riso National Laboratory, Roskilde, Denmark.

Reich, K. 2004. *Konstruktivistische Didaktik. Lehren und Lernen aus interaktionistischer Sicht*. Munich: Luchterhan, 2nd edition.

Reich, K. 2009. Constructivism: Diversity of approaches and connections with pragmatism. In Hickman, L. A.; Neubert, S.; and Reich, K., eds., *John Dewey Between Pragmatism and Constructivism*. Fordham University Press.

Rose, J. 2009. The early years: some comments on the origins and concepts of cybernetics. *Kybernetes* 38(1/2):20–24.

Rouse, W. B., and Morris, N. M. 1986. On looking into the black box: prospects and limits in the search for mental models. *Psychological Bulletin* 100(3):349–363.

Schmid, T. 2018. *Automatisierte Analyse von Impedanzspektren mittels konstruktivistischen maschinellen Lernens*. Ph.D. Dissertation, Leipzig.

Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; Dieleman, S.; Grewe, D.; Nham, J.; Kalchbrenner, N.; Sutskever, I.; Lillicrap, T.; Leach, M.; Kavukcuoglu, K.; Graepel, T.; and Hassabis, D. 2016. Mastering the game of go with deep neural networks and tree search. *Nature* 529:484–489.

Singh, A.; Thakur, N.; and Sharma, A. 2016. A review of supervised machine learning algorithms. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 1310–1315.

Stachowiak, H. 1973. *Allgemeine Modelltheorie*. Springer.

Veldhuyzen, W., and Stassen, H. G. 1977. The internal model concept: an application to modeling human control of large ships. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 19(4):367–380.

Wickens, C. D. 2000. *Engineering Psychology and Human Performance*. New Jersey: Prentice Hall, 3rd edition.

Yang, H.; Shao, L.; Zheng, F.; Wang, L.; and Song, Z. 2011. Recent advances and trends in visual tracking: A review. *Neurocomputing* 74(18):3823–3831.

You, J. 2015. Beyond the Turing test. *Science* 347(6218):116–116.

Zafeiriou, S.; Zhang, C.; and Zhang, Z. 2015. A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding* 138:1–24.