Symbolic Background Knowledge for Machine Learning

Benjamin Risse
Institute for GeoInformatics, University of Münster

Submission date: May 13th, 2022

Required funding: The requested funding are a six month post doc position (TV-L 13) for Reinhard Moratz. Reinhard Moratz is an expert for knowledge representations with two publications in the Artificial Intelligence Journal.

Abstract

The field of artificial intelligence can roughly be categorized into Machine Learning-based and Knowledge-based approaches. In the proposed incubator project we want to combine the strengths of both approaches. As a first step for such hybrid systems we want to develop a tool for using knowledge graphs to annotate training data with environmental features for making machine learning more reliable and transparent.

I. *Introduction*

The field of artificial intelligence can roughly be categorized into Machine Learning-based and Knowledge-based approaches (Humm 2021). Knowledge-based systems are more or less, directly or in an indirect way manually enabled to solve a certain task. In contrast machine-learning based systems are fitted to sample data sets via optimization and need less manual fine-tuning. With modern deep-learning based vision approaches features of the human computer vision like convolution style neural computations are integrated in such a way that the expressive power avoids underfitting. Overfitting is avoided by node copying (LeCun 1989). With node copying/sharing the degree of freedom of the model is constrained (e.g. Kroener and Moratz 1996) and in such a way less training data suffices. To better quantify overfitting, novel measures such as neural activation sparsity have been introduced to characterize the impact of certain topological network decision on generalization capabilities (Huesmann et al. 2021). In recent years not only subsymbolic AI made strong progress but also symbolic AI had substantial progress. For example Till Mossakowski and his colleagues could use theorem provers to verify the logical consistency of upper ontologies (Kutz and Mossakowski. 2011).. Knowledge graphs (Chen, Xiaojun et al, 2020) could support semantic-based search. The strengths of knowledge-based image processing have a chance to complement deep learning based machine vision in such a way that hybrid systems could have a higher reliability and also might be easier to adapt to new scenarios/contexts.

Our research goal is to use knowledge graphs to describe potential scenes to which image input data would correspond to and constraints over the instantiations of the potential scenes in the image data. For example a detection system for a certain bird species could have two very similar visual appearances of different species (e.g. juvenile bald eagle versus golden eagle in image data from Maine versus California). The location of the system would tell important information about the prior probability of the two species (e.g. there are no golden eagles in Maine). A hybrid symbolic and subsymbolic vision system which would have a substantially different location in training data versus the location of current operation could trigger a warning. As a first step with this incubator project we want to use knowledge graphs to describe the environmental features of the training data.

- Kutz, Oliver, and Till Mossakowski. "*A modular consistency proof for DOLCE*." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 25, no. 1, pp. 227-234. 2011.

- **Own previous work**: Kröner, Sabine, and Reinhard Moratz. "*Capacity of structured multilayer networks with shared weights*." In International Conference on Artificial Neural Networks, pp. 543-550. Springer, Berlin, Heidelberg, 1996.

- LeCun, Yann, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard, and Lawrence D. Jackel. "*Backpropagation applied to handwritten zip code recognition*." Neural computation 1, no. 4 (1989): 541-551.

- Chen, Xiaojun, Shengbin Jia, and Yang Xiang. "*A review: Knowledge reasoning over knowledge graph*." Expert Systems with Applications 141 (2020)

- **Own previous work**: Huesmann, Karim, Luis Garcia Rodriguez, Lars Linsen, and Benjamin Risse. "*The Impact of Activation Sparsity on Overfitting in Convolutional Neural Networks.*" In International Conference on Pattern Recognition, pp. 130-145. Springer, Cham, 2021.

- Humm, Bernhard G., et al. "*Machine intelligence today: applications, methodology, and technology*." Informatik Spektrum44.2 (2021): 104-114.

II. *Incubator Project description*

The goal of our incubator project is to design and build a semantic technology based annotation tool for training data for deep learning in computer vision. It shall be based on the KnowWhereGraph approach (Janowicz, Krzysztof, et al., 2022). The tool would use knowledge graph based information as meta data for annotating training data.
So for example the training data contains samples of bald eagles in Waterville, Maine. Then you can use the same format to annotate the environmental context of the working system. Then a future reasoning tool could automatically detect inconsistencies between training data and current input data and generate a warning. For example if the working system would operate in California then the warning "There are no golden eagles in Maine. But there are golden eagles in California" could be generated. The reasoning tool itself would not be part of the incubator project. But it would already be useful to have the means to annotate training data in a machine readable format starting as early as possible.
In the incubator project the Sensors, Observations, Sample, Actuator (SOSA) ontology (Zhu et al. 2021) would be augmented with finer grained spatial and temporal concepts from

Qualitative Spatio-Temporal Reasoning (Moratz, 2017). The Protégé Editor would be extended with a specific view tailored for annotating purposes. Sample data from several NFDI4Earth repositories would be annotated with this new tool and documentation would be made available. Experts in semantic technology (Krzystof Janowicz, Till Mossakowski) would be invited to review the extended ontology and the suggested annotation procedure.

- Janowicz, Krzysztof, et al. "*Know, Know Where, KnowWhereGraph: A densely connected, cross‐domain knowledge graph and geo‐enrichment service stack for applications in environmental intelligence*." AI Magazine 43.1 (2022): 30-39.

- **Own previous work**: Moratz, Reinhard. "*Qualitative Spatial Reasoning*." (2017): 1700-1707, Encyclopedia of GIS, Springer International Publishing.

- Zhu, Rui, Shirly Ambrose, Lu Zhou, Cogan Shimizu, Ling Cai, Gengchen Mai, Krzysztof Janowicz, Pascal Hitzler, and Mark Schildhauer. "*Environmental Observations in Knowledge Graphs*." In 2nd Workshop on Data and research objects management for Linked Open Science, pp. 1-11. 2021.

III. *Relevance for the NFDI4Earth*

We expect several projects from NFDI4Earth to benefit from our project. Currently, we have a cooperation with the Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF) e.V. (BMBF grant proposal submitted). We expect several other NFDI4Earth repositories could make use of our semantic annotation tool. In the last phase of our project we plan to cooperate with several NFDI4Earth projects and use their data to demonstrate the annotation methodology developed in our project. A workshop will be offered in which the participants learn how to use the developed annotation tool for their own repositories.

IV. *Deliverables*

Time	Work Package	Deliverable
1st-10th week	Addition of Qualitative Spatial and Temporal Concepts	Augmented SOSA ontology
11th-20th week	Extension of Protégé Editor with annotation view	Software modules in Github for the annotation module
21st-26th week	Annotation of several NFDI4Earth repositories and documentation for future users	User manual and documentation for annotation methodology