UNIVERSITY of GUELPH

COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

MSc Seminar

Wednesday September 11, 2019 at 1:00PM in Reynolds, Room 2224

Induction of Regular Languages using Recurrent Neural Networks

Labeeb Khan

Advisor: Dr. Stefan C Kremer

Advisory Committee: Dr. Luiza Antonie

ABSTRACT:

A finite-state automaton (FSA) is a state machine that can be in one of a number of states at a given time. The machine can transition between states in response to an input, known as a 'transition'. FSA can be defined by setting a number of states with an accept or reject label for each state, setting an initial state, and having transition conditions or rules for each state. There are two types of FSA, deterministic (DFA) and non deterministic (NFA). A deterministic finite state automaton (DFA) is a special case of a NFA where there exists a transition rule for each state with each input and an accept or reject label for each state. DFA and NFA recognize the set of regular languages, which are formal languages that can be expressed by a regular expression.

We can use a set of sample strings constructed from the DFA's alphabet, and label these strings using a target DFA to construct a dataset. This dataset can be represented by a trie-data structure called a prefix tree acceptor (PTA), which is a kind of search tree used to store an associate array where the keys are usually strings. We can use a merging strategy (depth-first search or breadth-first search) to merge the nodes of the PTA and collapse the data structure to a DFA. Different merging strategies will result in different DFA. Reconstructing the target DFA becomes a challenging combinatoric problem as there exists a large number of possible state-minimized automata generated by the minimization. Most state of the art approaches in DFA learning have leveraged a breadth first search (BFS) or depth first search (DFS) strategy that has proven to be successful on a number of problems. There exists a set of DFA learning problems in the Abadingo Competition from 1997 which the state of the art algorithms had failed to solve. Our goal is to create an algorithm that can learn an unknown target DFA given the dataset and apply the solution to the Abadingo problem sets. The seminar will introduce the problem of learning a DFA given a dataset. We will propose the creation and design of a learning algorithm that can learn the target DFA using the dataset. The algorithm will leverage recent optimization strategies used in neural networks like gradient descent to learn the target DFA. This approach avoids the traditional BFS / DFS strategy and searches the space of possible minimized automata to learn the target DFA. Each configuration of the model architecture can represent one or more minimized automata, and a parameter search can be conducted on the possible configurations to learn the best automaton that characterizes the target DFA.