School of **Computer Science**

MSC.CS Defence

Friday, December 18, 2020 at 1:00 pm on Teams

Inducing Grammar from Sparse Data Sets

Labeeb Khan

Chair: Dr. Gary Grewal
Advisory Member: Dr. Stefan Kremer
Advisory Member: Dr. Luiza Antonie
Non-Advisory Member: Dr. David Calvert

Abstract

This thesis proposes new tools and methods for identifying a regular
language based on a finite set of example strings. There have been a number of
different approaches proposed for grammar induction. Current state of the art
solutions leverage a depth-first search strategy, which is an evidence- based state
merging strategy or a breadth-first search strategy. We propose two solutions.
The first uses a simulation-based approach that samples the set
of deterministic finite state automata that represent the example strings and can
recognize the language generated by an unknown grammar. The second is
a genetic algorithm approach that generates and evolves a population of
automata to learn the example strings labelled by an unknown
grammar.