Available online at www.sciencedirect.com
ScienceDirect
Information and Computation 205 (2007) 242–261
Information and Computation
www.elsevier.com/locate/ic

Algorithmic complexity bounds on future prediction errors☆

Alexey Chernov [a,d], Marcus Hutter [a,c,*], Jürgen Schmidhuber [a,b]

[a] *IDSIA, Galleria 2, CH-6928 Manno-Lugano, Switzerland*
[b] *TU Munich, Boltzmannstr. 3, 85748 Garching, München, Germany*
[c] *RSISE/ANU/NICTA, Canberra, ACT 0200, Australia*
[d] *LIF, CMI, 39 rue Joliot Curie, 13453 Marseille cedex 13, France*

Received 18 July 2005; revised 12 September 2006
Available online 19 December 2006

Abstract

We bound the future loss when predicting any (computably) stochastic sequence online. Solomonoff finitely bounded the total deviation of his universal predictor M from the true distribution μ by the algorithmic complexity of μ. Here we assume that we are at a time $t>1$ and have already observed $x=x_1\cdots x_t$. We bound the future prediction performance on $x_{t+1}x_{t+2}\cdots$ by a new variant of algorithmic complexity of μ given x, plus the complexity of the randomness deficiency of x. The new complexity is monotone in its condition in the sense that this complexity can only decrease if the condition is prolonged. We also briefly discuss potential generalizations to Bayesian model classes and to classification problems.

Keywords: Kolmogorov complexity; Posterior bounds; Online sequential prediction; Solomonoff prior; Monotone conditional complexity; Total error; Future loss; Randomness deficiency

☆ This work was supported by SNF Grants 200020-107590/1, 2100-67712, and 200020-107616. A shorter version appeared in the proceedings of the ALT'05 conference [1].
* Corresponding author.
E-mail addresses: alexey@idsia.ch (A. Chernov), marcus@idsia.ch (M. Hutter), juergen@idsia.ch (J. Schmidhuber).
URLs: http://www.idsia.ch/~alexey (A. Chernov), http://www.idsia.ch/~marcus (M. Hutter), http://www.idsia.ch/~juergen (J. Schmidhuber).

0890-5401/$ - see front matter
doi:10.1016/j.ic.2006.10.004

1. Introduction

We consider the problem of online=sequential predictions. We assume that the sequences $x=x_1x_2x_3\cdots$ are drawn from some "true" but unknown probability distribution μ. Bayesians proceed by considering a class $\mathcal{M}$ of models=hypotheses=distributions, sufficiently large such that $\mu\in\mathcal{M}$, and a prior over $\mathcal{M}$. Solomonoff considered the truly large class that contains all computable probability distributions [2]. He showed that his universal distribution M converges rapidly to μ [3], i.e., predicts well in any environment as long as it is computable or can be modeled by a computable probability distribution (all physical theories are of this sort). $M(x)$ is roughly $2^{-K(x)}$, where $K(x)$ is the length of the shortest description of x, called the Kolmogorov complexity of x. Since K and M are incomputable, they have to be approximated in practice. See e.g., [4–7] and references therein. The universality of M also precludes useful statements about the prediction quality at particular time instances n [5, p. 62], as opposed to simple classes like i.i.d. sequences (data) of size n, where accuracy is typically $O(n^{-1/2})$. Luckily, bounds on the expected *total*=cumulative loss (e.g., number of prediction errors) for M can be derived [3,8–10], which is often sufficient in an online setting. The bounds are in terms of the (Kolmogorov) complexity of μ. For instance, for deterministic μ, the number of errors is (in a sense tightly) bounded by $K(\mu)$ which measures in this case the information (in bits) in the observed infinite sequence x.

What's new. In this paper we assume we are at a time $t>1$ and have already observed $x=x_1\cdots x_t$. Hence we are interested in the future prediction performance on $x_{t+1}x_{t+2}\cdots$, since typically we do not care about past errors. If the total loss is finite, the future loss must necessarily be small for large t. In a sense the paper intends to quantify this apparent triviality. If the complexity of μ bounds the total loss, a natural guess is that something like the conditional complexity of μ given x bounds the future loss. (If x contains a lot of (or even all) information about μ, we should make fewer (no) errors anymore.) Indeed, we prove two bounds of this kind but with additional terms describing structural properties of x. These additional terms appear since the total loss is bounded only in expectation, and hence the future loss is small only for "most" $x_1\cdots x_t$. In the first bound (Theorem 1), the additional term is the complexity of the length of x (a kind of worst-case estimation). The second bound (Theorem 7) is finer: the additional term is the complexity of the randomness deficiency of x. The advantage is that the deficiency is small for "typical" x and bounded on average (in contrast to the length). But in this case the conventional conditional complexity turned out to be unsuitable. So we introduce a new natural modification of conditional Kolmogorov complexity, which is monotone as a function of condition. Informally speaking, we require programs (=descriptions) to be consistent in the sense that if a program generates some μ given x, then it must generate the same μ given any prolongation of x. The new posterior bounds also significantly improve upon the previous total bounds.

Contents. The paper is organized as follows. Some basic notation and definitions are given in Sections 2 and 3. In Section 4 we prove and discuss the length-based bound Theorem 1. In Section 5 we show why a new definition of complexity is necessary and formulate the deficiency-based bound Theorem 7. We discuss the definition and basic properties of the new complexity in Section 6, and prove Theorem 7 in Section 7. We briefly discuss potential generalizations to general model classes $\mathcal{M}$ and classification in the concluding Section 8.