1 Examining Successful Learning Criteria

To make things a bit more concrete, let us consider the problem of learning a representation (grammar) of a set of strings for given alphabet Σ. Consider any $S \subseteq \Sigma^*$. A grammar $G$is a representation of S provided for any $s \in \Sigma^*$ the grammar correctly identifies whether $s \in S$.

1.1 Definition of identification in the limit from positive data

The box below precisely defines a learning criterion called identification in the limit from positive data [1, 2, 3]. Let us define the "evidence" when learning from positive data more precisely. A **positive presentation** of a stringset S is a function $t : \mathbb{N} \to S$ such that t is *onto*. Recall that a function $f : X \to Y$ is onto provided for every element y in its co-domain Y there is some element x in its domain X such that $f(x) = y$. Here, this means for every string $s \in S$, there is some $n \in \mathbb{N}$ such that $t(n) = s$.

Definition 1 (Identification in the limit from positive data)**.**

Algorithm A *identifies in the limit from positive data* a class of stringsets C provided

for all stringsets $S \in C$,
for all positive presentations t of S,
there is some number $n \in \mathbb{N}$ such that
for all $m > n$,
- the grammar output by A on t_m is the same as the the grammar output by A on t_n, and
- the language of the grammar output by A on t_m equals S.

This paradigm is also called **learning from text.**

Theorem 1. *For given k, the Strictly k-Local languages are identifiable in the limit from positive data.*

Theorem 2. *The class of finite languages are identifiable in the limit from positive data.*

Theorem 3. *No class of languages which properely contains the finite languages is identifiable in the limit from positive data.*

As a corollary from this last theorem, it follows that the following classes of languages are not identifiable in the limit: FO(<), MSO (<) (regular languages), context free languages, context-senstive languages.

1.2 Probably Approximately Correct

The PAC paradigm is a different measure of success [4, 5]. As before, consider the problem of learning sets of strings. If S is a set of strings and x a string, let $S(x)$ be true iff $x \in S$.

PAC assumes the data presentation is generated according to a probability distribution over Σ^*. Let S be a set of strings and let $EX(S, D)$ be a procedure (we will sometimes call it an *oracle*) that runs in unit time, and on each call returns a labeled example $(x, S(x)$ where x is drawn randomly and independently according to D.

If G is a grammar, then the distribution D provides a natural measure of **error** between G and the target stringset S. Letting $G(x)$ be a function which maps strings to True iff $x \in L(G)$, then define

$$error_D(G) = \Pr_{x \in D}[S(x) \neq G(x)].$$

Also, let $\texttt{size}(G)$ be a measure of the size of grammar G.

Finally, we can define the PAC criterion for learning.

Definition 2 (Probably Approximately Correct)**.**

Algorithm A PAC-learns a class of stringsets C provided

for all stringsets $S \in C$,

for all probability distributions D over X,

for all $\epsilon, \delta > 0$,

there exists $m = f(\texttt{size}(G), \frac{1}{\epsilon}, \frac{1}{\delta})$ such that

- f is a polynomial function, and
- after drawing m samples from $EX(S, D)$, the probability that L outputs a grammar G with $\texttt{error}_D(G) < \epsilon$ is at least $1 - \delta$.

Theorem 4. *For given k, the Strictly k-Local languages are PAC learnable.*

Theorem 5. *The class of finite languages is not PAC learnable.*

As a corollary from this last theorem, it follows that the following classes of languages are not PAC learnable: FO(<), MSO (<) (regular languages), context free languages, context-senstive languages.

References

[1] E.M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

[2] Daniel Osherson, Scott Weinstein, and Michael Stob. *Systems that Learn.* MIT Press, Cambridge, MA, 1986.

[3] Sanjay Jain, Daniel Osherson, James S. Royer, and Arun Sharma. *Systems That Learn: An Introduction to Learning Theory (Learning, Development and Conceptual Change).* The MIT Press, 2nd edition, 1999.

[4] L.G. Valiant. A theory of the learnable. *Communications of the ACM*, 27:1134–1142, 1984.

[5] Leslie Valiant. *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World.* Basic Books, 2013.