Evolutionary Language Learning

Ben Goertzel
September 20, 2011

I believe evolutionary learning is worth exploring as a methodology for computational language learning. I will describe here a new approach to automated language learning, which I believe would work much better than any method pursued so far. However, making it work would require a lot of effort and a lot of compute time.

A big advantage I see of the evolutionary learning approach is that it allows the combination, in one learning process, of multiple approaches to language learning, e.g.

- **intrinsic** (based on properties of the grammar, independent of any corpus)
- **unsupervised** (based on properties of the language evolved, evaluated via reference to a corpus of sentences known grammatical, versus a corpus of sentences thought with a high probability to be ungrammatical)
- **supervised** (based on a corpus of sentences whose correct parses are known)
- **badly supervised** (based on a corpus of sentences for which multiple parses are known, with some correct and some incorrect and no good knowledge of which are correct)
- **experiential** (based on relations between sentences and some extra-linguistic world, e.g. information on which pairs of sentences seem to mean the same thing in reference to some world)

That is, one can have an evolutionary learning process where

- The whole evolving population takes place relative to a certain fixed vocabulary of words (for starters at any rate – we could also deal with learning of new words, but it seems OK to defer that process for first experiments)

- The genotype is a lexicon: i.e. an individual in the evolving population is a data structure that associates, with each word in the vocabulary, a certain set of linguistic features

- Evaluation of a candidate lexicon L is done via a fitness function f of the form

  $$f(L) = w_0\, f_{intrinsic}(L) + w_1\, f_{unsup}(L) + w_2\, f_{sup}(L) + w_3\, f_{badsup}(L) + w_4\, f_{exp}(L)$$

with a component corresponding to each of the approaches to language learning mentioned above. (Of course, some other combination besides a weighted sum could be used. A sum is just posited as a first simple idea.)

There is the question of how to represent the lexicon. One approach would be to assume a dependency grammar format, and use a particular parser like the link grammar parser, or a word grammar parser. Then the features corresponding to each word would consist of a set of typed links (where the list of types is part of the genotype).

Given a commitment about the lexicon format, one can then articulate particularities of the components of the fitness function.

Intrinsic fitness should reward simplicity – e.g. penalize the lexicon for using too many link types, or having too large a total size (summed over all words).

Unsupervised fitness should reward the system for judging real sentences as grammatical, and random sentences (on the same word set) as ungrammatical.

Supervised fitness should reward the system for producing parses that can be mapped onto parses known correct. Assessing this may be costly because the link types learned by the system will not have the familiar names used in the parsed corpus. A search must be done to see: for all mappings of the learned link types onto the familiar link types in the parsed corpus, which one gives the closest match across the parsed corpus?

Badly supervised fitness can be estimated the same way as supervised fitness. This is intended to be used on a corpus produced by automated parsing with some sort of parse ranking. Maybe it should be merged with supervised fitness, or maybe separately weighted.

Embodied fitness could measure, for example, the degree to which sentences known to refer to related world-situations, are assigned parses with common substructures

Thus we have a language learning method that encompasses every methodology for language learning in one funky package.

The final ingredient is a good initial condition. Rather than starting with a random lexicon, we'd like to seed the population with a lexicon that is known to make some linguistic sense. For instance we could start with a stripped-down version of the link grammar lexicon, that avoids all the fine-grained link types and just keeps a few dozen main ones.

Using an intelligent initial condition like this requires an evolutionary learning algorithm that is capable of being guided by a smart initial population, rather than just scrambling it like GP does. MOSES has this property, so I'd suggest MOSES is a good candidate here.

Obviously, this kind of evolution would require massive compute time. However, MOSES is a distributed system already, so the software infrastructure is in place to enable this. With modern compute clouds it's certainly feasible to put hundreds of multi-core machines to work on something like this.

A final comment regards incremental learning. A language learned via MOSES, using a fitness function that has embodied experience as part of the fitness function, would clearly be naturally adaptable via further linguistic experience (embodied and otherwise). As new experience comes in, the lexicon would be updated via ongoing evolution. New word learning, language learning via logical inference and so forth, could be used to modify the lexicon, and the modified version could then be thrown back to MOSES for further refinement. So while the "evolutionary language learning" approach proposed here, on its own, is "narrow AI" ish, it is well suited to be used in an AGI context.