Learning from Noisy Data

Case Study: Word Order

Knowledge of canonical word order is acquired in infancy [1-4].

Possible mechanisms:

- Learning from noun phrase & verb distributions
- Bootstrapping from semantics [5] or prosody [6-7]

Problem: reliably identifying moved clause arguments takes time [4, 8-11]

1. What are you holding?
2. That’s the dog we like.
3. What are you holding?

Bayesian joint inference to select a canonical grammar + filter parameters

(1) What are you holding?
(2) That’s the dog we like.
(3) You’re holding a toy (by your sister).

- How do children avoid being misled by “noise” from non-canonical clause types? [5]

Proposal: Input Filtering

Expect that data are a noisy realization of a deterministic underlying system, and learn to filter noise
- Previously applied to learning of verb transitivity classes [12]
- Current work: this mechanism generalizes to more complex rule systems

Our Model

Observes strings of imperfectly-identified NPs and Vs, considers 4-way choice of canonical word order

- Grammar deterministically places subject before/after VP, object before/after V
- Some parts of strings are generated by “noise” processes: unknown grammatical phenomena that appear to insert, delete, or swap arguments

Bayesian joint inference to select a canonical grammar + filter parameters

- What do the data from the canonical grammar look like?
- What do the data from noise look like?
- What is the right division into signal vs. noise?

What does Filtering Look Like?

From strings of NPs and Vs, make a noisy guess about underlying tree structure

Decide which parts of guessed trees are signal, and which are noise, for an underlying word order grammar

- It is possible to learn how to divide up the data into signal vs. noise, without knowing ahead of time how much noise there is, or what its properties are

Toy Example

How might these data have arisen partially from a word order grammar distribution, and partially from the noise distribution?

Two solutions (of many):

- SVO
- SOV
- OVS
- VOS

Data: 45 strings

- NP V NP
- NP V V
- V NP

80% 20% 100% 0%

Noise Processes

S -> NP VP VP -> V (NP)

Variations

OVS

Simulations: Child-Directed English and French

50-sentence datasets sampled from Eve & Lyon (1984) and CHILDES corpora [13-14]

- Strings of NPs and Vs imperfectly identified from functional cues [15-17]
- Cues for NPs: is a full pronoun, or follows a determiner
- Cues for V: follows an auxiliary

From these strings, our model infers posterior probability distribution over underlying trees and word order grammars

Our Model

Leaver successfully assigns SVO highest posterior probability in both languages

- Even though data cannot be produced by any single word order grammar, without noise

Comparison: “Fully-Flexible” Learner

No 4-way choice of a canonical word order grammar:
- One grammar, all rules possible with some probability [18]

- Collapses distinction in our model between rules for canonical and non-canonical structures
- Learning canonical word order means identifying that some rules have probabilities near zero

Two variants: with and without an explicit bias to regularize (push probabilities towards zero-one) [18-21]

- Learner without bias to regularize infers distributions that mirror its noisy data
- Learner with bias to regularize gives high probability to several canonical word orders

Discussion

We find that input filtering can in principle enable acquisition of basic word order from noisy data

- From imperfectly-identified NP and V distributions alone, our model learns to separate evidence for canonical word order from the disturbing effects of “noise” processes
- It does so without knowing ahead of time what noise looks like, or how much there is

Restrictive options in the learner’s hypothesis space allow successful learning

- Each word order grammar allows only a certain combination of rules
- Preference emerges to use these when possible, rather than analyzing everything as noise

Provides a novel mechanism for regularization in grammar learning [18-21]

Simulations: Child-Directed English and French

- English
- French

0.36 np v
0.20 v
0.20 np v
0.36 np v
0.36 np v
0.20 v
0.20 np v
0.17 v
0.04 np v
0.03 np v
0.03 np v

0.48 np v
0.21 np v
0.13 np v
0.03 np v
0.03 np v

Mean posterior probability

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

Speech and language development [15] is driven by syntax acquisition [16] and processing [17]. Children’s syntactic representations are inferred from their lexical and syntactic capacities. Syntax is acquired through exposure to input, which contains noise [18] and variability [19]. Syntax acquisition is a process of learning both canonical and non-canonical structures [20]. Individuals vary in their ability to learn language, and the specific mechanisms underlying this variation are still not fully understood. This talk will present several research projects that explore different aspects of how children learn syntax and the role of input in this process. These projects include experiments on the role of prosody in syntax learning [21], and work on developing models of syntax acquisition [22].