

Generality's Price:

Inescapable Deficiencies in Machine-Learned Programs

John Case — University of Delaware
Keh-Jiann Chen — Academia Sinica
Sanjay Jain — Nat'l U. of Singapore
Wolfgang Merkle — Universität Heidelberg
James S. Royer — Syracuse University

Gold-Style Learning Theory

$$f(0), f(1), \dots \xrightarrow{\text{In}} M \xrightarrow{\text{Out}} p_0, p_1, \dots, | p_t, \dots$$

- Ref: S. Jain, D. Osherson, J. Royer, and A. Sharma, Systems That Learn, 2nd edition, MIT Press, 1999.
- Think of programs p_0, p_1, \dots as M 's succession of not-necessarily-rational beliefs about how to compute f based on successively more information about f .
- Hope: $\exists t$ such that p_t, p_{t+1}, \dots each do a good enough job computing f .
- Hope realized? Depends on M , f and ones exact criterion of success.
- Next: criteria of success, then examples.

Criteria of Success

$$f(0), f(1), \dots \xleftrightarrow{\text{In}} M \xleftrightarrow{\text{Out}} p_0, p_1, \dots, | p_t, \dots$$

- $N = \{0, 1, 2, 3, \dots\}$.
- Suppose $a \in N \cup \{*\}$. a for anomaly count.
- For $a = *$, a stands for finitely many.
- Suppose $\mathcal{F} \subseteq \mathcal{R}_{0,1}$, the class of all (total) computable functions : $N \rightarrow \{0, 1\}$.
- $\mathcal{F} \in \mathbf{Ex}^a \stackrel{\text{def}}{\Leftrightarrow} (\exists M)(\forall f \in \mathcal{F})$
 $[M \text{ fed } f(0), f(1), \dots, \text{ outputs } p_0, p_1, \dots \wedge$
 $(\exists t)[p_t = p_{t+1} = \dots \wedge p_t \text{ computes } f \text{ —}$
 $\text{except at up to } a \text{ inputs }]]$.
- $\mathcal{F} \in \mathbf{Bc}^a \stackrel{\text{def}}{\Leftrightarrow} (\exists M)(\forall f \in \mathcal{F})$
 $[M \text{ fed } f(0), f(1), \dots, \text{ outputs } p_0, p_1, \dots \wedge$
 $(\exists t)[p_t, p_{t+1}, \dots \text{ each computes } f \text{ — ex-}$
 $\text{cept at up to } a \text{ inputs}]]]$.

Examples

- For $k \geq 1$, $\mathcal{P}^k \stackrel{\text{def}}{=} \text{class all 0-1 val. funcs. comp. by (multi-tape) TMs in } O(n^k) \text{ time, w/ } n, \text{ length input. } \mathcal{P} \stackrel{\text{def}}{=} \bigcup \mathcal{P}^k.$
- Let slow be fixed slow grow. unbd. fun. $\in \mathcal{P}^1$, e.g., $\leq \text{Ackermann}^{-1}$. $\mathcal{Q}^k \stackrel{\text{def}}{=} \text{class all 0-1 val. funcs. comp. in } O(n^k \cdot \log(n) \cdot \text{slow}(n)) \text{ time. } \mathcal{P}^k \subset \mathcal{Q}^k \subset \mathcal{P}^{k+1}.$
- (\approx Gold'67) $\mathcal{P} \in \mathbf{Ex}^0$. $\mathcal{P}^k \in \mathbf{Ex}^0$ too (w/ each output conj. running in k -deg. poly time).
- (Case & Smith'78 + ...) $\mathbf{Ex}^0 \subset \mathbf{Ex}^1 \subset \mathbf{Ex}^2 \subset \dots \subset \mathbf{Ex}^* \subset \mathbf{Bc}^0 \subset \mathbf{Bc}^1 \subset \dots \subset \mathbf{Bc}^*.$
- (Harrington'83) $\mathcal{R}_{0,1} \in \mathbf{Bc}^*.$

Basic Notation

- $(\forall^\infty x)$ means for all but finitely many $x \in N$.
- $\mathcal{Z}^* \stackrel{\text{def}}{=} \{f \in \mathcal{R}_{0,1} \mid (\forall^\infty x)[f(x) = 0]\} (\subset \mathcal{P}^1)$.
- $\varphi_p^{\text{TM}} \stackrel{\text{def}}{=} \text{the partial computable function } : N \rightarrow N \text{ computed by Turing machine program (number) } p$.
- $\Phi_p^{\text{TM}}(x) \stackrel{\text{def}}{=} \text{the runtime of Turing machine program (number) } p \text{ on input } x, \text{ if } p \text{ halts on } x, \text{ and undefined, otherwise.}$
- $f[n] \stackrel{\text{def}}{=} \text{the sequence } f(0), \dots, f(n \Leftrightarrow 1)$.
- NEXT we'll look separately at complexity and information deficiencies in learned programs.
- Further examples re \mathcal{NP} vs \mathcal{P} , \mathcal{CF} vs \mathcal{REG} , ... in full conference paper.

Complexity Deficiency

- $M(f[n]) \stackrel{\text{def}}{=} M$'s output based only on $f[n]$. Can suppose without loss of generality $M(f[n])$ is always defined.

Proposition For each $k \geq 1$, $\exists M$ witnessing $\mathcal{P}^k \in \mathbf{Ex}^0$ such that $(\forall f)(\forall n)[\Phi_{M(f[n])}^{\text{TM}} \in O(|x|^k)]$.

By contrast:

Theorem [tightens Sipser, 1978] Suppose $k \geq 1$ and that M witnesses either $\mathcal{Q}^k \in \mathbf{Ex}^*$ or $\mathcal{Q}^k \in \mathbf{Bc}^0$ (special case: M witnesses $\mathcal{Q}^k \in \mathbf{Ex}^0$). Then: $(\exists f \in \mathcal{Z}^*)(\forall k\text{-degree polynomials } p)(\overset{\infty}{\forall} n)(\overset{\infty}{\forall} x)[\Phi_{M(f[n])}^{\text{TM}}(x) > p(|x|)]$.

Up gen. of M from \mathcal{P}^k to \mathcal{Q}^k entails run-times of M 's successful outputs on some $f \in \mathcal{Z}^*$ worse than any k -deg poly bnd.

See full conference paper for cases of \mathbf{Bc}^m , $m > 0$.

Information Deficiency

The next two results together nicely contrast with previous slide. Recall: its theorem featured criteria \mathbf{Ex}^* , \mathbf{Bc}^0 . $M(f)$ denotes M 's final output program, if any.

Theorem $\exists M$ outputting only total poly time conjs. & witnessing both $\mathcal{P} \in \mathbf{Bc}^*$ & $\mathcal{P}^1 \in \mathbf{Ex}^0$ such that $(\forall f \in \mathcal{P}^1)[\Phi_{M(f)}^{\mathbf{T}M} \in O(|x|)]$.

Theorem Suppose theory \mathbf{T} is any true, computably axiomatized extension of language of f.o. \mathbf{PA} (i.e., \mathbf{T} is safe, algorithmic info. extractor). Suppose $k \geq 1$ and M witnesses $\mathcal{Q}^k \in \mathbf{Bc}^*$. Then: $(\exists f \in \mathcal{Z}^*)(\forall^\infty n)[\mathbf{T} \not\ll \varphi_{M(f[n])}^{\mathbf{T}M}$ is computable* in $O(|x|^k)$ time $\gg]$.

While the learned programs for functions in \mathcal{Z}^* can have excellent run times, some of these learned programs will be informationally deficient — since we cannot prove from them even considerably weaker upper bounds on the run times of any total finite variants of the functions they compute.