Sample Compression for Multi-label Concept Classes

Rahim Samei
Ph.D. Student in Computer Science
University of Regina

Joint work with

Pavel Semukhin
Kurt Gödel Research Centre, Austria

Boting Yang and Sandra Zilles
Department of Computer Science
University of Regina

June 2014

Sample Compression Scheme (SCS)

Let n be a natural number and $C \subseteq \{0,1\}^n$.

What is an SCS for C? [Littlestone and Warmuth, 1986]

$c \in C$	X_1	X_2	X_3	X_4	X_5
c_1	0	0	0	0	0
c_2	0	0	0	1	0
c_3	0	1	0	0	1
c_4	1	1	0	1	1
c_5	1	0	1	0	1

$\mathbf{VCD(C) = 2}$

Sample $S = \{(X_2, 1), (X_4, 0), (X_5, 1)\}$

- Compression:
  $f(S) = S' = \{(X_2, 1), (X_5, 1)\} \subseteq S$
- Decompresssion (reconstruction):
  $g(S') = \{(X_1, 1), (X_2, 1), (X_3, 0), (X_4, 0), (X_5, 1)\} \supseteq S$
- Size of the scheme: Size of the **largest** $f(S)$

Sample Compression Scheme (SCS)

Conjecture [Littlestone and Warmuth, 1986; Floyd and Warmuth 1995]

If $\mathbf{VCD}(C) = d$ then C has an SCS of size d.

Proven so far when

- C is VCD-maximum (cardinality of C meets upper bound proven by Sauer, 1972) [Floyd and Warmuth 1995]
- $\mathrm{VCD}(C) = 1$ (because every such C is contained in a maximum class of $\mathrm{VCD}\ 1$) [Welzl and Woeginger, 1987]

Our Work:

Extending study of SCS to multi-label concept classes.

$c \in C$	X_1	X_2	X_3
c_1	0	0	0
c_2	0	1	1
c_3	2	1	0
c_4	1	1	2
c_5	3	2	0
c_6	0	2	2

How to define VCD for such classes?

Many suggestions in literature:

- Natarajan, 1989
- Pollard, 1990
- Gurvits, 1997

We studied a notion
that upper-bounds all existing ones.

Label Mapping and VCD^*

$c \in C$	X_1	X_2	X_3
c_1	2	0	0
c_2	0	1	1
c_3	1	1	2
c_4	3	2	2

$\xRightarrow{0,1\to 0 \text{ and } 2\to 1}$

$c \in C$	X_1	X_2	X_3
c_1	2	0	0
c_2	0	1	0
c_3	1	1	1
c_4	3	2	1

- Consider **all** combinations of label mappings on **all** columns
- The image of such mapping is a **binary** class
- $\mathrm{VCD}^*:=$ **maximum** VCD over all these binary classes

Analogue of **Sauer** bound on size of C with $\mathrm{VCD}^*(C) = d$ exists. [Gurvits 1997]

- C is VCD^*-maximum iff cardinality of C meets this bound

Results:

(1) Maximum classes of $\mathrm{VCD}^*\ d$ have SCS of size d

- extending Floyd and Warmuth's idea
- using a new technical result on structure of VCD^*-maximum classes

(2) Classes of $\mathrm{VCD}^*\ 1$ have SCS of size 1

- independent proof required (classes of $\mathrm{VCD}^*\ 1$ may not be contained in maximum classes of $\mathrm{VCD}^*\ 1$)

Remark: Meanwhile, we have extended these results to Graph dimension and found an analogous result to unlabeled SCS for maximum classes

Main Technical Result

Reduction of C w.r.t. X_4: all concepts on $\{X_1, X_2, X_3\}$ that have both extensions on X_4.

$c \in C$	X_1	X_2	X_3	X_4
c_1	0	0	0	0
c_2	0	0	0	1
c_3	1	1	0	1
c_4	1	1	1	0
c_5	0	1	1	0
c_6	0	1	1	1

$\xrightarrow[\text{w.r.t. } X_4]{\text{reduction}}$

$c' \in C'$	X_1	X_2	X_3
c'_1	0	0	0
c'_2	0	1	1

Floyd and Warmuth's scheme makes use of the fact that reductions of maximum classes are maximum.

Main Technical Result

Multi-label case: how to define reduction?

(a) concepts that have more than one extension

(b) concepts that have all possible extensions

$c \in C$	X_1	X_2	X_3
c_1	0	0	0
c_2	0	0	1
c_3	2	1	1
c_4	1	2	0
c_5	1	2	1
c_6	1	2	2
c_7	0	2	0

(a) reduction $\Longrightarrow$ w.r.t. X_3

$c' \in C'$	X_1	X_2
c'_1	0	0
c'_2	1	2

(b) reduction $\Longrightarrow$ w.r.t. X_3

$c'' \in C''$	X_1	X_2
c''_1	1	2

Crucial structure: For $\mathbf{VCD}^*$-**maximum** classes (a) and (b) are **equivalent**, i.e., each concept on $\{X_1, X_2\}$ has either a single extension on X_3 or all possible extensions on X_3.