Combinatorial proof of Muchnik's theorem

Alexander Shen

Abstract

A combinatorial proof of Muchnik's theorem about conditional descriptions (codes, see [2]) is given. It uses the notion of "online matching" in a bipartite graph that could be interesing in its own right.

1 Muchnik's theorem

An. Muchnik [2] has proved the following theorem:

Theorem. *Let a and b be two binary strings, $K(a) < n$ and $K(a|b) < k$. Then there exists a string p such that*

- $K(a|p, b) \leq O(\log n)$;
- $K(p) \leq k + O(\log n)$;
- $K(p|a) \leq O(\log n)$.

Here $K(u)$ stands for the Kolmogorov complexity of string u; conditional complexity of u when v is known is denoted by $K(u|v)$. The constants hidden in $O(\log n)$ do not depend on n, k, a, b.

Informally, theorem says that there exists a program p that transforms b to a, has the minimal possible complexity $K(a|b)$ (up to a logarithmic term) and, moreover, is easily obtained from a. (The last requirement is crucial, otherwise the statement becomes trivial and the shortest program that transforms b to a can be used.) We used a more abstract formulation to avoid references to programs.

In many cases statements about Kolmogorov complexity have combinatorial counterparts (and sometimes it is easy to show the equivalence between complexity and combinatorial statements). So it would be interesting to find a combinatorial counterpart of this theorem. Indeed, some equivalent statement (that involves some game on graphs) exists. (The general approach to translate statements in general recursion theory into the existence of a winning strategy in some game is explained in [1].) However, this game is rather complicated. But there exist a stronger and more natural statement about online matchingA which easily implies Muchnik's theorem and can be proved using the same ideas (slightly modified) that were used by Muchnik in his original proof.

2 On-line matchings

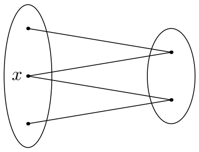
Consider a bipartite graph with left part L, right part R and set of edges $E \subset L \times R$. Let s be some integer. We are interested in the following property of the graph:

Dagstuhl Seminar Proceedings 06051
Kolmogorov Complexity and Applications
http://drops.dagstuhl.de/opus/volltexte/2006/625

for any subset L' of L of size at most s there exists a subset $E' \subset E$ that performs a bijection between L' and some $R' \subset R$.

A necessary and sufficient condition for this is provided by well known Hall theorem: for each set $L' \subset L$ of size $t \le s$ the set of all neighbors of elements of L' contains at least t elements.

However, this condition is not sufficient for the online version. In this version the adversary gives us elements of left part L (up to s elements) one by one. We should provide a counterpart for each given element x, i.e., to choose some its neighbor $y \in R$ not used before. This choice is final and cannot be changed later.



This is more difficult. For example, for the graph shown one can find a matching for each subset of size at most 2, but this cannot be done online (the difficulty arises if adversary starts with x).

Now we formulate combinatorial statement that implies Muchnik's theorem:

Combinatorial statement. *There exists a constant c with the following property: for any integers n and $k \le n$ there exists a bipartite graph $E \subset L \times R$ whose left part L has size 2^n, right part R has size $2^k n^c$, each vertex in L has at most n^c neighbors in R, and online matching is possible up to size 2^k.*

(So the size of the online matching is close to the size of R up to a polynomial factor, and the degree of all L-elements is polynomially bounded.)

3 Proof of Muchnik's theorem

Let us show how the combinatorial statement implies Muchnik's theorem. Note first that we may assume without loss of generality that *length* of string a (instead of its *complexity*) is less than n. Indeed, if we replace a by a shortest program that generates a, all complexities involving a change only by $O(\log n)$ term.

Then consider the graph E provided by the combinatorial statement. Its left part L is interpreted as the set of all strings of length less than n; therefore, string a is element of L. Let us enumerate strings x of length less than n such that $K(x|b) < k$. There are at most 2^k such strings, and a is one of them. So it is possible to find an online matching for them (in the order they appear during the enumeration). Let p be an element of R that correspongs to a in the matching.

Let us check that p satisfies all the conditions of Muchnik's theorem. First of all, note that the graph E can be chosen in such a way that its complexity is $O(\log n)$. The combinatorial statemenent guarantees that a graph with required properties exist. Knowing n and k, we can perform an exhaustive search until the first graph with these properties is found. This graph is a computable function of n and k, so its complexity does not exceed $K(n, k) = O(\log n)$.

If a is known (as well as n and k), then p can be specified by an ordinal number of p in the list of a-neighbors. This list contains at most n^c elements, so the ordinal number requires $O(\log n)$ bits. (And $K(n, k) = O(\log n)$ as we have already noted.)

To specify p without knowing a, we give its ordinal number in R, which is $k + O(\log n)$ bits long. (Again n and k are used, but this is another $O(\log n)$ bits.)

Finally, to reconstruct a from b and p, we start the process of enumerating all strings on lengths less than n that have conditional complexity (relative to b, which is known) less than k, and find R-counterparts for them using the online matching property, until p appears. Then a is p's L-counterpart.

(Formally speaking, for given n and k we should search for the graph G together with the online matching procedure, and then use the same procedure both for constructing p and for reconstructing a from b and p.)

4 On-line matchings exist

It remains to provide a proof for the combinatorial statement. This proof follows the original Muchnik's argument adapted for the combinatorial setting and consists of several steps.

Offline version

First, let us prove a weaker statement that deals only with offline matchings (or, better, with necessary and sufficient conditions for them provided by Hall theorem).

Offline version. *There exists a constant c with the following property: for any integers n and $k \le n$ there exists a bipartite graph $E \subset L \times R$ whose left part L has size 2^n, right part R has size $2^k n^c$, each vertex in L has at most n^c neighbors in R and or any subset $X \subset L$ of size $t \le 2^k$ the set $N(X)$ of all neighbors of all elements of X contains at least t elements.*

Let us prove this statement using the probabilistic argument. Probability distribution: n^c for each vertex $l \in L$ are chosen uniformly in R and are independent (for different l and for different neighbors of aa given l). In this way we obtain a (random) graph where all vertices in L have degree at most n^c (it could be less if two independent choices for a given vertex coincide).

We claim that this random graph has the required property with positive probability.

If it does not, there exists some set $X \subset L$ of some size $t \le n$ and some set Y of size less than t such that all neighbors of all elements of X belong to Y. For a fixed X and Y the probability of this event is bounded by

$$\left(\frac{1}{n^c}\right)^{tn^c}$$

since we made tn^c independent choices (n^c times for each of t elements) and the probability to get into Y for one choice is at most $1/n^c$ (the set Y covers at most $1/n^c$ fraction of R).

To get a bound for a probability of violating the requirement, we multiply this bound by the number of pairs X, Y. The set X can be chosen in at most $(2^n)^t$ different ways (for each element of t elements we have at most 2^n choices; actually the number is smaller since the order of elements does not matter), and for Y we have at most $(2^k n^c)^t$ choices. This is for a fixed t; we need then to add these bounds for all $t \le 2^k$. Therefore the total bound is

$$\sum_{t=1}^{2^k} \left(\frac{1}{n^c}\right)^{tn^c} (2^n)^t \left(2^k n^c\right)^t$$

This is a geometric series; the sum is less than 1 (which is our goal) if the base is small. Indeed, the base is

$$\left(\frac{1}{n^c}\right)^{n^c} (2^n)\left(2^k n^c\right) = \frac{2^{n+k}}{n^{c(n^c-1)}}$$

and $c = 2$ makes it small (it even tends to zero as $n \to \infty$). Offline version is proved.

Online modifications

Assume that graph $E \subset L \times R$ satisfies the conditions for the offline version (for given n and k). Now we use the same graph in online setting with the following straightforward ("greedy") strategy. When a new element $x \in L$ arrives, we check if it has neighbors that are not used yet. If yes, one of these neighbors is chosen to be a counterpart of x. If not, x is "rejected".

Before we explain what to do with the rejected elements, let us prove that at most half of 2^k given elements could be rejected. Assume that more than 2^{k-1} elements are rejected. Then less than 2^{k-1} elements are served and therefore less than 2^{k-1} elements of R are used as counterparts. But all neighbors of all rejected elements are used (since this is the only reason for rejection), and we get the contradiction with the condition $\#N(X) \geq \#X$ if X is the set of rejected elements.

Now we need to deal with rejected elements. They are forwarded to the "next layer" where the task is to find online matching for 2^{k-1} elements. If we can do this, then we combine both graphs using the same L and disjoint right parts R_1 and R_2. And the elements rejected at the first layer are satisfied at the second one. In other terms: (n,k) online problem is reduced to (n,k) offline problem and $(n,k-1)$ online problem. The latter can then be reduced to $(n,k-1)$ offline and $(n,k-2)$ online problems etc.

Finally we get k levels. Each level satisfies at least half of the requests and forwards the remaining half to the next layer. After k levels of filtering only one request can be left unserved, and one more layer is enough. (Note that we may use the same graph on all layers.)

More precisely, we have proved the following statement: *Let $E \subset L \times R$ be a graph that satisfies the conditions of the offline version for given n and k. Replace each element in R by $(k+1)$ copies, all connected to the same elements of L as before. Then new graph provides online matchings up to size 2^k.*

Note that this construction multiplies the size of R and the degree of vertices in L by $(k+1)$, so they remain polynomial in n.

The combinatorial statement is proven.

5 Questions

It would be interesting to find

- a more direct proof of combinatorial statement, without first proving its offline version;
- an explicit construction for graphs with the required property;
- what is the computational complexity of the problem "compute a maximal t such that given graph allows online matchings of size t";

• a combinatorial statement that would imply not only Muchnik theorem but also Wolf – Slepyan theorem (its counterpart in Shannon information theory);

• the connection between this result and known results about expander and extractor graphs.

Acknowledments

This article is based on the discussions and results reported at the Kolmogorov seminar (Moscow). The author is grateful to all participants of the seminar for many useful comments. I am thankful to Laboratoire d'Informatique Fondamentale de Marseille for hospitality and support.

References

[1] Мучник Ан. А. Об основных структурах дескриптивной теории алгоритмов. *Доклады Академии Наук СССР*, т. 285, № 2 (1985) 280–283. Translation: An.A. Muchnik, On basic structures of the descriptive theory of algorithms. *Soviet Math. Dokl.*, **32**, 671–674 (1985)

[2] An. Muchnik, Conditional complexity and codes, *Theoretical Computer Science*, v. 271, no. 1–2, p. 97–109 (2002).