Grounding Compositional Hypothesis Generation

Overview

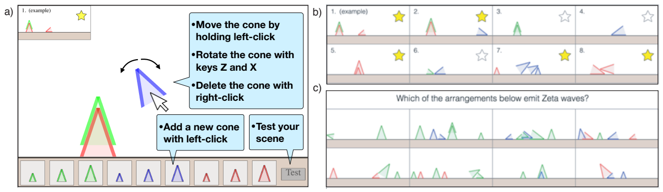
Many learning models assume a fixed hypothesis space, but this is not appropriate for most real-world learning

How do people come up with new theories and hypotheses?

- **Top down?** Can approximate Bayesian inference by sampling from compositional grammar prior expressing infinite class of possible hypotheses (cf, Piantadosi, Tenenbaum, & Goodman, 2016), but inefficient & costly
- **Bottom up?** We propose learners construct hypotheses semi-stochastically inspired by evidence, using grammar to describe observed features and relationships
- **Bottom up** generation more sample efficient + accounts better for human inferences

Task

Try it `https://github.com/neilbramley/discovery`



(a). 30 mTurkers construct and test "scenes" of simple objects called "cones"

(b). Try to infer the rule that makes some produce radiation (yellow stars).

(c). We probe learning through test choices, generalization and free description

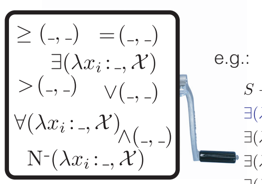
Test rules

	General	Specific	Example
1.	*Pair–value:*	There's a red $\exists(\lambda x_1: \; =(x_1, \text{red}, \text{color}), \mathcal{X})$	
2.	*Match:*	They're all the same size $\forall(\lambda x_1: \forall(\lambda x_2: \; =(x_1, x_2, \text{size}), \mathcal{X}), \mathcal{X})$	
3.	*Negation:*	Nothing is upright $\forall(\lambda x_1: \; \neg(=(x_1, \text{upright}, \text{orientation})), \mathcal{X})$	
4.	*Numerosity:*	There is exactly 1 blue $\text{exactly}(\lambda x_1: \; =(x_1, \text{blue}, \text{color}), 1, \mathcal{X})$	
5.	*Conjunct:*	There's something blue and small $\exists(\lambda x_1: \; \wedge(=(x_1, \text{blue}, \text{color}), =(x_1, 1, \text{size}), \mathcal{X})$	

Test rules continued...

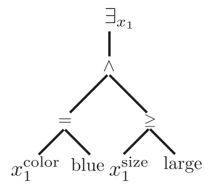
	General	Specific	Example
6.	*Disjunct:*	All are blue or small $\forall(\lambda x_1: \; \vee(=(x_1, \text{blue}, \text{color}), =(x_1, 1, \text{size}), \mathcal{X})$	
7.	*Relative property:*	A red is the largest piece $\exists(\lambda x_1: \forall(\lambda x_2: \wedge(=(x_1, \text{red}, \text{color}), >(x_1, x_2, \text{size})), \mathcal{X}), \mathcal{X})$	
8.	*General relation:*	Some pieces are touching $\exists(\lambda x_1: \exists(\lambda x_2: \Gamma(x_1, x_2, \text{contact}), \mathcal{X}), \mathcal{X})$	
9.	*Specific relation:*	A blue touches a red $\exists(\lambda x_1: \exists(\lambda x_2: \wedge(\wedge(=(x_1, \text{blue}, \text{color}), =(x_2, \text{red}, \text{color})), \Gamma(x_1, x_2, \text{contact})), \mathcal{X}), \mathcal{X})$	
10.	*Complex:*	Some pieces are stacked $\exists(\lambda x_1: \exists(\lambda x_2: \wedge(\wedge(\wedge(\wedge(\wedge(=(x_1, \text{upright}, \text{orientation}), =(x_1, \text{yes}, \text{grounded})), =(x_2, \text{upright}, \text{orientation})), =(x_2, \text{no}, \text{grounded})), =(x_1, x_2, \text{xpos}), \Gamma(x_1, x_2, \text{contact})), \mathcal{X}), \mathcal{X})$	

"Top down" Probabilistic Context-free Generation

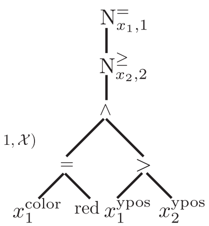


e.g.:

$S \rightarrow$
$\exists(\lambda x_1: A, \mathcal{X}) \rightarrow$
$\exists(\lambda x_1: B, \mathcal{X}) \rightarrow$
$\exists(\lambda x_1: H(B, B), \mathcal{X}) \rightarrow$
$\exists(\lambda x_1: \wedge(=(x_1, D1), I(x_1, D2)), \mathcal{X}) \rightarrow$
$\exists(\lambda x_1: \wedge(=(x_1, \text{blue}, \text{colour}), \geq(x_1, \text{medium}, \text{size})), \mathcal{X})$
There is a blue cone that is at least medium sized



$S \rightarrow$
$N^{=}(\lambda x_1: A, 1, \mathcal{X})$
$N^{=}(\lambda x_1: S, 1, \mathcal{X})$
$N^{=}(\lambda x_1: N^{\geq}(\lambda x_2, A, 2, \mathcal{X}), 1, \mathcal{X}):$
$N^{=}(\lambda x_1: N^{\geq}(\lambda x_2, B, 2, \mathcal{X}), 1, \mathcal{X}):$
$N^{=}(\lambda x_1: N^{\geq}(\lambda x_2, \wedge(B, B), 2, \mathcal{X}), 1, \mathcal{X}):$
$N^{=}(\lambda x_1: N^{\geq}(\lambda x_2: \wedge(=(x_1, D1), \; I(x_1, x_2, E2)), 2, \mathcal{X}), 1, \mathcal{X})$
$N^{=}(\lambda x_1: N^{\geq}(\lambda x_2: \wedge(=(x_1, \text{red}, \text{colour}), \; >(x_1, x_2, \text{ypos})), 2, \mathcal{X}), 1, \mathcal{X})$
One red cone is above at least two other cones



- Sample rules by combining primitives (∧, ∨, ≥, features, relations, etc) using probabilistic generative grammar
- Many sampled hypotheses contradictory or inconsistent with observations

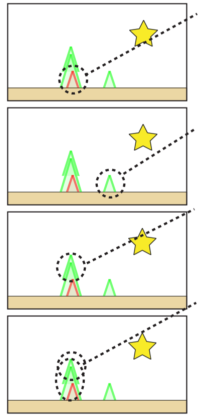
References

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*(2), 211.

Lewis, O., Perez, S., & Tenenbaum, J. (2014). Error-driven stochastic search for theories and concepts. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (Vol. 36). Cognitive Science Society Austin, TX.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological Review, 123*(4), 392.

"Bottom up" Instance-Driven Generation



red, colour $\rightarrow$
$=(x_1, \text{red}, \text{colour}) \rightarrow$
$\wedge(=(x_1, \text{red}, \text{colour}), =(x_1, \text{small}, \text{size})) \rightarrow$
$\exists(\lambda x_1: \; \wedge(=(x_1, \text{red}, \text{colour}), =(x_1, \text{small}, \text{size}), \mathcal{X})$
There is a small red cone

green, colour $\rightarrow$
$=(x_1, \text{green}, \text{colour}) \rightarrow$
$N^{\geq}(\lambda x_1: \; =(x_1, \text{green}, \text{colour}), 3, \mathcal{X})$
There are at least three green cones

medium, size $\rightarrow$
$\geq(x_1, \text{medium}, \text{size}) \rightarrow$
$N^{=}(\lambda x_1: \; x_1, \geq(\text{medium}, \text{size}), 2, \mathcal{X})$
Two cones of at least medium size

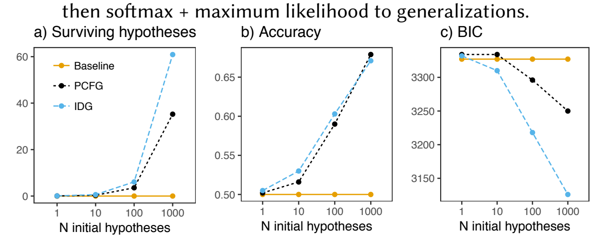
contact $\rightarrow$
$=(x_1, x_2, \text{contact}) \rightarrow$
$\wedge(\wedge(=(x_1, \text{green}, \text{colour}), =(x_2, \text{green}, \text{colour})), =(x_1, x_2, \text{contact})) \rightarrow$
$\exists(\lambda x_1: \; \exists(\lambda x_2: \; \wedge(\wedge(=(x_1, \text{green}, \text{colour}), =(x_2, \text{green}, \text{colour})), =(x_1, x_2, \text{contact})), \mathcal{X}), \mathcal{X})$
Two green cones touch

1. **Observe:** sample 1-2 cones + features from rule following scene
2. **Functionalize:** greedily sample true (in)equality statement about chosen cone(s)
3. **Extend:** conjunctively or disjunctively with some probability
4. **Quantify:** select true quantifiers

Results

Model	-LogL	BIC	τ	ES	N/30	Acc
Baseline	1663	3327	∞		17	0.500
PCFG	1594	3195	1.54	352	3	0.722
IDG	**1539**	**3085**	**1.01**	**610**	**10**	**0.702**

Comparison of inference sampling from probabilistic context free grammar (PCFG) or instance driven grammar (IDG) with feature weights fit to data then softmax + maximum likelihood to generalizations.



(a) Number of candidate hypotheses consistent with data for different N initial samples from PCFG or IDG (b) Average accuracy (c) BIC fit to data

Discussion

- Learners may *adapt* as well as *generate* hypotheses based on data e.g., augment disjunctively after false negative or conjunctively after false positive (cf. Lewis, Perez, & Tenenbaum, 2014).
- Learning benefits from "minimal" positive examples; may relate to positive testing behavior (Klayman & Ha, 1987).

Neil R. Bramley[1], Anselm Rothe[1], Joshua B. Tenenbaum[2], Fei Xu[3] & Todd M. Gureckis[1]

neil.bramley@nyu.edu

[1]New York University, [2]Massachusetts Institute of Technology, [3]University of California, Berkeley