Combining Deep Learning and Qualitative Spatial Reasoning to Learn Complex Structures from Sparse Examples with Noise

Nikhil Krishnaswamy, Scott Friedman, and James Pustejovsky

nkrisna@brandeis.edu • friedman@sift.net • jamesp@brandeis.edu

Introduction

• 17 staircases constructed by a naive user in a live multimodal interaction with an avatar
• Noisy, sparse samples
• Variant configurations
• Not isomorphic
• Satisfactory to at least one person
• Can an algorithm infer and reproduce commonalities?
• Blocks are interchangeable
• Direction-independent
• Stacks get progressively higher

“This is a staircase”

Data Gathering

• Study: Krishnaswamy and Pustejovsky (2018)
• Gesture and language interaction, definition of success up to subject
• Blocks world in 3D opens search space to all 3D variation
• Same label may have enormous search space of relations
• Difficulty using the system:
  • Hard to point accurately
  • User failure to discover gesture for actions

Learning Framework

First Move Selection
• MLP samples from training data
• 4x64 dense ReLU layers, RMSProp, sigmoid activation
• Input: 2 randomly chosen blocks; Output: relation

Reference Example Selection
• CNN predicts known sample from current configuration
• Highly inaccurate at start, less later

Next Move Prediction
• LSTM predicts moves to approach example
• 3x32 LSTM, RMSProp, Softmax over n timesteps (n = # relations defining example)
• Input: closest match of current state; Output: remaining relations

Results

Evaluator asked to score generated structure by how much (0-10) it resembles a staircase

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Avg. Score (μ)</th>
<th>Std. Dev. (σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chance</td>
<td>2.0375</td>
<td>1.0122</td>
</tr>
<tr>
<td>JD</td>
<td>4.3375</td>
<td>2.0387</td>
</tr>
<tr>
<td>LD</td>
<td>3.7688</td>
<td>2.1028</td>
</tr>
<tr>
<td>SPIRE</td>
<td>5.8313</td>
<td>2.7173</td>
</tr>
<tr>
<td>Comb.</td>
<td>4.7188</td>
<td>2.4309</td>
</tr>
</tbody>
</table>

Some desired inferences

Heuristics select best move toward example (from CNN) out of move options (from LSTM)
• Chance, Jaccard Distance, Levenshtein Distance, SPIRE graph matcher, combined (SPIRE + LD)

SPIRE computes state graph of relations that would hold after a move option, score maximal common subgraph (MCS) with the goal state, and chooses best