IFA
Institut für Arbeitsschutz der
Deutschen Gesetzlichen Unfallversicherung

Handout: Robustness assessment of deep neural networks

Digital manufacturing: free online seminars (IFA + IVSS)
Institute for Occupational Safety and Health of the German Social Accident Insurance
Section Intelligent Technical Systems and Working Environment

Complex software systems are deployed in multiple sectors where harmonised rules ought to ensure compliance with safety obligations. This presentation has shown how robustness methods can be implemented into the system lifecycle to facilitate the development and validation of machine learning components.

The general machine learning setting, where a specification is implicitly learned from data, was recapitulated. Computer vision methods based on deep learning (e.g., [1]) can be incorporated into industrial logic components. They offer many opportunities to increase occupational safety. However, risk factors associated with this technology must be adequately reduced.

Various aspects and examples of robustness were given. We identified a spectrum of perturbations reaching from the microscopic scale with a low correlation length to macroscopic ones where the semantic meaning is altered. It was illustrated how the overall generalisability of neural networks is linked to the performance near the boundary of the operational design domain (ODD). This relates to the bias-variance decomposition (e.g., [2]) and can be considered as a no-free-lunch theorem.

Because of those inherent limitations that purely data-driven approaches possess, the remaining aleatoric uncertainties (e.g., [3]) only can be further reduced on the system engineering level, e.g., with the installation of additional sensors to reduce ambiguities near the decision boundary. Such safety concepts should be considered an essential part of the system design and should be introduced in the development lifecycle early on.

Bibliography

[1] I. Goodfellow, Y. Bengio and Aaron Courville, Deep Learning, MIT Press, 2016.

[2] Wikipedia contributors, Bias-variance tradeoff, Wikipedia, accessed Dez. 2023.

[3] T. M. Cover, J. A. Thomas, Elements of Information Theory, Wiley, 2006.