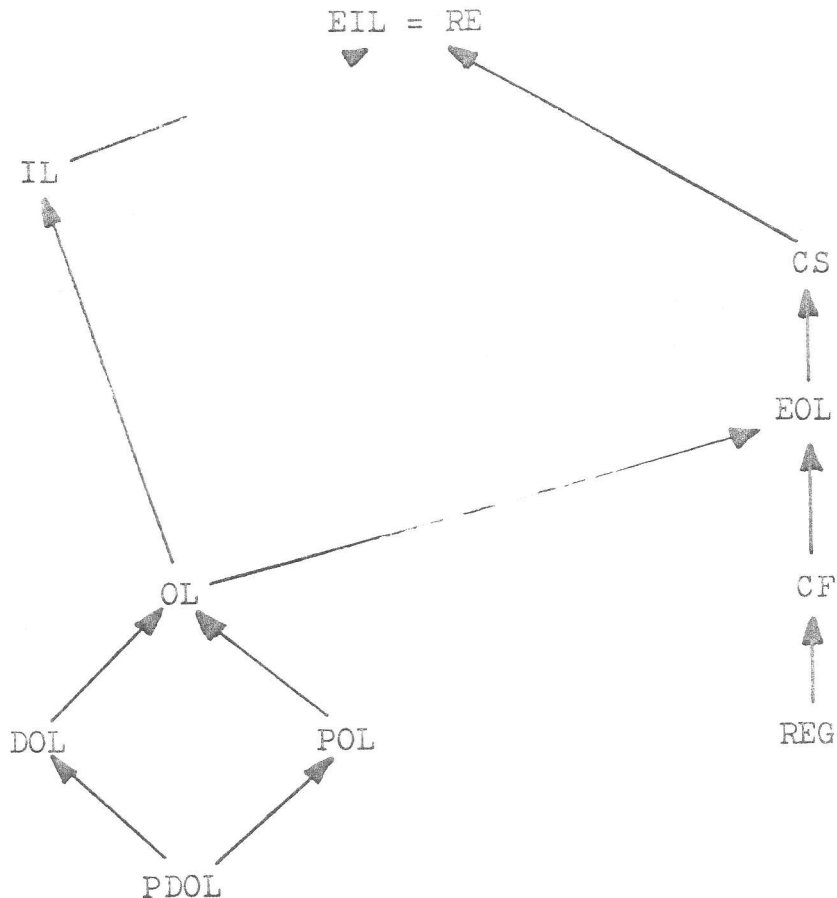


A THEORETICAL FRAME FOR THE GRAMMATICAL INFERENCE OF L-SYSTEMS

Wolfgang Coy
Universität Dortmund
Abteilung Informatik

Abstract

The grammatical inference problem is the problem to derive for a given set of data (a sample) a suitable grammar in a class of languages. Usually the sample is finite, while the grammar is thought to generate an infinite language. In a more logical context the same problem is known as "incomplete induction". Several notions of "grammatical inference" have been developed during the last years. Among the most promising are "identification in the limit by either full information or by positive information" (Gold 1967), where positive information means a sample which contains only examples of the language to be generated, while full information describes samples which contain also counter-examples (i.e. strings of the language's complement). Another useful notion is "strong approachability" as it is defined in Feldman's paper on decidability of grammatical inference. Identification demands that the learning algorithm will derive from the sample after some time a steady, correct hypotheses on the grammar to be learned, while strong approachability demands the elimination of false hypotheses during the learning process and the guess of a correct hypotheses for an infinite number of times while processing the (infinite) sample. Strong approachability is thus a weaker notion than identification. We apply these notions to the families of languages shown in the diagram. The basic families of languages in the diagram are recursive enumerable (RE), context-sensitive (CS), context-free (CF), regular (REG), and the L-languages OL and IL. The abbreviations D, P and E stand for deterministic, propagating (=non-erasing), and extended, respectively. We will show the following results, which constitute a frame for the further investigation of the grammatical inference problem for L-systems:



Relations between Chomsky-classes and L-systems. Lines mean proper inclusion.

- The language families DOL and PDOL are identifiable in the limit with positive information.
- All other families of languages are not identifiable by positive information alone.
- The language family EIL is not generally identifiable in the limit.
- The language families with solvable membership problem (word problem) are identifiable in the limit with full information. This includes CS and its subclasses EOL, OL, CF, DOL, POL, REG, and DPOL.
- The language family IL is identifiable in the limit under certain restrictions.

References

- (1) Gold, M., Language identification in the limit, Inf. & Control, 10:447-474, 1967.
- (2) Feldman, J., Some decidability results on grammatical inference and complexity, Inf. & Control, 20: 244-262, 1972.
- (3) Herman, G.T. & Rozenberg, G., Developmental systems and languages, North-Holland, Amsterdam, 1975.