Extreme-scale Neural Models

1. Physical Dynamics Model

   Language Model

2. Multimodal Script Knowledge

3. Causal Commonsense

   Language models $\neq$ knowledge models

   Social-interaction Commonsense

   Physical-entity Commonsense

   Event-centered Commonsense

4. Algorithm

Research Directions in the Era of Extreme-Scale Neural Models

1. Learning with **Interactions** in a 3D World
2. Learning from Complex **Multimodality**
3. Learning from Symbolic **Knowledge**
4. (Unsupervised) Inference-time **Algorithms**

===

"Smaller but better"

- **1 + 2 :** diversifying learning signal, emphasis on grounding, emphasis on the complexity of information inherent in the data, emphasis on learning knowledge about the world
- **3** : the importance of declarative knowledge as additional learning signal
- **4** : the importance of inference-time algorithms, large-scale seq-2-seq is not always the winning recipe

- cloze-like masks work
- No-cloze
- Uniform

Validation: from statistical to syntactic dependence

Hypothesis:

Learning statistical dependencies is useful because it correlates with useful linguistic structures such as syntax

This is consistent with earlier observations that MLMs seem extremely good at learning syntactic structure (compared to reasoning tasks).

How can we validate this hypothesis?

- Recover the statistical dependencies learned by BERT
- See whether these dependencies correspond to edges on a dependency parse