VRG19-008 - Knowledge-infused Deep Learning for Natural Language Processing

Abstract

The proposed research project with the title "Knowledge-infused deep learning for natural language processing" combines methods from computational linguistics and cutting-edge machine learning to make language analysis and information extraction robust and generally applicable. More specifically, the proposed research project aims at combining human knowledge (as captured in rule-based natural language processing) with statistical information from large amounts of data, with the goal to leverage the most useful properties of both sources of information. The goal is to develop statistical approaches that can generalize from the knowledge contained in the rules (using statistical pattern recognition techniques), while limiting the introduction of noise into the resulting model. The proposed research will provide solutions with the following practical benefits: 1. Powerful, modern artificial intelligence methods will be applicable to underserved domains and languages (not just driven by a narrow selection of annotated corpora for English). 2. While language and linguistics is a prototypical case for incorporating human insight into analysis, other fields where rule-based approaches have been applied traditionally can benefit from the methods developed in this project. 3. Human insight will be leveraged to provide explanations for automated predictions. Explainable machine learning is important from a societal and legal perspective. Objective 1: Data-driven generalization from rules. Recently, deep learning models have achieved impressive results on standard data sets and in domains for which there are large amounts of training data available. However, the resource requirements of deep learning algorithms restrict their usability for new tasks and domains: Annotations are lacking in many practical scenarios and practitioners have to resort to traditional rule-based algorithms that suffer from coverage issues. It is therefore very important to have more resource-effective means of annotating data (for new tasks, in different domains) that leverage pre-existing knowledge. One solution could be to use traditional rule-based systems to automatically annotate data, which in turn could be used to train machine learning systems. However, automatic annotation schemes may introduce biases. The proposed project will examine the best ways to mitigate these biases, generalize from underlying rules, and to make automatically annotated data maximally useful for modern machine learning in NLP. In order to solve this problem, we will develop algorithms that build on recent advances in machine learning, including generative adversarial networks, latent variable modeling, expectation regularization and other transfer learning techniques. Objective 2: Rule-backed explanations of machine-learning models. In order for users to gain trust into a machine learning system, it is important that the system gives an explanation together with a prediction. An explanation mechanism can give insight into how the system has combined several pieces of information and whether the applied reasoning mechanism is sound. Explanations for a prediction can also help to detect erroneous facts, which can then be corrected by an expert or user. The first steps to legally codify a 'right to explanation' of decisions resulting from automated processing of personal data (EU GDPR Recital 71) emphasize the societal importance of explanation mechanisms. Any explanation mechanism for deep learning models needs to approximate thousands or millions of abstract model parameters into a representation that is comprehensible by humans. Rule-based mechanisms that encode human knowledge would offer a unique opportunity for obtaining good automatic explanations. They can act as examples for explanations that relate data to outcomes and which can be understood by humans, and an algorithm can try to find other rules similarly suited for human comprehension. If a system is trained using human-designed rules (as suggested in Objective 1) additional relevant research questions can be posed, such as: Which rule was the most probable cause for the prediction? Which data points exemplify the rule best? Which mechanism was used to generalize from the underlying rule? Which features correlate most with the rule?

Scientific disciplines:
Artificial intelligence (25%) | Computational linguistics (25%) | Artificial neural networks (25%) | Knowledge engineering (25%)

Keywords:
Natural Language Processing; Neural Networks; Weak Supervision

---

VRG leader:	Benjamin Roth
Institution:	LMU Munich
Proponent:	Claudia Plant
Institution:	Data Science @ Uni Vienna

Status: Ongoing (01.09.2020 - 31.08.2028)

Further links to the persons involved and to the project can be found under
https://wwtf.at/funding/programmes/vrg/VRG19-008/