The Teaching Complexity of Erasing Pattern Languages With Bounded Variable Frequency

Ziyuan Gao

Department of Mathematics
National University of Singapore
10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore

8 August 2019

Pattern Languages

- pattern: nonempty string made up of symbols from two disjoint sets, Σ (alphabet of constant symbols) and X (countably infinite set of variables)
- e.g. $x_1 001 x_2 x_1 11 x_3 x_3$
- option to repeat variables corresponds to the use of *backreferences* in regular expressions (e.g. $\{ww : w \in \{0,1\}^*\}$ is generated by $(\langle x \rangle (0|1)^*)\backslash k\langle x \rangle)$
- proper subclass of *extended* regular expressions (regular expressions with backreferences)

Pattern Languages

- e.g. data entry systems
- author: $\langle$Angluin, D.$\rangle$;
- title: $\langle$Inductive Inference of Formal Languages From Positive Data$\rangle$;
- journal: $\langle$Information and Control, 45$\rangle$;
- year: $\langle$1980$\rangle$
- structure of records: author:$\langle x_1 \rangle$; title:$\langle x_2 \rangle$; journal:$\langle x_3 \rangle$; year: $\langle x_4 \rangle$

Pattern Languages

- the non-erasing pattern language generated by a pattern π ($L_{ne}(\pi)$) consists of all strings obtained by replacing all the variables in π with nonempty strings of constant symbols (Angluin, 1980)
- any two occurrences of the same variable are replaced with the same string
- erasing pattern language generated by π ($L_e(\pi)$): similar to non-erasing case, but variables may be replaced with the empty string ε (Shinohara, 1982)

Pattern Languages

- only erasing pattern languages will be considered here
- $L(\pi) \leftrightarrow L_e(\pi)$
- Π^z = class of patterns over some alphabet Σ of constant symbols and countably infinite alphabet $X = \{x_1, x_2, x_3, \ldots\}$ of variable symbols such that $|\Sigma|$ has exactly z distinct elements
- z may be infinity; corresponding class of patterns is denoted by Π^∞
- $\mathrm{Var}(\pi)$ = set of all distinct variables in π
- $\mathrm{Const}(\pi)$ = set of all constant symbols in π
- π is normalised iff its variables $x_1, \ldots, x_k$ occur in order of their first occurrences from left to right
- e.g. $\pi = x_1 a x_2 x_1 b x_3 x_4 a a x_2$

Pattern Languages: Examples

- $\Sigma = \{0, 1\}$
- $L(x_1) = \{\varepsilon, 0, 1, 01, \ldots\} = \Sigma^*$
- $L(x_1^2) = \{\varepsilon, 0^2, 1^2, 0101, \ldots\}$
- $L(x_1 x_2^2 x_1) = \{\varepsilon, 0^2, 1^2, 0101, \ldots, 01^20, 0^21^20^2, \ldots\}$
- note that $L(x_1^2) \subset L(x_1 x_2^2 x_1) \subset L(x_1)$
- if π is not a constant pattern, then $L(\pi)$ is infinite

Some Facts About Pattern Languages

- non-erasing pattern languages: natural example of a uniformly recursive class that is "learnable in the limit" (Angluin, 1980)
- a class of languages is learnable in the limit if there is a learning algorithm such that, given any infinite sequence of all positive examples for any language L in the class, the algorithm outputs a corresponding sequence of guesses for the target language (based on a representation system for the languages in the class) that converges to a fixed representation for L; this model is due to Gold (1967)
- membership problem for non-erasing (resp. erasing) pattern languages (given a string w and pattern π, is w in $L_{ne}(\pi)$ (resp. $L_e(\pi)$)?) is NP-complete, i.e. likely to be hard (Angluin, 1980; Jiang et al., 1994)

Some Facts About Pattern Languages

- inclusion problem for non-erasing (resp. erasing) pattern languages (given patterns π and τ, is $L_{ne}(\pi) \subseteq L_{ne}(\tau)$ (resp. $L_e(\pi) \subseteq L_e(\tau)$)?) is undecidable (Jiang et al., 1994, Bremer and Freydenberger, 2012)
- other classes of pattern languages (e.g. typed pattern languages (Wright, 1990)) and relational pattern languages (Geilke and Zilles, 2011)) have been studied

Teaching Dimension

- information complexity/teaching complexity of the erasing pattern languages
- teacher-directed learning (Goldman and Kearns (1995), Rivest and Yin (1995)); helpful teacher selects instances
- labelled example for a pattern π: (w, ℓ), w is a word over Σ and

$$\ell = \begin{cases} + & \text{if } w \in L(\pi); \\ - & \text{if } w \notin L(\pi). \end{cases}$$

- the teaching dimension (TD) of a pattern π w.r.t. a class $\Pi \ni \pi$ of patterns, denoted by $\mathrm{TD}(\pi, \Pi)$, is the minimum number of labelled examples needed to uniquely identify $L(\pi)$ from all languages generated by patterns in Π
- for any given π, the size of the smallest "hitting set" for the class of all symmetric differences between $L(\pi)$ and $L(\tau)$ ranging over all τ not equivalent to π

- TD of a class Π is the maximum value of $\mathrm{TD}(\pi, \Pi)$ over all $\pi \in \Pi$
- $\pi_1 = x_1$, $\pi_2 = x_1^2$, $\pi_3 = x_1 x_2^2 x_1$

Example

Let $\Sigma = \{0, 1\}$ and $\Pi = \{\pi_1, \pi_2, \pi_3\}$. Then $\mathrm{TD}(\pi_1, \Pi) = \mathrm{TD}(\pi_2, \Pi) = 1$ and $\mathrm{TD}(\pi_3, \Pi) = \mathrm{TD}(\Pi) = 2$.

Pattern	0	01^20	0^210^2	ε
x_1	$+$	$+$	$+$	$+$
x_1^2	$-$	$-$	$-$	$+$
$x_1 x_2^2 x_1$	$-$	$+$	$-$	$+$

- $\mathrm{TD}(\pi_3, \Pi) \geq 2 \because L(\pi_2) \subset L(\pi_3) \subset L(\pi_1)$
- i.e. need 1 positive example to distinguish π_3 from π_2 and 1 negative example to distinguish π_3 from π_1

Preference-Based Teaching Dimension

- in real-world learning scenarios, even the smallest possible teaching set for a given concept relative to some concept class may be impractically large
- sometimes more reasonable to teach concepts in a "nice" order
- first teach "easier" concepts, then proceed to "harder" concepts

Preference-Based Teaching Dimension

- more generally: given any class Π of patterns, consider all strict partial orders on Π (patterns that generate the same language are identified with each other)
- $\prec$: *strict partial order* on Π, i.e., $\prec$ is asymmetric and transitive.
- for every $\pi \in \Pi$, let $\Pi_{\prec\pi} = \{\pi' \in \Pi : \pi' \prec \pi\}$ be the set of patterns over which π is strictly preferred
- a *teaching set for* π *w.r.t.* $(\Pi, \prec)$ is defined as a teaching set for π w.r.t. $\Pi \setminus \Pi_{\prec\pi}$

Preference-Based Teaching Dimension

- $\mathrm{PBTD}(\pi, \Pi, \prec) := \inf\{|T| :$
  T is a teaching set for π w.r.t. $(\Pi, \prec)\} \in \mathbb{N}_0 \cup \{\infty\}$
- $\mathrm{PBTD}(\Pi, \prec) := \sup_{\pi \in \Pi} \mathrm{PBTD}(\pi, \Pi, \prec) \in \mathbb{N}_0 \cup \{\infty\}$ is called the *teaching dimension of* $(\Pi, \prec)$
- the *preference-based teaching dimension of* Π is defined by
  $\mathrm{PBTD}(\Pi) := \inf\{\mathrm{PBTD}(\Pi, \prec) :$
  $\prec$ is a strict partial order on $\Pi\}$

Example

Let $\Sigma = \{0, 1\}$ and $\Pi = \{\pi_1, \pi_2, \pi_3\}$, where $\pi_1 = x_1$, $\pi_2 = x_1^2$ and $\pi_3 = x_1 x_2^2 x_1$. Then $\text{PBTD}(\Pi) = 1$.

- define $\prec$ on Π by $\pi_3 \succ \pi_2 \succ \pi_1$
- $\pi_1 \rightsquigarrow (0^2 10^2, +)$, $\pi_2 \rightsquigarrow (01^2 0, -)$
- since $\pi_3 \succ \pi_2$ and $\pi_3 \succ \pi_1$, π_3 does not need to be distinguished from any other pattern in Π

Finite Distinguishability Problem

- fix an alphabet size z
- given a pattern $\pi \in \Pi^z$, is $\mathrm{TD}(\pi, \Pi^z) < \infty$ (i.e. is π finitely distinguishable w.r.t. Π)?

Definition

For any $\pi \in \Pi^z$, say that π is *simple block-regular* iff the following conditions are satisfied:

1. π is block-regular;
2. π does not contain any substring $\alpha \in \Sigma^+$ such that $|\alpha| \geq 2$;
3. π starts and ends with variables.

Every simple block-regular pattern is equivalent to a pattern π' of the shape $y_1 a_1 y_2 a_2 \ldots a_k\ y_{k+1}$, where $k \geq 0$, $a_1, a_2, \ldots, a_k \in \Sigma$ and $y_1, y_2, \ldots, y_{k+1}$ are $k+1$ distinct variables (Jain et al., 2010).

Finite Distinguishability Problem

Theorem (Bayeh et al., 2017)

Suppose $4 \leq z < \infty$ and $\pi \in \Pi^z$. Then π is finitely distinguishable w.r.t. Π^z iff π is simple block-regular.

- $\Pi^z_{k,m}$ = patterns in Π^z with at most k distinct variables, each of which occurs at most m times
- constant pattern 0: infinite TD w.r.t. whole class of patterns, but finite TD w.r.t. patterns in $\Pi^z_{\infty,m}$ for any fixed m
- would restricting the maximum number of variable repetitions in every pattern π belonging to a class Π of patterns to some fixed m make every pattern in Π "easier" to teach?
- $m \rightsquigarrow$ variable frequency of Π
- what is the relationship between the variable frequency of Π and TD(Π)?

Regular Pattern Languages

- $R\Pi^z$ = class of regular patterns (patterns in which every variable occurs at most once)

Theorem (Bayeh et al., 2017)

1. $\mathrm{TD}(R\Pi^1) = 3$.
2. For all $z \geq 2$ (including $z = \infty$), $\mathrm{TD}(R\Pi^z) \geq 5$.
3. For all $z \geq 7$ (including $z = \infty$), $\mathrm{TD}(R\Pi^z) = 5$.
4. For all $z \neq 2$, $\mathrm{PBTD}(R\Pi^z) = 2$.

Theorem

$2 \leq PBTD(R\Pi^2) \leq 4$.

Simple Block-Regular Pattern Languages

- $SR\Pi^z$ = class of simple block-regular patterns (patterns equivalent to those of the shape $x_1 a_1 x_2 \dots a_k x_{k+1}$, where $a_1, \dots, a_k \in \Sigma$ and $x_1, \dots, x_{k+1} \in X$).

Theorem

1. *For any $z \in \mathbb{N} \cup \{\infty\}$, $TD(SR\Pi^z) = 2$ and $PBTD(SR\Pi^z) = 1$.*
2. *$TD(SR\Pi^z, R\Pi^z) = 3$.*

Theorem

Suppose $z \in \mathbb{N} \cup \{\infty\}$ and $\pi = x_1 c_1 x_2 \dots c_{n-1} x_n$ for some $c_1, \dots, c_{n-1} \in \Sigma$ and $n \geq 2$. (i) If $z \in \{1, \infty\}$, then $TD(\pi, \Pi^z) \in \{1, 3\}$. (ii) If $2 \leq z < \infty$, then $TD(\pi, \Pi^z) = \Omega(|\pi|)$ and $TD(\pi, \Pi^z) = O(2^{|\pi|})$.

m-Quasi-Regular Pattern Languages

- $\mathrm{QR}\Pi^{z}_{\ell,m}$ = class of ℓ-variable patterns (patterns with at most k distinct variables) in which every variable occurs *exactly* m times
- Mitchell (1998) showed that for any $m \geq 1$, the class of m-quasi-regular pattern languages is learnable in the limit
- relation between limit learning and teaching: if every member of a uniformly recursive class has finite teaching dimension, then the class is learnable in the limit

m-Quasi-Regular Pattern Languages

Theorem

If $z = 1$, then $TD(QR\Pi^z_{\infty,m}) = 3$. If $z \geq 2$, then for every $\pi \in QR\Pi^z_{\infty,m}$,
$TD(\pi, QR\Pi^z_{\infty,m}) = O(2^{|\pi(\varepsilon)|} + D \cdot (|\pi(\varepsilon)| + D \cdot m)^{D \cdot m})$, *where*
$D := \max(\{(1/m) \cdot (2 \cdot |\pi| - |\pi(\varepsilon)|), 1 + (|\pi(\varepsilon)| + m + 4) \cdot |Var(\pi)|\})$.

- $QR\Pi^z_{\ell,m,cf}$ = class of ℓ-variable constant-free m-quasi-regular patterns

Theorem

For any $z \geq 1$, $TD(QR\Pi^z_{\infty,1,cf}) = PBTD(QR\Pi^z_{\infty,1,cf}) = 0$. Suppose $m \geq 2$. If $z = |\Sigma| \geq 4m^2 + 1$, then $PBTD(QR\Pi^z_{\infty,m,cf}) = 1$.

m-Regular Non-Cross Pattern Languages

- a *non-cross pattern* π is a constant-free pattern of the shape $x_0^{n_0} x_1^{n_1} \ldots x_k^{n_k}$, where $n_0, n_1, \ldots, n_k \in \mathbb{N}$
- e.g. $x_1, x_1^2 x_2^3, x_1^3 x_2^3 x_3^2$, etc.
- $\mathrm{NC}\Pi^z_{\infty,m}$ = class of all non-cross patterns π over any Σ with $|\Sigma| = z$ such that every variable of π occurs at most m times

Theorem (Bayeh et al., 2017)

For all $z \in \mathbb{N} \cup \{\infty\}$ and $\pi \in \mathrm{NC}^z_{\infty,\infty}$ with $L(\pi) \neq L(x_1)$, $\mathrm{TD}(\pi, \mathrm{NC}^z_{\infty,\infty}) = \infty$.

Theorem

For all $z \in \mathbb{N} \cup \{\infty\}$, $TD(NC\Pi^z_{\infty,1}) = PBTD(NC\Pi^z_{\infty,1}) = 0$.
Suppose $m \geq 2$.

1. *If* $z = 1$, *then* $TD(NC\Pi^z_{\infty,m}) = \Theta(m)$ *and* $PBTD(NC\Pi^z_{\infty,m}) = \Theta(m)$.
2. *For any* $n \in \mathbb{N}_0$, *let* $\omega(n)$ *denote the number of distinct prime factors of* n *and let* $\Pi(n)$ *denote the number of prime powers not exceeding* n. *If* $z \geq 2$, *then* $\max(\{\omega(n) : n \leq m\}) \leq TD(NC\Pi^z_{\infty,m}) \leq 2 + \Pi(m-1)$ *and* $PBTD(NC\Pi^z_{\infty,m}) = PBTD(NC\Pi^z) = 1$. *In particular,* $\max(\{\omega(n) : n \leq m\}) \leq TD(NC\Pi^z_{\infty,m}) <$ $O\left((m-1)^{\frac{1}{2}} \log(m-1)\right) + \dfrac{1.25506(m-1)}{\log(m-1)}$.

Remark

- establishing the exact TD of any given pattern in $\mathrm{NC\Pi}^{z}_{\infty,m}$ (for any fixed finite $z \geq 2$ and $m \geq 2$) seems to be quite difficult in general
- verifying a lower bound for $\mathrm{TD}(\pi, \mathrm{NC\Pi}^{z}_{\infty,m})$ for any $\pi \in \mathrm{NC\Pi}^{z}_{\infty,m}$ may be reduced to deciding a Π^0_2-sentence where the quantifier-free part of the sentence is a positive word equation (combinations of word equations using $\vee$ or $\wedge$)
- problematic because the set of all $\forall\exists$-sentences over positive word equations is undecidable (Ganesh, Minnes, Solar-Lezama and Rinard, 2012)

- $\Pi^z_{\ell,m}$ = class of ℓ-variable patterns π such that every variable occurs at most m times in π
- every $\pi \in \Pi^z_{\infty,m}$ is an *m-regular pattern*
- $\Pi^z_{\infty,m,cf}$ = class of all constant-free m-regular patterns

Theorem

$PBTD(\Pi^\infty) = 2$ and for any $m \geq 1$, $PBTD(\Pi^1_{\infty,m}) = \Theta(m)$.

Theorem

Suppose $m \geq 1$.

1. *$TD(\Pi^1_{\infty,m}) \leq 2^m + m + 1$ and for all $\pi \in \Pi^\infty_{k,m}$ with $k \geq 1$, $TD(\pi, \Pi^\infty_{\infty,m}) = O((D+1)^D)$, where $D := (4mk + |\pi| + 2) \cdot m$.*

Theorem

Suppose $m \geq 1$.

1. $TD(\Pi^1_{\infty,m}) \leq 2^m + m + 1$ *and for all* $\pi \in \Pi^\infty_{k,m}$ *with* $k \geq 1$, $TD(\pi, \Pi^\infty_{\infty,m}) = O((D+1)^D)$, *where* $D := (4mk + |\pi| + 2) \cdot m$.
2. *Let* $1\Pi^z_m$ *denote the class of patterns* π *over any alphabet of size* z *such that* π *contains at most one variable that occurs more than* m *times. Suppose* $\pi \in 1\Pi^z_m$. *If* $z \geq 4$, *then* $TD(\pi, 1\Pi^z_m) < \infty$ *only if* π *contains a variable that occurs more than* m *times or* $\pi \in SR\Pi^z$. *If* $z = \infty$, *then* $TD(\pi, 1\Pi^z_m) < \infty$ *if* π *contains a variable that occurs more than* m *times or* $\pi \in SR\Pi^z$.

Theorem

(Based on (Reidenbach, 2006)) Suppose $\pi = x_1^2 x_2^2 x_3^2$. *For any* $m \geq 4$, $TD(\pi, \Pi^2_{\infty,m,cf}) = \infty$.

	$z = 1$	$2 \leq z < \infty$	$z = \infty$
$\mathrm{SR\Pi}^z$	$\mathrm{TD} = 2$, $\mathrm{PBTD} = 1$	$\mathrm{TD} = 2$, $\mathrm{PBTD} = 1$	$\mathrm{TD} = 2$, $\mathrm{PBTD} = 1$
$\mathrm{QR\Pi}^z_{\infty,m}$	$\mathrm{TD} = 3$ $\mathrm{PBTD} = 2$	$(\forall\pi)[\mathrm{TD}(\pi, \Pi) < \infty]$ $\mathrm{PBTD} \geq 2$	$(\forall\pi)[\mathrm{TD}(\pi, \Pi) < \infty]$ $\mathrm{PBTD} = 2$
$\mathrm{NC\Pi}^z_{\infty,m}$	$\mathrm{TD}/\mathrm{PBTD} = \Theta(m), m \geq 2$, $\mathrm{TD}/\mathrm{PBTD} = 0, m = 1$	$\mathrm{TD} = o(m)$, $\mathrm{PBTD} = 1, m \geq 2$, $\mathrm{PBTD} = 0, m = 1$	$\mathrm{TD} = o(m)$, $\mathrm{PBTD} = 0, m = 1$
$\Pi^z_{\infty,m}$	$\mathrm{TD} = O(2^m)$ $\mathrm{PBTD} = \Theta(m)$	$(\exists\pi)[\mathrm{TD}(\pi, \Pi^2_{\infty,4,cf}) = \infty]$ $\mathrm{PBTD} \geq 2$	$(\forall\pi)[\mathrm{TD}(\pi, \Pi) < \infty]$ $\mathrm{PBTD} = 2$

Table : TD and PBTD of various pattern classes. In each entry, $m \geq 1$, the universal (resp. existential) quantifier is taken over all patterns belonging to the class in the corresponding row and Π refers to the class in the corresponding row.

Open Problems

- characterize the subclass of m-regular patterns that are finitely distinguishable over finite alphabets of size at least 3?
- $\mathrm{TD}(\Pi^{\infty}_{\infty,m}) = \infty$ for some $m \geq 2$?
- $\mathrm{TD}(\mathrm{QR}\Pi^{z}_{\infty,m}) = \infty$ for some finite $z \geq 2$ and $m \geq 1$?

Thank you.

Thank you.

Bibliography I

[1] Dana Angluin.
Finding patterns common to a set of strings.
Journal of Computer and System Sciences, 21: 46–62, 1980.

[2] Dana Angluin.
Inductive inference of formal languages from positive data.
Information and Control, 45(2): 117–135, 1980.

[3] Setsuo Arikawa, Satoru Miyano, Ayumi Shinohara, Satoru Kuhara, Yasuhito Mukouchi, Takeshi Shinohara.
A machine discovery from amino acid sequences by decision trees over regular patterns.
New Generation Computing, 11:361–375, 1993.

[4] Joachim Bremer and Dominik D. Freydenberger.
Inclusion problems for patterns with a bounded number of variables.
Information and Computation, 220:15–43, 2012.

[5] Thomas Erlebach, Peter Rossmanith, Hans Stadtherr, Angelika Steger and Thomas Zeugmann.
Learning one-variable pattern languages very efficiently on average, in parallel, and by asking queries.
Theoretical Computer Science, 261(1): 119–156, 2001.

[6] Dominik D. Freydenberger and Daniel Reidenbach.
Bad news on decision problems for patterns.
Information and Computation, 208:83–96, 2010.

[7] Dominik D. Freydenberger and Daniel Reidenbach.
Inferring descriptive generalisations of formal languages.
Journal of Computer and System Sciences, 79(5):622–639, 2013.

[8] Ziyuan Gao, Zeinab Mazadi, Regan Meloche, Hans U. Simon and Sandra Zilles.
Distinguishing pattern languages with membership examples.
Information and Computation, 256:348–371, 2017.

[9] Michael Geilke and Sandra Zilles.
Learning relational patterns.
Proceedings of the Twenty-second International Conference on Algorithmic Learning Theory (ALT 2011), pages 84–98, 2011.

[10] Sally A. Goldman and Michael J. Kearns.
On the complexity of teaching.
Journal of Computer and System Sciences, 50(1): 20–31, 1995.

[11] Tao Jiang, Arto Salomaa, Kai Salomaa and Sheng Yu.
Decision problems for patterns.
Journal of Computer and System Sciences, 50: 53–63, 1995.

[12] Stephen C. Kleene.
Representation of events in nerve nets and finite automata.
In: C. E. Shannon, J. McCarthy, and W. R. Ashby, editors,
Automata Studies, pages 3–42.
Princeton University Press, Princeton, NJ, 1956.

[13] M. Lothaire. Combinatorics on words. Addison-Wesley Publishing Company, Inc. (1983).

[14] Robert P. Nix.
Editing by example.
ACM Transactions on Programming Languages and Systems, 7: 600–621, 1985.

[15] Daniel Reidenbach.
A non-learnable class of E-pattern languages.
Theoretical Computer Science, 350: 91–102, 2006.

[16] Ayumi Shinohara and Satoru Miyano.
Teachability in computational learning.
New Generation Computing, 8(4): 337–347, 1991.

[17] Takeshi Shinohara.
Polynomial time inference of extended regular pattern languages.
RIMS Symposia on Software Science and Engineering, Lecture Notes in Computer Science Proceedings, Vol. 147, 115–127, 1983.

[18] Sandra Zilles, Steffen Lange, Robert Holte and Martin Zinkevich.
Models of cooperative teaching and learning.
Journal of Machine Learning Research, 12: 349–384, 2011.