Grounding NL Syntax in AI Planning

Mark Steedman

Nov 5th 2016
Outline

- Introduction
- I: Planning
- II: Action Representation
- III: From Collaborative Plan Search to Grammar
- Conclusion
Introduction

• There is a long tradition associating language and other serial cognitive behavior with an underlying motor planning mechanism (Piaget 1936, Lashley 1951, Miller et al. 1960).

• The evidence is evolutionary, neurophysiological, and developmental.

• It raises the possibility that language is much more directly grounded in embodied cognition than current linguistic theories of grammar suggest.

• The nature of human language and child language acquisition requires that this cognitive representation be symbolic and essentially homomorphic to language.
Introduction

• I’m going to argue that practically every aspect of language reflects this connection transparently.

The chances of inducing such representations directly from sensory motor data by sheer force of machine learning seem to be zero.

The world is not in fact its own best representation, pace Brooks, 1991.

• The talk discusses this connection in terms of deliberative planning as it is viewed in Robotics and AI, with some attention to applicable machine learning techniques (Steedman 2002a,b).
Introduction

• The paper will define a path between representations at the level of the sensory manifold and perceptron learning to the mid-level of plans and explanation-based learning, and on up to the level of language grammar and parsing model learning.

• At the levels of planning and linguistic representation, two simple but very general combinatory rule types, Composition (the operator $B$) and Type-Raising (the operator $T$) will appear repeatedly.

\[ B f g \equiv \lambda x. f(gx) \quad T a \equiv \lambda f. f a \]
I: Planning and Affordance

• Apes really can solve the monkeys and bananas problem, using tools like old crates to gain altitude in order to reach objects out of reach.

• Such planning involves
  – Retrieving appropriate actions from memory (such as piling boxes on top of one another, and climbing on them),
  – Sequencing them in a way that has a reasonable chance of bringing about a desired state or goal (such as having the bananas).
  – Remembering good plans.

 располагается

It is qualitatively different from Skinnerian shaping in animals like pigeons—cf. http://www.youtube.com/watch?v=mDntbGRPeEU
Figure 1: Köhler 1925
Figure 2: Köhler 1925
Planning and Affordance

- Köhler showed that, in apes at least, such search seems to be
  - reactive to the presence of the tool, and
  - forward-chaining, working forward from the tool to the goal, rather than backward-chaining (working from goal to tool).

- The first observation implies that actions are accessed via perception of the objects that mediate them—in other words that actions are represented in memory associatively, as properties of objects—in Gibson’s 1966 terms, as affordances of objects.

- The second observation suggests that in a cruel and nondeterministic world it is better to identify reasonably highly valued states that you have a reasonable chance of getting to than to optimize complete plans.
Planning and Affordance

• The problem of planning can therefore be viewed as the problem of Search for a sequence of actions or affordances \( \alpha, \beta, \text{etc.} \) in a “Kripke model”:

![S4/Kripke Model of Causal Accessibility Relation]

Figure 3: S4/Kripke Model of Causal Accessibility Relation

• Search is intrinsically recursive, and requires something like a Push-Down Automaton (PDA) to keep track of alternative paths to some limited depth.
II: Representing Actions

- We can think of actions as STRIPS operators or as finite-state transducers (FSTs) over (sparse) state-space vectors.

- FSTs are closed under composition, and can be learned as simple neural computational devices such as Perceptrons, the Associative Network or Willshaw Net (Willshaw 1981 cf. Marr 1969), or their modern equivalents.
III: From Planning to Language

- The STRIPS learner seems comparable to a Piagetian Stage VI child.

- How does the child get from seriation and affordance to language?

- *Seriation* of actions to form a plan is *Composition* of FSTs or functions of type $\text{state} \rightarrow \text{state}$

- The *Affordance* of a state is a function from all those actions that are possible in that state into their respective result states.

- States are defined by the objects they include, so this is like exchanging objects for *Type-Raised* functions that map states into other states resulting from actions on those objects.
From Planning to Language

• Thus, the affordance of a (state including a) box to an ape is a function from actions like the box falling, their climbing-on the box and their putting the box on another box into resulting states whose utility the ape can evaluate.

• The functions are of the following (Curried) types:
  
  – $\text{fall}_e \rightarrow_t$
  – $\text{climb-on}_e \rightarrow (e \rightarrow t)$
  – $\text{put-on}_e \rightarrow (e \rightarrow (e \rightarrow t))$

• Thus the ape’s concept of a box is a set of Type-Raised functions of type
  
  – $\text{box1}_{(e \rightarrow t) \rightarrow t}$
  – $\text{box2}_{(e \rightarrow (e \rightarrow t)) \rightarrow (e \rightarrow t)}$
  – $\text{box3}_{(e \rightarrow (e \rightarrow (e \rightarrow t))) \rightarrow (e \rightarrow (e \rightarrow t))}$
From Planning to Language

• We saw that (partially) searching the plan graph is an intrinsically recursive process.

• So we need at least a push-down automaton (PDA) to keep track of it.

⚠️ Is a PDA enough? (If it is, why don’t apes have language?)

• Not enough if the set of action- or plan- types is unbounded.
From Planning to Language

- Collaborative Plans involving other agents generate **types of unboundedly high valency**:
  - Help X to mind the baby
  - Promise Y to help X to mind the baby
  - Ask Z to promise Y to help X mind the baby.
  - etc.

Searching a graph with unboundedly many node-types needs a slightly more powerful automaton, the Embedded PDA (EPDA).

- Collaborative planning with other minds provides not only the only known motivation for language, but also the characteristic automaton that supports TAG and CCG.
Conclusion (I)

• The lexicon is the only locus of language specific information in the grammar.

• The universal projective syntactic component of natural language grammar is based on the combinators \( B, T \).

• These combinators are provided ready-made, by a sensory motor planning mechanism most of which we share with a number of animals.

• The problem of parsing is automata-theoretically equivalent to the problem of planning with other minds.

• The latter ability seems unique to humans.
Conclusion (II)

- The following components, which have been repeatedly evolved in the course of the last 200M years of vertebrate evolution may have recently combined in the essentially instantaneous recent emergence of language:

1. Reactive planning with non-recursive KR (finite-state) (pigeons);
2. Deliberative (forward-chaining, breadth-first) planning with non-recursive KR requiring composition and a (simulated) PDA (rats);
3. The same, with type-raising affordance (apes, killer-whales, corvines);
4. A PDA also supports recursive concepts in KR;
5. Plan inference with the specific recursive concepts that support human collaboration requires a (simulated) EPDA;
6. The EPDA immediately supports attested forms of Natural Language Grammar.
Conclusion (III)

This coming together in humans of components that have repeatedly been independently evolved in animal cognition may be more plausible in terms of evolutionary theory than the cataclysmic appearance of language-specific Merge of Berwick and Chomsky (2016).

- The problem of child language acquisition then reduces to the problem of learning (a) a lexicon, and (b) a parsing model, for a “supergrammar” of all categories and rules consistent with the (noisy) language specific data and the (ambiguous) contextual semantics (Abend et al., 2016).
References


