RUHR-UNIVERSITÄT BOCHUM

RUB

ORDER COMPRESSION SCHEMES

Malte Darnstädt, Thorsten Doliwa, Hans Ulrich Simon
and Sandra Zilles
Ruhr-University Bochum and University of Regina

24th International Conference on Algorithmic Learning Theory, Singapore
October 8, 2013

Outline

1. Sample compression schemes
2. Order Compression Schemes
3. Compression graphs
4. Connections to teaching
5. Results for special classes
6. Open questions

Outline

Sample compression schemes [Littlestone, Warmuth ’86]

- Binary concept class C over an instance space X

Example

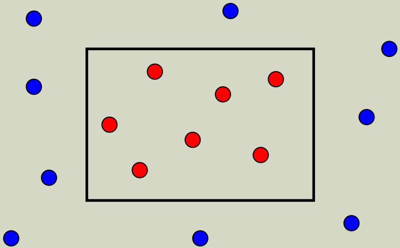
$X = \mathbb{R}^2$ $\quad C =$ orth. rectangles

Sample compression schemes [Littlestone, Warmuth ’86]

- Binary concept class C over an instance space X
- Finite sample S labeled by some concept $c \in C$

Example

$X = \mathbb{R}^2$ $C =$ orth. rectangles

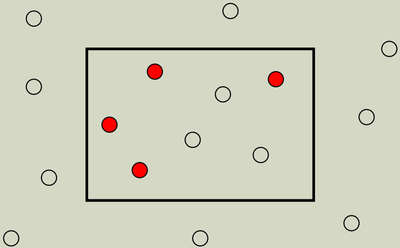


Sample compression schemes [Littlestone, Warmuth ’86]

- Binary concept class C over an instance space X
- Finite sample S labeled by some concept $c \in C$
- Compression function f
  $f(S) \subseteq S$

Example

$X = \mathbb{R}^2$ $C =$ orth. rectangles

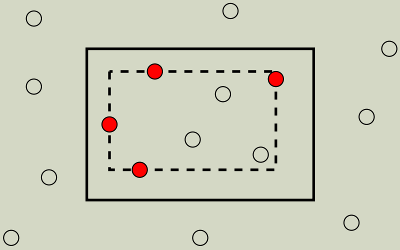


Sample compression schemes [Littlestone, Warmuth '86]

- Binary concept class C over an instance space X
- Finite sample S labeled by some concept $c \in C$
- Compression function f
  $f(S) \subseteq S$
- Reconstruction function g

Example

$X = \mathbb{R}^2$ $C =$ orth. rectangles

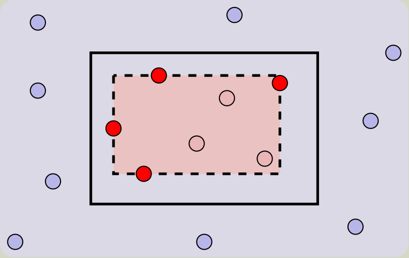


Sample compression schemes [Littlestone, Warmuth '86]

- Binary concept class C over an instance space X
- Finite sample S labeled by some concept $c \in C$
- Compression function f
  $f(S) \subseteq S$
- Reconstruction function g
  $g(f(S))$ consistent with S

Example

$X = \mathbb{R}^2$ $C =$ orth. rectangles

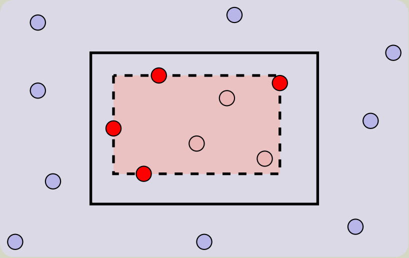


Sample compression schemes [Littlestone, Warmuth '86]

- Binary concept class C over an instance space X
- Finite sample S labeled by some concept $c \in C$
- Compression function f
  $f(S) \subseteq S$
- Reconstruction function g
  $g(f(S))$ consistent with S
- Hypothesis class H
  A scheme is called *proper* iff $H = C$

Example

$X = \mathbb{R}^2$ $C =$ orth. rectangles

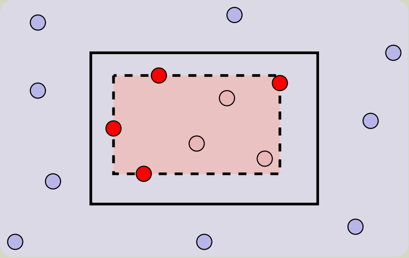


Sample compression schemes [Littlestone, Warmuth ’86]

- Binary concept class C over an instance space X
- Finite sample S labeled by some concept $c \in C$
- Compression function f
  $f(S) \subseteq S$
- Reconstruction function g
  $g(f(S))$ consistent with S
- Hypothesis class H
  A scheme is called *proper* iff $H = C$
- Size k: cardinality of the largest compression set $f(S)$

Example

$X = \mathbb{R}^2$ $C =$ orth. rectangles



Previous results [Floyd, Warmuth ’95]

Learnability

If C has a compression scheme of size k then C is PAC-learnable with a sample size of

$$O\left(\frac{k \cdot \log(1/\varepsilon) + \log(1/\delta)}{\varepsilon}\right)$$

Previous results [Floyd, Warmuth '95]

General lower bound

For any compression scheme for C holds

$$k \geq \mathrm{VCD}(C)/5$$

Upper bounds for special classes

Let C be a maximum or intersection-closed class. Then there exists a compression scheme of size

$$k = \mathrm{VCD}(C)$$

The sample compression conjecture

Conjecture [Floyd, Warmuth ’95; Warmuth ’03]

For all C exists a compression scheme of size

$$k \leq \mathrm{VCD}(C)$$

- We don’t even know if $k = O(\mathrm{VCD}(C))$
- Sufficient to prove the conjecture for finite classes [Ben-David, Litman ’98]

Outline

Order Compression Schemes

- Special case of compression schemes
- Less powerful, but easier to analyze
- Most finite classes C for which the conjecture is proven also have an Order Compression Scheme of size $\mathrm{VCD}(C)$
- A stepping stone for proving or disproving the conjecture

Teaching set

S is a *teaching set* for c with respect to C
$:\Leftrightarrow$ c is the only concept in C that is consistent with S

Teaching set

S is a *teaching set* for c with respect to C
$:\Leftrightarrow$ c is the only concept in C that is consistent with S

Order Compression Scheme (OCS)

Let C be a finite concept class and $H \supseteq C$ a finite hypothesis class.
Equip H with a total order, say $(h_1, h_2, \ldots, h_m)$.

Teaching set

S is a *teaching set* for c with respect to C
$:\Leftrightarrow$ c is the only concept in C that is consistent with S

Order Compression Scheme (OCS)

Let C be a finite concept class and $H \supseteq C$ a finite hypothesis class.
Equip H with a total order, say $(h_1, h_2, \dots, h_m)$.

1. Compression of S:
   Let t be the largest number such that h_t is consistent with S.
   Then $f(S)$ is a smallest subset of S that is a teaching set for h_t with respect to $\{h_t, \dots, h_m\}$.

Teaching set

S is a *teaching set* for c with respect to C
$:\Leftrightarrow$ c is the only concept in C that is consistent with S

Order Compression Scheme (OCS)

Let C be a finite concept class and $H \supseteq C$ a finite hypothesis class.
Equip H with a total order, say $(h_1, h_2, \dots, h_m)$.

1. Compression of S:
   Let t be the largest number such that h_t is consistent with S.
   Then $f(S)$ is a smallest subset of S that is a teaching set for h_t with respect to $\{h_t, \dots, h_m\}$.
2. Reconstruction of $f(S)$:
   Let t be the largest number such that h_t is consistent with $f(S)$.
   Then $g(f(S)) = h_t$.

Example: the class of singletons

Let $X = \{x_1, \ldots, x_n\}$ and $C_s = \{c_1, \ldots, c_n\}$ with

	x_1	x_2	x_3	$\cdots$	x_n
c_1	1	0	0	$\cdots$	0
c_2	0	1	0	$\cdots$	0
c_3	0	0	1	$\cdots$	0
$\vdots$				$\ddots$	
c_n	0	0	0	$\cdots$	1

Example: the class of singletons

Let $X = \{x_1, \ldots, x_n\}$ and $C_s = \{c_1, \ldots, c_n\}$.

OCS with $H = C_s$

W.l.o.g. let the ordering be $(c_1, c_2, \ldots, c_n)$.

If S contains a 1-labeled point it is compressed to this single point.

But $S = \{(x_2, 0), (x_3, 0), \ldots, (x_n, 0)\}$ is a teaching set for c_1 in H and no proper subset of S has this property.

$\Rightarrow$ The size of any proper OCS is $n - 1$.

Example: the class of singletons

Let $X = \{x_1, \ldots, x_n\}$ and $C_s = \{c_1, \ldots, c_n\}$.

OCS with $H = C_s \cup \{\text{all-0}\}$

Use the ordering $(c_1, c_2, \ldots, c_n, \text{all-0})$.

If S contains a 1-labeled point it is compressed to this single point; otherwise to the empty set.

$\Rightarrow$ The size of this improper OCS is 1.

Note: The all-zero hypothesis *must* be last (or second-last) in the ordering.

Example: the class of singletons

Let $X = \{x_1, \ldots, x_n\}$ and $C_s = \{c_1, \ldots, c_n\}$.

Contrast: general proper scheme of size 1

If S contains a 1-labeled point it is compressed to this single point; otherwise to a point $(x_i, 0)$ such that x_{i+1} is not in S.
Reconstruction works in the obvious way.

E.g. $S = \{(x_1,0),(x_2,0),(x_4,0)\} \quad f(S) = \{(x_2,0)\} \quad g(f(S)) = c_3$

$\Rightarrow$ The size of this compression scheme is 1.

The Order Compression Number

Order Compression Number

$$\mathrm{OCN}(C) := \min_{H \supseteq C} \; \min_{\text{order of } H} \; \text{size the OCS using } H \text{ for } C$$

The Order Compression Number

Order Compression Number

$$\mathrm{OCN}(C) := \min_{H \supseteq C} \; \min_{\text{order of } H} \; \text{size the OCS using } H \text{ for } C$$

Monotonicity

For all $C' \subseteq C$ and $X' \subseteq X$ holds

$$\mathrm{OCN}(C) \geq \mathrm{OCN}(C')$$

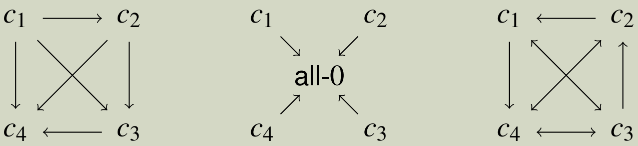
$$\mathrm{OCN}(C) \geq \mathrm{OCN}(C_{|X'})$$

Outline

The compression graph

- A general tool for sample compression schemes
- Especially useful for Order Compression Schemes

Example: the class of singletons (for $n = 4$)



Compression graph for a scheme (f, g)

Define $G(f,g) := (V,E)$, where

1. V is the set of hypotheses H
2. $(h_1, h_2) \in E$ iff there exists a sample S labeled by some $c \in C$, s.t. both h_1 and h_2 are consistent with $f(S)$ and $g(f(S)) = h_2$

Acyclic Compression Schemes

Theorem

$G(f,g)$ is acyclic $\iff$ (f,g) is an Order Compression Scheme

- Topological ordering of $G(f,g) \longleftrightarrow$ total order of H
- Technicality: normalization of f necessary
- E.g. the compression scheme for rectangles is an OCS (for finite C)

Outline

Teaching [Goldman, Kearns '95; Shinohara, Miyano '91]

Teaching	Sample Compression

Teaching [Goldman, Kearns ’95; Shinohara, Miyano ’91]

Teaching		Sample Compression
concept class C	$\longleftrightarrow$	concept class C

Teaching [Goldman, Kearns ’95; Shinohara, Miyano ’91]

Teaching		Sample Compression
concept class C	$\longleftrightarrow$	concept class C
teacher	$\longleftrightarrow$	compression function

Teaching [Goldman, Kearns ’95; Shinohara, Miyano ’91]

Teaching		Sample Compression
concept class C	$\longleftrightarrow$	concept class C
teacher	$\longleftrightarrow$	compression function
all elements of X labeled	$<$	subsets of X labeled

Teaching [Goldman, Kearns ’95; Shinohara, Miyano ’91]

Teaching		Sample Compression
concept class C	$\longleftrightarrow$	concept class C
teacher	$\longleftrightarrow$	compression function
all elements of X labeled	$<$	subsets of X labeled
learner	$\longleftrightarrow$	reconstruction function

Teaching [Goldman, Kearns ’95; Shinohara, Miyano ’91]

Teaching		Sample Compression
concept class C	$\longleftrightarrow$	concept class C
teacher	$\longleftrightarrow$	compression function
all elements of X labeled	$<$	subsets of X labeled
learner	$\longleftrightarrow$	reconstruction function
no "coding tricks"	$>$	relies heavily on such tricks

Teaching [Goldman, Kearns ’95; Shinohara, Miyano ’91]

Teaching		Sample Compression
concept class C	$\longleftrightarrow$	concept class C
teacher	$\longleftrightarrow$	compression function
all elements of X labeled	$<$	subsets of X labeled
learner	$\longleftrightarrow$	reconstruction function
no "coding tricks"	$>$	relies heavily on such tricks

$\uparrow$

relax this requirement

Recursive teaching [Zilles et al. ’08; Doliwa et al. ’10]

Teaching plan

A *teaching plan* for $C = \{c_1, \ldots, c_m\}$ is a sequence

$$\big((c_1, S_1), (c_2, S_2), \ldots, (c_m, S_m)\big)$$

such that S_i is a teaching set for c_i in $\{c_i, \ldots, c_m\}$.

The *order* of a teaching plan is the cardinality of the largest S_i.

Recursive Teaching Dimension

$$\begin{aligned} \mathrm{RTD}(C) &:= \text{minimum order over all teaching plans of } C \\ \mathrm{RTD}^*(C) &:= \max_{X' \subseteq X} \mathrm{RTD}(C_{|X'}) \end{aligned}$$

Implications for OCN

Lemma

$$\mathrm{OCN}(C) \geq \mathrm{RTD}(C)$$

Implications for OCN

Lemma

$$\mathrm{OCN}(C) \geq \mathrm{RTD}(C)$$

Theorem

$$\mathrm{OCN}(C) \geq \mathrm{RTD}^*(C)$$

Implications for OCN

Lemma

$$\mathrm{OCN}(C) \geq \mathrm{RTD}(C)$$

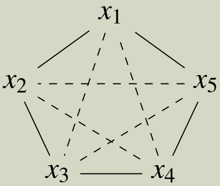
Theorem

$$\mathrm{OCN}(C) \geq \mathrm{RTD}^*(C)$$

Corollary

$$\mathrm{OCN}(C) \geq \mathrm{VCD}(C)$$

The class C_{MW}



$$\begin{aligned} \mathrm{VCD}(C_{MW}) &= 2 \\ \mathrm{RTD}(C_{MW}) &= 3 \\ \mathrm{OCN}(C_{MW}) &= 3 \end{aligned}$$

- Found by Manfred Warmuth
- The smallest class with the property $\mathrm{RTD}(C) > \mathrm{VCD}(C)$ [Doliwa et al. '10]
- Padding yields classes of arbitrary high VC dimension with $\mathrm{OCN}(C) = 3/2 \cdot \mathrm{VCD}(C)$
- C_{MW} has a *cyclic* compression scheme of size 2

Outline

Results for special classes

Theorem

For intersection-closed or maximum classes C holds

$$\mathrm{OCN}(C) = \mathrm{VCD}(C)$$

Check the compression graph of known compression schemes!

Results for special classes

Theorem

For intersection-closed or maximum classes C holds

$$\mathrm{OCN}(C) = \mathrm{VCD}(C)$$

Check the compression graph of known compression schemes!

- For intersection-closed classes: the standard scheme, using spanning sets [Floyd, Warmuth '95]
- For maximum classes: based on the Tail Matching Algorithm [Kuzmin, Warmuth '07]

Corollaries

Using results by [Ben-David, Litman ’98] and [Welzl, Wöginger ’87] we get immediately:

Corollary

For Dudley classes and classes of VC dimension 1 holds:

$$\mathrm{OCN}(C) = \mathrm{VCD}(C)$$

In the paper: also a result for nested differences of intersection-closed classes

Outline

Open questions

- How do OCN and VCD relate in general?
- How can one find the optimal hypothesis space H for a given class C?
- How can one find the optimal ordering of H?

Thank you!

Do you have any questions?