Learning Arithmetic from Handwritten Images with the Aid of Symbols

Daniel L. Silver, Ahmed Galila

Theory

Background:

- Many machine learning problems consider making one or more predictions based on a sequence of input data
- LSTM RNNs are capable of developing such models, most notably *seq2seq* models that develop a form of algorithm that inputs a sequence of values such as "3 + 2 =" to produce an output "5"
- Deep LSTM RNNs may increase the expressiveness of a model, but make the search for a global minimum more challenging - the complex topology of the network adds regions in the loss function that act as local minima

Definition:

A **symbol** is a noise-free, concise and consistent external representation of a concept that captures its key regularities or semantics and is meant to be shared between learning agents

0	0	0	0	1	1	1	1	1

Top: Example of a symbol representing the concept of the digit "5" Bottom: Noisy examples of the digit "5"

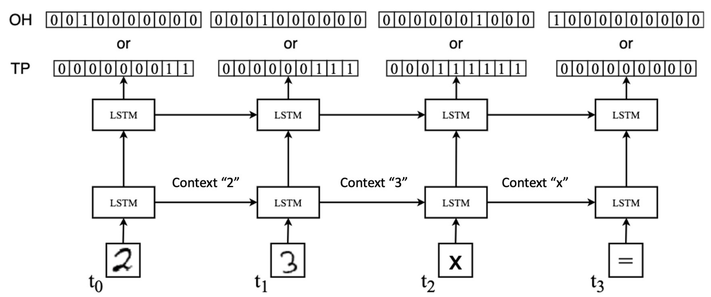
Hypothesis:

Similar to how individual human learners overcome the difficulties they encounter while learning an algorithm that manipulates a sequence of noisy examples of concepts, a RNN can better learn an algorithm by concurrently learning to classify each noisy example using a symbolic label of the concept

Approach

Objectives:

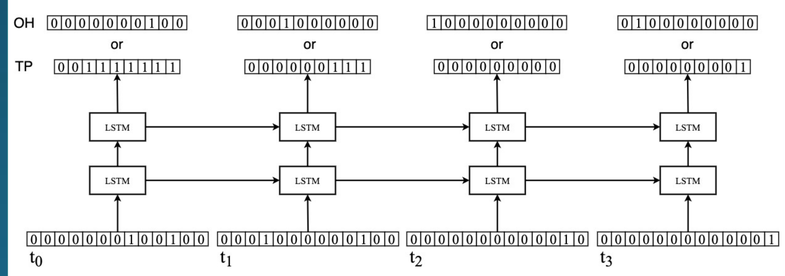
- Show that a RNN that is trained to map noisy examples of concepts to symbols while also learning an algorithm that manipulates those concepts performs significantly better than one trained without the use of symbols
- Show that, with properly encoded symbols, a RNN can develop an algorithm for arithmetic operations that generalizes to unseen combinations of digits



A sequential model learning to perform arithmetic (in this case, 2 x 3) using noisy inputs with one-hot or thermometer encoded symbols

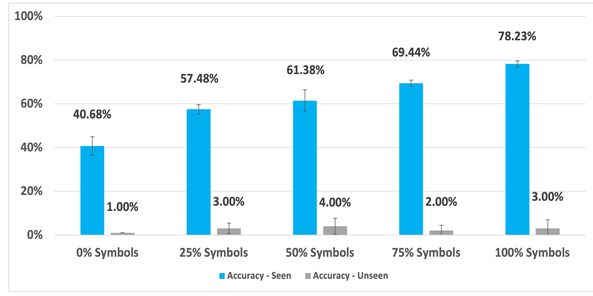
$$h_{ML} = \max_h p(C, D|h) = p(D|C, h).p(C|h)$$

A RNN that correctly maps the noisy digits to their symbolic class labels increases the likelihood that it will discover a hypothesis that can accurately do arithmetic using those digits

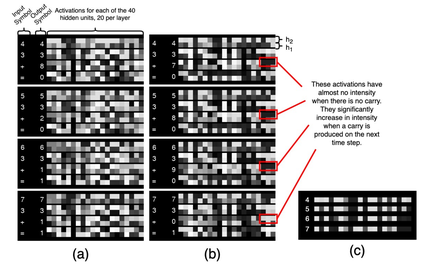


The constrained symbol-only architecture trained to add 7 and 3. The output is represented as a one-hot vector (OH) or as a thermometer (TP)

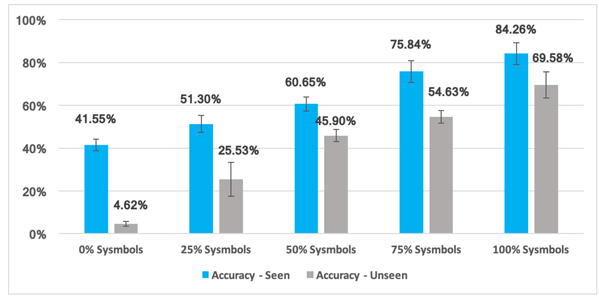
Results



Exp 1: Learning the arithmetic operations with differ percentage of noise



Exp 2: Encoding the semantics of digits



Exp 3: Learning the operations using thermometer encoded symbols

Conclusion:

- It is possible to develop a RNN model that can learn an algorithm from an impoverished set of examples more effectively in the presence of symbols than without; and
- Networks trained **using appropriate symbols** are capable of discovering an algorithm that generalizes better to all combinations of digits, instead of simply learning an input to output mapping function