Angluin's Characterization Theorem for Effective Identifiability

Nina Gierasimczuk Dick de Jongh

February 22th, 2013

1 Uniformly recursive families

Let φ be any fixed Gödel numbering of all partial recursive functions over $\mathbb{N}$. The family $(W_j)_{j\in\mathbb{N}}$ is defined as follows: for all $j \in \mathbb{N}$, $W_j = \{n \in \mathbb{N} \mid \varphi_j(n)\downarrow\}$. Let $(L_i)_{i\in\mathbb{N}}$ be a recursively enumerable class of recursively enumerable sets.

Definition 1 *A family of languages $(L_i)_{i\in\mathbb{N}}$ is* uniformly recursively enumerable *iff there is a partial recursive function $f : \mathbb{N} \times \mathbb{N} \to \{0, 1\}$ such that $L_j = \{w \in \mathbb{N} \mid f(j, w) = 1\}$ for all $j \in \mathbb{N}$.*

Definition 2 *A family of languages $(L_i)_{i\in\mathbb{N}}$ is* uniformly recursive *iff there is a total recursive function $f : \mathbb{N}\times\mathbb{N} \to \{0, 1\}$ such that $L_j = \{w \in \mathbb{N} \mid f(j, w) = 1\}$ for all $j \in \mathbb{N}$.*

Proposition 1 *If an infinite family of languages $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$ is uniformly recursive, then there is also a uniformly recursive enumeration $(L'_i)_{i\in\mathbb{N}}$ of $\mathcal{L}$ in which each language occurs exactly once ($L'_i = L'_j \Rightarrow i = j$).*

Proof Let $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$ be uniformly recursive. $L'_0 = L_0$. L'_{n+1} is the first L_j not in $\{L'_0, \ldots, L'_n\}$ such that: L_j differs from each of $L'_0, \ldots, L'_n$ on at least one of $0, \ldots, j$. That such an L_j exists follows from the infinity of $\mathcal{L}$. Note also that an L_j may be skipped for a while because it only differs from an earlier language with regard to some large numbers. Ultimately any difference will show up of course, and therefore all languages will be part of the enumeration in the long run. □

2 Angluin's characterization

Theorem 2 (Angluin's Theorem) *A uniformly recursive family $\mathcal{L}$ is effectively identifiable iff there exists a recursive function F such that for each i, n, $F(i, n)$ is a finite set D^n_i, with $m \leq n \Rightarrow D^m_i \subseteq D^n_i$, and the limit $D_i = \bigcup_{n\in\mathbb{N}} D^n_i$ is a telltale set for L_i.*

Proof
$\Rightarrow$: Assume that M is a recursive learner that identifies $\mathcal{L}$. We have to define, for each i, a recursive sequence of increasing finite sets D_i^n that in the limit converges to a finite telltale set D_i for L_i.

Let $x_1, x_2, x_3, \ldots$ effectively enumerate L_i. We follow the idea of the locking sequence theorem to find a locking sequence (which by now we know to exist!), and then use the fact that the elements of a locking sequence form a telltale set. We follow the same procedure in always looking for an extension with a different value, and produce in that way longer and longer sequences σ_n. This time it will be an unending search, but from a certain point an extension with a different value will no longer be found so that we stabilize on a certain finite sequence, and thereby on a fixed finite set. A second difference will be that, since we want to be recursive we cannot look through all sequences at each stage; we simply restrict the search at stage n to sequences of length $\leq n$ of the elements $x_1, x_2, \ldots, x_n$, in the long run all sequences will be checked in this way.

- *Stage* 1. We start with the string $\langle x_1 \rangle$. Note that in the former proof $\langle x_1 \rangle$ denoted a sequence, in this proof it is the number coding the sequence. Now, if for some τ of length 1 containing only x_1 (there aren't too many of those!), $M(\langle x_1 \rangle^\wedge \tau) \neq M(\langle x_1 \rangle)$, take $\sigma_1 = \langle x_1 \rangle^\wedge \tau$ for the first of such sequences, otherwise take $\sigma_1 = \langle x_1 \rangle$.
- *Stage* $n+1$. Assume σ_n has been constructed in stage n. Now, if there is a sequence τ of length $\leq n+1$ containing only $x_1, x_2, \ldots, x_{n+1}$ such that $M(\sigma_n{}^\wedge \tau) \neq M(\sigma_n)$, take $\sigma_{n+1} = \sigma_n{}^\wedge \tau {}^\wedge \langle x_1, \ldots, x_{n+1} \rangle$ for the first one of such sequences, otherwise take $\sigma_{n+1} = \sigma_n$.

In each case let D_i^n be $\mathit{content}(\sigma_n)$. From the proof of the locking sequence theorem it is clear that this procedure cannot keep producing new sequences, from a given stage on the σ_i and therefore the D_i^n will remain the same. From the point that that is the case onwards we have a locking sequence σ_n for L_i and therefore $D_i^n = D_i$ will be a telltale set for L_i.

$\Leftarrow$: Assume a two-place recursive function D_i^n is given with the right properties. We define a learner in the following manner:

$M(\sigma) = \mu i \leq lh(\sigma)(D_i^{lh(\sigma)} \subseteq \mathit{content}(\sigma) \subseteq L_i)$ if such i exists, otherwise $= \mu i\, (\mathit{content}(\sigma) \subseteq L_i)$.

Let t be a text for L_i and $L_j \neq L_i$ for $j < i$. It is sufficient to show that, for n large enough, $M(t[n]) = i$. Fix $n \geq i$ such that $D_i^n \subseteq D_i \subseteq \mathit{content}(t[n])$. Then i satisfies the relevant conditions, and there are only finitely many $j < i$ satisfying the conditions. The proof now goes in exactly the same way as for the non-effective version of Angluin's theorem. $\square$

Note that in the above proof learner M is always defined if the text t is for some $L_i \in \mathcal{L}$, i.e., M is always defined if we have index-preserving identification.

In that case M is also consistent.

There are some remarks here.

1. The proof of $\Rightarrow$ above does not at all depend on the space of the hypotheses made by the recursive learner M, it always leads to the definition of a telltale set in the limit by a recursive function.
2. The proof of $\Leftarrow$ shows that we have such a telltale set in the limit by a recursive function, then we use only the numbers of the uniform enumeration in defining the learner, so $\mathcal{L}$ is exactly identifiable.

From this we can conclude that if A uniformly recursive family is identifiable, it is exactly identifiable.

References

Angluin, D. (1980). Inductive inference of formal languages from positive data. *Information and Control*, 45(2):117–135.