László Grad-Gyenge

EIT Digital Budapest DTC

PhD topic: Grammar Induction

PhD supervisor: Dr. András Kornai, SZTAKI

Contact: laszlo.grad-gyenge@creo.hu



'Besides my PhD studies, EIT Digital improves my skill to see my research area through the eyes of a startupper. EIT Digital also provides access to the valuable network of the European startup scene such as researchers, innovators and funding partners.'

Achievements & further plans

His research area is **natural language processing**. The more specific field he works in is **grammar induction**. He is familiar with the mathematical tools serving this area, such as probabilistic methods, neural networks, deep learning, machine learning, representation learning and mathematical optimization. **State-ofthe-art** semantic and representation methods focus on mapping words to a vector space using various models. His goal is to generalize this representation in the same space to the level of phrases or sentences. The method is based on the simultaneous learning of the representation and the transformation. The transformation combines the representation of the words (or phrases) with the representation of their phrase.

At the moment, I am working on 7 startups. In my work I utilize my experience in research areas such as natural language processing, recommender systems and machine learning.

EIT Digital DTC educational phase at Spring semester of 2016:

Reserach topic

Grammar induction is one of the core problems of natural language processing. Its goal is to reveal the linguistic structure behind human languages. The linguistic capabilities of humans are helped by the contextual situation and the additional meta-communication channels. Humans are able to acquire linguistic knowledge via mostly unsupervised learning. Grammar induction examines the structure of human language



whose structure is distributed in various representations. Grammar induction tries to model hierarchical information in human language. By evolving grammar induction methods, human communication can be encoded in a more essential form. The aim of the research is to enhance current grammar induction methods. Currently, finite state automata, formal grammars (context free grammars), probabilistic models, conditional random fields, deep learning techniques, sparse representation techniques (lasso,



spike and slab coding) are applied for grammar inference. To accomplish this task, various machine learning methods are used and information representation methods are being developed, with a special focus on graph related methods. Current grammar induction methods rarely use a knowledge base to enhance inference quality. The aim of the research is to develop grammar induction methods which operate with the help of a knowledge graph or a semantic network.



Digital Doctoral school

doctoralschool.eitdigital.eu