Learning optionality and repetition

Meaghan Fowlie
UCLA Linguistics

Adjuncts are syntactic elements that are optional, transparent to selection, and often, though not always, repeatable. Classic examples are adjectives and adverbs. How do learners learn optionaility? How do they learn repeatability? I explore a variety of learners' approaches to optionality and repeatability, including both human learners and learning algorithms.

Mathematically, a learner is a function from an input text to a grammar. The kinds of patterns in the input that the learner is sensitive to depends on the assumptions that the particular learner makes about the nature of the language.

In learnability theory, learners of Regular languages are much better understood than those for languages higher on the Chomsky hierarchy (Chomsky 1959). Since human languages are known to be Mildly Context-Sensitive (Joshi 1985), learning algorithms for such languages are clearly more relevant to actual human language learning; however, as research into such learners is still in its infancy, and since our understanding of Regular learners has driven higher-level learners (see for example Clark, Eyraud, & Habraud (2008)'s substitutable CF learner and Yoshinaka (2008)'s *k,l*-substitutable CF learner, which are extentions of Angluin (1982)'s 0- and *k*-reversible learners to the context free level), I will look at learners low on the Chomsky hierarchy as well. I provide here three examples.

N-gram learners:

An n-gram learner, for some *n*, learns languages defined entirely by good substrings of length *n*. It simply memorises all n-grams it encounters, and accepts/generates strings that contain only n-grams from the list it memorised. An n-gram learner generalises repetition from A^n to A^*: that is, if the input includes strings in which contain from 0 to *n* *A*'s in a row in the same context, the grammar it learns will accept strings with A^* in that context.

0-reversible learner:

Angluin (1982)'s 0-reversible learner generises directly from optionality to indefinite repeatability. The learner merges states with any suffix in common. Thus if there is an optional element *A*; i.e. two input strings st and sAt, for string s and t, the prefixes s and sA share the suffix t. They are therefore merged, forming a loop. For example, the grammar in Figure 1 will have states 1 and 2 merged, yielding the grammar in Figure 2.

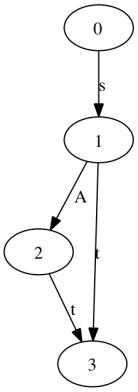

Figure 1: Grammar before states 1 and 2 are merged

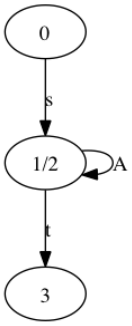

Figure 2: Grammar after states 1 and 2 are merged

A relationship between arbitrarily repeatable elements and optional elements is arguably desirable in general. Another way of saying that an element may or may not occur in a situation (in a context or after a state) is to say that the situation is the same whether the element has occurred or not. If this is a situation in which element A may occur, and if the situation is the same once A has occurred, then

A may occur again, leaving us in the same situation. For example, the X-bar rule N'→(A) N' means that whether or not A occurs to the left of this N', the result is another N'. This allows indefinite repetition of A. Similarly, in Figure 2, once we get to state 1/2 we remain in state 1/2 no matter how many times *A* occurs, until *t* occurs.

Clark & Thollard:

Clark & Thollard (2004) describe a PAC (Probably Approximately Correct) learner of probabilistic finite state languages. The learner is similar to Angluin's 0-reversible learner, except that the criterion for merging states is stricter: the similarity of the suffix sets of two states must be within a pre-determined margin for them to be merged. One suffix in common is not enough.

This learner can learn repeatability if the input is representative of the probabilities in the generating grammar. It can also generalise from optionality to repeatability, but only if the repeated element is rare enough in the input and the threshold for similarity is set high enough.

Artificial language learning

For some definition of learn, humans learn natural language. Our learning models are not yet adequate to describe real human language acquisition, since humans learn mildly context sensitive languages. Even when we do have a learner for that level of complexity, we still won't know if it bears any resemblance to how humans actually learn.

A goal of artificial grammar/language learning experiments is to probe what people can learn and what kinds of generalisations they tend to make; i.e. what properties the human learner has. I have in progress such a study looking at whether people generalise from an element occurring 0, 1, or 2 times in a context to it occurring indefinitely many times in that context. For example, if people hear AC, ABC, $ABBC$, do they also accept $ABBBC$?

Preliminary results indicate that such generalisation is definitely possible. About half the pilot participants accepted such generalisations and about half rejected them.

References

Angluin, D. 1982. Inference of reversible languages. *Journal of the ACM (JACM)* 29(3):741–765.

Chomsky, N. 1959. On certain formal properties of grammars. *Information and control* 2(2):137–167.

Clark, A., and Thollard, F. 2004. Pac-learnability of probabilistic deterministic finite state automata. *The Journal of Machine Learning Research* 5:473–497.

Clark, A.; Eyraud, R.; and Habrard, A. 2008. A polynomial algorithm for the inference of context free languages. In *Grammatical Inference: Algorithms and Applications*. Springer. 29–42.

Joshi, A. 1985. How much context-sensitivity is necessary for characterizing structural descriptions. In Dowty, D.; Karttunen, L.; and Zwicky, A., eds., *Natural Language Processing: Theoretical, Computational and Psychological Perspectives*. New York: Cambridge University Press. 206–250.

Yoshinaka, R. 2008. Identification in the limit of k, l-substitutable context-free languages. In *Grammatical Inference: Algorithms and Applications*. Springer. 266–279.