Refutably Probably Approximately Correct Learning

Matsumoto, Satoshi
Research Institute of Fundamental Information Science Kyushu University

Shinohara, Ayumi
Research Institute of Fundamental Information Science Kyushu University

出版情報：RIFIS Technical Report. 87, 1994-04-22. Research Institute of Fundamental Information Science, Kyushu University
バージョン：
権利関係：

RIFIS-TR-CS-87

RIFIS Technical Report

Refutably Probably Approximately Correct Learning

Satoshi Matsumoto
Ayumi Shinohara

April 22, 1994

Research Institute of Fundamental Information Science
Kyushu University 33
Fukuoka 812, Japan
E-mail: matumoto@rifis.kyushu-u.ac.jp Phone: 092-641-1101 ex. 4459

Refutably Probably Approximately Correct Learning

Satoshi Matsumoto Ayumi Shinohara

Research Institute of Fundamental Information Science
Kyushu University 33, Fukuoka 812, Japan
e-mails:{matumoto, ayumi}@rifis.kyushu-u.ac.jp

We propose a notion of the *refutably PAC learnability*, which formalizes the refutability of hypothesis spaces in the PAC learning model. Intuitively, the refutable PAC learnability of a concept class $\mathcal{F}$ requires that the learning algorithm should refute $\mathcal{F}$ with high probability if a target concept can not be approximated by any concept in $\mathcal{F}$ with respect to the underlying probability distribution. We give a general upper bound of $O((1/\varepsilon + 1/\varepsilon')\log(|\mathcal{F}_n|/\delta))$ on the number of examples required for refutably PAC learning of $\mathcal{F}$. Here, ε and δ are the standard accuracy and confidence parameters, and ε' is the refutation accuracy. We also define the *strongly refutably PAC learnability* by introducing the refutation threshold. We prove a general upper bound of $O((1/\varepsilon^2 + 1/\varepsilon'^2)\log(|\mathcal{F}_n|/\delta))$ for strongly refutably PAC learning of $\mathcal{F}$. These upper bounds reveal that both the refutably PAC learnability and the strongly refutably PAC learnability are equivalent to the standard PAC learnability within the polynomial size restriction.

1 Introduction

In the standard PAC learning model due to Valiant [Val84] and most of its variants [BEHW89, Nat91], a target concept is assumed to be in a hypothesis space. In these models, a learning algorithm has only to find a hypothesis which is consistent with given examples. There have been some studies [Hau89, KSS92, KS91, Yam90] which weakened the assumption. However, their main subjects are to find the best approximation in the hypothesis space, and they have paid little attention to determine whether or not the hypothesis space is suitable to approximate the target concept.

As a practical application of PAC learning, we developed a machine learning system which finds a motif from given positive and negative strings [AKM+92, AMS+93, SSS+93], and made some experiments on amino acid sequences. In particular, we applied it to the following two problems. One is the transmembrane domain identification, which is rather an easy problem. The other is the protein secondary structure prediction, which is one of the most challenging problem in Molecular Biology. Our learning system succeeded in discovering some simple and accurate motifs for the transmembrane domain sequences in very short time. On the other hand, it has failed to find a rule to predict the secondary structures of proteins with high accuracy. Thus we have suspected that the representation is not suitable for the secondary structure prediction problem. Nevertheless, we did not have any criterion to terminate the learning algorithm even if there remains no possibility to find any good hypotheses. We need to refute all hypotheses in the current hypothesis space before trying some other space.

The refutability of the whole space of hypotheses was originally introduced by Mukouchi and Arikawa [MA93] in the framework of inductive inference. It is a essence of a logic of

machine discovery.

In this paper, we formalize the refutability of hypothesis spaces in the PAC learning model. We propose a notion of the *refutably PAC learning*. In this model, a learning algorithm tries to find a good approximation for a target concept with respect to the underlying probability distribution, in the same way as the standard PAC learning model. Additionally, the learning algorithm is required to refute the hypothesis space with high probability, if the target concept cannot be approximated by any concept in the hypothesis space. We also define the *strongly refutably PAC learning* by introducing the refutation threshold.

We prove general upper bounds of the number of examples which are required for both the refutably PAC learning and the strongly refutably PAC learning. These upper bounds implies that the polynomial-sample refutable PAC learnability and strongly refutably PAC learnability are equivalent to the standard polynomial-sample PAC learnability within the polynomial size restriction.

2 Refutably PAC Learnability

Let $X = \Sigma^*$ be the set of all strings on a finite alphabet Σ. We call X a *learning domain*. X_n denotes the set of all strings of length n or less for $n \geq 1$. A *concept* f is a subset of X. A *concept class* is a nonempty set $\mathcal{F} \subseteq 2^X$. For a concept $f \in \mathcal{F}$ and an integer $n \geq 1$, we denote the *n-th subclass of* $\mathcal{F}$ by $\mathcal{F}_n = \{f \cap X_n \mid f \in \mathcal{F}\}$. Let I_f be the indicator function for f on X, that is, $\mathrm{I}_f(x) = 1$ if $x \in f$ and $\mathrm{I}_f(x) = 0$, otherwise. An *example* on $x \in X$ for a concept f is a pair $\langle x, \mathrm{I}_f(x)\rangle$. If $\mathrm{I}_f(x) = 1$, $\langle x, \mathrm{I}_f(x)\rangle$ is a *positive* example; otherwise, it is a *negative* example.

Let $\mathcal{F}$ be a concept class on X. For any integer $n \geq 1$, we define the dimension of n-th subclass by $\dim \mathcal{F}_n = \log_2 |\mathcal{F}_n|$. We say that concept class $\mathcal{F}$ is the *polynomial dimension* if there is a polynomial function $p(n)$ with $\dim \mathcal{F}_n \leq p(n)$ for any $n \geq 1$.

Let g be a concept class and f be a target concept. For a probability distribution P, we define $\mathbf{er}_{P,f}(g) = P(g \triangle f)$, where $f \triangle g$ denotes the symmetric difference $f \cup g - f \cap g$. We call $\mathbf{er}_{P,f}(g)$ the error of g for f with respect to P. We define $opt(P, \mathcal{F}) = \min_{g \in \mathcal{F}} \mathbf{er}_{P,f}(g)$. We remark that if the target concept f is in $\mathcal{F}$, then $opt(P, \mathcal{F}) = 0$ for any probability distribution P.

Now we define a notion of refutably PAC learnability. Intuitively, we expect the following algorithm $\mathcal{A}$ for a concept class $\mathcal{F}$. If $opt(P, \mathcal{F}) = 0$, then $\mathcal{A}$ finds *good* approximation $h \in \mathcal{F}$ for a target concept f. Otherwise, $\mathcal{A}$ refutes $\mathcal{F}$. However, if $opt(P, \mathcal{F})$ is very close to 0, it is hard for the learning algorithm to determine $opt(P, \mathcal{F}) = 0$ or not. Thus we relax the requirement by introducing the *refutation accuracy* ε'. That is, $\mathcal{A}$ refutes $\mathcal{F}$ if $opt(P, \mathcal{F}) \geq \varepsilon'$. The formal definition is as follows.

Definition 1. Let $\mathcal{F}$ be a concept class on X. An algorithm $\mathcal{A}$ is a *refutably PAC learning algorithm* for $\mathcal{F}$ if

(a) $\mathcal{A}$ takes ε, ε', δ and n ($0 < \varepsilon, \varepsilon', \delta < 1$, $n \in \mathbf{N}^+$) as inputs.

(b) $\mathcal{A}$ may call EXAMPLE, which returns examples for some concept $f \subseteq X$. Note that f is called a *target concept*. The examples are chosen randomly according to an arbitrary and unknown probability distribution P on X_n. Note that *the concept f is not necessarily in the concept class $\mathcal{F}$.*

(c) $\mathcal{A}$ satisfies the following conditions for any concept $f \subseteq X$ and any probability distribution P on X_n:

(i) If $opt(P, \mathcal{F}) \geq \varepsilon'$, then $\mathcal{A}$ refutes the hypothesis class $\mathcal{F}$ with probability at least $1 - \delta$.

(ii) If $opt(P, \mathcal{F}) = 0$, then $\mathcal{A}$ outputs a *hypothesis* $h \in \mathcal{F}$ which satisfying $P(f \triangle h) < \varepsilon$ with probability at least $1 - \delta$.

We set up a complexity measure for learning algorithms to measure the number of examples required by the algorithm as a function of the various parameters.

Definition 2. Let $\mathcal{A}$ be a learning algorithm for a concept class $\mathcal{F}$. The *sample complexity* of $\mathcal{A}$ is the function $s : \mathbf{R} \times \mathbf{R} \times \mathbf{R} \times \mathbf{N} \to \mathbf{N}$ such that $s(\varepsilon, \varepsilon', \delta, n)$ is the maximum number of calls of EXAMPLE by $\mathcal{A}$, where the maximum is taken over all runs of $\mathcal{A}$ on inputs ε, ε', δ and n, with the target concept f ranging over all $f \subseteq X$ and the probability distribution P ranging over all distribution on X_n. If no finite maximum exists, $s(\varepsilon, \varepsilon', \delta, n) = \infty$

The sample complexity of algorithm $\mathcal{A}$ is the number of examples which is required by $\mathcal{A}$ as a function of the input parameters. If this function is bounded by a polynomial in $\frac{1}{\varepsilon}, \frac{1}{\varepsilon'}, \frac{1}{\delta}$ and n, we consider the learning task to be feasible.

Definition 3. A concept class $\mathcal{F}$ is said to be *polynomial-sample refutably learnable* if there exists a polynomial p and a refutably PAC learning algorithm $\mathcal{A}$ for $\mathcal{F}$ with sample complexity $p(\frac{1}{\varepsilon}, \frac{1}{\varepsilon'}, \frac{1}{\delta}, n)$.

Now we show an upper bound of the sample complexity for refutably PAC learnability.

Theorem 1. Let $\mathcal{F}$ be a concept class. Then there exists a refutably PAC learning algorithm for $\mathcal{F}$ with sample complexity

$$p\left(\frac{1}{\varepsilon}, \frac{1}{\varepsilon'}, \frac{1}{\delta}, n\right) = \left\lceil \left(\frac{1}{\varepsilon} + \frac{1}{\varepsilon'}\right) \log \frac{|\mathcal{F}_n|}{\delta} \right\rceil .$$

Proof. Algorithm $\mathcal{A}_1$ below is a refutably learning algorithm for $\mathcal{F}$.

Learning Algorithm $\mathcal{A}_1$
input: ε, ε', δ, n;
begin
  let $m = \left\lceil \left(\frac{1}{\varepsilon} + \frac{1}{\varepsilon'}\right) \log \frac{|\mathcal{F}_n|}{\delta} \right\rceil$;
  make m calls of EXAMPLE;
  let S be the set of examples seen;
  if there exists a concept $g \in \mathcal{F}$ that is consistent with S **then**
    begin
      pick a concept $h \in \mathcal{F}$ that is consistent with S;
      output h;
    end
  else
    refute the concept class $\mathcal{F}$;
end

We estimate the number of examples from which the algorithm $\mathcal{A}_1$ refutes the concept class $\mathcal{F}$ with probability at least $1-\delta$.

Suppose that $opt(P,\mathcal{F}) \geq \varepsilon'$. By the definition of $\mathcal{A}_1$, we may consider only a probability distribution P on X_n. Then, without loss of generality, we can assume that a concept class is the n-th subclass $\mathcal{F}_n$. If the algorithm $\mathcal{A}_1$ outputs some concept g, all examples produced by EXAMPLE is consistent with the concept g. By the supposition, $P(g \triangle f) \geq \varepsilon'$ for any concept $g \in \mathcal{F}_n$. Then, the probability that any call of EXAMPLE will produce an example consistent with g is at most $(1-\varepsilon')$. Hence, the probability that m calls of EXAMPLE will produce examples all consistent with g is at most $(1-\varepsilon')^m$. Now, there are at most $|\mathcal{F}_n|$ choices for g. We will make m sufficiently large to bound the probability $|\mathcal{F}_n|(1-\varepsilon')^m$ by δ.

Using the approximation $(1-\varepsilon')^m \leq e^{-m\varepsilon'}$,

$$|\mathcal{F}_n| e^{-m\varepsilon'} \leq \delta$$

Simplifying, we obtain the following inequation:

$$m \geq \frac{1}{\varepsilon'} \log \frac{|\mathcal{F}_n|}{\delta}.$$

If the condition(ii) in Definition 1 holds, then we may refer [Nat91]. □

Corollary 1. If a concept class $\mathcal{F}$ is of polynomial dimension, then $\mathcal{F}$ is polynomial-sample refutably learnable.

3 Strongly Refutably PAC learnability

In a practical setting, it is unusual that there exists a concept $g \in \mathcal{F}$ with $P(g \triangle f) = 0$. As long as the minimum error $opt(P,\mathcal{F})$ is small enough, it is desirable that a learning algorithm should produce some approximation instead of refuting $\mathcal{F}$. For this purpose, we introduce a new parameter η $(0 \leq \eta < 1)$, which is a *refutation threshold*. The formal definition is as follows.

Definition 4. Let $\mathcal{F}$ be a concept class on X. An algorithm $\mathcal{A}$ is a *strongly refutably PAC learning algorithm* for $\mathcal{F}$ if

(a) $\mathcal{A}$ takes ε, η, ε', δ and n $(0 < \varepsilon, \varepsilon', \delta < 1,\ 0 \leq \eta < 1,\ n \in \mathbf{N}^+)$ as inputs.

(b) $\mathcal{A}$ may call EXAMPLE, which returns examples for some concept $f \subseteq X$. Note that f is called a *target concept*. The examples are chosen randomly according to an arbitrary and unknown probability distribution P on X_n.

(c) $\mathcal{A}$ satisfies the following conditions for any concept $f \subseteq X$ and any probability distribution P on X_n:

- **(i)** If $opt(P,\mathcal{F}) \geq \eta + \varepsilon'$, then $\mathcal{A}$ refutes the hypothesis class $\mathcal{F}$ with probability at least $1-\delta$.
- **(ii)** If $opt(P,\mathcal{F}) \leq \eta$, then $\mathcal{A}$ outputs a *hypothesis* $h \in \mathcal{F}$ which satisfying $P(f \triangle h) < \eta + \varepsilon$ with probability at least $1-\delta$.

We define the sample complexity of strongly refutably PAC learning algorithm in the same way as Definition 2, with the refutation threshold η ranging over all $\eta \in [0,1)$.

The following lemma is important in Theorem 2.

Lemma 1. [AL88] If $0 \leq p \leq 1, 0 \leq r \leq 1$, and m is any positive integer then

$$\sum_{k=\lceil m(p+r) \rceil}^{m} \binom{m}{k} p^k (1-p)^{m-k} \leq e^{-2r^2 m},$$

and

$$\sum_{k=0}^{\lfloor m(p-r) \rfloor} \binom{m}{k} p^k (1-p)^{m-k} \leq e^{-2r^2 m}.$$

Theorem 2. Let $\mathcal{F}$ be a concept class. Then, there exists a strongly refutably PAC learning algorithm for $\mathcal{F}$ with sample complexity

$$p\left(\frac{1}{\varepsilon}, \frac{1}{\varepsilon'}, \frac{1}{\delta}, n\right) = \left\lceil \left(\frac{2}{\varepsilon^2} + \frac{2}{\varepsilon'^2}\right) \log \frac{2|\mathcal{F}_n|}{\delta} \right\rceil.$$

Proof. Algorithm $\mathcal{A}_2$ below is a strongly refutably PAC learning algorithm for $\mathcal{F}$.

Learning Algorithm $\mathcal{A}_2$
input: ε,ε',δ,η,n;
begin
let $m = \left\lceil \left(\frac{2}{\varepsilon^2} + \frac{2}{\varepsilon'^2}\right) \log \frac{2|\mathcal{F}_n|}{\delta} \right\rceil$;
let $\kappa = \min\{\varepsilon, \varepsilon'\}$;
make m calls of EXAMPLE;
let S be the sequence of examples seen;
if there exists a concept $g \in \mathcal{F}$ such that the number of examples in S that is inconsistent with g is at most $\left\lfloor m\left(\eta + \frac{1}{2}\kappa\right)\right\rfloor$ **then**
begin
pick a concept $h \in \mathcal{F}$ such that the number of examples in S that is inconsistent with h is at most $\left\lfloor m\left(\eta + \frac{1}{2}\kappa\right)\right\rfloor$;
output h;
end
else
refute the concept class $\mathcal{F}$;
end

We estimate the number of examples from which the algorithm $\mathcal{A}_2$ refutes the concept class $\mathcal{F}$ with probability at least $1 - \delta$.

Suppose that $opt(P, \mathcal{F}) \geq \eta + \varepsilon'$. By the definition of $\mathcal{A}_2$, we may consider only a probability distribution P on X_n. Then, without loss of generality, we can assume that a concept class is the n-th subclass $\mathcal{F}_n$. For any concept $g \in \mathcal{F}_n$, let $\nu_g = P(f \triangle g)$. By the condition (i) in Definition 4, we see that $\nu_g \geq \eta + \varepsilon'$. If the algorithm $\mathcal{A}_2$ outputs a concept g, the number of examples that is inconsistent with a target concept f is at most $\lfloor m(\eta + \frac{1}{2}\kappa) \rfloor$. Since the probability that the concept g is inconsistent with a target concept f is ν_g, the probability that the algorithm $\mathcal{A}_2$ outputs a concept g is at most

$$\sum_{i=0}^{\lfloor m\left(\eta + \frac{1}{2}\kappa\right) \rfloor} \binom{m}{i} {\nu_g}^i (1-\nu_g)^{m-i}.$$

Then

$$\begin{aligned}\sum_{i=0}^{\lfloor m(\eta+\frac{1}{2}\kappa)\rfloor}\binom{m}{i}{\nu_g}^i(1-\nu_g)^{m-i} &\leq \sum_{i=0}^{\lfloor m(\eta+\frac{1}{2}\varepsilon')\rfloor}\binom{m}{i}{\nu_g}^i(1-\nu_g)^{m-i}\\ &\leq \sum_{i=0}^{\lfloor m(\nu_g-\frac{1}{2}\varepsilon')\rfloor}\binom{m}{i}{\nu_g}^i(1-\nu_g)^{m-i}\\ &\leq e^{-2m\left(\frac{1}{2}\varepsilon'\right)^2}\end{aligned}$$

Now, there are at most $|\mathcal{F}_n|$ choices for g. We will make m sufficiently large to bound this probability $|\mathcal{F}_n|e^{-2m\left(\frac{1}{2}\varepsilon'\right)^2}$ by δ. Simplifying, we obtain the following inequation:

$$m \geq \frac{2}{\varepsilon'^2}\log\frac{|\mathcal{F}_n|}{\delta}.$$

If the condition (ii) in Definition 4 holds, we can show that if $m \geq \left(\frac{2}{\varepsilon'^2}+\frac{2}{\varepsilon^2}\right)\log\frac{2|\mathcal{F}_n|}{\delta}$ then the probability that the algorithm outputs a concept $g \in \mathcal{F}$ with $P(f\triangle g) < \eta+\varepsilon$ is greater than $1-\delta$ in the same way. □

Corollary 2. If a concept class $\mathcal{F}$ is of polynomial dimension, then $\mathcal{F}$ is polynomial-sample strongly refutably learnable.

4 Conclusion

We have formalized the refutability of hypothesis space in the PAC-learning model. We have also proved general upper bounds of the sample complexity both for refutably PAC learnability and for strongly refutable PAC learnability.

We will discuss time complexity in future works.

Acknowledgment

The authors would like to thank Prof. Setsuo Arikawa for helpful discussions.

References

[AKM+92] S. Arikawa, S. Kuhara, S. Miyano, A. Shinohara, and T. Shinohara. A learning algorithm for elementary formal systems and its experiments on identification of transmembrane domains. In *Proc. 25th Hawaii International Conference on System Sciences, Vol. I*, pp. 675–684, 1992.

[AL88] Dana Angluin and Philip Laird. Learning from noisy examples. *Machine Learning*, Vol. 2, No. 4, pp. 343–370, 1988.

[AMS+93] S. Arikawa, S. Miyano, A. Shinohara, S. Kuhara, Y. Mukouchi, , and T. Shinohara. A machine discovery from amino acid sequences by decision trees over regular patterns. *New Generation Computing*, Vol. 11, No. 3,4, pp. 361–375, 1993.

[BEHW89] A. Blumer, A. Ehrenfeucht, D. Haussler, and M.K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, Vol. 36, No. 4, pp. 929–965, 1989.

[Hau89] D. Haussler. Generalizing the PAC model: Sample size bounds from metric dimension-based uniform convergence results. In *Proceedings of the 2nd Annual Workshop on Computational Learning Theory*, pp. 385, 1989.

[KS91] Michael J. Kearns and Robert E. Schapire. Efficient distribution-free learning of probabilistic concepts. In *Proceedings of the 31st Annual Symposium on Foundations of Computer Science*, pp. 382–391, 1991.

[KSS92] Michael J. Kearns, Robert E. Schapire, and Linda M. Sellie. Toward efficient agnostic learning. In *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, pp. 341–352, 1992.

[MA93] Yasuhito Mukouchi and Setsuo Arikawa. Inductive inference machines that can refute hypothesis spaces. *4th International Workshop, ALT '93*, pp. 123–136, 1993.

[Nat91] Balas K. Natarajan. *MACHINE LEARNING A Theoretical Approach.* Morgan Kaufmann, 1991.

[SSS+93] S. Shimozono, A. Shinohara, T. Shinohara, S. Miyano, S. Kuhara, and S. Arikawa. Finding alphabet indexing for decision trees over regular patterns: an approach to bioinformatical knowledge acquisition. In *Proc. 26th Annual Hawaii International Conference on System Sciences, Vol. I*, pp. 763–773, 1993.

[Val84] L. G. Valiant. A theory of the learnable. *communications of the acm*, Vol. 27, No. 11, pp. 1134–1142, 1984.

[Yam90] Kenji Yamanishi. A learning criterion for stchastic rules. In *Proceedings of the 3rd Annual Workshop on Computational Learning Theory*, pp. 67–81, 1990.