Recursive Neural Networks for Learning Logical Semantics

Samuel R. Bowman[*†]
sbowman@stanford.edu

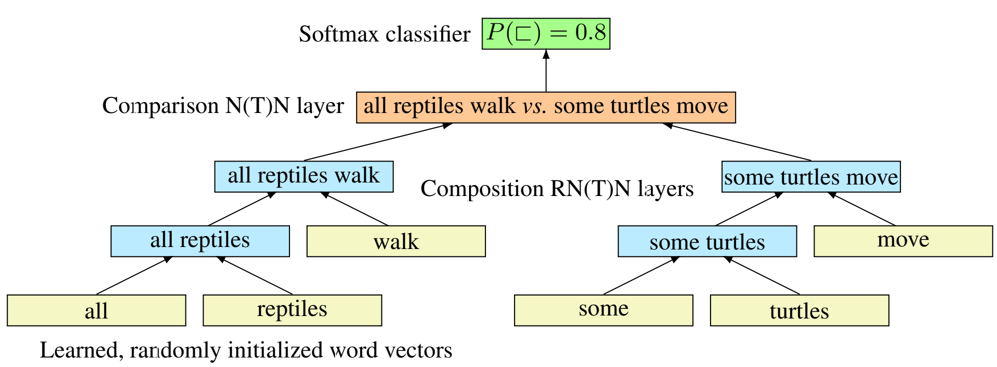
Christopher Potts[*]
cgpotts@stanford.edu

Christopher D. Manning[*†‡]
manning@stanford.edu

[*]Stanford Linguistics
[†]Stanford NLP Group
[‡]Stanford Computer Science

Supervised recursive neural network models (RNNs) for sentence meaning have been successful in an array of sophisticated language tasks, but it remains an open question whether they can learn compositional semantic grammars that support logical deduction. We address this question directly by for the first time evaluating whether each of two classes of neural model — plain RNNs and recursive neural tensor networks (RNTNs) — can correctly learn relationships such as entailment and contradiction between pairs of sentences, where we have generated controlled data sets of sentences from a logical grammar. Our first experiment evaluates whether these models can learn the basic algebra of logical relations involved. Our second and third experiments extend this evaluation to complex recursive structures and sentences involving quantification. We find that the plain RNN achieves only mixed results on all three experiments, whereas the stronger RNTN model generalizes well in every setting and appears capable of learning suitable representations for natural language logical inference.

Recursive neural network models



In our experiments, we train pairs of recursive (tree structured) neural network models [1] which are joined together with a shared top layer that generates features for a classifier. The classifier predicts the logical relation that holds between the sentences represented by the two trees. For an activation function, we use either a plain NN layer or a tensor combination layer.

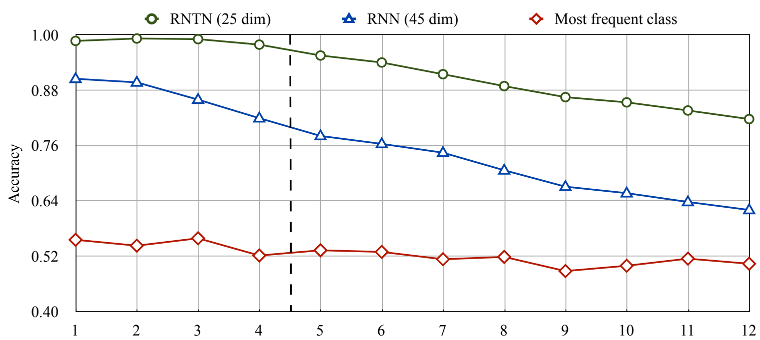
Reasoning about semantic relations

Our models classify into the seven relations from [2], shown in the table below, which define possible relationships between pairs of terms or sentences of natural language in terms of their set-theoretic denotations. If any model is to learn the behavior of a relational logic like the one presented here from a finite amount of data, it must learn to deduce new relations from already seen relations. Our first experiment evaluates the ability of our models to do this over pairs of atomic symbols in a large corpus of artificial data. The model is trained on examples like $\{a \sqsubset b, b \wedge c\}$, and tested on examples that follow from them, like $\{a \mid c\}$.

Name	Symbol	Set-theoretic definition	Example
entailment	$x \sqsubset y$	$x \subset y$	*turtle, reptile*
reverse entailment	$x \sqsupset y$	$x \supset y$	*reptile, turtle*
equivalence	$x \equiv y$	$x = y$	*couch, sofa*
alternation	$x \mid y$	$x \cap y = \emptyset \wedge x \cup y \neq \mathcal{D}$	*turtle, warthog*
negation	$x \,^\wedge\, y$	$x \cap y = \emptyset \wedge x \cup y = \mathcal{D}$	*able, unable*
cover	$x \smile y$	$x \cap y \neq \emptyset \wedge x \cup y = \mathcal{D}$	*animal, non-turtle*
independence	$x \# y$	(else)	*turtle, pet*

Recursive structure in propositional logic

Our second experiment introduces compositionality to our examples, training on short statements of propositional logic, like *not* $a \sqsupset (a$ *and* $b)$. We train our models on only pairs of statements with up to four symbols, but observe that the RNN performs reasonably both on those and on much longer test pairs.



Reasoning with natural language quantifiers and negation

For our third experiment, we generate pairs of sentences in which each sentence contains one quantifier, and any of a small set of common nouns, as in the example *(no warthogs) move* $\sqsubset$ *(no (not reptiles)) swim*. The parentheses indicate the tree structure for each sentence as it will be used by the model. We defined several different types of train–test split for this experiment. A tuned RNTN model performed either well ($> 85\%$ accuracy) or perfectly on all of them, while a tuned RNN did not break 80% in any setting.

References

[1] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of EMNLP*, pages 1631–1642, 2013.

[2] Bill MacCartney and Christopher D. Manning. An extended model of natural logic. In *Proceedings of IWCS*, pages 140–156, 2009.