Deep Learning and Functional Safety

Architekturoptimierung unterstützt Funktionale Sicherheit

Dr. Ulrich Bodenhausen, Ulrich Bodenhausen AI Coaching;
Vector Consulting Services GmbH

This paper describes the challenges in arguing safety of systems using Deep Learning Neural Networks. The market potential of safety critical products using AI is very attractive and Deep Learning Neural networks have proven strengths to provide important functionality. Challenges remain in the understanding and further optimization of generalization capability and in the improvement of verification/validation methodology. Algorithmic optimization of the architecture of Deep Learning Neural Network can be used beneficially to reduce residual risk of functional insufficiencies. Additionally, it can also be used to improve analyzability by construction of architectures with required observation points.

AI Market Potential and Safety

Market analysis reports predict significant growth of revenue for products and services using Artificial Intelligence (AI). Bank of America Merrill Lynch analyze the Robot and AI market development. A categorization of their market forecast reveals that approx. 35% of robots and artificial intelligence market are safety critical products.

BCC Research expects autonomous robot category to account for the largest share - 23 percent - of annual market growth until 2024 and thus to dominate the smart machine market. Tractica Research predicts a growth rate (CAGR) of 57% and GII Research estimate the artificial intelligence market to have cumulated growth rate of 54% from 2015 to 2020.

Allover, the growth potential seems to be very attractive. There are two implications from this: From the perspective of safety critical products it is likely that AI will become a very important driver for safety critical products. From the perspective of AI it is likely that Functional Safety will become an important requirement for large portion of AI based products. This paper aims to bridge the gap between both perspectives.

AI Market Potential and Embedded Software

Autonomous vehicles account for a large share within this predicted market growth of AI. However, there are many other applications, where an AI-based algorithm can potentially be used to augment an established embedded functionality by an intelligent usage of available or new sensor data. In case of safety critical systems, the safety case needs to be extended to cover the added sensors and AI engine.

Machine Learning (ML) and Real-World Applications

For the context of this paper, Machine Learning is defined as an approach using inductive learning techniques for system design, where the run-time system uses the results of a learning process to perform algorithmic operations (e.g. running a Deep Learning Neural Network having precomputed weights).

Machine Learning algorithms have been developed and applied for several decades. In case of application to real world data, the following design goals have been pursued:

- The run-time systems performs with low deviation from desired system functionality in the defined application context
- The functionality is robust against variances (i.e. rotation, size, shape, color, …)
- The performance is uniform across classes of inputs (i.e. critical classes do not yield unacceptable poor performance)
- The system is robust against differences between its training and testing data sets

One class of Machine Learning approaches is Neural Networks (NN). A NN consists of simple processing units ("nodes") which are connected via weights. In the generic approach each node of one layer of nodes is connected to each node of the next layer. A NN is a Deep Learning NN, if there are several hidden layers between input and output for internal processing (see Figure 1).

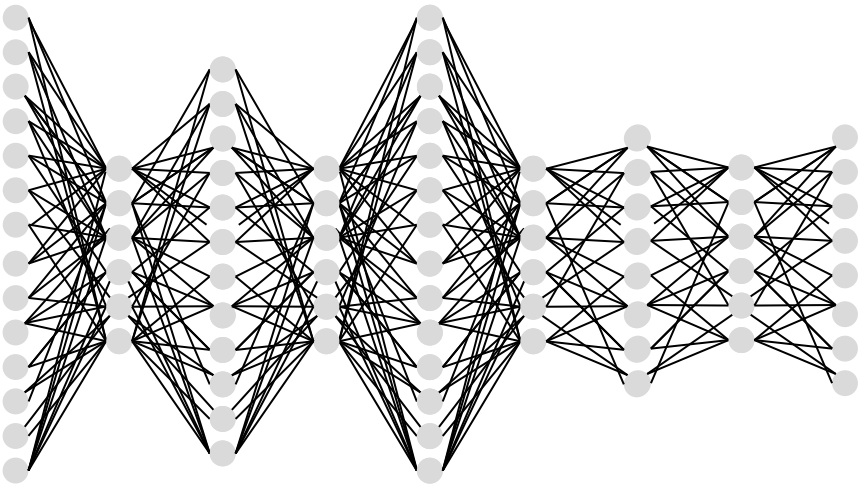

Figure 1: Generic (Deep Learning) Neural Network. Each node is connected to each node of next layer. Deep Learning Neural Network refers to several internal (hidden) layers between input and output for internal processing.

In practice over several decades of enhancement of Machine Learning, the approach of Deep Learning Neural Network NN and Deep Learning Convolutional NN have proven to be very powerful approaches to Machine Learning in real-world applications. How is this measured? The only way the get a relevant measure on how well a learning system is performing is to divide the available data in two sets, a set of training cases and a set of testing cases (see figure 2). The set of training cases (the green set in figure 2) is used by the learning algorithm to compute the parameters (i.e. weights of the interconnections in figure 1). A good performance (small deviation of the intended system functionality) on the set of training data means that the learning algorithm is at least able to memorize the set of training data.

The set of test data (the orange set of test cases in figure 2) is then used to determine how well the system is performing on cases that have not been used during the training. If the system behaves according to the intended functionality on this set, then the learning algorithm must have extracted something form the training data that goes beyond pure memorization: It must have extracte features that allow the system to perform with the intended functionality, even on cases the system has never seen before. This capability is called “generalization capability”.

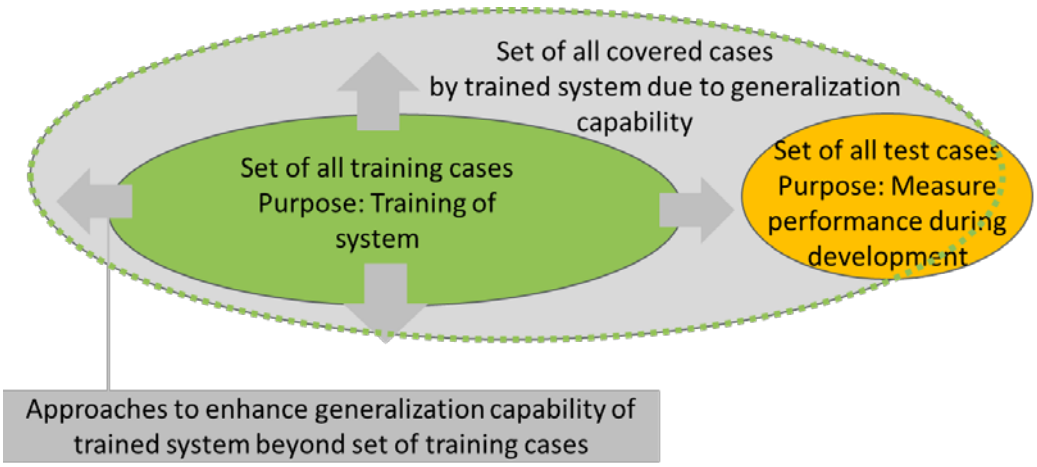

Figure 2: Illustration of generalization capability: The system is trained on training data (green set). Due to generalization capability and approaches to enhance the generalization capability further, the system behaves with the intended functionality on many more cases (gray set), including many cases from the test set (orange set of da.

Under certain conditions Deep Learning NN have the proven ability to develop strong generalization capability. These conditions have recently been analyzed in detail [1] with the result that NN can generalize surprisingly well. Another outcome is that changes in architecture resulted in bigger improvements in generalization capability than other approaches, i.e. regularization techniques. Changes of architecture can be done manually (with very high effort from experts) or by automatic optimization algorithms. The research on algorithmic optimization of NN architectures started around 1991 and has advanced since then significantly. Recent work [2] shows impressive results: An approach where a recurrent NN as controller was used for search of NN architectures resulted in a very high generalization performance that was better than the best human-invented architectures available at that time.

Machine Learning/Deep Learning and Safety Standards

First sources of information about Deep Learning and Safety are established and upcoming Safety Standards. ISO26262-1 provides a very detailed methodology, but is not adapted to Machine Learning/Deep Learning. Examples:

- Development and verification methods within part 6 does not address that in a NN the functionality is embedded in highly dimensional data matrices.

- Verification methods such as coding guidelines, white-box code coverage provide no relevant insights into learned NN structures.

The upcoming update ISO26262-2 is not released yet. It needs to be discussed when final version is available.
Safety of the Intended Functionality (SOTIF) ISO/WD PAS 21448 is currently under development. It needs to be discussed when final version is available.

Spearhead of Safety Critical ML Technology: Validation of Highly Autonomous Vehicles

In addition to established and upcoming Safety Standards it is helpful to look at practices for Highly Automated Vehicles (HAV) as spearhead of ML technology in safety critical applications:

- What is open to the public is that extensive road tests are performed - hundreds of Mio. of miles. The major players are in competition concerning the number of accumulated miles of road test. The limitations of number of road miles as validation criterion has been discussed frequently (Example: [5]).
- What is done on top of road testing is in many cases kept as proprietary approaches, which are not open for public evaluation yet.

Proposals for an advancement of methodology have been made (Examples: [3], [5]). A very systematic approach is to propose an assurance case structure for HAV containing ML subsystems [3]. Besides other important elements it states goals to be met by the ML system:

- Acceptable residual risk of functional insufficiencies
- Function is robust against distributional shift
- Function exhibits uniform behavior over critical classes
- Function is robust against differences between training and execution platform
- Function is robust against changes in system context
- Operational context is well defined and reflected in training data

Fortunately, the five functional goals align very well with the design goals of ML for real-world applications. Deep Learning NN have proven to be very successful to match these goals. So Deep Learning NN should be a very beneficial choice when realizing an AI-based safety critical system. The question remains on the adequate safety evidence that the goals are met.

Deep Learning Neural Networks in Safety Analyses

One important concept of safety cases is an analysis and review of structures, code and test results. Fortunately, the analysis of internal representations in NN has drawn some attention and research is done in this field. However, there is still a considerably higher expert knowledge with Deep Learning Methods needed to analyze internal structures than to use NN libraries for training.

Analysis by reconstruction of images from different layers of well-known reference NN architectures have been investigated (Example: [4]). The results are insightful for

the data scientist, but it is at least surprising for the safety engineer to see what the results of comprehensive analysis of internal NN structures are. From the use case of a safety engineer, internal representation of input images are opaque and not yet useful in safety analyses.

Challenges and Solution Approaches

As seen in the last paragraph, the internal structure of Deep Learning NN is inherently hard to analyze, especially in Deep Learning NN layers with full connectivity. One possible attempt to overcome this challenge is to apply methods to enforce an analyzable structure in the NN architecture. This approach goes in line with the proposal of inserted “test points” to improve observability and to explicitly test the system is doing the right thing for the right reason [5]. This can be implemented in the training process manually by experienced data scientists, or via an automatic architecture optimization process.

In addition, the safety engineers face the challenge to prove that the residual risk of functional insufficiencies is low enough. This challenge must be addressed with two types of measures:

- Improvement of generalization capability of the learning approach. Besides generic methods like data augmentation, weight decay and dropout it is important to optimize the architecture. Optimization of architecture are usually done manually, but can also be done automatically (Example: [2]).
- Improvement of verification/validation methodology: Clarification of goals of testing, testing according to these goals, validation of residual risks, i.e. unexpected scenarios and environment, degraded infrastructure, enhancement of “analyzability” for safety analyzes by observation points.

Conclusions and Future Work

AI will be an important driver for safety critical products and Safety will be an important requirement for a significant portion of AI products. This means that both disciplines must work closely together to benefit from each other.

The update ISO 26262-2 as well as SOTIF ISO/WD PAS 21448 are currently in preparation. They are needed to provide standards, but many additional ideas and details of methods will be needed to complete the proof of safety for AI based systems.

One big challenge is the understanding and further optimization of generalization capability of the learning approach of Depp Learning NN. Another challenge is the improvement of verification/validation methodology. Both need further advancement and the close collaboration of AI experts with Safety experts to succeed with business in this very promising market.

References:

[1] Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O., „Understanding Deep Learning Requires Rethinking Generalization“, ICLR 2017, arXiv:1611.03530 [cs.LG]

[2] Zoph, B., “Learning Transferable Architectures for Scalable Image Recognition”, arXiv:1707.07012v4 [cs.CV] 11 Apr 2018

[3] Burton, S., Gauerhof, L., Heinzemann, C., “Making the Case for Safety of Machine Learning in Highly Automated Driving”, SAFECOMP 2017

[4] Dosovitskiy , A., Brox, T. “Inverting Visual Representations with Convolutional Networks”, CoRR, vol. abs/1506.02753, 2015

[5] Koopman, P., Wagner, M., “Toward a Framework for Highly Automated Vehicle Safety Validation”, SAE World Congress 2018

Author

Dr. Ulrich Bodenhausen got a PhD in Machine Learning from KIT. Besides application of neural networks to speech, handwriting and gesture recognition at Carnegie Mellon University and KIT, he gathered experience in strategic business consulting, strategic business planning at a 1st tier automotive supplier and was responsible for innovation management, development processes and knowledge management at 1st tier automotive supplier.
He is now managing a team of consultants for transition programs and optimization of product development (e.g. agile methods, development methods, functional safety, crisis management) at Vector Consulting Services. He is also exploring AI-related business opportunities.

Contact

Internet: www.grow-with-ai.com
Email: info@grow-with-ai.com

Internet: www.vector.com
Email: Ulrich.Bodenhausen@vector.com