

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

Nikhil Krishnaswamy, Scott Friedman, and James Pustejovsky  
[nkrishna@brandeis.edu](mailto:nkrishna@brandeis.edu) • [friedman@sift.net](mailto:friedman@sift.net) • [jamesp@brandeis.edu](mailto:jamesp@brandeis.edu)



Brandeis University



## 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
- Extracted qualitative relations between blocks in structure (RCC8, RCC-3D, TPCC, QSRLib)

## Learning Framework

### First Move Selection

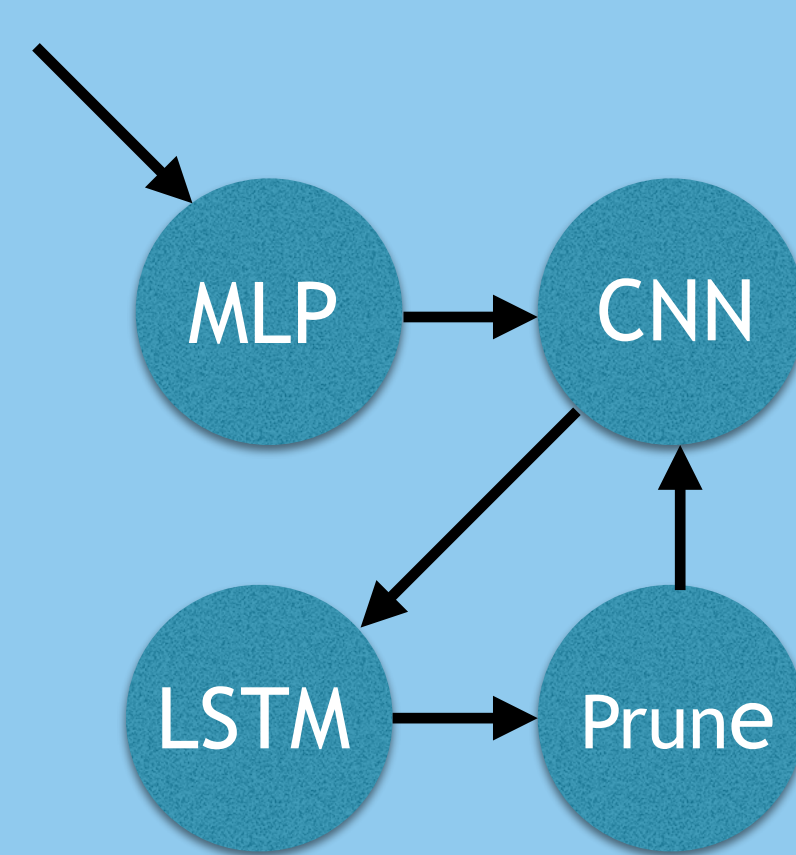
- MLP samples from training data
- 4x64 dense ReLU layers, RMSProp, sigmoid activation
- Input: 2 randomly chosen blocks; Output: relation

### 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



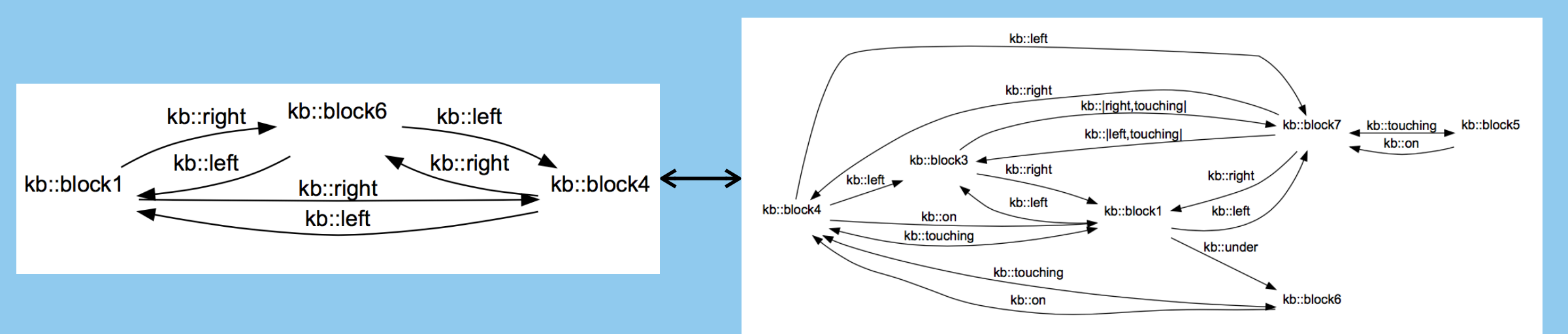
```

right block7 block1      right,touching block6 block7
touching block3 block1  right block5 block1
left block1 block5      under,touching, support block7 block5
left block1 block7      under,touching, support block1 block3
under,touching, support block3 block4  touching block5 block7
touching block6 block5  right block5 block3
under block1 block4     block7 <359.883; 1.222356; 359.0561>
touching block4 block3  block1 <0; 0; 0>
left block3 block5     block6 <0.1283798; 359.5548; 0.9346825>
left block1 block6     block3 <0; 0; 0>
left,touching block7 block6  block5 <0; 0; -2.970282E-08>
right block6 block1     block4 <0; 0; 0>
    
```

Example relation set

## Heuristics and Graph Matching

- 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)



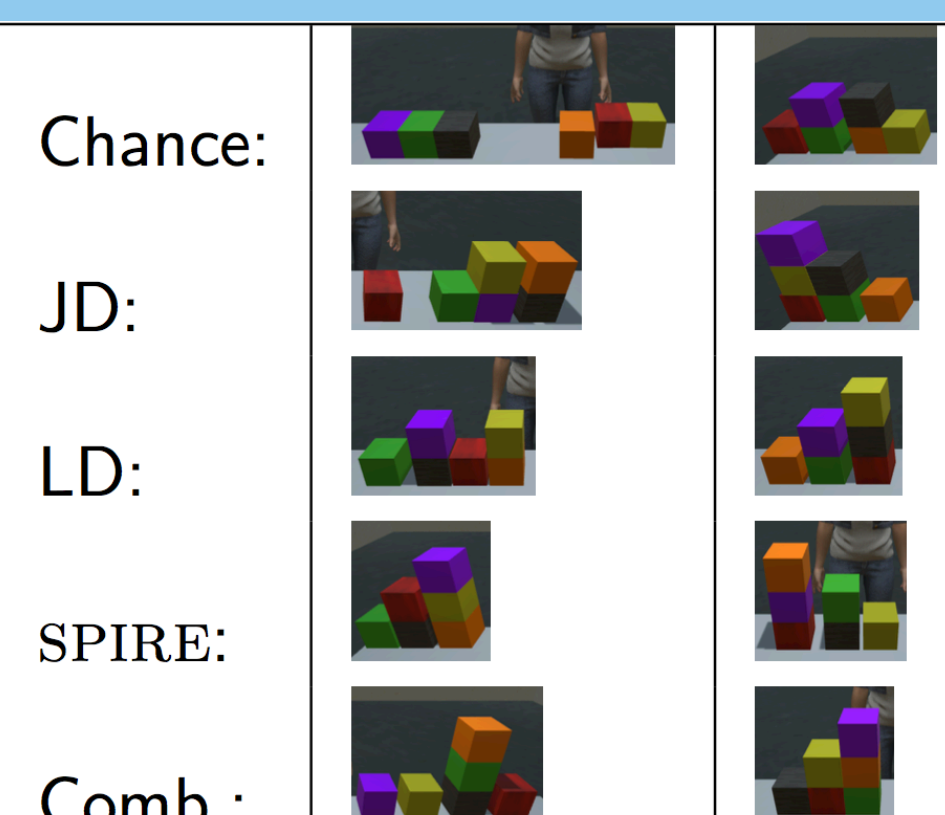
SPIRE computes state graph of relations that would hold after a move option, score maximal common subgraph (MCS) with the goal state, and chooses best

## Results

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

Heuristic	Avg. Score ( $\mu$ )	Std. Dev. ( $\sigma$ )
Chance	2.0375	1.0122
JD	4.3375	2.0387
LD	3.7688	2.1028
SPIRE	5.8313	2.7173
Comb.	4.7188	2.4309

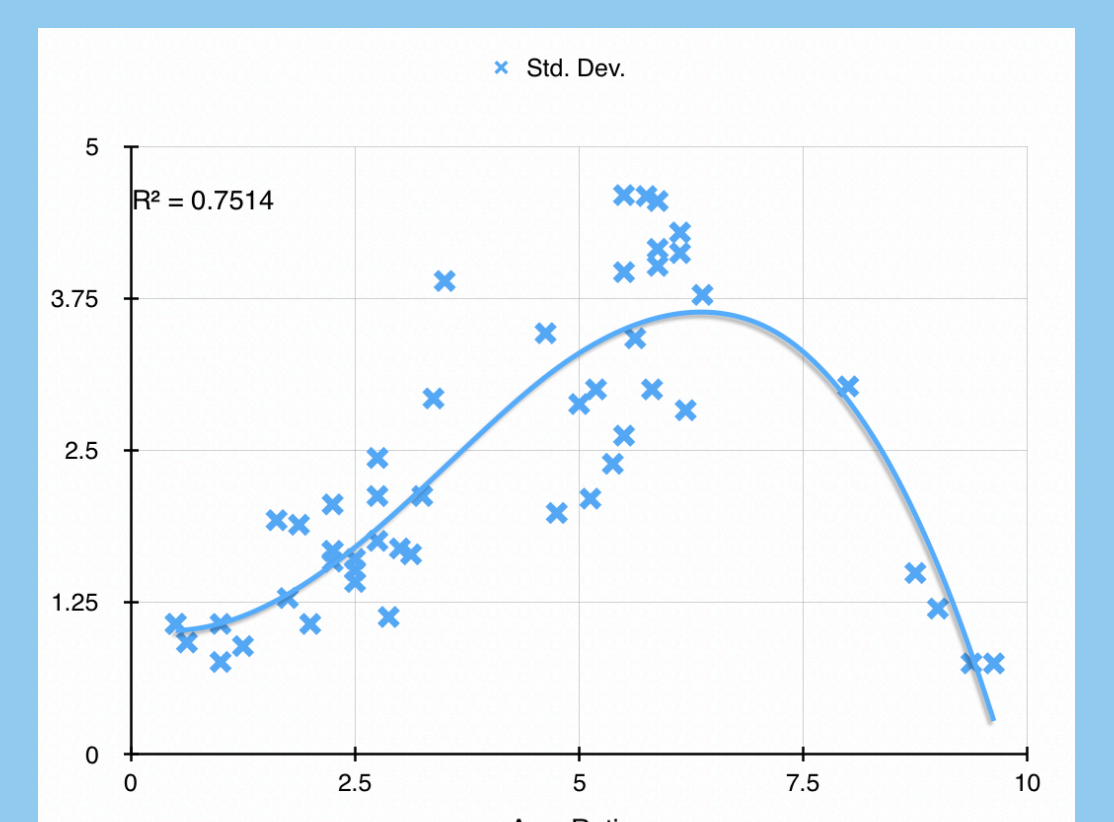
Evaluator judgment of structure quality



Median + best structures



Some desired inferences



Avg. score vs. std. dev.