Learning Generalized Scoring Rules

Nika Haghtalab
Computer Science Department
Carnegie Mellon University
nika@cmu.edu

Abstract

Generalized Scoring Rules (GSRs), which were recently introduced by Xia and Conitzer (2008), form a broad category that includes most of the commonly used voting rules. In this work, we examine the PAC-learnability of GSRs based on some of their important properties: degree of consistency and order. We first improve existing results on the connection between these properties. We then examine the learning theoretic measures of the class of GSRs. Furthermore, we show that if the class of anonymous and consistent voting rules is efficiently learnable, then so is the class of GSRs of a given degree of consistency.

1 Introduction

Consider a setting where an entity, called the *designer*, has some voting rule in mind. Given any voting profile, the designer can compute the winner of that profile (perhaps with considerable computation time). Our goal is to find a voting rule that is "similar" to the designer's rule, through observing his choice of winners on randomly selected sample profiles. In this work, we seek to answer *whether this goal is achievable?* and *what are the requirements, in terms of computational and sample complexities, for achieving this goal?*

To better grasp the motivation of this work consider the following situation. The designer has a voting rule in mind that reflects societal or personal values. However, he does not know how to formalized this rule - in his perspective, this rule is his gut feeling. Or, the designer is able to formalized the rule, but his formalization results in a long list of cases and exceptions. So, it would take a considerable effort to run this election repeatedly in large scale, or even to communicate the rule (perhaps with the purpose of allowing a third-party to run the election). Hence, there is a need to find a "concise" representation of the designer's rule, not by looking at his long list of cases and exceptions, but by observing the outcome of his voting rule on sample profiles.

The choice of the class of voting rules, from which our predictions are made, is essential to this problem. There are many classes of common voting rules, each attempting to aggregate preferences as to make the public happy in some way. These rules have different properties and result in very different outcomes. So, *which class of voting rules should we use?* In this paper, we use a recently introduced class of voting rules called *Generalized Scoring Rules* (GSRs) (Xia and Conitzer, 2008). This class includes most common voting rules and is rich enough to capture the complexities of the designer's voting rule, yet, it is concisely representable.

Our Model: We assume that there is an underlying distribution on profiles and their respective winners (determined by the designer's rule) that is fixed but unknown. We assume that we can ask the designer to determine the winners of a number of profiles chosen i.i.d from the underlying distribution. Our objective is to learn a GSR that with high probability has a low expected error. To do so, we will build on existing definitions and results from Learning Theory and specifically PAC-learning. We introduce this model in detail in later sections.

Our contribution is in three parts:

- We examine GSRs and two of their important properties, the degree of consistency (t) and order (d). We study some common voting rules and show how they can be represented as GSRs.
- We prove a connection between the order and degree of consistency of GSRs. Our result improves previous bounds presented by Xia and Conitzer (2009).
- We use learning theoretic measures to study the learnability of GSRs in terms of the number of candidates and the degree of consistency. We prove that if the class of consistent GSRs (restriction of GSRs) is efficiently PAC-Learnable, then so is the class of all GSRs of a given degree of consistency. Moreover, we derive the sample complexity needed to learn GSRs.

This paper is organized as follows. In Section 2, we discuss some previous work related to our paper. In Section 3, we introduce definitions and notions that will be used in the remainder of the paper. In Section 4, we prove upper bounds on the values of the properties of the GSRs. In Section 5, we introduce the learning model that is used in this paper and prove learnability results.

2 Related Work

This paper can be seen as an extension to the work of Procaccia et al. (2009, 2007), which examined the learnability of Scoring Rules and Voting Trees. They showed that Scor-

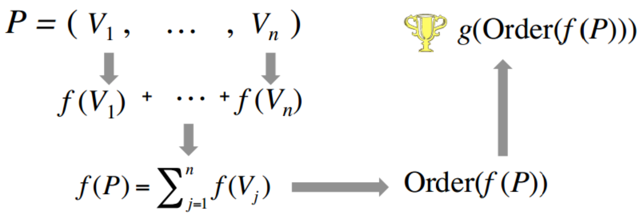


Figure 1: An illustration of the definition of GSRs presented by Xia (2013)

ing Rules are efficiently PAC-learnable, but, Voting Trees require exponentially many samples in the general case. In this work, we study the learnabiliy of a much larger class of voting rules, which includes the previously studied rules. Moreover, we study the learnability of GSRs based on their degree of consistency (which was not done in previous work), so negative results pertaining to Voting Trees do not extend to our work.

Although there is limited work on using learning techniques in voting theory, learning has been used in other economic settings. Kalai (2003) explored the learnability of rationalized choice functions. Lahaie and Parkes (2004) used learning algorithms as a means to preference elicitation in combinatorial auctions. In another paper, Beigman and Vohra (2006) used PAC-learnability to predict the demand at unobserved prices after seeing demand and price samples.

3 Preliminaries

Basics of Voting

Let $C = \{1, \ldots, m\}$ be a set of alternatives. Let $L(C) = \{l_1, \ldots, l_{m!}\}$ be the set of all linear orders over C. A *vote*, V, is a member of $L(C)$. An *n-voter profile* includes n votes and is represented by $P = (V_1, \ldots, V_n)$ where for all $i \in [n]$, $V_i \in L(C)$. Set of all profiles on C are shown by $\mathcal{P}(C)$.

A *voting rule*, $r : \mathcal{P}(C) \to C$, is a function that takes as an input any profile and returns a winner for that profile. Rule r is called *anonymous* if it does not consider the identity of the voter. An anonymous voting rule is *homogenous* if for any profile P and $k > 0$, $r(P) = r(kP)$, where kP is a profile where each vote of P is repeated k times. Rule r is *locally consistent* on a set of profiles $\mathcal{P}$, if for any two profiles $P, P' \in \mathcal{P}$, if $r(P) = r(P')$ then $P \cup P' \in \mathcal{P}$ and $r(P \cup P') = r(P) = r(P')$. A rule r is called *t-consistent*, if there is a partition of $\mathcal{P}(C)$ to t classes such that r is locally consistent on each class. A voting rule satisfies *Finite Local Consistency (FLC)* if it is t-consistent for some natural number t. The *degree of consistency* of a rule is the smallest t for which it is t-consistent.

Generalized Scoring Rules

In this section, we first define a broad category of voting rules that was first introduced byXia and Conitzer (2008).

We also mention important results pertaining to these rules that are essential to the development of our work.

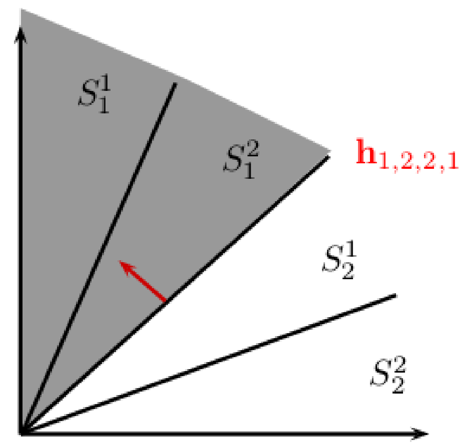


Figure 2: A decomposition of the $\mathbb{Q}_{\geq 0}^{m!}$ to mt convex cones for $t = m = 2$. The shaded area corresponds to the $\mathbf{h}_{1,2,2,1}$ half-space.

Definition 1 (Generalized Scoring Rules). *Let $k \in \mathbb{N}$, $f : L(C) \to \mathbb{R}^k$, and $g : \mathbb{R}^k \to C$, where g depends only on the sorted order of the input. Then a* generalized scoring rule *or GSR for short, $GSR(f, g)$, takes as input a preference profile $P = (V_1, \ldots, V_n)$, and returns $g(SortedOrder(\sum_{i=1}^{n} f(V_i)))$. We say that $GSR(f, g)$ is of dimension k (See Figure 1).*

Define the *order* of a voting rule r to be the smallest integer d for which f and g exist, such that $GSR(f, g) = r$ and $GSR(f, g)$ has dimension d.

In this paper, we frequently use a geometrical representation of GSR from the work of Xia and Conitzer (2009). Next, we define this representation. Let r be a voting rule that satisfies anonymity and FLC, then r is anonymous and homogeneous. Hence, there is a one-to-one map between the set of all profiles and the set of points in $\mathbb{N}^{m!}$ as follows: Any profile $P = \sum_{x=1}^{m!} w_x l_x$, where w_x is the number of voters in profiles P who voted l_x. We use vector $\mathbf{p} = (w_1, \ldots, w_{m!})$ to represent this profile. Since, r is homogeneous, the domain of r can be transformed to $\mathbb{Q}_{\geq 0}^{m!}$. Xia and Conitzer (2009) showed that any GSR, r, represents a decomposition of $\mathbb{Q}_{\geq 0}^{m!}$ to convex cones S_i^j for $i \in [t], j \in [m]$, such that for any profile $\mathbf{p} \in S_i^j$, $r(\mathbf{p}) = j$. The converse also holds.

In this work, we introduce the use of half-spaces to better benefit from the previous representation of GSRs. Any convex cone is the intersection of a set of half-spaces. Let $\mathbf{h}_{i_1,j_1,i_2,j_2}$ be the half-space that separates $S_{i_1}^{j_1}$ from $S_{i_1}^{j_1}$ (See Figure 2). In other words, for any $\mathbf{p}_1 \in S_{i_1}^{j_1}$, $\langle \mathbf{h}_{i_1,j_1,i_2,j_2}, \mathbf{p}_1 \rangle \geq 0$ and for any $\mathbf{p}_2 \in S_{i_2}^{j_2}$, $\langle \mathbf{h}_{i_1,j_1,i_2,j_2}, \mathbf{p}_2 \rangle < 0$. With a slight abuse of notation, we use h_{i_1,j_1,i_2,j_2} to refer to the set of points such that $\langle \mathbf{h}_{i_1,j_1,i_2,j_2}, \mathbf{p}_1 \rangle \geq 0$. Then $S_i^j = \bigcap_{\substack{i' \neq i \text{ or} \\ j' \neq j}} h_{i,j,i',j'}$.

4 Properties of GSRs

In this section, we first derive bounds on the order and consistency of some well-known voting rules, including Voting Trees and Copeland. We also establish a relationship between the order and consistency of GSRs in general.

Bounds for Common Voting Rules

In this section, we prove upper bounds on the order and degree of consistency of some common voting rules. The summary of results is shown in the following table.[1]

Rule	Copeland	Maximin	Voting Tree
t	$3^{\frac{m(m-1)}{2}}$	$(m-1)^{(m-1)}$	$3^{\frac{m(m-1)}{2}}$
d	$\binom{m}{2}+1$	$m(m-1)$	$\binom{m}{2}+1$

For a vote $V \in L(C)$, let $c_i \succ_V c_j$, if c_i appears before c_j in V (c_i is preferred to c_j). For a profile $P = (V_1, \dots, V_n)$, let $b_P(c_i, c_j) = |\{k : c_i \succ_{V_k} c_j\}|$, i.e. number of voters who prefer c_i to c_j. We use $c_i \succ_P c_j$ to indicate that the majority of voters prefer c_i to c_j, i.e. $b_P(c_i, c_j) > \frac{|P|}{2}$.

Next we define some common voting rules:

Definition 2. *(Copeland) The c_i's* Copeland *score increases by 1 for every candidate c_j that $c_i \succ_P c_j$ and decreases by 1 for every candidate c_j that $c_j \succ_P c_i$. The Copeland voting rule returns the candidate with highest Copeland score.*

Definition 3. *(Maximin) Let candidate c_i's* Maximin *score be $\min_{c_j \neq c_i} b_P(c_i, c_j)$. The Maximin voting rule returns the candidate with the highest Maximin score.*

Definition 4. *(Voting Tree) Any* Voting Tree *is represented by a binary tree, whose nodes are marked by a comparison (c_i, c_j) and its leaves are marked with alternatives. One starts at the root and at each node takes the left child if $c_i \succ_P c_j$ and the right child otherwise. The Voting Rule returns the candidate at the leaf where this process terminates.*

Lemma 1. *If $c_i \succ_P c_j$ and $c_i \succ_{P'} c_j$, then $c_i \succ_{P \cup P'} c_j$. Similarly, if $c_i =_P c_j$ and $c_i =_{P'} c_j$, then $c_i =_{P \cup P'} c_j$.*

Proof. Simple arithmetic. The prof is eliminated in the interest of space. □

Theorem 1. *The degree of consistency of any voting tree is at most $3^{\frac{m(m-1)}{2}}$.*

Proof. Let $O(C)$ be the set of pairwise orderings over C. For any $i \neq j$, $c_i \succ_P c_j$, $c_i \prec_P c_j$, or $c_i =_P c_j$. Therefore, there are at most $|O| = 3^{\binom{m}{2}}$ pairwise orderings over the set of candidates. For each $o \in O$, let $\mathcal{P}_o = \{P :$ $\forall i, j,$ pairwise election of a_i and a_j in P is the same as $o\}$. For any $P, P' \in \mathcal{P}_o$, the pairwise election between any two candidates in profiles P and P' are both equal to o. Using lemma 1, the pairwise election between the candidates is the same for $P \cup P'$. Therefore $P \cup P' \in \mathcal{P}_o$. Moreover, the winner of each voting tree is uniquely determined by the pairwise election between the candidates (assuming that the same tie-breaking mechanism is used), so $r(P) = r(P') = r(P \cup \mathcal{P}')$ and r is locally consistent over $\mathcal{P}_o$. Therefore the degree of consistency of any voting tree is at most $|O| = 3^{\frac{m(m-1)}{2}}$. □

Theorem 2. *The order of the Copeland voting rule is at most $\binom{m}{2} + 1$.*

Proof. Let $d = \binom{m}{2} + 1$. Define f as follows. For any $i < j$, let $f_{i,j}(V) = 1$ if $c_i \succ_V c_j$, and 0 otherwise. Let $f_d = \frac{1}{2}$. Let $f(P) = \sum_i f(V_i)$. First, for each $i < j$, let g add 1 to the score of i, if $f(P)_{i,j} > f(P)_d$, 1 to the score of j if $f(P)_{i,j} < f(P)_d$, and 0.5 to the score of i and j otherwise. Note that $f(P)_{i,j} > f(P)_d$ if and only if $f(P)_{i,j} > \frac{n}{2}$, which holds when $c_i \succ_P c_j$. Therefore, this score represents the Copeland score of each candidate. g then selects the candidate with the highest score as the winner. $GSR(f, g)$ implements the Copeland rule and is of order d. Hence, the order of Copeland is at most $d = \binom{m}{2} + 1$. □

Theorem 3. *The order of any voting tree is at most $\binom{m}{2} + 1$.*

Proof sketch: Build f similar to the construction of Theorem 2. Let g run the tree tournament such that for each $i < j$, c_i beats c_j if and only if $f(P)_{i,j} > f(P)_d$. □

Theorem 4. *The order the Maximin voting rule is at most $m(m-1)$.*

Proof. For any ordered pair (i, j), let $f_{(i,j)}(V) = 1$ if $c_i \succ_V c_j$, and 0 otherwise. Let $f(P) = \sum_i f(V_i)$. Then $f(P)_{(i,j)} = b_P(c_i, c_j)$ is the number voters who prefer c_i to c_j. Let $g = \arg\max_{c_i} \min_{j \neq i} f(P)_{(i,j)}$. $GSR(f, g)$ implements the Maximin rule and is of order $m(m-1)$. The order of the Maximin rule is at most $m(m-1)$. □

Relation between Order and Consistency

In this section, we will derive an upper bound on the order of any generalized scoring rule. Our result, proved in Theorem 5, improves on the existing bound of Xia and Conitzer (2009) of $d \leq \binom{m}{2}\binom{t}{2}m! + 1$.

Theorem 5. *Any generalized scoring rule on m alternatives that is t-consistent has order*

$$d \leq \binom{mt}{2} + 1$$

Proof. We need to prove that for any t-consistent GSR, r, there are functions $f : L(C) \to \mathbb{R}^d$ and $g : \mathbb{R}^d \to C$, such that $r = GSR(f, g)$ and $d \leq \binom{mt}{2} + 1$.

Let $P = \{V_1, \dots, V_n\}$ be any profile and $\mathbf{p} \in \mathbb{Q}_{\geq 0}^{m!}$ be its corresponding vector form. For each $i \in [n]$, let $\mathbf{v_i}$ be V_i's corresponding vector form, if it were to be a profile of a single vote V_i. Note that $\sum_i \mathbf{v_i} = \mathbf{p}$. Define function $f(V) = (0, \mathbf{h}_{i_1,j_1,i_2,j_2}(\mathbf{v}) : i_1, i_2 \in [t], j_1, j_2 \in [m])$. Let $f(P) = \sum_i f(V_i)$. Using the definition of half-spaces and

[1] Xia and Conitzer (2008) mention the upper bounds on the degree of consistency of Maximin and Copeland.

the fact that inner product is linear the (i_1, j_1, i_2, j_2) coordinate of $f(P)$ is:

$$\begin{aligned} f(P)_{i_1,j_1,i_2,j_2} &= \sum_{i=1}^{n} \langle \mathbf{h}_{i_1,j_1,i_2,j_2}, \mathbf{v_i} \rangle \\ &= \langle \mathbf{h}_{i_1,j_1,i_2,j_2}, \sum_{i=1}^{n} \mathbf{v_i} \rangle \\ &= \langle \mathbf{h}_{i_1,j_1,i_2,j_2}, \mathbf{p} \rangle \end{aligned}$$

So, $f(P)$ is a vector whose (i_1, j_1, i_2, j_2) coordinate is non-negative if and only if $\mathbf{p}$ belongs to the half-space separating $S_{i_1}^{j_1}$ from $S_{i_2}^{j_2}$. Therefore, $\mathbf{p} \in S_i^j$ if and only if $f(P)_{i,j,i',j'} \geq 0$ for all $i' \neq i$ and $j \neq j'$.

Now let g be a function that takes $f(P)$ and for any i_1, j_1, i_2, j_2 uses the $\mathrm{sort}(f_0, f(P)_{i_1,j_1,i_2,j_2})$ to compute whether the (i_1, j_1, i_2, j_2) coordinate is non-negative. Then g returns j for which there exists $i \in [t]$ such that $f(P)_{i,j,i',j'}$ is non-negative. This is well-defined, because the convex cones are a partition of the space, so any point $\mathbf{p}$ belongs to exactly one of them (let ties be broken in an arbitrary but fixed method, for example in favor of smaller j). Hence, $r = GSR(f, g)$.

Note that in the above construction we use $mt(mt-1)$ half-spaces. However, $\mathbf{h}_{i,j,i',j'}$ is the complement half-space of $\mathbf{h}_{i',j',i,j}$. So it suffices to use $\binom{mt}{2}$ half-spaces. Hence, r is a GSR with order at most $\binom{mt}{2} + 1$. □

5 Learning GSR

In this section, we will state important definitions and results from Learning Theory and we prove bounds on the learnability of GSRs.

Background on Learning Theory

In the PAC model, given a domain set X, a set of possible labels Y, and an unknown but fixed distribution D over $X \times Y$, the learner's objective is to find a labeling function (a.k.a. hypothesis) $f : X \to Y$ from a class of hypotheses $\mathcal{F}$, such that $err_D(f) = \Pr_{(x,y)\sim D}[f(x) \neq y]$ is small. To this end, the learner observes a set of samples $\{(x_1, y_1), \ldots, (x_s, y_s)\}$ that is drawn i.i.d. from distribution D.

A class of hypothesis $\mathcal{F}$ is called PAC-learnable, if for every ϵ and δ, there exists $s_{\epsilon,\delta}$, such that there is a learning algorithm that takes any sample set of size at least $s_{\epsilon,\delta}$, runs in $poly(\frac{1}{\epsilon}, \frac{1}{\delta})$, and returns a hypothesis f^* that satisfies $\Pr[err_D(f^*) > \epsilon] \leq \delta$. In other words, it is possible to see enough samples and predicts a hypothesis that is *Probably Approximately Correct*. A class is efficiently PAC-learnable, if the learning algorithm is also polynomial in its input.

Since the learner can only observe the set of samples and is otherwise blind to the underlying distribution, he can choose the hypothesis that has the least error on the sample set, hoping that this hypothesis will have a small error on the distribution D. This learning paradigm is called *Empirical Risk Minimization*, or ERM for short. Under specific conditions, it is known that ERM is the algorithm that can produce a *Probably Approximately Correct* hypothesis and be the evidence that a class of hypotheses is PAC-learnable (Natarajan, 1991). First, we introduce important notations for the statement of this result.

An important measure of the richness of a hypothesis class is the number of different labellings it can produce on a set of points. The next definition captures this measure.

Definition 5 (Growth Function)**.** *For any hypothesis class $\mathcal{F}$ of $\{0, \ldots, m\}$-valued functions and any natural number s, the growth function is the maximum number of different values that any s elements can take. In other words,*

$$\pi_{\mathcal{F}}(s) = \max_{S, |S|=s} |\{f(S) : \forall f \in \mathcal{F}\}|$$

Definition 6 (Natarajan Shattering)**.** *Let $\mathcal{F}$ be a class of hypotheses from X to Y. We say that $\mathcal{F}$* shatters *$S \subseteq X$ if there exists $f_1, f_2 \in \mathcal{F}$, such that for all $x \in S$, $f_1(x) \neq f_2(x)$, and for any $S' \subseteq S$ there exists $f \in \mathcal{F}$ such that $f(x) = f_1(s)$ for $x \in S'$ and $f(x) = f_2(x)$ otherwise.*

Definition 7 (Natarajan Dimension)**.** *The Natarajan dimension of a hypothesis class $\mathcal{F}$, denoted by* $\text{N-Dim}(\mathcal{F})$*, is the greatest integer such that there exists a set of that cardinality that is shattered by $\mathcal{F}$.*

Based on these definitions, it is easy to see that if $\pi_{\mathcal{F}}(s) < 2^s$, then $\text{N-Dim}(\mathcal{F})$ is less than s. In other words, $\text{N-Dim}(\mathcal{F}) \leq min\{s : \pi_{\mathcal{F}}(s) < 2^s\}$. Another important connection between the N-Dim and the growth function is captured in the well-known Sauer's lemma. Next, we introduce a generalization of this lemma for the $\{0 \ldots, m\}$-valued functions.

Lemma 2 (Generalized Sauer's Lemma)**.** *(Haussler and Long, 1995) For any $\{0, \ldots, m\}$-valued hypothesis class $\mathcal{F}$, let $d = \text{N-Dim}(\mathcal{F})$, then*

$$\pi_{\mathcal{F}}(s) \leq \sum_{i}^{d} \binom{s}{i} \binom{m+1}{2}^i \leq \left[\frac{es}{d} \binom{m+1}{2} \right]^d$$

An important aspect of Natarajan Dimension (and other generalizations of VC- dimension) is its effect on learnability of a class of hypothesis. The next theorem addresses the sample complexity requirement of learning a class of hypotheses based on its Natarjan Dimension.

Theorem 6. *(Ben-David et al., 1995) Let $\mathcal{F}$ be a class of $\{0, \ldots, m\}$-valued hypotheses. Then the PAC-learning sample complexity of $\mathcal{F}$ is*

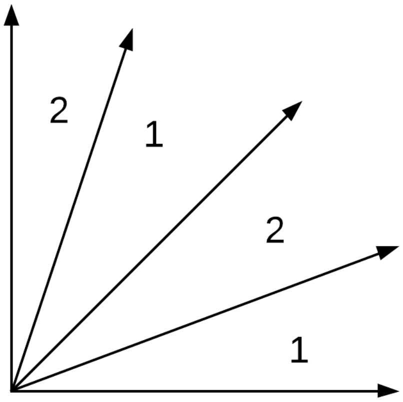
$$s_{\epsilon,\delta} \in O\left(\frac{\text{N-Dim}(\mathcal{F}) \cdot \log m \cdot \log \frac{1}{\epsilon} + \log \frac{1}{\delta}}{\epsilon} \right)$$

Moreover, this bound is achieved by any ERM.

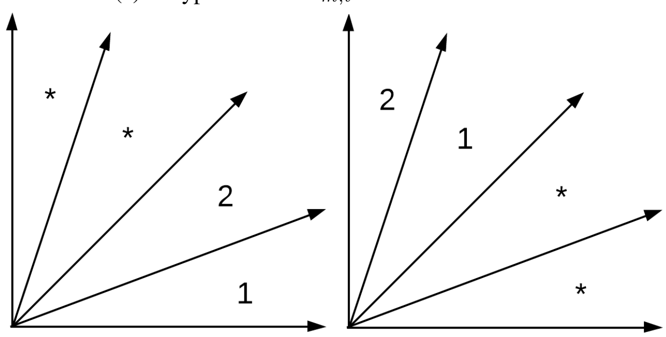
Learnability of GSRs

In this subsection, we prove upper bounds on the Natarajan Dimension and sample complexity of the class of GSRs of a given degree of consistency. We show that if the class of 1-consistent GSRs is efficiently PAC-learnable, so is the class of t-consistent GSRs.

Let $\mathcal{F}_{m,t}$ be a class of hypotheses, where each hypothesis is a t-consistent GSR for m alternatives. In other words,



(a) A hypothesis in $\mathcal{F}_{m,t}$ for $m = t = 2$.



(b) Two hypotheses in $\mathcal{G}'$ that will add up to a hypothesis in $\mathcal{F}$.

Figure 3: Illustrating how any hypothesis in $\mathcal{F}_{m,t}$ is equal to the union ($\bigsqcup$) of t hypothesis in $\mathcal{G}'$.

each hypothesis is a decomposition of $\mathbb{Q}_{\geq 0}^{m!}$ to mt convex cones with at most t cones marked with a given label in $[m]$. Let $\mathcal{G}_{a,b}$ be a class of hypothesis, where each hypothesis is a decomposition of $\mathbb{Q}_{\geq 0}^{a}$ to b convex cones, marking each cone with a unique value in $[b]$. The next lemmas show the connection between hypothesis class $\mathcal{F}$ and $\mathcal{G}$.

Lemma 3. *For any m and t, $\pi_{\mathcal{F}_{m,t}}(s) \leq [\pi_{\mathcal{G}_{m!,mt}}(s)]^t$.*

Proof. Define $\mathcal{G}'_{m!,mt}$ as a class of hypotheses, where each hypothesis is a decomposition of $\mathbb{Q}_{\geq 0}^{a}$ to mt convex cones, marking m cones with unique labels from $[m]$ and the remaining cones with label $*$. Clearly, $\pi_{\mathcal{G}'_{m!,mt}}(s) \leq \pi_{\mathcal{G}_{m!,mt}}(s)$.

Define operation $\bigsqcup$ between hypotheses as follows: For $u = \bigsqcup_{i\in[k]} g'_i$, u is defined such that $u(\mathbf{p}) = c$ if there is i, such that $g'_i(\mathbf{p}) = c$ and for all $j \neq i$, $g'_i(\mathbf{p}) \in \{c, *\}$. Otherwise, $u(\mathbf{p}) = *$. Define $\mathcal{U}^t_{m!,mt}$ to be the class of hypotheses formed from the union of any t hypotheses of $\mathcal{G}'$. More concretely, $\mathcal{U}^t_{m!,mt} = \{u \equiv \bigsqcup_{i\in[k]} g'_i : \forall i, g'_i \in \mathcal{G}'_{m!,mt}\}$. Since each member of $\mathcal{U}^t_{m!,mt}$ is deterministically derived from t members of $\mathcal{G}'_{m!,mt}$, we have that $\pi_{\mathcal{U}^t_{m!,mt}}(s) \leq [\pi_{\mathcal{G}'_{m!,mt}}(s)]^t$.

For any $f \in \mathcal{F}$, with convex cones S^j_i for $i \in [t]$ and $j \in [m]$, let g'_k be the convex decomposition of the space to mt convex cones such that S^i_k for all $i \in [m]$ maps to i and all other convex cones map to $*$ (as in figure 3). Then, for all k, $g'_k \in \mathcal{G}'_{m!,mt}$ and $f \equiv \bigsqcup_{i\in[t]} g'_i$. So, $\pi_{\mathcal{F}_{m,t}}(s) \leq \pi_{\mathcal{U}^t_{m!,mt}}(s)$.

Using the above three inequalities, we have, $\pi_{\mathcal{F}_{m,t}}(s) \leq [\pi_{\mathcal{G}_{m!,mt}}(s)]^t$. □

Lemma 4. *For any m and t, $\pi_{\mathcal{G}_{m!,mt}}(s) \leq \pi_{\mathcal{F}_{mt,1}}(s)$.*

Proof. Let $g \in \mathcal{G}_{m!,mt}$ be a decomposition of $\mathbb{Q}_{\geq 0}^{m!}$ to mt convex cones, $S_1, \ldots, S_{mt}$. It suffices to show that there is a decomposition (or its corresponding half-spaces) of $\mathbb{Q}_{\geq 0}^{(mt)!}$ to mt cones that create the same labelling as g on a set of s points.

Let $\mathbf{p}_1, \ldots, \mathbf{p}_s$ be the set of s vectors in $\mathbb{Q}_{\geq 0}^{m!}$ (corresponding to profiles) that accept the most number of labellings among any s points using $\mathcal{G}_{m!,mt}$. Let $\mathbf{h}_{i,j}$ be the half-space in $\mathbb{Q}_{\geq 0}^{m!}$ that separates S_i from S_j. Then for all i, $S_i = \bigcap_{j\neq i} h_{i,j}$. Define $\overline{\mathbf{p}_i} = (\mathbf{p}_i, 0, \ldots, 0)$ and $\overline{\mathbf{h}_{i,j}} = (\mathbf{h}_{i,j}, 0, \ldots, 0)$ to be the *extended* vectors in $\mathbb{Q}_{\geq 0}^{(mt)!}$ whose first $m!$ components are the same as $\mathbf{p}_i$ and $\mathbf{h}_{i,j}$, respectively, and the rest are zeros. For every point $\overline{\mathbf{p}_i}$ and half-spaces $\overline{\mathbf{h}_{i,j}}$,

$$\langle \overline{\mathbf{h}_{j,k}}, \overline{\mathbf{p}_i} \rangle = \langle \mathbf{h}_{j,k}, \mathbf{p}_i \rangle$$

Therefore, the labelling on $\mathbf{p}_i$ using half-space $\mathbf{h}_{j,k}$ is the same as labelling of $\overline{\mathbf{p}_i}$ using half-space $\overline{\mathbf{h}_{j,k}}$. Moreover, these extended half-spaces also define a decomposition of the space to mt convex cones (the proof is omitted in the interest of space). So, the convex-cone decomposition corresponding to the extended half-spaces is a member of $\mathcal{F}_{mt,1}$. Therefore, $\pi_{\mathcal{G}_{m!,mt}}(s) \leq \pi_{\mathcal{F}_{mt,1}}(s)$. □

Theorem 7. *For any m and t, let $d = \text{N-Dim}(\mathcal{F}_{mt,1})$, then*

$$\pi_{\mathcal{F}_{m,t}}(s) \leq \left[\frac{es}{d}\binom{mt+1}{2}\right]^{dt}$$

Proof. Using Lemmas 3 and 4 and Sauer's Lemma (Lemma 2) we have that

$$\pi_{\mathcal{F}_{m,t}}(s) \leq [\pi_{\mathcal{F}_{mt,1}}(s)]^t \leq \left[\frac{es}{d}\binom{mt+1}{2}\right]^{dt}$$

□

Theorem 8. *For any m and t, let $d = \text{N-Dim}(\mathcal{F}_{mt,1})$, then*

$$\text{N-Dim}(\mathcal{F}_{m,t}) \leq mdt^2 \log dt + 2dt \log mt$$

Proof. We know that if a hypothesis class shatters a set of size s then its growth function is more than 2^s. Let $s = mdt^2 \log dt + 2dt \log mt$. If we show that $\mathcal{F}_{m,t}$ can not shatter a set of size s then its Natarajan dimension is less than s. For $mt \geq 2$

$$\begin{aligned}
s &= mdt^2 \log dt + 2dt \log mt \\
&> dt \log 3mdt^2 + 2dt \log mt \\
&> dt[\log(mdt^2 \log dt + 2dt \log mt)] + 2dt \log mt \\
&> dt \log\left(s\binom{mt+1}{2}\right) \\
&> \log \pi_{\mathcal{F}_{m,t}}(s)
\end{aligned}$$

This results in $2^s > \pi_{\mathcal{F}_{m,t}}(s)$, so $\text{N-Dim}(\mathcal{F}_{m,t}) \leq s$. $\square$

Theorem 9. *Let* $\text{N-Dim}(\mathcal{F}_{mt,1}) = d$*, then the sample complexity of* $\mathcal{F}_{m,t}$ *is*

$$s_{\epsilon,\delta} \in O\left(\frac{mdt^2 \log dt \cdot \log m \cdot \log \frac{1}{\epsilon} + \log \frac{1}{\delta}}{\epsilon}\right)$$

Moreover, this bound is achieved by any ERM.

Proof. The proof is done by direct application of theorem 6 and theorem 8. $\square$

Theorem 9 shows that any potential exponential lower bound (inefficiency) for learning $\mathcal{F}_{m,t}$ would be caused by high Natarajan Dimension of $\mathcal{F}_{mt,1}$. So, if $\text{N-Dim}(\mathcal{F}_{mt,1}) \in poly(m,t)$, so is $\text{N-Dim}(\mathcal{F}_{m,t})$. In other words, the degree of consistency of a class of rules is not by itself a source of difficulty in learning.

Corollary 1. *If* $\mathcal{F}_{mt,1}$ *is efficiently PAC-learnable, so is* $\mathcal{F}_{m,t}$*.*

It is interesting to note that if we restrict our attention to GSRs that, in addition to satisfying anonymity and FLC, also satisfy *neutrality*, then $\text{N-Dim}(\mathcal{F}_{mt,1}) \leq mt$ (Procaccia et al., 2009). So, it may be possible to show that the class of all t-consistent GSRs that are neutral is PAC-learnable in polynomial time. An interesting direction for future work is to examine $\text{N-Dim}(\mathcal{F}_{mt,1})$ in the general setting and define restrictions (possibly weaker than neutrality) that can cause $\text{N-Dim}(\mathcal{F}_{mt,1})$ to be polynomial in m and t.

6 Conclusions

In this work, we examined the learnability of GSRs We studied two of their important properties, degree of consistency and order. We examined upper bounds on the order and degree of consistency of three common voting rules: Copeland, Maximin, and Voting Trees. We also showed a connection between these two values in the general case. Our result, showing that $d \leq \binom{mt}{2} + 1$, improved on the previous bound of Xia and Conitzer (2009).

We also used the Natarajan Dimension to examine how order and degree of consistency of a class of GSRs is related to its ability to express different winners for different profiles. We established upper bounds on the sample complexity and Natarajan Dimension of GSRs which is polynomial in the degree of consistency, number of candidates, and the Natarajan Dimension of a class of GSRs that are 1-consistent. In other words, if the set of anonymous and consistent voting rules is efficiently PAC-learnable, so is the class of GSRs of a given consistency level.

One direction for future work is to examine the Natarajan Dimension of the class of anonymous and consistent voting rules (or $\mathcal{F}_{m,1}$). As we mentioned previously, restricting this class further to voting rules that are neutral results in polynomial Natarajan Dimension. One can investigate other weaker assumptions that can ensure that $\text{N-Dim}(\mathcal{F}_{m,1}) \in poly(m)$.

References

Beigman, E., and Vohra, R. 2006. Learning from revealed preference. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, 36–42. ACM.

Ben-David, S.; Cesabianchi, N.; Haussler, D.; and Long, P. M. 1995. Characterizations of learnability for classes of $\{0, ..., n\}$-valued functions. *Journal of Computer and System Sciences* 50(1):74–86.

Haussler, D., and Long, P. M. 1995. A generalization of sauer's lemma. *Journal of Combinatorial Theory, Series A* 71(2):219 – 240.

Kalai, G. 2003. Learnability and rationality of choice. *Journal of Economic Theory* 113(1):104–117.

Lahaie, S. M., and Parkes, D. C. 2004. Applying learning algorithms to preference elicitation. In *Proceedings of the 5th ACM conference on Electronic commerce*, 180–188. ACM.

Natarajan, B. K. 1991. *Machine Learning: A Theoretical Approach*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Procaccia, A. D.; Zohar, A.; Peleg, Y.; and Rosenschein, J. S. 2007. Learning voting trees. In *In Proceedings of the 22nd Conference on Artificial Intelligence (AAAI*, 110–115.

Procaccia, A. D.; Zohar, A.; Peleg, Y.; and Rosenschein, J. S. 2009. The learnability of voting rules. *Artificial Intelligence* 173(12):1133–1149.

Xia, L., and Conitzer, V. 2008. Generalized scoring rules and the frequency of coalitional manipulability. In *Proceedings of the 9th ACM conference on Electronic commerce*, 109–118. ACM.

Xia, L., and Conitzer, V. 2009. Finite local consistency characterizes generalized scoring rules. In *IJCAI*, 336–341.

Xia, L. 2013. Generalized scoring rules: A framework that reconciles borda and condorcet. *SIGecom Exch.* 12(1):42–48.