GENERALISED ENTROPIES AND ASYMPTOTIC COMPLEXITIES OF LANGUAGES

Yuri Kalnishkan, Michael V. Vyugin, and Vladimir Vovk

Computer Learning Research Centre and Department of Computer Science,
Royal Holloway, University of London,
Egham, Surrey, TW20 0EX, United Kingdom

ABSTRACT

The talk explores connections between asymptotic complexity and generalised entropy. Asymptotic complexity of a language (a language is a set of finite or infinite strings) is a way of formalising the complexity of predicting the next element in a sequence: it is the loss per element of a strategy asymptotically optimal for that language. Generalised entropy extends Shannon entropy to arbitrary loss functions; it is the optimal expected loss given a distribution on possible outcomes. It turns out that the set of tuples of asymptotic complexities of a language w.r.t. different loss functions can be described by means of generalised entropies corresponding to the loss functions.

1. INTRODUCTION

The complete version of this paper has been accepted to *Information and Computation*. An earlier version [1] appeared in conference proceedings.

We consider the following on-line learning scenario: given a sequence of previous outcomes $x_1, x_2, \ldots, x_{n-1}$, a prediction strategy is required to output a prediction γ_n for the next outcome x_n.

We assume that outcomes belong to a finite *outcome space* Ω. Predictions may be drawn from a compact *prediction space* Γ. A loss function $\lambda : \Omega \times \Gamma \to [0, +\infty]$ is used to measure the discrepancy between predictions and actual outcomes; it is assumed to be continuous. The triple $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$ describing the prediction environment is called a game.

The performance of a strategy $\mathfrak{S}$ on a finite string $\boldsymbol{x} = (x_1 x_2 \ldots, x_n)$ is measured by the cumulative loss $\mathrm{Loss}_{\mathfrak{S}}(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda(x_i, \gamma_i)$. Different aspects of this prediction framework have been extensively studied; see [2] for an overview.

One is tempted to define complexity of a string as the loss of an optimal strategy so that elements of "simple" strings $\boldsymbol{x}$ are easy to predict and elements of "complicated" strings are hard to predict and large loss is incurred. However this intuitive idea is difficult to implement formally because it is hard to define an optimal strategy. If $\boldsymbol{x}$ is fixed, the strategy can be tailored to suffer the minimum possible loss on $\boldsymbol{x}$ (0 for natural loss functions such as square, absolute, or logarithmic). If there is complete flexibility in the choice of $\boldsymbol{x}$, i.e., "anything can happen", then every strategy can be tricked into suffering large loss and being greatly outperformed by some other strategy on some sequences $\boldsymbol{x}$.

One approach to this problem is predictive complexity introduced in [3] and studied in [4, 5, 6]. This approach replaces strategies by the class of semi-computable superloss processes. Under certain restrictions on Γ and λ this class has a natural optimal element. Predictive complexity of a finite string is defined up to a constant and is similar in many respects to Kolmogorov complexity; predictive complexity w.r.t. the logarithmic loss function equals the negative logarithm of Levin's a priori semi-measure.

This paper takes a different approach and introduces asymptotic complexity, which is in some respects easier and more intuitive. It is defined for languages (infinite sets of finite strings and sets of infinite sequences) and it equals the asymptotically optimal loss per element. This idea leads to several versions of complexity that behave slightly differently. An important advantage of this approach is that asymptotic complexity exists for all loss functions λ thus eliminating the question of existence, still partly unsolved for predictive complexity. One can consider effective and polynomial-time versions of asymptotic complexity by restricting oneself to computable or polynomial-time computable strategies. The existence of corresponding asymptotic complexities follows trivially.

In this paper we study the following question. Let $\mathfrak{G}_k = \langle \Omega, \Gamma_k, \lambda_k \rangle$, $k = 1, 2, \ldots, K$, be games with the same finite set of outcomes Ω. How do asymptotic complexities of a same set of finite or infinite sequences of elements of Ω compare? We answer this question by describing the set

$$(\mathrm{AC}_1(L), \mathrm{AC}_2(L), \ldots, \mathrm{AC}_K(L)) \subseteq \mathbb{R}^K \ ,$$

where AC_k is an asymptotic complexity w.r.t. $\mathfrak{G}_k$ and L ranges over all non-trivial languages. The set turns out to have a simple geometric description in terms of the generalised entropy studied in [7]. The set depends on the type of asymptotic complexity and may be different for different complexities [1].

For the Shannon entropy there are many results connecting it with complexity and Hausdorff dimension; see,

[1] Note that the statement of the main theorem in the conference version [1] of this paper was inaccurate in this respect. A corrected journal version will appear soon

e.g., Theorem 2.8.1 in [8] and [9]. This paper directly generalises the main result of [10].

The set depends on the type of asymptotic complexity and may be different for different complexities [2].

2. ASYMPTOTIC COMPLEXITY

2.1. Finite Sequences

Let $L \subseteq \Omega^*$ be a set of finite strings. We call the values

$$\overline{\mathrm{AC}}(L) = \inf_{\mathfrak{A}} \limsup_{n\to+\infty} \max_{\boldsymbol{x}\in L\cap\Omega^n} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x})}{n} \ , \qquad (1)$$

$$\underline{\mathrm{AC}}(L) = \inf_{\mathfrak{A}} \liminf_{n\to+\infty} \max_{\boldsymbol{x}\in L\cap\Omega^n} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x})}{n} \qquad (2)$$

the *upper* and *lower asymptotic complexity* of L w.r.t. the game $\mathfrak{G}$. We use subscripts for AC to specify a particular game if it is not clear from the context.

In this paper we are concerned only with infinite sets of finite sequences and asymptotic complexity of a finite or an empty language $L \subseteq \Omega^*$ is undefined. Thus by assumption there are strings of infinitely many lengths in L.

Still there may be no strings of a certain length in L. Let us assume that the limits in (1) and (2) are taken over the subsequence $n_1 < n_2 < \ldots$ of values such that $L \cap \Omega^{n_i} \neq \varnothing$.

2.2. Infinite Sequences

There are two natural ways to define complexities of nonempty languages $L \subseteq \Omega^\infty$.

First we can extend the notions we have just defined. Indeed, for a nonempty set of infinite sequences consider the set of all finite prefixes of all its sequences. The language thus obtained is infinite and has upper and lower complexities. For the resulting complexities we shall retain the notation $\overline{\mathrm{AC}}(L)$ and $\underline{\mathrm{AC}}(L)$. We refer to these complexities as *uniform*.

The second way is the following. Let

$$\overline{\overline{\mathrm{AC}}}(L) = \inf_{\mathfrak{A}} \sup_{\boldsymbol{x}\in L} \limsup_{n\to+\infty} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x}|_n)}{n} \ ,$$

$$\underline{\underline{\mathrm{AC}}}(L) = \inf_{\mathfrak{A}} \sup_{\boldsymbol{x}\in L} \liminf_{n\to+\infty} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x}|_n)}{n} \ .$$

We refer to this complexity as *non-uniform*.

The concept of asymptotic complexity generalises certain complexity measures studied in the literature. The concepts of predictability and dimension studied in [10] can be easily reduced to asymptotic complexity: the dimension is the lower non-uniform complexity w.r.t. a multidimensional generalisation of the logarithmic game and predictability equals $1 - \underline{\underline{\mathrm{AC}}}$, where $\underline{\underline{\mathrm{AC}}}$ is the lower non-uniform complexity w.r.t. a multidimensional generalisation of the absolute-loss game.

[2] Note that the statement of the main theorem in the conference version of this paper was inaccurate in this respect. A corrected journal version will appear soon

3. OTHER DEFINITIONS

3.1. Entropy

Let $\mathbb{P}(\Omega)$ be the set of probability distributions on Ω of size M. The set Ω is finite and we can identify $\mathbb{P}(\Omega)$ with the standard $(M-1)$-simplex

$$\mathbb{P}_M = \left\{ \left(p^{(0)}, p^{(1)}, \ldots, p^{(M-1)}\right) \in [0,1]^M \mid \sum_{i=0}^{M-1} p^{(i)} = 1 \right\} \ .$$

Generalised entropy $H : \mathbb{P}(\Omega) \to \mathbb{R}$ is the infimum of expected loss over $\gamma \in \Gamma$, i.e., for

$$p^* = \left(p^{(0)}, p^{(1)}, \ldots, p^{(M-1)}\right) \in \mathbb{P}(\Omega)$$

we have

$$H(p^*) = \min_{\gamma\in\Gamma} \mathbf{E}_{p^*} \lambda(\omega, \gamma) = \min_{\gamma\in\Gamma} \sum_{i=0}^{M-1} p^{(i)} \lambda(\omega^{(i)}, \gamma) \ .$$

Since $p^{(i)}$ can be 0 and $\lambda(\omega^{(i)}, \gamma)$ can be $+\infty$, we need to resolve an ambiguity. Let us assume that in this definition $0 \times (+\infty) = 0$.

3.2. Sublattices and Subsemilattices

A set $\mathcal{M} \subseteq \mathbb{R}^K$ is a *sublattice* of $\mathbb{R}^K$ if for every $x, y \in M$ it contains their coordinate-wise greatest lower bound $\min(x, y)$ and least upper bound $\max(x, y)$. Clearly, a sublattice of $\mathbb{R}^K$ contains the coordinate-wise maximum and minimum of any finite subset. Similarly, a set $\mathcal{M} \subseteq \mathbb{R}^K$ is an *upper subsemilattice* if for every $x, y \in \mathcal{M}$ it contains their smallest upper bound $\max(x, y)$; a set $M \subseteq \mathbb{R}^K$ is a *lower subsemilattice* if for every $x, y \in \mathcal{M}$ it contains their largest lower bound $\min(x, y)$. In this paper we mostly use upper subsemilattices and therefore sometimes omit the word "upper" in what follows.

A *sublattice closure* of a set $\mathcal{M} \subseteq \mathbb{R}^K$ is the smallest sublattice containing $\mathcal{M}$. Respectively, an *upper subsemilattice closure* of a set $\mathcal{M} \subseteq \mathbb{R}^K$ is the smallest upper semilattice containing $\mathcal{M}$ and a *lower subsemilattice closure* of a set $\mathcal{M} \subseteq \mathbb{R}^K$ is the smallest lower subsemilattice containing $\mathcal{M}$. The sub(semi)lattice closure of $\mathcal{M}$ exists and it is the intersection of all sub(semi)lattices containing $\mathcal{M}$. The sublattice closure contains the subsemilattice closures because each sublattice is a subsemilattice.

Note that the definitions are coordinate-dependent.

3.3. Weak Mixability

The results of this paper are valid for the so called weakly mixable games defined in [11]. A game $\mathfrak{G}$ is weakly mixable if for every two prediction strategies $\mathfrak{S}_1$ and $\mathfrak{S}_2$ there is a prediction strategy $\mathfrak{S}$ such that

$$\mathrm{Loss}_{\mathfrak{S}}(\boldsymbol{x}) \le \min\left(\mathrm{Loss}_{\mathfrak{S}_1}(\boldsymbol{x}), \mathrm{Loss}_{\mathfrak{S}_2}(\boldsymbol{x})\right) + \alpha(|\boldsymbol{x}|) \quad (3)$$

for all finite strings $\boldsymbol{x}$, where $|\boldsymbol{x}|$ is the length of $\boldsymbol{x}$ and $\alpha(n) = o(n)$ as $n \to \infty$. It is shown in [11] that weak

mixability is equivalent to the convexity of the set of superpredictions w.r.t. $\mathfrak{G}$. In particular, if Γ is convex and λ is convex in predictions, weak mixability holds.

3.4. Effective Versions of Complexities

One can restrict the range of possible strategies to computable or polynomial-time computable and obtain effective and polynomial-time versions of the asymptotic complexities.

The concept of a computable strategy requires clarification. We will give a definition along the lines of [12]; see also [13, Sections 7 and 9.4].

A *dyadic* rational number is a number of the form $m/2^n$, where m is an integer and n is a positive integer. We call a triple $\langle b, \boldsymbol{x}, \boldsymbol{y}\rangle$, where $b \in \mathbb{B}$ is a bit and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{B}^*$ are binary strings, a *representation of a dyadic number* d if $\boldsymbol{x}$ is the binary representation of a nonnegative integer $m > 0$, $\boldsymbol{y}$ is the binary representation of a nonnegative integer $n > 0$, and b represents a sign s (assume that $s = 1$ if $b = 1$ and $s = -1$ if $b = 0$) so that $d = sm/2^n$.

For every $x \in \mathbb{R}$ define a set CF_x of dyadic Cauchy sequences exponentially converging to x, i.e., functions ϕ_x from non-negative integers to dyadic numbers such that $|\phi_x(n) - x| \le 2^{-n}$ for all n. Any element of CF_x can be thought of as a dyadic representation of x.

Let Ω be a finite set. A function $f : \Omega^* \to \mathbb{R}$ is computable if there is a Turing machine that given a finite string $\boldsymbol{x} = x_1 x_2 \dots x_m \in \Omega^*$ and non-negative integer precision n outputs a representation of a dyadic number d such that $|f(\boldsymbol{x}) - d| \le 2^{-n}$. In other words, for every $\boldsymbol{x} \in \Omega^*$ the machine calculates a function from $\mathrm{CF}_{f(\boldsymbol{x})}$. If there is a polynomial $p(\cdot,\cdot)$ such that the machine always finishes work in $p(m, n)$, we say that f is polynomial-time computable. A function $f = (f_1, f_2, \dots, f_k) : \Omega^* \to \mathbb{R}^k$ is (polynomial-time) computable if all its components $f_1, f_2, \dots, f_k$ are (polynomial-time) computable.

A function $f : M \to \mathbb{R}$, where $M \subseteq \mathbb{R}$, is computable if there is an oracle Turing machine that given a non-negative integer precision n (as a binary string) and an oracle evaluating some $\phi_x \in \mathrm{CF}_x$ outputs a representation of a dyadic number d such that $|f(x) - d| \le 2^{-n}$. If there is a polynomial $p(\cdot)$ such that the machine finishes work in $p(n)$ for all $x \in M$, we say that f is polynomial-time computable. Intuitively a machine can at any moment request a dyadic approximation of x up to 2^{-m} and get it in no time. Computable and polynomial-time computable functions on $M \subseteq \mathbb{R}^k$ and $M \times \Omega^*$ to $\mathbb{R}$ and $\mathbb{R}^m$ are defined in a similar fashion.

We call a game $\mathfrak{G} = \langle \Omega, \Gamma, \lambda\rangle$ *(polynomial-time) computable* if $\Gamma \subseteq \mathbb{R}^k$ is a closure of its interior and the function $e^{-\lambda(\omega,\gamma)}$ is (polynomial-time) computable. Note that we do not postulate computability of λ itself because if would have implied boundedness of λ. A *(polynomial-time) computable strategy* w.r.t. $\mathfrak{G}$ is a (polynomial-time) computable function $\Omega^* \to \Gamma$.

3.5. Computability and Weak Mixability

A (polynomial-time) computable game $\mathfrak{G}$ will be called *(polynomial-time) computable very weakly mixable* if for all (polynomial-time) computable strategies $\mathfrak{S}_1$ and $\mathfrak{S}_2$ and $\varepsilon > 0$ there is a (polynomial-time) computable strategy $\mathfrak{S}$ such that

$$\mathrm{Loss}_{\mathfrak{S}}(\boldsymbol{x}) \le \min\left(\mathrm{Loss}_{\mathfrak{S}_1}(\boldsymbol{x}), \mathrm{Loss}_{\mathfrak{S}_2}(\boldsymbol{x})\right) + \varepsilon|\boldsymbol{x}| + \alpha_\varepsilon(|\boldsymbol{x}|)$$

for all finite strings $\boldsymbol{x}$, where $\alpha_\varepsilon(n) = o(n)$ as $n \to \infty$.

It is not easy to formulate a simple criterion of computable mixability. The following rather general condition is sufficient. If a game $\mathfrak{G}$ is (polynomial-time) computable, the prediction space Γ is convex, and the loss function $\lambda(\omega, \gamma)$ is convex in the second argument, then $\mathfrak{G}$ is (polynomial-time) computable weakly mixable.

If we add the requirement of boundness of λ, we can achieve an effective version of (3), but this is not necessary for the purpose of this paper.

4. MAIN RESULT

Consider $K \ge 1$ games $\mathfrak{G}_1, \mathfrak{G}_2, \dots, \mathfrak{G}_K$ with the same finite set of outcomes Ω. Let H_k be $\mathfrak{G}_k$-entropy for $k = 1, 2, \dots, K$. The $\mathfrak{G}_1/\mathfrak{G}_2/\dots/\mathfrak{G}_K$*-entropy set* is the set $\{(H_1(p), H_2(p), \dots, H_K(p)) \mid p \in \mathbb{P}(\Omega)\} \subseteq \mathbb{R}^K$. The convex hull of the $\mathfrak{G}_1/\mathfrak{G}_2/\dots/\mathfrak{G}_K$-entropy set is called the $\mathfrak{G}_1/\mathfrak{G}_2/\dots/\mathfrak{G}_K$*-entropy hull*.

Theorem 1. *If games $\mathfrak{G}_1, \mathfrak{G}_2, \dots, \mathfrak{G}_K$ ($K \ge 1$) have the same finite outcome space Ω and are weakly mixable, then the sublattice closure of the $\mathfrak{G}_1/\mathfrak{G}_2/\dots/\mathfrak{G}_K$-entropy hull coincides with the following sets (here AC_k is asymptotic complexity w.r.t. $\mathfrak{G}_k$, $k = 1, 2, \dots, K$):*

$$\left\{ (\underline{\mathrm{AC}}_1(L), \underline{\mathrm{AC}}_2(L), \dots, \underline{\mathrm{AC}}_K(L)) \mid L \subseteq \Omega^* \text{ and } L \text{ is infinite} \right\};$$

$$\left\{ (\underline{\mathrm{AC}}_1(L), \underline{\mathrm{AC}}_2(L), \dots, \underline{\mathrm{AC}}_K(L)) \mid L \subseteq \Omega^\infty \text{ and } L \ne \varnothing \right\};$$

$$\left\{ \left(\underline{\underline{\mathrm{AC}}}_1(L), \underline{\underline{\mathrm{AC}}}_2(L), \dots, \underline{\underline{\mathrm{AC}}}_K(L)\right) \mid L \subseteq \Omega^\infty \text{ and } L \ne \varnothing \right\};$$

the upper subsemilattice closure of the $\mathfrak{G}_1/\mathfrak{G}_2/\dots/\mathfrak{G}_K$-entropy hull coincides with the following sets:

$$\left\{ \left(\overline{\mathrm{AC}}_1(L), \overline{\mathrm{AC}}_2(L), \dots, \overline{\mathrm{AC}}_K(L)\right) \mid L \subseteq \Omega^* \text{ and } L \text{ is infinite} \right\};$$

$$\left\{ \left(\overline{\mathrm{AC}}_1(L), \overline{\mathrm{AC}}_2(L), \ldots, \overline{\mathrm{AC}}_K(L)\right) \mid L \subseteq \Omega^\infty \text{ and } L \neq \varnothing \right\} ;$$

$$\left\{ \left(\overline{\overline{\mathrm{AC}}}_1(L), \overline{\overline{\mathrm{AC}}}_2(L), \ldots, \overline{\overline{\mathrm{AC}}}_K(L)\right) \mid L \subseteq \Omega^\infty \text{ and } L \neq \varnothing \right\} .$$

If the games $\mathfrak{G}_1, \mathfrak{G}_2 \ldots, \mathfrak{G}_K$ are (polynomial-time) computable very weakly mixable, the same holds for effective and polynomial-time complexities.

The conference version [1] of the paper incorrectly claimed that all the sets of complexity tuples coincide with the upper subsemilattice closure of the entropy hull. This is not true because upper subsemilattice closure of the entropy hull may be different from the sublattice closure.

5. RECALIBRATION LEMMA

The key element of the proof is the following lemma:

Lemma 1. *Let $\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_K$ be prediction strategies for weakly mixable games $\mathfrak{G}_1, \mathfrak{G}_2 \ldots, \mathfrak{G}_K$ with the same set of outcomes Ω of size M. Then for every weakly mixable game $\mathfrak{G}$ and $\varepsilon > 0$ there is a prediction strategy $\mathfrak{S}$ and a function $f : \mathbb{N} \to \mathbb{R}$ such that $f(n) = o(n)$ as $n \to \infty$ and for every finite string $\boldsymbol{x} \in \Omega^*$ there are distributions $p_1, p_2, \ldots, p_N \in \mathbb{P}_M$ and $q = (q_1, q_2, \ldots, q_N) \in \mathbb{P}_N$ such that*

1. *for all $k = 1, 2, \ldots, K$ if H_k is the generalised entropy w.r.t. $\mathfrak{G}_k$ then*

$$\sum_{i=1}^{N} q_i H_k(p_i) \leq \frac{\mathrm{Loss}_{\mathfrak{A}_k}^{\mathfrak{G}_k}(\boldsymbol{x})}{|\boldsymbol{x}|} + \varepsilon \ ;$$

2. *if H is the generalised entropy w.r.t. $\mathfrak{G}$ then*

$$\mathrm{Loss}_{\mathfrak{S}}^{\mathfrak{G}}(\boldsymbol{x}) \leq |\boldsymbol{x}| \left(\sum_{i=1}^{N} q_i H(p_i) + \varepsilon \right) + f(|\boldsymbol{x}|) \ .$$

The idea behind the lemma can be described informally as follows. Consider a predictor outputting, say, the likelihood of a rain. Suppose that by analysing its past performance we have found a pattern of the following kind. Whenever the predictor outputs the value of 70%, it actually rains in 90% of cases. We can thus improve the predictor by *recalibrating* it: if we see the prognosis of 70%, we replace it by 90%. Generally speaking, we may observe that whenever a predictor outputs a prediction γ_1, a more appropriate choice would be γ_2. By outputting γ_1, the predictor signals us about a specific state of the nature; however, γ_2 is a better prediction for this state. The loss per element of the optimised strategy is close to the generalised entropy w.r.t. some distribution and this leads to the first part of the lemma.

The intuitive interpretation of the seond part is as follows. Predictions of (discretised) strategies allow us to split a string to several (generally speaking, not contiguous) substrings. The strategies tell us nothing of the behaviour of outcomes within the substrings so we can assume that inside each substring the outcomes are i.i.d. (independent identically distributed) and construct a new strategy exploiting this. The loss per element of the new strategy will be a convex combination of entropies w.r.t. the distributions of outcomes from the substrings and the new strategy will perform better or nearly as well as the original strategies.

6. REFERENCES

[1] V. Vovk Y. Kalnishkan and M.V. Vyugin, "Generalised entropy and asymptotic complexities of languages," in *20th Annual Conference on Learning Theory, COLT 2007*. 2007, pp. 293–307, Springer.

[2] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*, Cambridge University Press, 2006.

[3] V. Vovk and C. J. H. C. Watkins, "Universal portfolio selection," in *Proceedings of the 11th Annual Conference on Computational Learning Theory*. 1998, pp. 12–23, ACM Press.

[4] Y. Kalnishkan, "General linear relations among different types of predictive complexity," *Theoretical Computer Science*, vol. 271, pp. 181–200, 2002.

[5] Y. Kalnishkan, V. Vovk, and M. V. Vyugin, "Loss functions, complexities, and the Legendre transformation," *Theoretical Computer Science*, vol. 313, no. 2, pp. 195–207, 2004.

[6] Y. Kalnishkan, V. Vovk, and M. V. Vyugin, "How many strings are easy to predict?," *Information and Computation*, vol. 201, no. 1, pp. 55–71, 2005.

[7] P. D. Grünwald and A. P. Dawid, "Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory," *The Annals of Statistics*, vol. 32, no. 4, pp. 1367–1433, 2004.

[8] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, Springer, 3rd edition, 2008.

[9] B. Ya. Ryabko, "Noiseless coding of combinatorial sources, hausdorff dimension, and Kolmogorov complexity," *Problems of Information Transmission*, vol. 22, no. 3, pp. 170–179, 1986.

[10] L. Fortnow and J. H. Lutz, "Prediction and dimension," *Journal of Computer and System Sciences*, vol. 70, no. 4, pp. 570–589, 2005.

[11] Y. Kalnishkan and M. V. Vyugin, "The weak aggregating algorithm and weak mixability," *Journal of Computer and System Sciences*, vol. 74, no. 8, pp. 1228–1244, 2008.

[12] Ker-I Ko, *Complexity theory of real functions*, Birkhäuser, 1991.

[13] K. Weihrauch, *Computable Analysis*, Springer, 2000.