INDUCTIVE INFERENCE FROM NEGATIVE DATA

Shinohara, Takeshi
Computer Center, Kyushu University

INDUCTIVE INFERENCE FROM NEGATIVE DATA

By

Takeshi SHINOHARA*

Abstract

Inductive inference of a language L from negative data is the one based only on words not in L. In other words it is the inference of the complement L^c from positive data. This paper describes the relation between inferability from positive data and that from negative data. It is not the case that the inferability of a class of languages always guarantees the inferability from negative data. Some nontrivial classes are inferable both from positive data and from negative data.

1. Preliminaries

We start with a brief review of inductive inference from positive data according to Angluin [1, 2].

Let Σ be a finite alphabet of symbols. The set of all finite strings over Σ is denoted by Σ^*. A *language* is a subset of Σ^*.

DEFINITION 1. A class of lanuages $\mathcal{L} = L_1, L_2, \cdots$ is said to be an *indexed family of recursive languages* if there exists a computable function f such that

$$f(i, x) = \begin{cases} 1, & \text{if } x \text{ is in } L_i; \\ 0, & \text{otherwise.} \end{cases}$$

From here on, the classes of languages are assumed to be indexed families of recursive languages.

DEFINITION 2. A *positive presentation* of a nonempty language L is an infinite sequence $s = s_1, s_2, \cdots$ such that the set of all strings in s is equal to L.

DEFINITION 3. A *negative presentation* of a language L not equal to Σ^* is an infinite sequence $s = s_1, s_2, \cdots$ such that the set of all strings in s is euqal to the complement $\Sigma^* - L$.

DEFINITION 4. *Inference machine* M is an effective procedure that requests input from time to time and produces output from time to time. M *on input sequence* $s = s_1, s_2, \cdots$ *converges* to g_0 iff the sequence of outputs $g = g_1, g_2, \cdots$ produced by M when elements in s are successively given to M is either a finite sequence ending with g_0 or an infinite sequence such that all but finitely many elements in g are equal to g_0.

DEFINITION 5. A class of languages $\mathcal{L} = L_1, L_2, \cdots$ is said to be *inferable from positive data (inferable from negative data)* if there exits an inference machine M such

---

* Computer Center, Kyushu University 91, Fukuoka, 812 Japan.

that M on input s converges to i for any i and any positive (negative) presentation s of L_i.

The following theorem essentially characterizes the classes inferable from positive data.

THEOREM 1. [2]   *A class $\mathcal{L}$ of nonempty languages is inferable from positive data if and only if there exists an effective procedure P which enumerates elements in T_i for any index i such that*

   1) *T_i is finite,*
   2) *$T_i \subset L_i$, and*
   3) *$T_i \subset L_j \Rightarrow \neg(L_j \subsetneqq L_i)$ for all j.*

The right part of Theorem 1 is called Condition 1.

COROLLARY 2. [2]   *If $\mathcal{L}$ is a class of languages such that $C(F) = \{L \in \mathcal{L} \mid F \subset L\}$ is finite for any finite language F then $\mathcal{L}$ is inferable from positive data.*

The assumption part of Corollary 2 is called Condition 2. Although Condition 2 is not necessary for inferability from positive data, it is convenient because it is independent of indexing. In fact, many classes containing the pattern languages of Angluin are shown to be inferable from positive data by using Corollary 2.

2. Inductive Inference from Negative Data

Let $L^c = \Sigma^* - L$ and $\mathcal{L}^c = L_1^c, L_2^c, \cdots$. Then, clearly from definitions, a class $\mathcal{L}$ of languages not equal to Σ^* is inferable from negative data if and only if $\mathcal{L}^c$ is inferable from positive data. Thus we can apply Theorem 1 or Corollary 2 to the class of the complements to show that a class is inferable from negative data.

The first question we may ask is whether a class of languages which is inferable from positive data is still inferable from negative data. The following theorem answers in the negative to this question.

THEOREM 3.   *$\mathcal{F} = \{L \subset \Sigma^* \mid L$ is nonempty and finite$\}$ is inferable from positive data but not inferable from negative data.*

PROOF.   It is clear that $\mathcal{F}$ is inferable from positive data.

Assume that $\mathcal{F}$ is inferable from negative data. Then $\mathcal{F}^c$ is inferable from positive data. By Theorem 1, for any $L_i^c = \Sigma^* - L_i$, there exits T_i such that

   1) T_i is finite,
   2) $T_i \subset L_i^c$, and
   3) $T_i \subset L_j^c \Rightarrow \neg(L_j^c \subsetneqq L_i^c)$ for all j.

Let T_i be any set satisfying both 1) and 2). Then $L_i^c - T_i$ is nonempty. For any element x in $L_i^c - T_i$, $L_i \cup \{x\}$ is finite and there exits j such that

$$L_j = L_i \cup \{x\}, \quad \text{and} \quad T_i \subset L_j^c \subsetneqq L_i^c.$$

Therefore T_i can not satify 3).   □

The next question should be this. Does there exist a class that is inferable both from positive data and from negative data? The answer is clearly "yes". Any class which consists of finitely many languages is trivially inferable both from positive data and from negative data. As already mentioned, many classes inferable from positive

data satisfy Condition 2. The following theorem shows that such classes are also inferable from negative data.

THEOREM 4. *A class of languages not equal to Σ^* which satisfies Condition 2 is inferable from negative data.*

PROOF. Let $\mathcal{L}$ be any class of languages which satisfies Condition 2. It sufficies to show that $\mathcal{L}^c$ satisfies Condition 1, that is, there exists an effective procedure which enumerates the elements in T_i for any i such that

1) T_i is finite,
2) $T_i \subset L_i^c$, and
3) $T_i \subset L_j^c \Rightarrow \neg(L_j^c \subsetneqq L_i^c)$ for all j.

Consider the following procedure, where we regard additions to the set A as the enumeration by this procedure.

> **begin**
> $u := $ any element in L_i
> $A := $ empty
> **loop**
> > **begin**
> > find (j, x) such that
> > > $u \in L_j$ and
> > > $A \subset L_j^c$ and
> > > $x \in L_i^c - L_j^c$
> > **if** found **then** $A := A \cup \{x\}$
> > **end**
>
> **end**

Since we can find out all the pairs (j, x) of index j and string x, the procedure is obviously effective.

For input i, let T_i be the set enumerated by this procedure in the limit. It is clear that T_i satisfies 2).

Now we show that 1) holds for T_i. Let u be the element in L_i selected at the first stage in the procedure. Then the number of languages in $\mathcal{L}$ containing u is finite, because $\mathcal{L}$ satisfies Condition 2, that is,

$$C(\{u\}) = \{\mathcal{L} \in L \mid u \in L\}$$

is finite. The number of indexes j such that $u \in L_j$ may be naturally infinite. However, once a pair (j, x) is found, the string x is added to A. Then as L_j^c does not include A, any pair (j', x) with $L_{j'} = L_j$ can not be found. Therefore T_i is finite.

Finally we show that 3) is satisfied. Assume that 3) is violated. Then, for some index j,

$$T_i \subset L_j^c \quad \text{and} \quad L_j^c \subsetneqq L_i^c (\Rightarrow L_i \subsetneqq L_j).$$

Since u belongs to L_i,

$$u \in L_j.$$

Therefore a pair (j, x) should be found for some string x not in T_i, and then x must be enumerated. This contradicts our assumption that T_i is the set enumerated by the

procedure in the limit.   □

3.   Inference from Positive Data with One Negative Data

In this section, we discuss the problem which directly motivates the present work. Does inference from positive data with one negative data have more power than inference from positive data? Hence if we can solve this problem, we can reveal how the negative data contributes to the inference.

DEFINITION 6.   A *positive presentation with one negative data* of a language L is an infinite sequence s_1, s_2, $\cdots$ such that $s_1 \notin L$ and the set of all strings in s but s_1 is just L. Inference by using positive presentations with one negative data is called *inference from positive data with one negative data*.

It is well-known that inference only from positive data has strictly less power than inference from positive and negative data [3]. Another natural question is whether the differnce in the number of available negative data makes a hierarchy of inference powers.

Assume that there exists a class $\mathcal{L}$ of languages which satisfies Condition 2 but is not inferable from negative data. Then the class $\mathcal{L}^c$ of the complements is not inferable from positive data but it is inferable from positive data with one negative data, because the number of languages in $\mathcal{L}^c$ which does not contain a string given by negative data is finite. Unfortunately our main result (Theorem 4) says that any class of languages satisfying Condition 2 is always inferable from negative data. Thus our problems on the contribution of negative data to inference remains open.

References

[1]   ANGLUIN, D.: *Finding Patterns Common to a Set of Strings.* Proceedings, 11th Annual Symposium on Theory of Computing, (1979), 130–141.
[2]   ANGLUIN, D.: *Inductive Inference of Formal Languages from Positive Data.* Inform. Contr., **45** (1980), 117–135.
[3]   GOLD, E. M.: *Language Identification in the Limit.* Inform. Contr., **10** (1967), 447–474.