Recursion-Theoretic Ranking and Compression

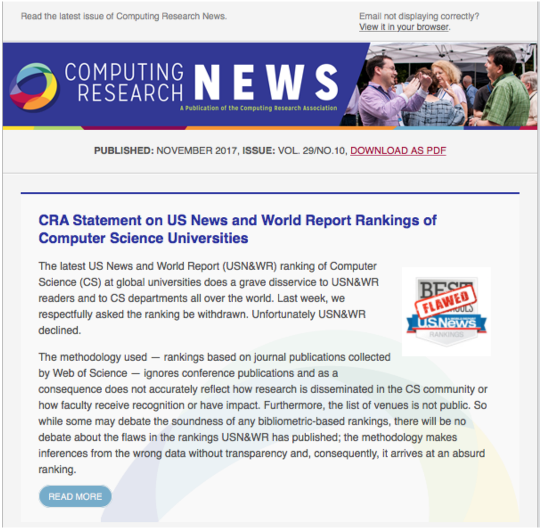
Lane A. Hemaspaandra and Dan Rubery
Department of Computer Science
University of Rochester

November 17, 2017

NYCAC 2017, CUNY Graduate Center, November 17, 2017

Happy (Day Before Day Before) Birthday, Eric!

Is This Talk on the Currently Hottest Topic in the Academic Computer Science World? You Be the Judge!



Read the latest issue of Computing Research News.

Email not displaying correctly? View it in your browser.

COMPUTING RESEARCH NEWS

A Publication of the Computing Research Association

PUBLISHED: NOVEMBER 2017, **ISSUE:** VOL. 29/NO.10, DOWNLOAD AS PDF

CRA Statement on US News and World Report Rankings of Computer Science Universities

The latest US News and World Report (USN&WR) ranking of Computer Science (CS) at global universities does a grave disservice to USN&WR readers and to CS departments all over the world. Last week, we respectfully asked the ranking be withdrawn. Unfortunately USN&WR declined.

The methodology used — rankings based on journal publications collected by Web of Science — ignores conference publications and as a consequence does not accurately reflect how research is disseminated in the CS community or how faculty receive recognition or have impact. Furthermore, the list of venues is not public. So while some may debate the soundness of any bibliometric-based rankings, there will be no debate about the flaws in the rankings USN&WR has published; the methodology makes inferences from the wrong data without transparency and, consequently, it arrives at an absurd ranking.

READ MORE

"Anyone with knowledge of CS research will see these rankings for what they are—nonsense—and ignore them. But others may be seriously misled. … We urge the community to ignore the USN&WR rankings of Computer Science."—From the CRA Statement

Overview

1. Introduction
2. Definitions
3. Hard Rankable and Compressible Sets
4. Rankability and Compressibility for RE Sets
5. Rankability and Compressibility for coRE Sets
6. And There Is More...

Introduction

- [Compression] We are looking at which sets A can have "the air crushed out of them" by different classes of functions. This means we are speaking (in some sense) of a bijection between A and Σ^*.
- [Ranking] We also want to know for which sets we can "crush the air out" while still respecting the order of elements in A.
- We will view this from a computability perspective, finding which sets can be compressed/ranked by recursive or partial recursive functions.

Why? After all, programmers are not clamoring to have recursion-theoretic perfect, minimal hash functions for infinite sets. But the goal here is learning more about the structure of sets, and the nature of—or in some cases the impossibility of—compression by total and partial recursive functions. In particular, what sets and classes can we show to have, or lack, such compression and ranking functions?

Introduction

- [Compression] We are looking at which sets A can have "the air crushed out of them" by different classes of functions. This means we are speaking (in some sense) of a bijection between A and Σ^*. (Opposite to the traditional direction of notion transfer, we are studying the *r.f.t.* analogue of a notion from *complexity*, namely, the P-compressible sets of Goldsmith, Hemachandra, and Kunen, 1992.)
- [Ranking] We also want to know for which sets we can "crush the air out" while still respecting the order of elements in A. (This was first considered in complexity theory by Allender, 1985, and Goldberg and Sipser, 1985. The latter for example showed that even sets in P can have ranking functions that are complete for $\#\mathrm{P}$.)
- We will view this from a computability perspective, finding which sets can be compressed/ranked by recursive or partial recursive functions. (The existing r.f.t. notions of regressive sets, retraceable sets, and isolic reductions are the closest notions in r.f.t., but in the paper we prove them to much differ from our notions.)

Introduction

In some sense, we are simply looking at minimal, perfect hash functions... for infinite sets... in the recursion-theoretic realm.

Definitions

For a set $A \subseteq \Sigma^*$, and a function f, possibly partial, we say that f is a *compression function* for A if:

- $domain(f) \supseteq A$,
- $f(A) = \Sigma^*$, and
- f is injective on A, i.e., for any $x, y \in A$, if $x \neq y$, then $f(x) \neq f(y)$.

Given a class of (possibly partial) functions $\mathcal{F}$ mapping Σ^* to Σ^*, typically $\mathcal{F}_{REC}$ or $\mathcal{F}_{PR}$, A is $\mathcal{F}$*-compressible* if there is a function $f \in \mathcal{F}$ such that f is a compression function for A.

Note that on $\overline{A}$ the compression function can do whatever warms its (possibly evil) heart, as long as doing so doesn't invalidate its membership in $\mathcal{F}$. It can (if $\mathcal{F}$ allows) diverge. Or, for example, 1776 or an infinite number of members of $\overline{A}$ can map to the same string in Σ^* (which necessarily will also be mapped to by exactly one element of A).

Definitions (cont.)

We will also overload our definition a little, saying:
$\mathcal{F}$-compressible $= \{A \mid A \text{ is } \mathcal{F}\text{-compressible}\}$.

For each set of languages $C \subseteq 2^{\Sigma^*}$, we will say that C *is $\mathcal{F}$-compressible* if $(\forall A \in C)[A \text{ infinite} \implies A \text{ is } \mathcal{F}\text{-compressible}]$.

Note: No finite set can be compressible, since finite sets are not big enough to "cover" Σ^*. When we want to denote the variant of our compression classes that for free just tosses in all the finite sets, we'll denote that by adding a prime: *$\mathcal{F}$-compressible$'$*. (So, as a heads-up, note that a prime throws in the finite sets, but also due to the above things of the form "[class] is $\mathcal{F}$-compressible" are definitionally building them in whenever that particular locution is used.)

Definitions (cont.)

Ranking is a special case of compression that respects lexicographic order. For a set $A \subseteq \Sigma^*$, and a function f, possibly partial, f is a *ranking function* for A if:

- *domain*$(f) \supseteq A$ and
- if $x \in A$, then $f(x) = \|A^{\leq x}\|$ (that is—via implicit coercion—if x is the i^{th} string in A, then $f(x)$ is the i^{th} string in Σ^*).

$\mathcal{F}$-rankable is defined analogously to the compression case.

$\mathcal{F}$-rankable $= \{A \mid A \text{ is } \mathcal{F}\text{-rankable}\}$.

For $C \subseteq 2^{\Sigma^*}$, C is said to be *$\mathcal{F}$-rankable* if $(\forall A \in C)[A \text{ is } \mathcal{F}\text{-rankable}]$.

Basic Inclusions

- $REC \subseteq \mathcal{F}_{REC}$-rankable $\subseteq \mathcal{F}_{PR}$-rankable.
- $REC \subseteq \mathcal{F}_{REC}$-compressible$'$ $\subseteq \mathcal{F}_{PR}$-compressible$'$ (and $\mathcal{F}_{REC}$-compressible $\subseteq \mathcal{F}_{PR}$-compressible).
- For any $\mathcal{F}$, $\mathcal{F}$-rankable $\subseteq \mathcal{F}$-compressible$'$.
- RE is $\mathcal{F}_{PR}$-compressible. (This claim/proof are examples of the "[class] is" type of thowing in of the finite sets.) If $A \in$ RE is infinite, take a machine that enumerates A without repetitions. The compression function f maps the i^{th} output of the enumerator to the i^{th} string in Σ^*. The compression function will not halt on strings in $\overline{A}$, but this is allowed.

Hard Rankable and Compressible Sets

If foo is a reduction type (in our case, recursive 1-tt reductions) such that $\equiv_{foo}$ is an equivalence relation, then each equivalence class of that relation is said to be a *foo degree*.

Theorem

Every 1-tt degree (except that of the recursive sets) contains:

- *A set that is $\mathcal{F}_{REC}$-rankable.*
- *A set that is $\mathcal{F}_{REC}$-compressible but not $\mathcal{F}_{PR}$-rankable.*

(Note: $A \leq_{\text{1-}tt} B$ if A can be decided using at most one query about membership in B. A and B are in the same 1-tt degree exactly if $A \leq_{\text{1-}tt} B$ and $B \leq_{\text{1-}tt} A$.)

Hard Rankable and Compressible Sets

Let $s_0, s_1, s_2, \ldots$ enumerate all strings in Σ^* in lexicographic order. Then for any nonrecursive language A, the language

$$B_1 = \{s_{2i} \mid s_i \in A\} \cup \{s_{2i+1} \mid s_i \notin A\}$$

is 1-tt equivalent to A, and is $\mathcal{F}_{REC}$-rankable by the function f defined by:

$$f(s_{2i}) = f(s_{2i+1}) = s_i.$$

Hard Rankable and Compressible Sets

The set:

$$B_2 = \{s_{4i} \mid i \geq 0\} \cup \{s_{4i+1} \mid s_i \in A\} \cup \{s_{4i+2} \mid i \geq 0\} \cup \{s_{4i+3} \mid s_i \notin A\}$$

is 1-tt equivalent to A, and is $\mathcal{F}_{REC}$-compressible.
The compression function f maps:

- $f(s_{4i}) = s_{3i}$
- $f(s_{4i+1}) = s_{3i+1}$
- $f(s_{4i+2}) = s_{3i+2}$
- $f(s_{4i+3}) = s_{3i+1}$.

Hard Rankable and Compressible Sets

The set:

$$B_2 = \{s_{4i} \mid i \geq 0\} \cup \{s_{4i+1} \mid s_i \in A\} \cup \{s_{4i+2} \mid i \geq 0\} \cup \{s_{4i+3} \mid s_i \notin A\}$$

is not $\mathcal{F}_{PR}$-rankable, however.
Suppose B_2 were $\mathcal{F}_{PR}$-rankable with ranking function g. Then $s_i \in A$ if and only if $g(s_{4i+2}) - g(s_{4i}) = 2$. Since g must halt on inputs in B_2, this procedure will always halt, and hence B_2 is recursive. This contradicts our assumption that A was nonrecursive, since $A =_{tt} B$.

Hard Rankable and Compressible Sets

Theorem

Every 1-tt degree (except that of the recursive sets) contains:

- *A set that is $\mathcal{F}_{REC}$-rankable.*
- *A set that is $\mathcal{F}_{REC}$-compressible but not $\mathcal{F}_{PR}$-rankable.*

Corollary

There exist sets that are not in the arithmetical hierarchy, but that are $\mathcal{F}_{REC}$-rankable (and thus are certainly also $\mathcal{F}_{REC}$-compressible).

Rankability and Compressibility for RE Sets

Theorem

$RE \cap \mathcal{F}_{PR}$-*rankable* $= REC$.

Rankability and Compressibility for RE Sets

Theorem

$RE \cap \mathcal{F}_{PR}$-rankable $= REC$.

The $\supseteq$ direction is immediate.
For the $\subseteq$ direction, consider a set $A \in RE \cap \mathcal{F}_{PR}$-rankable. If A is finite, then it is recursive, so assume A infinite. Since A is r.e., take an enumerator machine E for A, and a $\mathcal{F}_{PR}$ ranking function f for A. Then the following procedure will decide if $x \in A$. Run the enumerator E. Each time E enumerates a string, check if it just enumerated x, in which case $x \in A$. Otherwise, check if E has enumerated two strings y, y' with $y < x < y'$, and $f(y') - f(y) = 1$ (or has enumerated a string y', $x < y'$, with $f(y') = 1$, i.e., $f(y') = \epsilon$). In this case, we know $x \notin A$, since E has enumerated adjacent elements rank-wise within A that "bracket" x in Σ^* (or that E has enumerated that a string larger than x is the first string in the A).
Note: This is essentially building an in-order enumerator for A.

Rankability and Compressibility for RE Sets

Theorem

$RE \cap \mathcal{F}_{REC}\text{-compressible}' = REC$.

(The ranking theorem we did on the previous slides was about $\mathcal{F}_{PR}$-rankability. So, can our above theorem perhaps be improved to: $RE \cap \mathcal{F}_{PR}$-compressible $= REC$? No; since every infinite RE set is $\mathcal{F}_{PR}$-compressible, every set in $RE - REC$ is a counterexample to such an improvement.)

Rankability and Compressibility for RE Sets

Theorem

$RE \cap \mathcal{F}_{REC}$-*compressible*$' = REC$.

The $\supseteq$ direction is immediate.
For the $\subseteq$ direction, since finite sets are all recursive, consider an infinite set $A \in RE \cap \mathcal{F}_{REC}$-compressible$'$. Let f be the $\mathcal{F}_{REC}$ compression function. Then note that $\overline{A} = \{y \mid \exists x \in A, x \neq y, f(x) = f(y)\}$. This is r.e., so A is co-r.e., and therefore recursive.

Rankability and Compressibility for RE Sets

To summarize, sets in $RE - REC$ are (infinite, obviously, and):

- $\mathcal{F}_{PR}$-compressible.
- Not $\mathcal{F}_{REC}$-compressible.
- Not $\mathcal{F}_{PR}$-rankable.

$\Delta_2^0 \not\subseteq \mathcal{F}_{PR}$-compressible$'$

Theorem

$\Delta_2^0 \not\subseteq \mathcal{F}_{PR}$-compressible$'$.

Note: Δ_2^0 is the languages that are recursive in the halting problem.

$\Delta^0_2 \not\subseteq \mathcal{F}_{PR}$-compressible$'$

Theorem

$\Delta^0_2 \not\subseteq \mathcal{F}_{PR}$-compressible$'$.

Fix an enumeration of Turing machines $M_1, M_2, M_3, \ldots$, and view each as computing a partial recursive function $\phi_1, \phi_2, \phi_3, \ldots$ We will construct, by diagonalization, an infinite set $A \in \Delta^0_2$, that is not $\mathcal{F}_{PR}$-compressible. This will be done in stages, defining sequences A_i and w_i. A will be defined as $\bigcup_{i \geq 0} A_i$. We will ensure:

- $A_i \subseteq A_{i+1}$,
- $A^{<w_i} = A_i^{<w_i}$, and
- A_i ensures ϕ_i cannot be a compression function for A.

$\Delta_2^0 \not\subseteq \mathcal{F}_{PR}$-compressible′

Theorem

$\Delta_2^0 \not\subseteq \mathcal{F}_{PR}$-compressible′.

We start with $A_0 = \emptyset$, and $w_0 = \epsilon$. Starting with stage 1, we do the following procedure:

- Check if ϕ_i fails to be injective on $A_{i-1} \cup (\Sigma^*)^{\geq w_{i-1}}$. This is an r.e. query, since we are asking if there exists $x, y \in domain(\phi_i)$ with $\phi_i(x) = \phi_i(y)$. If we have two such strings, we can put x, y into A_i, and ensure that ϕ_i will not be injective on A, and therefore cannot be a compression function. Set $w_i = max(x, y, w_{i-1}) + 1$ and $A_i = A_{i-1} \cup \{x, y, w_{i-1}\}$.

$\Delta_2^0 \not\subseteq \mathcal{F}_{PR}$-compressible$'$

Theorem

$\Delta_2^0 \not\subseteq \mathcal{F}_{PR}$*-compressible*$'$.

- If ϕ_i is injective, we can easily make sure it is not surjective by attempting to remove a single element from the domain. We query if $\phi_i((\Sigma^*)^{\geq w_{i-1}}) \neq \emptyset$. If so, we can take a single element $x \in domain(\phi_i) \cap (\Sigma^*)^{\geq w_{i-1}}$. Set $A_i = A_{i-1} \cup \{x+1\}$ and $w_i = x + 2$. Now ϕ_i will not be surjective on A, since we know there is no y with $\phi_i(x) = \phi_i(y)$, so $\phi_i(x) \notin \phi_i(A)$.

$\Delta_2^0 \not\subseteq \mathcal{F}_{PR}$-compressible′

Theorem

$\Delta_2^0 \not\subseteq \mathcal{F}_{PR}$-compressible′.

- In the final case, $\phi_i((\Sigma^*)^{\geq w_{i-1}}) = \emptyset$. In this case, ϕ_i is only defined on finitely many inputs, so it cannot be a compression function to begin with. We set $A_i = A_{i-1} \cup \{w_{i-1}\}$ and $w_i = w_{i-1} + 1$.

Note that we added at least one element to A in each stage, so A is infinite, yet stage i ensured ϕ_i cannot be an $\mathcal{F}_{PR}$-compression function for A. So $\Delta_2^0 \not\subseteq \mathcal{F}_{PR}$-compressible′.

Rankability and Compressibility for coRE Sets

Theorem

Suppose A is co-r.e. and has an infinite r.e. subset. If A is $\mathcal{F}_{PR}$-rankable, then A is recursive.

Rankability and Compressibility for coRE Sets

Theorem

Suppose A is co-r.e. and has an infinite r.e. subset. If A is $\mathcal{F}_{PR}$-rankable, then A is recursive.

Let B be an infinite r.e. subset of A, and g a $\mathcal{F}_{PR}$-ranking function for A. Let M accept $\overline{A}$. We will give a procedure that determines if $x \in A$. By dovetailing, identify an element $y \in B$ with $x \leq y$. Then $g(y)$ exists since $y \in B \subseteq A$, and we can dovetail M on all strings $\leq y$. We know that $g(y)$ strings $\leq y$ are in A, and $y - g(y)$ are in $\overline{A}$. So once M accepts $y - g(y)$ strings, the remaining strings are all in A. Since $x \leq y$, we will have determined if $x \in A$.

Rankability and Compressibility for coRE Sets

We say A is a *cylinder* if $A \equiv_{iso} B \times \Sigma^*$.
The major property we will use is that if A is a cylinder, and $L \leq_m A$, then $L \leq_1 A$. This follows from

- If f many-one reduces L to A, then $g(x) = \langle f(x), x \rangle$ one-to-one reduces L to $A \times \Sigma^*$ (well, itself bijected back into Σ^* via any nice, fixed, standard pairing function).
- $A \times \Sigma^* \equiv_{iso} B \times \Sigma^* \times \Sigma^* \equiv_{iso} B \times \Sigma^* \equiv_{iso} A$.

Rankability and Compressibility for coRE sets

Theorem

Suppose A is co-r.e. and a nonempty cylinder. Then A is $\mathcal{F}_{REC}$-compressible.

Let E enumerate $\overline{A}$ without repetition, and $s_0, s_1, s_2, \ldots$ enumerate Σ^* in lexicog. order. We will assume $\overline{A}$ is infinite, since otherwise A is infinite and recursive, and hence $\mathcal{F}_{REC}$-compressible. Define the language $B = \{\langle x, \epsilon\rangle \mid x \in A\} \cup \{\langle x, s_i\rangle \mid i \geq 1 \wedge x \text{ is the } i^{th} \text{ string enumerated by } E\}$. Then B is $\mathcal{F}_{REC}$-compressible by projection onto the first coordinate. $A \leq_1 B$ by mapping x to $\langle x, \epsilon\rangle$. And $B \leq_m A$ by:

- Map $\langle x, \epsilon\rangle$ to x.
- For $\langle x, s_i\rangle$, check if x is the i^{th} string enumerated by E. If so, output a fixed string in A, otherwise output a fixed string not in A.

Since A is a cylinder, $B \leq_1 A$, and hence $B \equiv_{iso} A$ by the Myhill Isomorphism Theorem (which in one version states that $(\forall A, B)[A \equiv_1 B \iff A \equiv_{iso} B]$). Since B is $\mathcal{F}_{REC}$-compressible, so is A.

Rankability and Compressibility for coRE sets

Since all coRE-complete sets are nonempty co-r.e. cylinders, we have the following corollary to the results just stated on slides 27 (keeping in mind that all nonempty cylinders even have infinite recursive subsets) and 29.

Corollary

If A is coRE-complete, then

- *A is $\mathcal{F}_{REC}$-compressible, and*
- *A is not $\mathcal{F}_{PR}$-rankable.*

Rankability and Compressibility for coRE sets

Two slides ago we had this result.

Theorem

Suppose A is co-r.e. and a nonempty cylinder. Then A is $\mathcal{F}_{REC}$-compressible.

Fix a standard, nice indexing (naming scheme)—ϕ_1, ϕ_2, ϕ_3, $\cdots$—for the partial recursive functions. A set A is an *index set* exactly if there exists a (possibly empty) collection F' of partial recursive functions such that $A = \{i \mid \phi_i \in F'\}$. Since all index sets are cylinders, we have the following corollary.

Corollary

All co-r.e. index sets except $\emptyset$ are $\mathcal{F}_{REC}$-compressible.

And There Is More...

Our two papers on this, both available at arXiv.org, also explore other aspects of the recursion-theoretically compressible and rankable sets, such as: compression onto targets other than Σ^*; the fact that our results relativize; how compressibility interacts with (recursive) honesty and the semi-recursive sets; and the closure properties and the nonclosure properties of our classes with respect to boolean and other operations (e.g., none of $\mathcal{F}_{REC}$-rankable, $\mathcal{F}_{REC}$-compressible$'$, $\mathcal{F}_{PR}$-rankable, or $\mathcal{F}_{PR}$-compressible$'$, is closed under intersection).

Thank you for your time!