A compositional object-based approach to learning physical dynamics

Summary

1. factorization of physical scene into composable obj-based representation
2. a NN architecture whose compositional structure factorizes obj dynamics into pairwise interactions.
3. The approach can generalize to variable no of obj count and infer latent properties of obj such as mass

There are 4 main components:

- Obj-based representation
- Pairwise factorization
- Context Selection
- function Composition

Obj-based Representation

1. The state vector comprises:
   - extrinsic properties (position, velocity, etc)
   - intrinsic properties (mass, obj type, etc)
   - global properties (gravitional, frictional, etc)
2. Presumably, those properties are given.

Pair-wise factorization

1. Let a particular obj be the focus obj f and all other obj in the scene be context obj c.
2. An encoder receives the state o_f and o_c for each c and output an embedding.
3. The sum of the embedding and the focus obj's post state are used as input to the decoder function.
4. The decoder then predicts the focus obj's velocity $v_f^{[t+1]}$.

Context Selection

1. Each (o_f, o_c) pair is selected to be in the set of neighbors of f by a neighborhood masking function $\mathbb{1}[\|p_c - p_f\| < N(o_f)]$