Jointly learning motion verbs and frame semantics from natural language and grounded scenes

Jon Gauthier

Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

Jiayuan Mao

Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

Tianmin Shu

Massachusetts Institute of Technology, Cambrdige, Massachusetts, United States

Roger Levy

Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

Josh Tenenbaum

MIT, Cambridge, Massachusetts, United States

Abstract

We propose a computational model of verb learning implemented as a probabilistic compositional semantic parser, that jointly learns individual verb meanings and overarching associations between syntactic verb frames and compositional semantic predicates from distant supervision on grounded natural language data. In tandem, we present a new corpus for training and evaluating grounded language learning models, containing natural language descriptions of scenes generated in a rich environment that simulates realistic interactions between animate agents and physical objects. We demonstrate how the model can acquire interpretable correspondences between syntactic frames by incrementally parsing individual sentences, evaluating candidate verb meanings on grounded scenes, and investigate how the models acquired frame semantics priors generalize to support efficient inferences about the meanings of novel verbs on a few shot learning task.