On Universal Computably Enumerable Prefix Codes

Cristian S. Calude and Ludwig Staiger

2007

1. Characterisation of universal c.e. prefix codes
2. Properties of universal c.e. prefix codes

Motivation

The prefix-free computable set $S = \{1^{\log n} 0 \mathrm{bin}(n) : n \geq 0\}$ codes every integer $n \geq 0$ with a string of length $2 \log n + 1$ bits (here $\mathrm{bin}(n)$ is the binary representation of $n+1$ without the leading 1).

There is prefix-free computable set T that codes every integer $n \geq 0$ with a string of length $\log n + 2 \log n \log + 1$ bits.

Is there a best way for representing integers with prefix-free computable codes?

Introduction 1

Fix an alphabet $X = \{0, \ldots, r-1\}$ and denote by X^* the set of finite strings (words) on X.

If v is a prefix of w we write $v \sqsubseteq w$. A set $V \subseteq X^*$ is prefix-free or a prefix code if for all $v, w \in V$ with $v \sqsubseteq w$ we have $v = w$.

A *self-delimiting Turing machine* (shortly, a *machine*) is a Turing machine C processing binary strings such that its program set (domain) $\mathrm{dom}(C) = \{\pi \mid \pi \in X^* \wedge \mid C(\pi) \text{ halts}\}$ is prefix-free.

A prefix code is computably enumerable (c.e.) iff it is the domain of a machine.

Introduction 2

We can effectively construct a machine U (called universal) such that for every machine C, there exists a constant k (depending only on U and C) such that for every string $\pi \in \mathrm{dom}(C)$ there exists a string $\pi' \in \mathrm{dom}(U)$ such that $U(\pi') = C(\pi)$ and $|\pi'| \leq |\pi| + k$.

A prefix-universal machine U is a machine such that for every machine C there exists a string w (depending only on U and C) such that for every string $\pi \in \mathrm{dom}(C)$ we have $U(w\pi) = C(\pi)$.

Every prefix-universal machine is universal, but the converse is not true.

Main questions

A c.e. prefix code is universal if it can code any non-negative integer in an optimal way up to a additive constant.

- Characterise universal c.e. prefix codes.
- How large are universal codes?

A characterisation of universal c.e. prefix codes 1

Theorem

Let $V \subseteq X^*$ be a c.e. prefix code. Then, the following statements are equivalent:

1. *The set V is a universal c.e. prefix code.*
2. *For every partial computable one-one function $g : \mathbb{N} \to X^*$ having a prefix-free range, there exist a partial computable one-one function $f : \mathbb{N} \to X^*$ and a constant $k \in \mathbb{N}$ such that*
   a. $f(\mathrm{dom}(f)) \subseteq V$,
   b. $\mathrm{dom}(g) \subseteq \mathrm{dom}(f)$ *and* $|f(n)| \leq |g(n)| + k$, *for every* $n \in \mathrm{dom}(g)$.

A characterisation of universal c.e. prefix codes 2

Theorem (continuation)

3. *For every computable one-one function* $g : \mathbb{N} \to X^*$ *having a prefix-free range, there exist a computable one-one function* $f : \mathbb{N} \to X^*$ *and a constant* $k \in \mathbb{N}$ *such that*
   a. $f(\mathbb{N}) \subseteq V$,
   b. $|f(n)| \leq |g(n)| + k$, *for every* $n \in \mathbb{N}$.
4. *For every c.e. prefix code* $D \subseteq X^*$ *there exist a partial computable one-one function* $\varphi : X^* \to X^*$ *and a constant* $k \in \mathbb{N}$ *such that:*
   a. $D \subseteq \mathrm{dom}(\varphi)$, $\varphi(D) \subseteq V$, *and*
   b. $|\varphi(u)| \leq |u| + k$, *for every* $u \in \mathrm{dom}(\varphi)$.

A characterisation of universal c.e. prefix codes 3

For the case $V = \mathrm{dom}(U)$, where U is a universal machine we have:

Corollary

For every c.e. prefix code $D \subseteq X^*$ and every universal machine U there are a one-one partial computable one-one function $\varphi : X^* \to X^*$ and a constant $k \in \mathbb{N}$ such that:

a. $D \subseteq \mathrm{dom}(\varphi)$, $\varphi(D) \subseteq \mathrm{dom}(U)$,

b. $|\varphi(u)| \leq |u| + k$, *for all* $u \in D$, *and*

c. $U(\varphi(u)) = u$, *for all* $u \in D$.

A characterisation of universal c.e. prefix codes 4

Theorem

Let $V \subseteq X^*$ be a c.e. prefix code. Then, the following statements are equivalent:

1. *There exists a prefix-universal machine U such that $V = \mathrm{dom}(U)$.*
2. *For every c.e. prefix code $D \subseteq X^*$ there exists a string $w \in X^*$ such that $wD = V \cap wX^*$.*

Noncomputability

Theorem

No universal c.e. prefix code $V \subseteq X^*$ is computable.

Maximality

A prefix code V is maximal if for every prefix code W, $V \subseteq W$ implies $W = V$.

Lemma

If $V \subseteq X^*$ is a c.e. maximal prefix code, then V is computable.

Corollary

No universal c.e. prefix code is a maximal prefix code.

Not every computable prefix code is contained in a computable maximal prefix code.

Sparsity

A set $W \subseteq X^*$ is called sparse if there is a polynomial $p(n)$ such that $|W \cap X^n| \leq p(n)$ for every $n \in \mathbb{N}$.

Theorem

Let $W \subseteq X^*$ be the complement of c.e. subset of X^*, and let W be non-sparse. Then, for all $0 \leq \alpha < 1$ we have: $\sum_{w \in W} r^{-\alpha \cdot H_U(w)} = \infty$.

Corollary

Let U be a universal prefix machine, let $W \subseteq X^*$ be computably enumerable or a non-sparse complement of a computably enumerable language and let $D = \{\pi : \pi \in \mathrm{dom}(U) \wedge U(\pi) \in W\}$. Then $\sum_{w \in D} r^{-\alpha \cdot |w|} = \infty$ for all $\alpha < 1$.

Entropy 1

The unique value $\mathsf{H}_W \in [0,1]$ such that $\sum_{w \in W} r^{-\alpha \cdot |w|}$ converges for all $\alpha > \mathsf{H}_W$ is known as the *entropy* of the language W. It can be calculated by the formula:

$$\mathsf{H}_W = \limsup_{n \to \infty} \frac{\log_r (|W \cap X^n| + 1)}{n}.$$

Theorem

Let $V \subseteq X^*$ be a universal c.e. prefix code. Then $\mathsf{H}_V = 1$.

Entropy 2

Theorem

Let $V \subseteq X^*$ be a universal c.e. prefix code. Then, the lower entropy of V is 1:

$$\liminf_{n\to\infty} \frac{1}{n} \log_r\left(\left|V \cap X^{\leq n}\right| + 1\right) = 1.$$