E. Mark Gold - Language Identification in the Limit (1967)

Tom Rochette <tom.rochette@coreteks.org>

November 2, 2024 — 36c8eb68

0.1 Context

0.2 Learned in this study

0.3 Things to explore

1 Overview

- Time will be taken to be quantized and start at a finite time:
  - t = 1, 2, …
- At each time t the learner is presented with a unit of information i_t concerning the unknown language L.
- At each time t the learner is to make a guess g_t of a name of L (a grammar) based on the information it has received through time t. Thus the learner is a function G which takes strings of units of information into names:
  - $g_t = G(i_1, ..., i_t)$
- L will be said to be identified in the limit if, after some finite time, the guesses are all the same and are a name (grammar) of L
- A class of languages will be called identifiable in the limit with respect to a given language learnability model if there is an effective learner, i.e., an algorithm for making guesses, with the following property:
  - Given any language of the class and given any allowable training sequence for this language, the language will be identified in the limit
- Three conclusions from the anomalous model:
  - It shows that restrictions on the order of presentation of elements of the text can greatly increase the learnability power of this method of information presentation
  - A text only presents the learner with positive instances, namely, elements of the language, whereas an informant presents both positive and negative instances. Therefore, one would expect the informant to be more powerful. However, "anomalous text" is more powerful than any of the "informant" models, which shows that one must carefully consider the details of the learnability model
  - "Anomalous text" shows that the choice of naming relation can make a difference since, in this case, the generator-naming relation is far more powerful than tester
- Based on the study of child learning of language, it is believed that it is possible to learn the syntax of a natural language solely from positive instances
- Identification by enumeration refers to the following guessing rule: Enumerate the class of objects in any way, perhaps with repetitions.
- The following guessing algorithm shows that the class of languages of finite cardinality is identifiable in the limit from an arbitrary text:
  - Guess the unknown language to solely consist of the strings generated so far by the text (lookup table)

2   See also

3   References

- Gold, E. Mark. "Language identification in the limit." Information and control 10.5 (1967): 447-474.
- https://en.wikipedia.org/wiki/Language_identification_in_the_limit