Recursive Teaching Dimension Versus VC Dimension

Zehua Lai
2014012668

Ziheng Huang
2014012678

He Li
2014012684

Abstract

In this work we study the quantitative relation between the recursive teaching dimension (RTD) and the VC dimension (VCD) of concept classes of finite sizes. The RTD of a concept class is a combinatorial complexity measure characterized by the worst-case number of examples necessary to identify a concept according to the recursive teaching model. An open problem 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., 2013) of that class. In this article, we will review related works in this topic and discuss about the challenges in fully solving the open problem.

1. Introduction

Sample complexity is one of the central concepts in machine learning. It is the amount of data needed to achieve a desired learning accuracy. For example, in PAC-learning, sample complexity is characterized by the VC dimension (VCD) of the concept class (Vapnik and Chervonenkis, 1971). PAC-learning is a passive learning model. In this model, the role of the teacher is limited to providing labels to data randomly drawn from the underlying distribution. Different from PAC-learning, there are important models in which teacher involves more actively in the learning process. For example, in the classical teaching model (Goldman and Kearns, 1995), the teacher chooses a set of labeled examples so that the learner, after receiving the examples, can distinguish the target concept from all other concepts in the concept class. In this model, the key complexity measure of a concept class is the teaching dimension, which is defined as the worst-case number of examples needed to be selected by the teacher.

One of the most important questions in computational learning theory is how such information complexity parameters relate to each other. As Simon and Zilles (2015) posed, an open problem is: is the RTD upper-bounded by a function that grows only linearly in the VCD?

2. Definitions and Notation

X denotes a finite set Sometimes, for brevity, we let $X = [n] := \{1, 2, ..., n\}$. A concept C is a subset of X. $\mathcal{C}$ denotes a concept class over X, which is a subset of $2^{|X|}$. For $X' \subseteq X$, we define $\mathcal{C}_{|X'} := \{C \cap X' | C \in \mathcal{C}\}$. We treat concepts interchangeably as subsets of X and as 0,1-valued functions on X. A labeled example is a pair (x, l) with $x \in X$ and $l \in \{0, 1\}$. A labeled example (x, l) is consistent with a concept C if $\chi(x \in C) = l$. If S is a set of labeled examples, we define $X(S) = \{x \in X | (x, *) \in S\}$.

$A \subseteq X$ is said to be shattered by $\mathcal{C}$ if $\mathcal{C}_{|A} = 2^{|A|}$. VCD($\mathcal{C}$) denotes the VC-dimension of a concept class $\mathcal{C}$, which is the maximum size of a shattered subset of X.

A teaching set for a concept $C \in \mathcal{C}$ is a set S of labeled examples such that C, but no other concept in $\mathcal{C}$, is consistent with S. Let $\mathcal{TS}(C; \mathcal{C})$ denote the family of teaching sets for $C \in \mathcal{C}$, let $TS(C; \mathcal{C})$ denote the size of the smallest teaching set for $C \in \mathcal{C}$.

Definition 1 (Zilles et al. 2013) *A teaching plan P for $\mathcal{C}$ is a sequence $((C_1, S_1), ..., (C_N, S_N))$, such that $\mathcal{C} = \{C_1, ..., C_n\}$ and $\forall t = 1, 2, ..., N$, $S_t \in \mathcal{TS}(C_t; \{C_t, ..., C_N\})$. The quantity $ord(P) := max_{t=1,...,N}|S_t|$ is called the order of the teaching plan P. Finally, we define recursive teaching dimension RTD($\mathcal{C}$) := $min\{ord(P)|$ P is a teaching plan for $\mathcal{C}\}$.*

It is clear that RTD($\mathcal{C}$) is monotonic. If $\mathcal{C}' \subseteq \mathcal{C}$, RTD($\mathcal{C}'$) $\leq$ RTD($\mathcal{C}$). We have several propositions which can be used as equivalent definitions for recursive teaching dimension.

Proposition 2 *P is called a canonical teaching plan for $\mathcal{C}$, if for any $i, j \in \{1, ..., N\}$, $i < j$, we have $TS(C_i; \{C_i, ..., C_N\}) \leq TS(C_j; \{C_i, ..., C_N\})$. For a canonical teaching plan P^*, RTD($\mathcal{C}$) = $ord(P^*)$.*

Proof For a canonical teaching plan P^*, it is clear that RTD($\mathcal{C}$) $\leq ord(P^*)$. We need to prove RTD($\mathcal{C}$) $\geq ord(P^*)$. Let $P^* = ((C_1, S_1), ..., (C_N, S_N))$, and another teaching plan $P = ((C_1', S_1'), ..., (C_N', S_N'))$. Choose the minimal j such that $|S_j| = TS(C_j; \{C_j, ..., C_N\}) = ord(P^*)$. Choose the minimal i such that $C_i' \in \{C_j, ..., C_N\}$. So, $C_i' = C_k$ for some k. By the definition of canonical teaching plan, $TS(C_k; \{C_j, ..., C_N\}) \geq TS(C_j; \{C_j, ..., C_N\}) \geq ord(P^*)$. $ord(P) \geq TS(C_i'; \{C_i', ..., C_N'\}) \geq TS(C_k; \{C_j, ..., C_N\}) \geq ord(P^*)$ hence finish the proof. ■

Proposition 3 *Let $TS_{min}(\mathcal{C}) := min_{C \in \mathcal{C}} TS(C; \mathcal{C})$. Then RTD($\mathcal{C}$) = $max_{\mathcal{C}' \subseteq \mathcal{C}} TS_{min}(\mathcal{C}')$*

Proof For a canonical teaching plan P, we have $TS_{min}(\mathcal{C}) = TS(C_1; \mathcal{C})$. It follows that RTD($\mathcal{C}$) = $max\{TS(C_1; \mathcal{C})$, RTD($\mathcal{C} \backslash C_1$)$\}$ = $max\{TS_{min}(\mathcal{C})$, RTD($\mathcal{C} \backslash C_1$)$\}$. RTD($\mathcal{C}$) $\leq max_{\mathcal{C}' \subseteq \mathcal{C}} TS_{min}(\mathcal{C}')$ follows inductively. As for the reverse direction, let $\mathcal{C}' \subseteq \mathcal{C}$ be the maximizer of TS_{min}. We have RTD($\mathcal{C}$) $\geq$ RTD($\mathcal{C}'$) $\geq TS_{min}(\mathcal{C}') = max_{\mathcal{C}' \subseteq \mathcal{C}} TS_{min}(\mathcal{C}')$ ■

The open problem under consideration is that whether RTD can be linear bounded by VCD.

3. RTD for Special Classes

The following families of concept classes $\mathcal{C}$ are known to have recursive teaching dimension of size equal or smaller than VCD($\mathcal{C}$).

Definition 4 *A concept class $\mathcal{C}$ of VC-dimension d over X of cardinality n is called maximum classes if $\mathcal{C}$ achieved the Sauer's bound(Sauer, 1972), $|\mathcal{C}| = \sum_{i=0}^{d} C_n^i$.*

Definition 5 *A concept class $\mathcal{C}$ is called intersection-closed if the intersection of any two concepts from $\mathcal{C}$ is itself a concept in $\mathcal{C}$ as well.*

Definition 6 *Let $\mathcal{F}$ be a vector space of real-valued functions over some domain X and $h : X \to R$. For every $f \in \mathcal{F}$, let $C_f(x) = 1$ if $h(x) + f(x) > 0$, else $C_f(x) = 0$. Then $D_{\mathcal{F},h} = \{C_f | f \in \mathcal{F}\}$ is called a Dudley class. The VC-dimension of $D_{\mathcal{F},h}$ is equal to the dimension of the vector space $\mathcal{F}$.*

In Darnstdt et al. (2016), it has been proved that

Theorem 7 *For a concept class $\mathcal{C}$, if it is a maximum class, intersection-closed class or of VC-dimension 1, then RTD($\mathcal{C}$) = VCD($\mathcal{C}$). If it is a Dudley class, then RTD($\mathcal{C}$) $\leq$ VCD($\mathcal{C}$).*

Above result involved explicit construction of a teaching plan or some compression schemes. However, for general concept classes, a explicit construction may be hard to find. For the Dudley class case, Ben-David and Litman (1998) prove that Dudley classes of dimension k are embeddable in maximum classes of VC-dimension k. It is not known to date whether every concept class C of VC-dimension d can be embedded into a maximum concept class of VC-dimension $O(d)$. Indeed, finding such an embedding is considered as a promising method for solving the open problem. Liu and Zhu (2016) also analysed some special settings and gave the teaching dimension for ridge regression, support vector machines, and logistic regression.

Another worth noting approach for analysing teaching complexity is the algebraic method. Samei et al. (2014) used algebraic method to proved several important propositions of RTD and VCD, including the Sauer's bound for RTD and the properties of VCD and RTD-maximum class. Considering the consistency of this method in proving the Sauer's bound for both RTD and VCD, it is a promising method to find more connections between the two complexity concepts.

4. Quadratic Upper Bound of RTD

In the paper of Chen et al. (2016), they proved the first bound for RTD of general concept class. Let $f(d) = max_{\mathcal{C}:VCD(\mathcal{C})\leq d} TS_{min}(\mathcal{C})$. They used the method in Kuhlmann (1999) and Moran et al. (2015) to prove that $f(d) \leq 2^d(d-1)+1+f(d-1)$ and $f(d) \leq 2^{d+1}(d-2)+d+4$ by induction.

Hu et al. (2017) generalized this method to analysed (x, y)-class. A concept class $\mathcal{C}$ is an (x, y)-class if for any $A \subseteq X$, $|A| \leq x$, we have $|\mathcal{C}_{|A}| \leq y$. Following theorem can be proved:

Theorem 8 *For any positive integer x, y, z such that $y \leq 2x - 1$ and $z \leq 2y + 1$, the following inequality holds:*

$$f(x + 1, z) \leq f(x, y) + \left\lceil \frac{(y + 1)(x - 1) + 1}{2y - z + 2} \right\rceil$$

Noting that when $x = d-1, y = 2^{d-1}-1, z = 2^d - 1$, the inequality coincides with $f(d) \leq 2^d(d-1) + 1 + f(d-1)$, so it is a generalization of the latter inequality. The essential finding is the following: for a concept class with VC-dimension d, it it not only a $(d+1, 2^{d+1}-1)$-class, but also a $(n, \sum_{i=0}^{d} C_n^i)$-class. Furthermore, to optimize the induction steps from $f(1,1)$ to $f(x,y)$, we can show that a exponential growth is the best choice by simple calculation. $f(x, \lfloor a^x \rfloor)$ where $a = 1.71757$ and $x = 4.71607d$ will give us the quadratic bound. In fact, the exponential bound in Chen et al. (2016)'s method has not really been improved, we just found a clever way to ignore it. We may still need a new combinatorial technic to fully solve the problem.

References

Shai Ben-David and Ami Litman. Combinatorial variability of vapnik-chervonenkis classes with applications to sample compression schemes. *Discrete Applied Mathematics*, 86(1): 3–25, 1998.

Xi Chen, Yu Cheng, and Bo Tang. On the recursive teaching dimension of vc classes. In *Advances in Neural Information Processing Systems*, pages 2164–2171, 2016.

Malte Darnstdt, Thorsten Kiss, Hans Ulrich Simon, and Sandra Zilles. Order compression schemes. *Theoretical Computer Science*, 620(C):73–90, 2016.

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

Lunjia Hu, Ruihan Wu, Tianhong Li, and Liwei Wang. Quadratic upper bound for recursive teaching dimension of finite vc classes. *arXiv preprint arXiv:1702.05677*, 2017.

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

Ji Liu and Xiaojin Zhu. The teaching dimension of linear learners. *Journal of Machine Learning Research*, 17(162):1–25, 2016.

Shay Moran, Amir Shpilka, Avi Wigderson, and Amir Yehudayoff. Compressing and teaching for low vc-dimension. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 40–51. IEEE, 2015.

Rahim Samei, Pavel Semukhin, Boting Yang, and Sandra Zilles. Algebraic methods proving sauer's bound for teaching complexity. *Theoretical Computer Science*, 558:35–50, 2014.

N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory*, 13(1): 145–147, 1972.

Hans U Simon and Sandra Zilles. Open problem: Recursive teaching dimension versus vc dimension. In *Conference on Learning Theory*, pages 1770–1772, 2015.

V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, 17(2):264–280, 1971.

Sandra Zilles, Steffen Lange, Robert Holte, and Martin Zinkevich. Models of cooperative teaching and learning. *Journal of Machine Learning Research*, 12(12):349–384, 2013.