

Accepted Manuscript

Hierarchical design of fast Minimum Disagreement algorithms

Malte Darnstädt, Christoph Ries, Hans Ulrich Simon

PII: S0304-3975(17)30852-6
DOI: <https://doi.org/10.1016/j.tcs.2017.11.022>
Reference: TCS 11394

To appear in: *Theoretical Computer Science*



Please cite this article in press as: M. Darnstädt et al., Hierarchical design of fast Minimum Disagreement algorithms, *Theoret. Comput. Sci.* (2018), <https://doi.org/10.1016/j.tcs.2017.11.022>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Hierarchical Design of Fast Minimum Disagreement Algorithms

Malte Darnstädt^a, Christoph Ries^a, Hans Ulrich Simon^a

^a*Fakultät für Mathematik, Ruhr-Universität Bochum, D-44780 Bochum*

Abstract

We compose a toolbox for the design of Minimum Disagreement algorithms. This box contains general procedures which transform (without much loss of efficiency) algorithms that are successful for some d -dimensional (geometric) concept class \mathcal{C} into algorithms which are successful for a $(d + 1)$ -dimensional extension of \mathcal{C} . An iterative application of these transformations has the potential of starting with a base algorithm for a trivial problem and ending up at a smart algorithm for a non-trivial problem. In order to make this working, it is essential that the algorithms are not proper, i.e., they return a hypothesis that is not necessarily a member of \mathcal{C} . However, the “price” for using a super-class \mathcal{H} of \mathcal{C} is so low that the resulting time bound for achieving accuracy ε in the model of agnostic learning is significantly smaller than the time bounds achieved by the up-to-date best (proper) algorithms.

We evaluate the transformation technique for $d = 2$ on both artificial and real-life data sets and demonstrate that it provides a fast algorithm, which can successfully solve practical problems on large data sets.

1. Introduction

In this paper, we are concerned with the Minimum Disagreement problem (sometimes also called Maximum Weight problem) associated with a family \mathcal{C} of sets over some ground set \mathcal{X} : given a sequence $S = [(x_1, w_1), \dots, (x_n, w_n)] \in (\mathcal{X} \times \mathbb{R})^n$ of points in \mathcal{X} along with their weights, find a set $C \in \mathcal{C}$ whose total weight $W_S(C) := \sum_{i:x_i \in C} w_i$ is as large as possible. Note that $W_S(C)$ is maximized iff

$$E_S(C) := \sum_{i:w_i > 0, x_i \notin C} w_i - \sum_{i:w_i < 0, x_i \in C} w_i \quad (1)$$

is minimized. In statistical learning theory, $E_S(C)$ is called the *empirical error of C on S* , and this term plays a central role, especially in the model of agnostic learning [1].

Email addresses: malte.darnstaedt@rub.de (Malte Darnstädt), christoph.ries@rub.de (Christoph Ries), hans.simon@rub.de (Hans Ulrich Simon)

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات