HOList: An Environment for Machine Learning of Higher-Order Theorem Proving

Kshitij Bansal, Sarah M. Loos, Markus N. Rabe, Christian Szegedy, Stewart Wilcox

Google Research
Can we create a human level AI to reason about mathematics?
HOList
An Environment for Machine Learning of Higher-Order Theorem Proving

- HOList provides a simple API for ML researchers and theorem prover developers to experiment with using machine learning for mathematics.
- We use deep networks trained on an existing corpus of human proofs to guide the prover.
- We can improve our results by adding synthetic proofs (generated from supervised models and verified correct by the prover) to the training corpus.
APIs for Theorem Prover Developers and ML Researchers

Proof Assistant

Proof Search

One goal/subgoal to prove
One proof step:
Tactic application, relevant premises

Subgoals or *proved*

Ranking of tactics and premises

One goal/subgoal to prove

Machine Learning
## Results - Supervised Learning on Human Proofs

<table>
<thead>
<tr>
<th>Model</th>
<th>Percent of Validation Theorems Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong>: ASM_MESON_TAC</td>
<td>6.10%</td>
</tr>
<tr>
<td>ASM_MESON_TAC + WaveNet premise selection</td>
<td>9.20%</td>
</tr>
<tr>
<td>Wavenet</td>
<td>31.72%</td>
</tr>
<tr>
<td>Deeper WaveNet</td>
<td>32.65%</td>
</tr>
<tr>
<td>Wider WaveNet</td>
<td>27.60%</td>
</tr>
</tbody>
</table>
## Results - Supervised Learning on Human Proofs

<table>
<thead>
<tr>
<th>Method</th>
<th>Percent of Validation Theorems Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong>: ASM_MESON_TAC</td>
<td>6.10%</td>
</tr>
<tr>
<td>ASM_MESON_TAC + WaveNet premise selection</td>
<td>9.20%</td>
</tr>
<tr>
<td>Wavenet</td>
<td>31.72%</td>
</tr>
<tr>
<td>Deeper WaveNet</td>
<td>32.65%</td>
</tr>
<tr>
<td>Wider WaveNet</td>
<td>27.60%</td>
</tr>
</tbody>
</table>
# Results - Supervised Learning on Human Proofs

<table>
<thead>
<tr>
<th>Model</th>
<th>Percent of Validation Theorems Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong>: ASM_MESON_TAC</td>
<td>6.10%</td>
</tr>
<tr>
<td>ASM_MESON_TAC + WaveNet premise selection</td>
<td>9.20%</td>
</tr>
<tr>
<td>Wavenet</td>
<td>31.72%</td>
</tr>
<tr>
<td><strong>Deeper WaveNet</strong></td>
<td><strong>32.65%</strong></td>
</tr>
<tr>
<td>Wider WaveNet</td>
<td>27.60%</td>
</tr>
</tbody>
</table>
## Results - Prover in the loop

<table>
<thead>
<tr>
<th>Loop Type</th>
<th>Percent Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavenet Loop</td>
<td>36.30%</td>
</tr>
<tr>
<td>- Trained on loop output</td>
<td>36.80%</td>
</tr>
<tr>
<td>Tactic Dependent Loop</td>
<td>38.90%</td>
</tr>
</tbody>
</table>

The graph shows the percent proved over loop rounds for different loops.

- **Loop on subgoals**
- **[GenTypes/Vars] Loop Tac Dep**
- **Loop**
- **Loop tactic dependent**

*Note: The data for Tactic Dependent Loop is highlighted.*
APIs for Theorem Prover Developers and ML Researchers

- **Proof Assistant**
  - One goal/subgoal to prove
  - One proof step:
    - Tactic application, relevant premises
  - Subgoals or *proved*

- **Proof Search**
  - Ranking of tactics and premises
  - One goal/subgoal to prove

- **Machine Learning**
  - One goal/subgoal to prove
APIs for Theorem Prover Developers and ML Researchers

**Prover**

HOL-Light

Input:
- Load premises
- Apply a tactic to a goal

Output:
- Open goals left to prove

**Proof Search**

- Manages the state of the proof search tree.
- Allows arbitrary nodes to be explored.

**Supervised Learning**

Training Data:
*TF Examples* from Human & Synthetic Proofs

Features:
- Goal (or subgoal)

Labels:
- Tactic applied
- Premises used
deephol.org

- Code is available on GitHub
- Training data
  - 30K theorems and definitions
    In the areas of: topology, multivariate calculus, real and complex analysis, geometric algebra, measure theory, etc., as well as the formal proof of the Kepler Conjecture.
  - 375K human proof steps
  - 830K synthesized proof steps
- Trained model checkpoints
- Docker images for the proof assistant and proof search