

質問学習における学習可能性の統一的特徴づけ

3T-10

林 洋祐[†] 松本 哲志[‡] 篠原 歩[†] 竹田 正幸[†][†]九州大学大学院システム情報科学研究科 [‡]東海大学理学部

1. はじめに

大量のデータからそれを説明する規則を計算機によって見つけ出すためのアプローチの1つとして、機械学習による規則性の発見が有用であると考えられる。Angluinは、目標概念に関する質問とその応答から、概念クラス中の目標概念を厳密に同定する質問学習のモデルを提案した¹⁾。このモデルにおいて、さまざまな概念に対する学習アルゴリズムを開発する研究が数多くなされ、一方で多項式回の質問から学習できる概念クラスを特徴づける研究も進められている。特に所属性、等価性の2種類の質問に関しては、それぞれを単独で用いた場合^{2, 3, 4, 6)}、および両方を組み合わせた場合^{6, 7)}について、その同値条件が個別に調べられている。

本論文では、任意の形式の質問の、任意の組合せを用いた質問学習において、その学習可能性を統一的に特徴づける組合せ的性質を示す。直観的には、学習の容易さは「良い質問」の存在に依存すると考えられる。良い質問とは、いかなる返答に対しても仮説空間が十分に絞り込めるものを言う。我々はこれを δ -goodな質問として定式化し、目標概念のサイズが既知である場合には、 δ -goodな質問の存在が概念クラスの学習可能性の必要十分条件を与えることを示す。また、目標概念のサイズが未知の場合に対しては、現在の仮説が正しいことを学習者自身が質問によって検証する必要があるが、我々はこれを一意特定質問集合として定式化し、上記の δ -goodな質問の存在と合わせて、学習可能性の必要十分条件を与えることを示す。このことによって、質問の形式や、その組み合わせ方によらず、多項式回の質問を用いて学習可能な概念クラスを統一的に完全に特徴づけることができる。この特徴づけは、等価性質問のみを使うことを考えた場合には、Angluinの示した approximate fingerprint property²⁾と等価となる。

Uniform Characterizations of Learnabilities via Polynomial Numbers of Queries.

Y. Hayashi[†], S. Matsumoto[‡], A. Shinohara[†] and M. Takeda[†]
[†]Department of Informatics, Kyushu University, Fukuoka, 812-8581 Japan

[‡]Faculty of Science, Tokai University, Kanagawa, 259-1291 Japan

2. 定義

Σ を文字の有限集合とし、 Σ^* を Σ 上の有限長の文字列すべての集合、 Σ^n を Σ 上のちょうど n 文字からなる文字列すべての集合とする。さらに文字列 $w \in \Sigma^*$ に対して $\|w\|$ をその文字列長とし、集合 S に対して $|S|$ はその要素数を表す。

Σ^* の部分集合を概念と呼び、 $C \subseteq 2^{\Sigma^*}$ を概念クラスと呼ぶ。概念を文字列で表現する方法を仮定し、その長さを概念のサイズと言う。本論文では、 $c \subseteq \Sigma^n$ となる概念のみを対象とし、 $C_n = \{c \in C \mid c \subseteq \Sigma^n\}$ 、 $C_{m,n} = \{c \in C_n \mid c \text{の表現長は高々 } m\}$ と定義する。さらに概念 c に対して、質問 σ をおこなった時の c に関する答えの集合を $c[\sigma]$ と表す。例えば、所属性質問 σ_1 に対しては、 $c[\sigma_1]$ は{"yes"}または{"no"}であり、等価性質問 σ_2 に対しては、 $c[\sigma_2]$ は{"yes"}または反例全体からなる集合となる。

定義 1. Q を質問の集合とする。あるアルゴリズム A と多項式 $p(\cdot, \cdot)$ が存在して次が成り立つ時、概念クラス C は Q を使って多項式質問で学習可能であると言う。

任意の自然数 m, n と未知の目標概念 $c^* \in C_{m,n}$ に対して、まず A が n を入力として受けとり、 Q 中から教師に質問を行ない答をもらう。質問を繰り返し、最終的に A は停止した後、 c^* と等価な概念の表現を出力する。このとき要する質問と答の長さの合計が高々 $p(m, n)$ である。

3. Good Queries

この章では、目標概念のサイズ m が既知である場合の学習可能性について考える。また次章も含めて、ここで与える定理の証明については文献⁵⁾を参照されたい。

定義 2. 概念クラス T と質問 σ およびそれに対する答え α に対して、 $Cons(T, \sigma, \alpha)$ を T 中で σ と α に無矛盾な概念の集合とする。すなわち、

$$Cons(T, \sigma, \alpha) = \{h \in T \mid \alpha \in h[\sigma]\}.$$

定義 3. 質問 σ が概念クラス T に対して δ -goodであるとは、任意の答え α に対して次の式が成り立つことをいう。

$$|Cons(T, \sigma, \alpha)| \leq (1 - \delta)|T|.$$

Algorithm LEARNER1 (m, n : 自然数)

Given Q : 利用可能な質問集合

begin

$H := C_{m,n}$;

while $|H| \geq 2$ do

H に対して $1/q(m, n)$ -good である質問 $\sigma \in Q$
を見つける;

α を σ に対する答とする;

$H := \text{Cons}(H, \sigma, \alpha)$

endwhile;

if $|H| = 1$ then H 中の唯一の仮説 h を出力

else “ $C_{m,n}$ 中に目標概念は存在しない” と出力

end.

図 1 アルゴリズム LEARNER1 (m は既知)

定理 1. 学習者が目標概念のサイズ m を知っているものと仮定する。このとき、概念クラス C が多項式質問で学習可能であることは、次の条件が成り立つことと同値である。

ある多項式 $p(\cdot, \cdot), q(\cdot, \cdot)$ が存在し、任意の自然数 m, n と任意の $T \subseteq C_{m,n}$ ($|T| \geq 2$) に対して、 $\frac{1}{q(m,n)}$ -good である質問 $\sigma \in Q$ ($\|\sigma\| \leq p(m, n)$) が存在する。

この証明に用いる学習アルゴリズムを図 1 に示す。

4. 一意特定質問集合

この章では、目標概念のサイズ m が未知である場合について扱う。

定義 4. 概念集合 T と概念 $c \in T$ に対して、次の条件を満たす質問の集合 Q を一意特定質問集合と言う。

$$\{h \in T \mid h[\sigma] = c[\sigma] \text{ for all } \sigma \in Q\} = \{c\}.$$

すなわち、 Q 中の質問とその答えによって、 c は T の中から一意に特定される。

定理 2. 概念クラス C が Q を使って多項式質問で学習可能であることは、ある多項式 $p(\cdot, \cdot), q(\cdot, \cdot), r(\cdot, \cdot)$ が存在し、任意の自然数 m, n に対して、以下の条件がともに成り立つことと同値である。

- (1) 任意の $T \subseteq C_{m,n}$ ($|T| \geq 2$) に対して、 $\frac{1}{q(m,n)}$ -good である質問 $\sigma \in Q$ ($\|\sigma\| \leq p(m, n)$) が存在する。
- (2) 任意の概念 $c \in C_n$ に対して、 C_n 中での c に対する一意特定質問集合 $Q \subseteq Q$ ($\|Q\| \leq r(m, n)$) が存在する。

この証明に用いる学習アルゴリズムを図 2 に示す。

Algorithm LEARNER2 (n : 自然数)

Given Q : 利用可能な質問集合

begin

$m = 1$;

repeat

Q を使って LEARNER1(m, n) を模倣する;

if LEARNER1 が仮説 h を出力 then

Q を C_n 中での h に対する特定質問集合とする;

if h が Q 中のすべての質問に対する答えに

無矛盾である then

h を出力し、停止

$m = m + 2$

forever

end.

図 2 アルゴリズム LEARNER2 (m は未知)

5. おわりに

本論文では、 $c \subseteq \Sigma^n$ となる概念クラスのみを対象とし、多項式質問で学習可能な概念クラスの特徴づけを行なった。 Σ^* の部分集合である一般の概念に対して学習可能な概念クラスの特徴づけを行なうことが今後の課題である。

参考文献

- 1) D. Angluin. Queries and concept learning. *Machine Learning*, 2:319-342, 1988.
- 2) D. Angluin. Negative results for equivalence queries. *Machine Learning*, 5:121-150, 1990.
- 3) R. Gavaldà. On the power of equivalence queries. In *Proceedings of the 1st European Conference on Computational Learning Theory*, pp. 193-203, 1993.
- 4) S. Goldman and M. Kearns. On the complexity of teaching. *Journal of Computer and System Sciences*, 50:20-31, 1995.
- 5) Y. Hayashi, S. Matsumoto, A. Shinohara, and M. Takeda. Uniform characterizations of polynomial-query learnabilities. In *Proceedings of the First International Conference on Discovery Science*, 1998. to appear.
- 6) T. Hegedüs. Generalized teaching dimensions and the query complexity of learning. In *Proceedings of the 8th Annual Conference on Computational Learning Theory*, pp. 108-117, 1995.
- 7) L. Hellerstein, K. Pillaipakkamnatt, V. Raghavan, and D. Wilkins. How many queries are needed to learn? *Journal of the ACM*, 43(5):840-862, 1996.