U-Shaped Learning May Be Necessary

Lorenzo Carlucci — Univ. of Delaware

John Case, Presenter* — Univ. of Delaware

Sanjay Jain — Nat'l Univ. of Singapore

Frank Stephan — Nat'l Univ. of Singapore

*Contact email: case@cis.udel.edu

- U-Shaped Learning Behavior: Learn, Unlearn, Relearn. Occurs in child development re, e.g., verb regularization [PM91, MPU+92, TA02] & understanding of various (Piaget-like) conservation principles [SS82], e.g., temperature & weight conservation & interaction bet. object tracking/object permanence.

- Irregular Verb Example: Child first uses spoke, correct past tense of irregular verb speak. Then child overregularizes incorrectly using speaked. Lastly, child returns to using spoke.

- Concern Prior Literature: How model U-shaped learning? E.g., lang. learn., by gen. rules vs. tables of exceptions [Bow82]?

- Our Further Interest: Is U-shaped learning an unnecessary accident of human evolution or is U-shaped learning advantageous in that some classes of tasks can be learned in U-shaped way, but not otherwise? I.e., are some classes of tasks learnable only by returning to abandoned correct, learnable behavior?

Formal Language Learning

- We examine prior question re necessity of U-shaped learning in context of formal (computational) language learning theory [Gol67, JORS99].

- Without loss of generality and for mathematical convenience, all languages L will be $\subseteq N = \{0, 1, 2, \ldots\}$.

- T is a text for $L \stackrel{\text{def}}{\Leftrightarrow} \{T(0), T(1), \ldots\} = L$. Suppose T is a text for a language L.

$$T(0), T(1), \ldots \stackrel{\textbf{In}}{\Longleftrightarrow} M \stackrel{\textbf{Out}}{\Longleftrightarrow} p_0, p_1, \ldots, | \, p_t, \ldots$$

- M above is a machine (i.e., algorithmic device).

- $p_0, p_1, \ldots$ above are programs/grammars for generating languages — L or other language(s).

$$T(0), T(1), \ldots \overset{\textbf{In}}{\Leftrightarrow\!\!\!\rightarrow} M \overset{\textbf{Out}}{\Leftrightarrow\!\!\!\rightarrow} p_0, p_1, \ldots, \mid p_t, \ldots$$

Following are criteria for: some M is <u>successful</u> at learning every (task) L in a class of languages $\mathcal{L}$. Suppose $b \in (N^+ \cup \{*\})$, where $N^+ = \{1, 2, \ldots\}$ & $x \leq *$ means $x < \infty$.

- $\mathcal{L} \in \mathbf{TxtEx}$ [Gol67]: $(\exists M)(\forall L \in \mathcal{L})(\forall T \text{ for } L)(\exists t)[p_t = p_{t+1} = \cdots \wedge p_t$ generates/enumerates L]. E.g., class $\mathcal{F}$ of all finite languages $\in \mathbf{TxtEx}$ [Gol67].

- $\mathcal{L} \in \mathbf{TxtBc}$ [CL82, OW82, Wex82]: $(\exists M)(\forall L \in \mathcal{L})(\forall T \text{ for } L)(\exists t)$ $[p_t, p_{t+1}, \ldots$ each generates/enumerates L]. E.g., $\mathcal{K} = \{K \cup \{x\} \mid x \in N\} \in (\mathbf{TxtBc} \Leftrightarrow \mathbf{TxtEx})$, where K is the diagonal halting problem.

- $\mathcal{L} \in \mathbf{TxtFex}_b$ [OW82, Cas99]: $(\exists M)(\forall L \in \mathcal{L})(\forall T \text{ for } L)(\exists t)$ $[p_t, p_{t+1}, \ldots$ each generates/enumerates L $\wedge$ $\mathrm{card}(\{p_t, p_{t+1}, \ldots\}) \leq b$]. E.g., $\mathbf{TxtFex}_1 = \mathbf{TxtEx}$ & $\mathcal{K} \notin \mathbf{TxtFex}_b$.

$T(0), T(1), \ldots \overset{\textbf{In}}{\Leftrightarrow\!\!\rightarrow} M \overset{\textbf{Out}}{\Leftrightarrow\!\!\rightarrow} p_0, p_1, \ldots, | \, p_t, \ldots$

- $W_p \overset{\text{def}}{=}$ language generated/enumerated by program/grammar p. Informally: W_p is the [summary of the] behavior of p.

- **Theorem** [Cas99] Let $\langle \cdot, \cdot \rangle$ computably map $N \times N$ 1-1, onto N. $\forall^\infty z$ means for all but finitely many $z \in N$. Suppose $n \in N^+$. Let $\mathcal{L}_n =$ the set of all ∞L such that

  $$(\exists e_1, \ldots, e_n)[W_{e_1} = \ldots = W_{e_n} = L \wedge (\forall^\infty \langle x, y \rangle \in L)[y \in \{e_1, \ldots, e_n\}]].$$

  Let $\mathcal{L}_* = \cup_{n \in N^+} \mathcal{L}_n$.
  Then $\mathcal{L}_{n+1} \in (\mathbf{TxtFex}_{n+1} \Leftrightarrow \mathbf{TxtFex}_n) \wedge$ $\mathcal{L}_* \in (\mathbf{TxtFex}_* \Leftrightarrow \cup_{n \in N^+} \mathbf{TxtFex}_n)$.

- Suppose $\mathbf{C} \in \{\mathbf{TxtFex}_b, \mathbf{TxtBc}\}$. Then, $\mathcal{L} \in \mathbf{NonUC}$: $(\exists M$ witnessing $\mathcal{L} \in \mathbf{C})(\forall L \in \mathcal{L})(\forall T$ for $L)(\forall i, j, k \mid i < j < k)[W_{p_i} = W_{p_k} = L \Rightarrow W_{p_j} = W_{p_i}]$. Non U-shaped learners never abandon correct behaviors $\in \mathcal{L}$ and return to them.

Prior U-Shaped Learning Results

- **Proposition** [SC03]
  $\mathcal{K} \in (\mathbf{NonUTxtBc} \Leftrightarrow \mathbf{TxtFex}_b)$.

- A proof in [FJO94] is easily modified to show:

  Corollary [SC03]
  $(\mathbf{TxtBc} \Leftrightarrow \mathbf{NonUTxtBc}) \neq \emptyset$.

  Hence, for $\mathbf{TxtBc}$, U-shaped learning is necessary — for full learning power.

- **Theorem** [SC03]
  $\mathbf{NonUTxtEx} = \mathbf{TxtEx}$.

  Hence, for $\mathbf{TxtEx}$, U-shaped learning is not necessary — for full learning power.

Our New Results

- **Corollary** Suppose $2 \leq b \leq b'$. Then any M witnessing $\mathcal{L}_b \in \mathbf{TxtFex}_{b'}$ necessarily employs U-shaped learning on $\mathcal{L}_b$. However,

- **Theorem** $\mathbf{TxtFex}_2 \subset \mathbf{NonUTxtBc}$.
Hence, the cases where $\mathbf{TxtFex}_2$-learning necessitates U-shaped learning are circumventable by removing the bound on the number of successful programs. However,

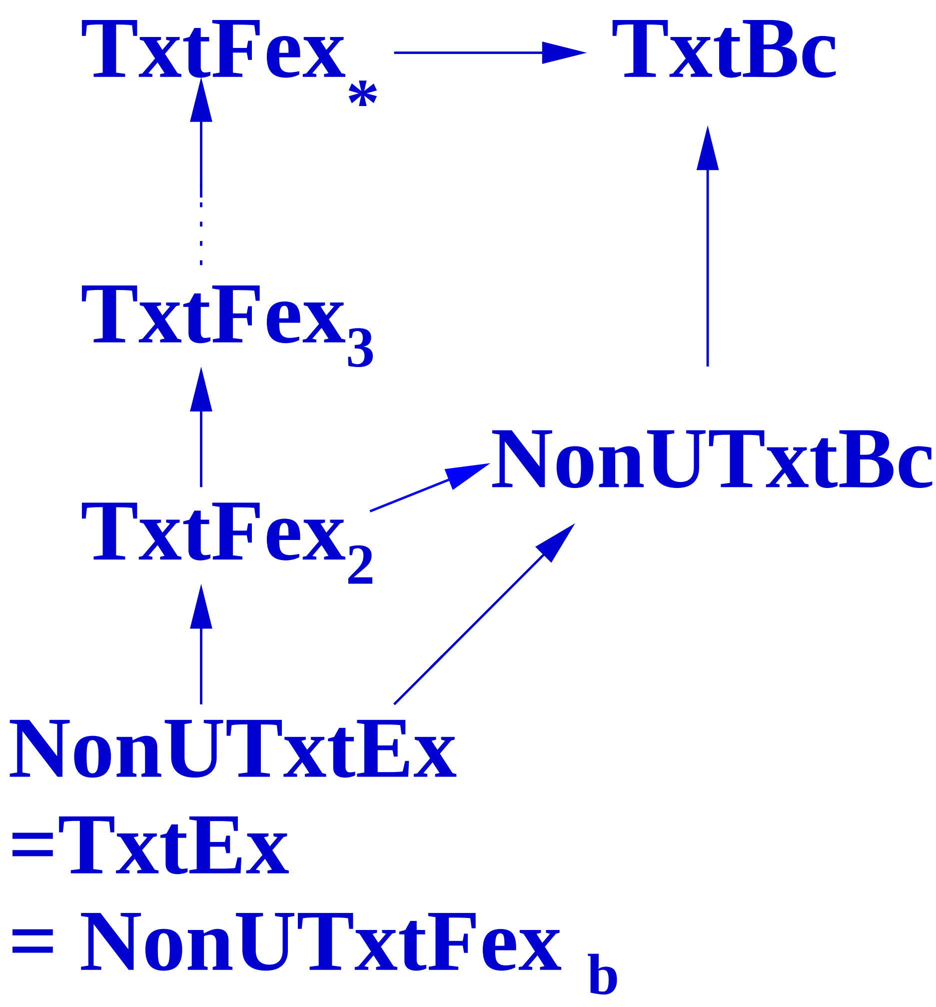
- **Theorem** $(\exists \mathcal{L} \in \mathbf{TxtFex}_3 \subset \mathbf{TxtFex}_4 \subset \ldots \subset \mathbf{TxtBc})$ $(\forall M$ witnessing $\mathcal{L} \in \mathbf{TxtBc})$ $[M$ must employ U-shaped learning on $\mathcal{L}]$.
Hence, there is no escaping the necessity of U-shaped learning for this $\mathcal{L} \in \mathbf{TxtFex}_3$.

- Proof of previous theorem intriguingly features learning finite tables vs. gen. rules, but does not feature learning incorrect gen. rule followed by correct gen. rule augmented by finite table. Corollary above follows from prior results & following theorem — proved by a counting arg.

  Theorem $\mathbf{NonUTxtFex}_b = \mathbf{TxtEx}$.

Summary

The transitive closure of the following inclusions (→) hold AND no other inclusions hold.



Summary Continued

- From prior work, U-shaped learning is not needed for $\mathbf{TxtEx}$ learning, i.e., for learning ONE successful program in the limit.

- For $b \geq 2$, $\mathcal{L}_2 \in \mathbf{TxtFex}_2$ cannot be $\mathbf{NonUTxtFex}_b$ learned, i.e., it can't be learned with $\leq b$ successful programs in the limit without U-shaped learning on $\mathcal{L}_2$.

- However, any class in $\mathbf{TxtFex}_2$ can be $\mathbf{NonUTxtBc}$ learned, i.e., learned with no bound on how many successful programs in the limit and without employing U-shaped learning.

- Some $\mathcal{L} \in \mathbf{TxtFex}_3$ cannot be $\mathbf{NonUTxtFex}_3$ learned, i.e., it can't be learned with ≤ 3 successful programs in the limit without employing U-shaped learning on $\mathcal{L}$, AND $\mathcal{L}$ requires U-shaped learning even with no bound on how many successful programs in the limit are allowed. Does the class of tasks humans must learn to be competitive in the genetic marketplace, like this $\mathcal{L}$, necessitate U-shaped learning?

References

[Bow82] M. Bowerman. Starting to talk worse: Clues to language acquisition from children's late speech errors. In S. Strauss and R. Stavy, editors, *U-Shaped Behavioral Growth*. Academic Press, New York, 1982.

[Cas99] J. Case. The power of vacillation in language learning. *SIAM Journal on Computing*, 28(6):1941–1969, 1999.

[CL82] J. Case and C. Lynes. Machine inductive inference and language identification. In M. Nielsen and E. Schmidt, editors, *Proceedings of the 9th International Colloquium on Automata, Languages and Programming*, volume 140 of *Lecture Notes in Computer Science*, pages 107–115. Springer-Verlag, Berlin, 1982.

[FJO94] M. Fulk, S. Jain, and D. Osherson. Open problems in Systems That Learn. *Journal of Computer and System Sciences*, 49(3):589–604, December 1994.

[Gol67] E. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

[JORS99] S. Jain, D. Osherson, J. Royer, and A. Sharma. *Systems that Learn: An Introduction to Learning Theory*. MIT Press, Cambridge, Mass., second edition, 1999.

[MPU+92] G. Marcus, S. Pinker, M. Ullman, M. Hollander, T.J. Rosen, and F. Xu. *Overregularization in Language Acquisition*. Monographs of the Society for Research in Child Development, vol. 57, no. 4. University of Chicago Press, 1992. Includes commentary by H. Clahsen.

[OW82] D. Osherson and S. Weinstein. Criteria of language learning. *Information and Control*, 52:123–138, 1982.

[PM91] K. Plunkett and V. Marchman. U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. *Cognition*, 86(1):43–102, 1991.

[SC03] F. Stephan and J. Case. Decisive versus U-shaped learning. In *Proceedings of the 12th International Congress of Logic, Methodology, and Philosophy of Science*, 2003.

[SS82] S. Strauss and R. Stavy, editors. *U-Shaped Behavioral Growth*. Developmental Psychology Series. Academic Press, NY, 1982.

[TA02] N.A. Taatgen and J.R. Anderson. Why do children learn to say broke? A model of learning the past tense without feedback. *Cognition*, 86(2):123–155, 2002.

[Wex82] K. Wexler. On extensional learnability. *Cognition*, 11:89–95, 1982.