Refutably Probably Approximately Correct Learning

Satoshi Matsumoto and Ayumi Shinohara

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

Abstract. We propose a notion of the refutably PAC learning, which formalizes the refutability of hypothesis spaces in the PAC learning model. Intuitively, the refutably PAC learning for 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 \( O((1/\varepsilon + 1/\varepsilon') \ln(|\mathcal{F}^{[n]}|/\delta)) \) on the number of examples required for refutably PAC learning of \( \mathcal{F} \). Here, \( \varepsilon \) and \( \delta \) are the standard accuracy and confidence parameters, and \( \varepsilon' \) is the refutation accuracy. Furthermore we also define the strongly refutably PAC learning by introducing the refutation threshold. We prove a general upper bound of \( O((1/\varepsilon^2 + 1/\varepsilon'^2) \ln(|\mathcal{F}^{[n]}|/\delta)) \) for strongly refutably PAC learning of \( \mathcal{F} \). These upper bounds reveal that both the refutably learnability and the strongly refutably learnability are equivalent to the standard learnability within the polynomial size restriction. We also define the polynomial-time refutably learnability of a concept class, and characterize it.

1 Introduction

In the standard PAC learning model due to Valiant [12] and most of its variants [4, 10], 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 [5, 8, 7, 13] 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 [2, 3, 11], 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 hypothesis space is not suitable for the secondary structure prediction problem. Nevertheless, we have no 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. If the learning algorithm can tell us that there are no target concept in the hypothesis space which explains a given sample, we may give a new other hypothesis space.

The refutability of the hypothesis space was originally introduced by Mukouchi and Arikawa [9] in the framework of inductive inference. It is an 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. Furthermore 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 reveal that the polynomial-sample refutably learnability and polynomial-sample strongly refutably learnability are equivalent to the standard polynomial-sample learnability.

We also formalize a notion of the polynomial-time strongly refutably learnability. In order to characterize it, we propose a random polynomial-time refutably hypothesis finder. We show that the polynomial-time strongly refutably learnability of a concept class \( \mathcal{F} \) in representation \( R \) is equivalent to the existence of a random polynomial-time hypothesis finder for \( \mathcal{F} \) in \( R \) under some conditions.

2 Preliminaries

This section briefly summarizes the PAC-learnability due to Valiant [12].

Let \( X = \Sigma^* \) be the set of all strings over a finite alphabet \( \Sigma \). We call an element of \( X \) a word and \( X \) a learning domain. \( X_n \) denotes the set of all strings of length at most \( n \) for \( n \geq 1 \). A concept \( f \) is a subset of \( X \). A concept class is a set \( \mathcal{F} \subseteq 2^X \). Let \( I_f \) be the indicator function for \( f \), that is, \( I_f(x) = 1 \) if \( x \in f \) and \( I_f(x) = 0 \), otherwise. An example for a concept \( f \) on \( x \in X \) is a pair \( \langle x, I_f(x) \rangle \). If \( I_f(x) = 1 \), \( \langle x, I_f(x) \rangle \) is a positive example; otherwise, it is a negative example. A sample of \( f \) is a nonempty sequence of examples for \( f \). The size of a sample is the number of examples it contains. The length of a sample is the total length of the strings in the sample. We say that a concept \( g \) is consistent with a sample \( \langle x_1, a_1 \rangle, \cdots, \langle x_m, a_m \rangle \) if \( I_g(x_i) = a_i \) for all \( 1 \leq i \leq m \). For a sample \( S \) and a concept \( g \in \mathcal{F} \), let \( d(g, S) \) denote the number of examples in \( S \) which are inconsistent with \( g \). For convenience, we assume hereafter that a polynomial