JMLR: Workshop and Conference Proceedings vol 40:1–3, 2015

Open Problem: Recursive Teaching Dimension Versus VC Dimension

Hans Simon HANS.SIMON@RUB.DE
Fakultät für Mathematik, Ruhr-Universität Bochum, D-44780 Bochum, Germany

Sandra Zilles ZILLES@CS.UREGINA.CA
Department of Computer Science, University of Regina, Regina, SK, Canada S4S 0A2

Abstract

The Recursive Teaching Dimension (RTD) of a concept class $\mathcal{C}$ is a complexity parameter referring to the worst-case number of labelled examples needed to learn any target concept in $\mathcal{C}$ from a teacher following the recursive teaching model. It is the first teaching complexity notion for which interesting relationships to the VC dimension (VCD) have been established. In particular, for finite maximum classes of a given VCD d, the RTD equals d. To date, there is no concept class known for which the ratio of RTD over VCD exceeds $3/2$. However, the only known upper bound on RTD in terms of VCD is exponential in the VCD and depends on the size of the concept class. We pose the following question: is the RTD upper-bounded by a function that grows only linearly in the VCD?

Answering this question would further our understanding of the relationships between the complexity of teaching and the complexity of learning from randomly chosen examples. In addition, the answer to this question, whether positive or negative, is known to have implications on the study of the long-standing open sample compression conjecture, which claims that every concept class of VCD d has a sample compression scheme in which samples for concepts in the class are compressed to subsets of size no larger than d.

1. Introduction

In Machine Learning, an important aspect of efficient algorithms is the amount of data that needs to be processed for successful learning. We will refer to the worst case number of information items (in our context, labelled examples) needed to identify any concept in a given concept class as the *information complexity* of the class.

In PAC-learning, for example, the information complexity is upper-bounded and lower-bounded by functions that are linear in the VC dimension (VCD) of the concept class (Blumer et al., 1989). In most models of query learning, the number of queries asked or the number of prediction mistakes made may define the notion of information complexity. In the original model of teaching (Goldman and Kearns, 1995; Shinohara and Miyano, 1991), information complexity is measured by the worst-case number of examples required to distinguish a target concept from all other concepts in the class; the latter is called the teaching dimension (TD) of the class.

One of the most fundamental questions in computational learning theory is how such information complexity parameters relate to each other. The open problem formulated below is whether any linear function of the VCD of a finite concept class gives an upper bound on the recursive teaching dimension Zilles et al. (2011) of that class.

2. Problem Formulation

Let X be a finite set and $\mathcal{C}$ a concept class over X, i.e., concepts are subsets of X. A labelled example is a pair $(x, l) \in X \times \{0, 1\}$. $\mathrm{VCD}(\mathcal{C})$ denotes the VC dimension of a concept class $\mathcal{C}$. A set S of labelled examples is consistent with a concept C, if $(x, l) \in S$ implies $C(x) = l$. (Here we treat a concept as a binary function on X.) A *teaching set* for a concept $C \in \mathcal{C}$ is a set of labelled examples that is consistent with C, but with no other concept in $\mathcal{C}$. The quantity $\mathrm{TD}(C; \mathcal{C})$ refers to the size of the smallest teaching set for $C \in \mathcal{C}$. Then $\mathrm{TD}(\mathcal{C}) = \max_{C \in \mathcal{C}} \mathrm{TD}(C; \mathcal{C})$ is called the *teaching dimension* of $\mathcal{C}$ (Goldman and Kearns, 1995; Shinohara and Miyano, 1991). By $\mathrm{TD}_{min}(\mathcal{C}) = \min_{C \in \mathcal{C}} \mathrm{TD}(C; \mathcal{C})$ we refer to the *best-case teaching dimension* of $\mathcal{C}$.

The recursive teaching dimension (RTD) refers to the size of the largest teaching set used in the so-called recursive teaching protocol (Zilles et al., 2011). It is computed by recursively removing from the given concept class all concepts whose teaching dimension with respect to the remaining concepts is smallest. The largest value of these smallest teaching dimensions encountered in the recursive process is then the RTD of the concept class.

Definition 1 (Zilles et al. (2011)) *The* teaching hierarchy of $\mathcal{C}$ *is the sequence* $(\mathcal{C}_1^{min}, \ldots, \mathcal{C}_z^{min})$ *defined as follows. Let* $\mathcal{C}_1 = \mathcal{C}$ *and let, for all* $t \in \{1, \ldots, z\}$,

- $\mathcal{C}_t^{min} = \{C \in \mathcal{C}_t \mid \mathrm{TD}(C; \mathcal{C}_t) = \mathrm{TD}_{min}(\mathcal{C}_t)\}$, *and*
- $\mathcal{C}_{t+1} = \mathcal{C}_t \setminus \mathcal{C}_t^{min}$.

Here z *is the smallest value such that* $\mathcal{C}_{z+1} = \emptyset$. *The recursive teaching dimension of* $\mathcal{C}$, *denoted by* $\mathrm{RTD}(\mathcal{C})$, *is then defined as* $\mathrm{RTD}(\mathcal{C}) = \max_{1 \le t \le z} \mathrm{TD}_{min}(\mathcal{C}_t)$.

For example, if $\mathcal{C}$ is the class of all singletons and the empty concept over $X = \{1, \ldots, n\}$, then $\mathrm{RTD}(\mathcal{C}) = \mathrm{VCD}(\mathcal{C}) = 1$. With respect to $\mathcal{C}$, the teaching dimension of each singleton is 1, while that of the empty concept is n. However, the teaching hierarchy of $\mathcal{C}$ contains the class of all singletons as $\mathcal{C}_1^{min}$, thus leaving only the empty concept C_0 at the second stage, which then has a teaching dimension of 0 with respect to $\mathcal{C}_2 = \{C_0\}$. We ask the following question:

> For finite concept classes $\mathcal{C}$, is $\mathrm{RTD}(\mathcal{C}) \in O(\mathrm{VCD}(\mathcal{C}))$, i.e., does the RTD of a finite concept class never exceed its VCD by more than a constant factor c?

It has been noted (Doliwa et al., 2014) that this question is equivalent to the following question: For finite concept classes $\mathcal{C}$, is $\mathrm{TD}_{min}(\mathcal{C}) \in O(\mathrm{VCD}(\mathcal{C}))$, i.e., does the best-case teaching dimension of a finite concept class never exceed its VCD by more than a constant factor c?

3. Relevant Known Results and Potential Implications of Solving the Problem

This problem was first addressed by Kuhlmann (1999), who showed that there are finite concept classes $\mathcal{C}$ with $\mathrm{RTD}(\mathcal{C}) = \frac{3}{2}\mathrm{VCD}(\mathcal{C})$. The smallest such class was found by Manfred K. Warmuth and published by Doliwa et al. (2014). To date, no finite class $\mathcal{C}$ with $\mathrm{RTD}(\mathcal{C}) > \frac{3}{2}\mathrm{VCD}(\mathcal{C})$ has been found.

It is known that any maximum class $\mathcal{C}$, i.e., a class $\mathcal{C}$ whose size equals the Sauer bound, fulfills $\mathrm{RTD}(\mathcal{C}) = \mathrm{VCD}(\mathcal{C})$, and so does every class of VCD 1 (Doliwa et al., 2014). For intersection-closed classes, VCD upper-bounds RTD (Doliwa et al., 2014). The only general upper bound on RTD in terms of the VCD that can be found in the literature was proven by Moran et al. (2015); it is exponential in VCD and depends on the size of the concept class.

What if RTD is linearly bounded by the VCD? This would, first of all, imply a close relationship between learning from randomly chosen examples and learning from teachers, in terms of information complexity. The structural properties that make a concept class hard or easy to learn from randomly chosen examples would have implications on the difficulty of teaching the class by yielding an upper bound on the number of examples needed in the worst case. Second, any relationship between RTD and VCD may bring new insights into the long-standing sample compression conjecture which claims that every concept class has a sample compression scheme of size (linear in) its VCD (Warmuth, 2003). In the proof that $\mathrm{RTD} = \mathrm{VCD}$ for maximum classes, the recursive teaching examples form compression sets equal to the ones of size at most VCD resulting from Rubinstein and Rubinstein's (2012) corner-peeling. Further relationships between RTD and sample compression were established in the study of order compression schemes (Darnstädt et al., 2013).

What if RTD is not linearly bounded by the VCD? One possible approach to proving the sample compression conjecture would be to show that every finite concept class can be embedded into a maximum class with only linear increase in the VCD. As Doliwa et al. (2014) observed, this approach will be fruitless if RTD is not linearly bounded in VCD.

Hence, we believe that answering the question formulated above will entail substantial progress in the study of teaching models, information complexity in general, as well as sample compression.

References

Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *JACM*, 36(4):929–965, 1989.

Malte Darnstädt, Thorsten Doliwa, Hans Ulrich Simon, and Sandra Zilles. Order compression schemes. In *ALT*, pages 173–187, 2013.

Thorsten Doliwa, Gaojian Fan, Hans Ulrich Simon, and Sandra Zilles. Recursive teaching dimension, VC-dimension and sample compression. *J. Mach. Learn. Res.*, 15:3107–3131, 2014.

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

Christian Kuhlmann. On teaching and learning intersection-closed concept classes. In *EuroCOLT*, pages 168–182, 1999.

Shay Moran, Amir Shpilka, Avi Wigderson, and Amir Yehudayoff. Teaching and compressing for low VC-dimension. *CoRR*, abs/1502.06187, 2015.

Benjamin I. P. Rubinstein and J. Hyam Rubinstein. A geometric approach to sample compression. *J. Mach. Learn. Res.*, 13:1221–1261, 2012.

Ayumi Shinohara and Satoru Miyano. Teachability in computational learning. *New Gen. Comput.*, 8:337–348, 1991.

Manfred K. Warmuth. Compressing to VC dimension many points. In *COLT*, pages 743–744, 2003.

Sandra Zilles, Steffen Lange, Robert Holte, and Martin Zinkevich. Models of cooperative teaching and learning. *J. Mach. Learn. Res.*, 12:349–384, 2011.