Language Learning from a Computational Viewpoint

Karen Seidel
Hasso-Plattner-Institute

Hiddensee, August 7th 2017

Gold's Model

Gold's Model

IN

$16, 256, 16, 4$

Gold's Model

IN

$16, 256, 16, 4$

Gold's Model

IN $\longrightarrow$

$16, 256, 16, 4$

OUT $\longrightarrow$

"powers of 2"

Gold’s Model

IN →

$16, 256, 16, 4,$
$4, 200, 24, 96$

OUT →

“powers of 2”

Gold's Model

IN →

$16, 256, 16, 4,$
$4, 200, 24, 96$

OUT →

"powers of 2"
"multiples of 4"

Gold's Model

IN $\longrightarrow$

$16, 256, 16, 4,$
$4, 200, 24, 96,$
$6, 8, 146, 212$

OUT $\longrightarrow$

"powers of 2"
"multiples of 4"

Gold's Model

IN $\longrightarrow$

$16, 256, 16, 4,$
$4, 200, 24, 96,$
$6, 8, 146, 212$

OUT $\longrightarrow$

"powers of 2"
"multiples of 4"
"multiples of 2"

Gold's Model

IN →

$16, 256, 16, 4,$
$4, 200, 24, 96,$
$6, 8, 146, 212,$
$\vdots$

OUT →

"powers of 2"
"multiples of 4"
"multiples of 2"
$\vdots$

Gold’s Model

IN $\longrightarrow$

$16, 256, 16, 4,$
$4, 200, 24, 96,$
$6, 8, 146, 212,$
$\vdots$

OUT $\longrightarrow$

“powers of 2”
“multiples of 4”
“multiples of 2”
$\vdots$

M

Gold's Model

IN $\longrightarrow$

$16, 256, 16, 4,$
$4, 200, 24, 96,$
$6, 8, 146, 212,$
$\vdots$

OUT $\longrightarrow$

"powers of 2"
"multiples of 4"
"multiples of 2"
$\vdots$

For every $t \in \mathbb{N}$:

$$\xrightarrow{\sigma_t := u_0 u_1 u_2 \ldots u_{t-1}} \quad M \quad \xrightarrow{h_t}$$

Gold's Model

IN → **OUT**

$16, 256, 16, 4,$
$4, 200, 24, 96,$
$6, 8, 146, 212,$
$\vdots$

"powers of 2"
"multiples of 4"
"multiples of 2"
$\vdots$

For every $t \in \mathbb{N}$:

$\sigma_t := u_0 u_1 u_2 \ldots u_{t-1}$ → M → h_t

Criteria for Success

For every presentation $u_0, u_1, \ldots, u_t, \ldots$ of L we gain

$$\xrightarrow{\sigma_0, \sigma_1, \ldots, \sigma_t, \ldots} \boxed{M \quad \xrightarrow{h_0, h_1, \ldots, h_t, \ldots}}$$

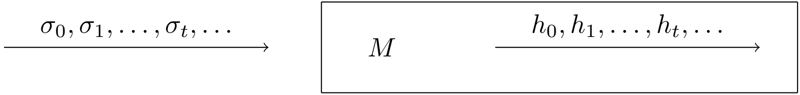
Definition

Let $b \in \mathbb{N}_{\geqslant 1} \cup \{*, \infty\}$. M $\mathbf{TxtFex}_b$-*learns* a language L if for every presentation T of L, i.e., $T \in \mathbf{Txt}(L)$, there is t_0 such that for all $t \geqslant t_0$ the hypothesis $h_t = M(T[t])$ is correct and $|\{M(T[t]) \mid t \geqslant t_0\}| \leqslant b$.
$b = 1$ is denoted by $\mathbf{TxtEx}$ (Gold) and $b = \infty$ by $\mathbf{TxtBc}$.

We are interested in collections of languages $\mathcal{L}$ such that there is a single M which $\mathbf{TxtFex}_b$-learns *every* language in $\mathcal{L}$.

Criteria for Success

For every presentation $u_0, u_1, \ldots, u_t, \ldots$ of L we gain



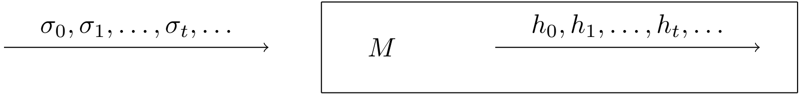
Definition

Let $b \in \mathbb{N}_{\geqslant 1} \cup \{*, \infty\}$. M $\mathbf{TxtFex}_b$-*learns* a language L if for every presentation T of L, i.e., $T \in \mathbf{Txt}(L)$, there is t_0 such that for all $t \geqslant t_0$ the hypothesis $h_t = M(T[t])$ is correct and $|\{M(T[t]) \mid t \geqslant t_0\}| \leqslant b$.
$b = 1$ is denoted by $\mathbf{TxtEx}$ (Gold) and $b = \infty$ by $\mathbf{TxtBc}$.

We are interested in collections of languages $\mathcal{L}$ such that there is a single M which $\mathbf{TxtFex}_b$-learns *every* language in $\mathcal{L}$.

Criteria for Success

For every presentation $u_0, u_1, \ldots, u_t, \ldots$ of L we gain



Definition

Let $b \in \mathbb{N}_{\geqslant 1} \cup \{*, \infty\}$. M $\mathbf{TxtFex}_b$-*learns* a language L if for every presentation T of L, i.e., $T \in \mathbf{Txt}(L)$, there is t_0 such that for all $t \geqslant t_0$ the hypothesis $h_t = M(T[t])$ is correct and $|\{M(T[t]) \mid t \geqslant t_0\}| \leqslant b$.
$b = 1$ is denoted by $\mathbf{TxtEx}$ (Gold) and $b = \infty$ by $\mathbf{TxtBc}$.

We are interested in collections of languages $\mathcal{L}$ such that there is a single M which $\mathbf{TxtFex}_b$-learns *every* language in $\mathcal{L}$.

Criteria for Success

For every presentation $u_0, u_1, \ldots, u_t, \ldots$ of L we gain

$\sigma_0, \sigma_1, \ldots, \sigma_t, \ldots$ $\longrightarrow$ M $\longrightarrow$ $h_0, h_1, \ldots, h_t, \ldots$

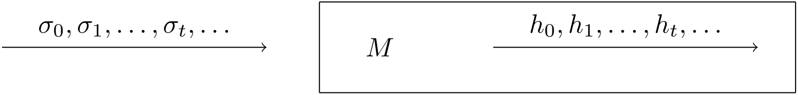
Definition

Let $b \in \mathbb{N}_{\geqslant 1} \cup \{*, \infty\}$. M $\mathbf{TxtFex}_b$*-learns* a language L if for every presentation T of L, i.e., $T \in \mathbf{Txt}(L)$, there is t_0 such that for all $t \geqslant t_0$ the hypothesis $h_t = M(T[t])$ is correct and $|\{M(T[t]) \mid t \geqslant t_0\}| \leqslant b$. $b = 1$ is denoted by $\mathbf{TxtEx}$ (Gold) and $b = \infty$ by $\mathbf{TxtBc}$.

We are interested in collections of languages $\mathcal{L}$ such that there is a single M which $\mathbf{TxtFex}_b$-learns *every* language in $\mathcal{L}$.

Criteria for Success

For every presentation $u_0, u_1, \ldots, u_t, \ldots$ of L we gain



Definition

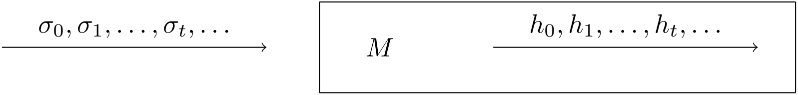
Let $b \in \mathbb{N} \cup \{*, \infty\}$. $[\mathbf{TxtFex}_b]$ denotes the set of all $\mathcal{L}$ for which there exists M which $\mathbf{TxtFex}_b$-learns *every* language in $\mathcal{L}$.

Theorem (Case99)

For all $b \in \mathbb{N}_{\geqslant 1}$ *holds* $[\mathbf{TxtFex}_b] \subsetneq [\mathbf{TxtFex}_{b+1}]$ *and moreover* $\bigcup_{b \geqslant 1} [\mathbf{TxtFex}_b] \subsetneq [\mathbf{TxtFex}_*]$.

Criteria for Success

For every presentation $u_0, u_1, \ldots, u_t, \ldots$ of L we gain



Definition

Let $b \in \mathbb{N} \cup \{*, \infty\}$. $[\mathbf{TxtFex}_b]$ denotes the set of all $\mathcal{L}$ for which there exists M which $\mathbf{TxtFex}_b$-learns *every* language in $\mathcal{L}$.

Theorem (Case99)

For all $b \in \mathbb{N}_{\geqslant 1}$ *holds* $[\mathbf{TxtFex}_b] \subsetneq [\mathbf{TxtFex}_{b+1}]$ *and moreover* $\bigcup_{b \geqslant 1} [\mathbf{TxtFex}_b] \subsetneq [\mathbf{TxtFex}_*]$.

Constraints on Learnability and Characterizations

Theorem (Subset Principle)

Let $b \in \mathbb{N}_{\geqslant 1} \cup \{*, \infty\}$ and M $\mathbf{TxtFex}_b$-learning the language L. Then there exists $D \subseteq L$ finite such that no proper subset L' of L with $D \subseteq L'$ is $\mathbf{TxtFex}_b$-learned by M.

Corollary

The set of all regular languages is not TxtBc-learnable.

Remark

Angluin's Tell-Tale-Condition is an important generalized version of the theorem for uniformly decidable collections of languages.

Constraints on Learnability and Characterizations

Theorem (Subset Principle)

Let $b \in \mathbb{N}_{\geqslant 1} \cup \{*, \infty\}$ and M $\mathbf{TxtFex}_b$-learning the language L. Then there exists $D \subseteq L$ finite such that no proper subset L' of L with $D \subseteq L'$ is $\mathbf{TxtFex}_b$-learned by M.

Corollary

The set of all regular languages is not $\mathbf{TxtBc}$-learnable.

Remark

Angluin's Tell-Tale-Condition is an important generalized version of the theorem for uniformly decidable collections of languages.

Insensitivity to Presentation Order

Definition

M is *set-driven* ($\mathbf{Sd}$) if for any two finite sequences with the same content it yields the same hypothesis.

Definition

$[\mathbf{SdTxtFex}_b]$ denotes the set of all $\mathcal{L}$ for which there exists a set-driven M $\mathbf{TxtFex}_b$-learning every language in $\mathcal{L}$.

The following result positively answers the question whether $[\mathbf{SdTxtEx}] \subsetneq [\mathbf{TxtEx}]$ carries over to $b > 1$.

Theorem (KSS17)

For all $b \in \mathbb{N}_{\geqslant 1} \cup \{*, \infty\}$ *holds* $[\mathbf{SdTxtFex}_b] \subsetneq [\mathbf{TxtFex}_b]$.

Insensitivity to Presentation Order

Definition

M is *set-driven* ($\mathbf{Sd}$) if for any two finite sequences with the same content it yields the same hypothesis.

Definition

$[\mathbf{SdTxtFex}_b]$ denotes the set of all $\mathcal{L}$ for which there exists a set-driven M $\mathbf{TxtFex}_b$-learning every language in $\mathcal{L}$.

The following result positively answers the question whether $[\mathbf{SdTxtEx}] \subsetneq [\mathbf{TxtEx}]$ carries over to $b > 1$.

Theorem (KSS17)

For all $b \in \mathbb{N}_{\geqslant 1} \cup \{*, \infty\}$ *holds* $[\mathbf{SdTxtFex}_b] \subsetneq [\mathbf{TxtFex}_b]$.

Insensitivity to Presentation Order

Definition

M is *set-driven* ($\mathbf{Sd}$) if for any two finite sequences with the same content it yields the same hypothesis.

Definition

$[\mathbf{SdTxtFex}_b]$ denotes the set of all $\mathcal{L}$ for which there exists a set-driven M $\mathbf{TxtFex}_b$-learning every language in $\mathcal{L}$.

The following result positively answers the question whether $[\mathbf{SdTxtEx}] \subsetneq [\mathbf{TxtEx}]$ carries over to $b > 1$.

Theorem (KSS17)

For all $b \in \mathbb{N}_{\geqslant 1} \cup \{*, \infty\}$ *holds* $[\mathbf{SdTxtFex}_b] \subsetneq [\mathbf{TxtFex}_b]$.

Empirical (Non-)U-Shaped Learning

- childrens understanding of conservation-principles (similar to Piaget) and verb regularization follow the pattern

learn | unlearn | relearn

- whether U-Shapes in learning are necessary is a major open question

Empirical (Non-)U-Shaped Learning

- childrens understanding of conservation-principles (similar to Piaget) and verb regularization follow the pattern

learn | unlearn | relearn

- whether U-Shapes in learning are necessary is a major open question

Formal (Non-)U-Shaped Learning

Definition

M behaves *non-U-shaped* ($\mathbf{NU}$) on a presentation of a language L if whenever h_t and h_{t+n} are correct descriptions of L, then for all $i \leqslant n$ the description h_{t+i} is also correct.

Definition

M $\mathbf{TxtNUFex}_b$*-learns* a language L if it $\mathbf{TxtFex}_b$-learns L and additionally behaves non-U-shaped on every presentation of L.

$[\mathbf{TxtNUFex}_b]$ denotes the set of all $\mathcal{L}$ for which there exists M $\mathbf{TxtNUFex}_b$-learning every language in $\mathcal{L}$.

Formal (Non-)U-Shaped Learning

Definition

M behaves *non-U-shaped* ($\mathbf{NU}$) on a presentation of a language L if whenever h_t and h_{t+n} are correct descriptions of L, then for all $i \leqslant n$ the description h_{t+i} is also correct.

Definition

M $\mathbf{TxtNUFex}_b$*-learns* a language L if it $\mathbf{TxtFex}_b$-learns L and additionally behaves non-U-shaped on every presentation of L.

$[\mathbf{TxtNUFex}_b]$ denotes the set of all $\mathcal{L}$ for which there exists M $\mathbf{TxtNUFex}_b$-learning every language in $\mathcal{L}$.

Results Regarding $\mathbf{NU}$-Learning

Theorem (CCJS08)

For all $b \in \mathbb{N}_{\geqslant 1} \cup \{*\}$ *holds* $[\mathbf{TxtNUFex}_b] = [\mathbf{TxtEx}]$.

Theorem (BCMSW05)

We have $[\mathbf{TxtEx}] \subsetneq [\mathbf{TxtNUBc}]$.

Theorem (FJO94)

We have $[\mathbf{TxtNUBc}] \subsetneq [\mathbf{TxtBc}]$.

Further results from CCJS08:

- $[\mathbf{TxtFex}_2] \subseteq [\mathbf{TxtNUBc}]$
- $[\mathbf{TxtFex}_3] \backslash [\mathbf{TxtNUBc}] \neq \varnothing$
- $[\mathbf{TxtNUBc}] \backslash [\mathbf{TxtFex}_*] \neq \varnothing$

Results Regarding $\mathbf{NU}$-Learning

Theorem (CCJS08)

For all $b \in \mathbb{N}_{\geqslant 1} \cup \{*\}$ *holds* $[\mathbf{TxtNUFex}_b] = [\mathbf{TxtEx}]$.

Theorem (BCMSW05)

We have $[\mathbf{TxtEx}] \subsetneq [\mathbf{TxtNUBc}]$.

Theorem (FJO94)

We have $[\mathbf{TxtNUBc}] \subsetneq [\mathbf{TxtBc}]$.

Further results from CCJS08:

- $[\mathbf{TxtFex}_2] \subseteq [\mathbf{TxtNUBc}]$
- $[\mathbf{TxtFex}_3] \backslash [\mathbf{TxtNUBc}] \neq \varnothing$
- $[\mathbf{TxtNUBc}] \backslash [\mathbf{TxtFex}_*] \neq \varnothing$

Results Regarding $\mathbf{NU}$-Learning

Theorem (CCJS08)

For all $b \in \mathbb{N}_{\geqslant 1} \cup \{*\}$ *holds* $[\mathbf{TxtNUFex}_b] = [\mathbf{TxtEx}]$.

Theorem (BCMSW05)

We have $[\mathbf{TxtEx}] \subsetneq [\mathbf{TxtNUBc}]$.

Theorem (FJO94)

We have $[\mathbf{TxtNUBc}] \subsetneq [\mathbf{TxtBc}]$.

Further results from CCJS08:

- $[\mathbf{TxtFex}_2] \subseteq [\mathbf{TxtNUBc}]$
- $[\mathbf{TxtFex}_3] \backslash [\mathbf{TxtNUBc}] \neq \varnothing$
- $[\mathbf{TxtNUBc}] \backslash [\mathbf{TxtFex}_*] \neq \varnothing$

Results Regarding $\mathbf{NU}$-Learning

Theorem (CCJS08)

For all $b \in \mathbb{N}_{\geqslant 1} \cup \{*\}$ *holds* $[\mathbf{TxtNUFex}_b] = [\mathbf{TxtEx}]$.

Theorem (BCMSW05)

We have $[\mathbf{TxtEx}] \subsetneq [\mathbf{TxtNUBc}]$.

Theorem (FJO94)

We have $[\mathbf{TxtNUBc}] \subsetneq [\mathbf{TxtBc}]$.

Further results from CCJS08:

- $[\mathbf{TxtFex}_2] \subseteq [\mathbf{TxtNUBc}]$
- $[\mathbf{TxtFex}_3] \setminus [\mathbf{TxtNUBc}] \neq \emptyset$
- $[\mathbf{TxtNUBc}] \setminus [\mathbf{TxtFex}_*] \neq \emptyset$

Results Regarding $\mathbf{NU}$-Learning

Theorem (CCJS08)

For all $b \in \mathbb{N}_{\geqslant 1} \cup \{*\}$ *holds* $[\mathbf{TxtNUFex}_b] = [\mathbf{TxtEx}]$.

Theorem (BCMSW05)

We have $[\mathbf{TxtEx}] \subsetneq [\mathbf{TxtNUBc}]$.

Theorem (FJO94)

We have $[\mathbf{TxtNUBc}] \subsetneq [\mathbf{TxtBc}]$.

Further results from CCJS08:

- $[\mathbf{TxtFex}_2] \subseteq [\mathbf{TxtNUBc}]$
- $[\mathbf{TxtFex}_3] \backslash [\mathbf{TxtNUBc}] \neq \varnothing$
- $[\mathbf{TxtNUBc}] \backslash [\mathbf{TxtFex}_*] \neq \varnothing$

Results Regarding NU-Learning

Theorem (CCJS08)

For all $b \in \mathbb{N}_{\geqslant 1} \cup \{*\}$ *holds* $[\mathbf{TxtNUFex}_b] = [\mathbf{TxtEx}]$.

Theorem (BCMSW05)

We have $[\mathbf{TxtEx}] \subsetneq [\mathbf{TxtNUBc}]$.

Theorem (FJO94)

We have $[\mathbf{TxtNUBc}] \subsetneq [\mathbf{TxtBc}]$.

Further results from CCJS08:

- $[\mathbf{TxtFex}_2] \subseteq [\mathbf{TxtNUBc}]$
- $[\mathbf{TxtFex}_3] \setminus [\mathbf{TxtNUBc}] \neq \varnothing$
- $[\mathbf{TxtNUBc}] \setminus [\mathbf{TxtFex}_*] \neq \varnothing$

Map for $\mathbf{TxtNU}$-Learning

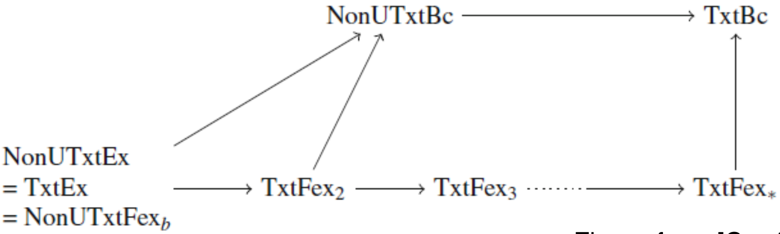


Figure from [Cas16]

Theorem (KSS17)

For all $b \in \mathbb{N}_{\geqslant 1} \cup \{*, \infty\}$ *holds* $[\mathbf{SdTxtNUFex}_b] \subsetneq [\mathbf{TxtNUFex}_b]$.

Map for $\mathbf{TxtNU}$-Learning

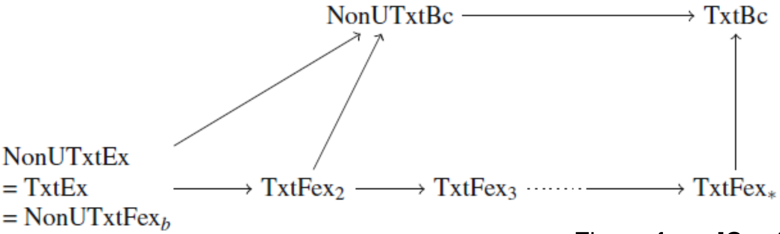


Figure from [Cas16]

Theorem (KSS17)

For all $b \in \mathbb{N}_{\geqslant 1} \cup \{*, \infty\}$ *holds* $[\mathbf{SdTxtNUFex}_b] \subsetneq [\mathbf{TxtNUFex}_b]$.

Further Remarks

- Presenting positive and negative examples for L to the learner is covered by investigations on informants.
- Abundance of information and limited memory have been modelled by iterative learners and other attempts.
- It has been shown that NU is not a restriction for iterative learners for $b = 1$. The behavior for $b > 1$ needs further investigation.
- Complexity considerations question whether the TM-based attempt is appropriate for modelling learning.

Further Remarks

- Presenting positive and negative examples for L to the learner is covered by investigations on informants.
- Abundance of information and limited memory have been modelled by iterative learners and other attempts.
- It has been shown that $\mathbf{NU}$ is not a restriction for iterative learners for $b = 1$. The behavior for $b > 1$ needs further investigation.
- Complexity considerations question whether the TM-based attempt is appropriate for modelling learning.

Further Remarks

- Presenting positive and negative examples for L to the learner is covered by investigations on informants.
- Abundance of information and limited memory have been modelled by iterative learners and other attempts.
- It has been shown that $\mathbf{NU}$ is not a restriction for iterative learners for $b = 1$. The behavior for $b > 1$ needs further investigation.
- Complexity considerations question whether the TM-based attempt is appropriate for modelling learning.

Further Remarks

- Presenting positive and negative examples for L to the learner is covered by investigations on informants.
- Abundance of information and limited memory have been modelled by iterative learners and other attempts.
- It has been shown that $\mathbf{NU}$ is not a restriction for iterative learners for $b = 1$. The behavior for $b > 1$ needs further investigation.
- Complexity considerations question whether the TM-based attempt is appropriate for modelling learning.

Learning with Automata

Definition

An automatic learner M uses two automatic functions hyp and mem which produce sequences $h_0, \ldots, h_t, \ldots$ interpreted as its hypotheses and $m_0, \ldots, m_t, \ldots$ standing for the information stored in M's memory via $h_{t+1} = \mathrm{hyp}(u_t, m_t)$ and $m_{t+1} = \mathrm{mem}(u_t, m_t)$.

Learning with Automata

Definition

An automatic learner M uses two automatic functions hyp and mem which produce sequences $h_0, \dots, h_t, \dots$ interpreted as its hypotheses and $m_0, \dots, m_t, \dots$ standing for the information stored in M's memory via $h_{t+1} = \mathrm{hyp}(u_t, m_t)$ and $m_{t+1} = \mathrm{mem}(u_t, m_t)$.

Thanks for listening! 

References

- Ganesh Baliga, John Case, Wolfgang Merkle, Frank Stephan, and Rolf Wiehagen, *When unlearning helps*, Information and Computation **206** (2008), no. 5, 694–709.
- John Case, *The power of vacillation in language learning*, SIAM Journal on Computing **28** (1999), no. 6, 1941–1969.
- ______, *Gold-style learning theory*, Topics in Grammatical Inference, Springer, 2016, pp. 1–23.
- Lorenzo Carlucci, John Case, Sanjay Jain, and Frank Stephan, *Non-u-shaped vacillatory and team learning*, Journal of Computer and System Sciences **74** (2008), no. 4, 409–430.
- Mark Fulk, Sanjay Jain, and Daniel N Osherson, *Open problems in "systems that learn"*, Journal of Computer and System Sciences **49** (1994), no. 3, 589–604.
- Timo Kötzing, Martin Schirneck, and Karen Seidel, *Normal forms in semantic language identification*, unpublished, 2017.