関数の極限ラムダ計算による学習

Learning Lambda-Terms in Limit

赤間 陽二 †
Yohji Akama
† 東北大学理学研究科数学専攻
Mathematical Institute, Tohoku University
akama@math.tohoku.ac.jp

計算可能関数を学習の対象とする学習理論を手本にして, ラムダ式を学習の対象とする学習理論を構築を試みた. (1)Blum の抽象計算量をラムダ計算に導入し, (2) 計算可能関数の学習理論におけるクラス $\mathcal{NV}$ の $\mathcal{EX}$ がラムダ計算にも道理にかなった方法で定義できることを認識し, (3) ラムダ計算における $\mathcal{NV}$ を, ラムダ計算における従順性の概念により特徴づけ, (4) ラムダ計算における $\mathcal{NV}$ と $\mathcal{EX}$ がそれぞれ合併定理と否合併定理を満たすことをを証明した

1 はじめに

コンピュータ/ネットワークの重要なシステム設定の多くは小さなスクリプトであり, それらを自動的に学習するシステムは有用性が高い. また, コンピュータ/ネットワークの管理においては単純な情報の連続であるログの傾向からシステムの anomaly を検出するが, そこにも学習理論が有用ではないかと考えられる [1].

学習の対象とするスクリプトとして, 関数型プログラムで型のないものを考えた場合, それにより若干の実際的な貢献ができないかと思い, 関数型プログラムの理論的基盤であるラムダ式を学習の対象とすることを考えた. 計算可能関数を学習の対象とする従来の学習理論が素直に展開できることがわかったので報告する.

まずラムダ計算の独自の性質である range property に基づき, 学習の対象とするラムダ式は入力が本質的に非負整数に限ることを導いた. 次ぎに, 計算可能関数の学習理論で援用される Blum の抽象計算量を型のないラムダ計算に移入した. そこではラムダ式の計算量として正規化列の長さの最小値を考えた (2 節).

次ぎに, 学習理論における基本的な学習可能なオブジェクトのクラスである $\mathcal{NV}$ と $\mathcal{EX}$ をラムダ計算に移入し, 移入した後でもそれらの基本的な性質である合併定理と非合併定理が成立することを証明した (3 節, 4 節). 関数の学習理論における $\mathcal{NV}$ と $\mathcal{EX}$ の定義および (否) 合併定理の証明は文献 [10] に基づいた.

1.1 関連研究

プログラムを学習するために, 型つきラムダ式の高階 unification や matching を考えるアプローチは萩谷などが有名であり, 独自のユーリスティクスの考案が盛んであった. 例えば, 萩谷はさらに Curry-Howard isomorphism と組み合わせて Proving-by-example を提唱した [7]. 一方, 本稿では, 学習対象として, Blum の抽象計算量と相性のよい型のないラムダ式を選んだことにより階層を見る端緒を与えている.

形式証明の試験を企図した林 [8] は, 極限計算可能関数の学習理論における重要性を認識し, 形式証明の実現可能性解釈を極限計算可能関数によってを行ない, 形式証明の試験を学習理論を用いようとしている.

学習理論をソフトウェア工学や形式手法に援用する試みは多く, 本稿の動機もそこに近い.

[1] では学習理論における極限操作に関心を絞り, ラムダ計算／プログラム理論的な考察を与え, 極限ラムダ計算を提案している.

また, [2] において $\forall x.\exists y.\, P(x,y)$ の半構成的証明から, monotone modified realizability interpretation により, x に対する y の上界を与える計算可能関数 $f(x)$ やラムダ式が得られるが, f の度合を調べるために, Blum の抽象計算量理論における “従順さ” や

[1]実際, CFengine はシステムの望ましい状態を記述した小さなスクリプトをコンピュータ/ネットワークに (確率的に) 学習させるというアプローチを取り, 大規模なシステムの管理に使われている

“穏健さ” で計れると考えた.

2 準備

Λ^0 で自由変数のない型の無いラムダ式 (閉ラムダ式) 全体を表すものとする. 非負整数 i のチャーチニューメラルを $\underline{i} \in \Lambda^0$ で表す.

1 引数計算可能部分関数の枚挙 $\phi_0, \phi_1, \ldots$ に対応して, 閉ラムダ式 $\mathbf{E}$ が存在して

$$\forall M \in \Lambda^0 . \, \mathbf{E} \llcorner M \lrcorner \twoheadrightarrow_{\beta\eta} M$$

であることを思い出す ([9],[3, Theorem 8.1.6]). ここで $\llcorner M \lrcorner$ は M のゲーデル数を表すチャーチニューメラルである.

また再帰定理に対応して

$$\forall F \, \exists e \in \mathbb{N} \forall x \in \mathbb{N}. \, \mathbf{E} \ \underline{e} \ \underline{x} =_{\beta\eta} F \ \underline{e} \ \underline{x} \tag{1}$$

が成立することに注意する.

関数の学習理論で推定の対象とするのは,

$$R = \{ f : \mathbb{N} \to \mathbb{N} \mid f \text{ は 1 引数計算可能関数 }^2 \}$$

であったが, 本稿では推定の対象とするラムダ式は,

定義 2.1

$$\mathcal{R} \ := \ \{ N \in \Lambda^0 \mid \forall i \in \mathbb{N}. \, N \ \underline{i} \text{ は正規形を持つ} \}.$$

とする. すなわち、推定の対象は関数型プログラム M で, 勝手な非負整数を入力されると停止する高階関数と考えられる. 入力を非負整数に限ったことにより, 一般性は著しく失われないと考えるが [3], 将来は入力として, 予め定めたデータ構造の勝手なデータとしたい.

付言すると「全ての閉じた正規形 L に対して NL は正規形を持つ」とすると, NL は同一の正規形を持ってしまうため, 意味のある議論が展開できなくなる. それは range property [3, p. 517] による: 「全ての $N \in \Lambda^0$ に対して, 商集合 $\{NL \mid L \in \Lambda^0\}/=_{\beta\eta}$ は単元集合か商集合 $\{L \mid L \in \Lambda^0\}/=_{\beta\eta}$ のいずれかである.」

計算可能関数の学習理論においては, 関数 $f \in R$ に対して機械に $f(0), f(1), \ldots$ を入力するにつれて非負整数列を出力し続け, その出力列の漸近的挙動や極限が f の性質を表す. それに対応して記法を導入する: $F \in \mathcal{R}$ に対して $\hat{F} \in \mathcal{R}$ が存在して, 勝手な非負整数 i までの提示をリストとして返す:

$$\hat{F} \ \underline{i} \ =_{\beta\eta} \ [F \ \underline{0}, \ F \ \underline{1}, \ldots, F \ \underline{i-1}]$$

ここで閉ラムダ式を用いて, リスト, リストの長さを返す関数 $\underline{length}$, また, $i \in \mathbb{N}$ 番目の要素をリスト x から取り出す関数 $\underline{nth}$ を作ることができる.

$\sigma \in R$ による閉ラムダ式の効果的 “列挙” を

$$\mathcal{S}_\sigma := \{ N \in \Lambda^0 \mid N \ =_{\beta\eta} \ \mathbf{E} \ \underline{\sigma(i)} \ (\exists i \in \mathbb{N}) \}$$

で表す.

2.1 Blum の抽象計算量

Blum の抽象計算量は部分計算可能関数のためのものであり, Blum の加速定理等を導き (例えば [4]), 一方で学習理論で援用される. 本稿では型の無いラムダ計算のためのものに拡張する.

まず $\beta\eta$-reduction tree の幅優先探索により次の定理が成立する:

定理 2.2 閉ラムダ式 **cost** が存在して全ての非負整数 i, j と全ての $M \in \Lambda^0$ に対して次が成立する.

$M \ \underline{i}$ は長さ j 以内の $\beta\eta$-正規化列 を持つならば,

$$\mathbf{cost} \ \llcorner M \lrcorner \ \underline{i} \ \underline{j} \ =_{\beta\eta} \ \underline{1}.$$

さもなければ $\mathbf{cost} \ \llcorner M \lrcorner \ \underline{i} \ \underline{j} \ =_{\beta\eta} \ \underline{0}$.

$M \in \Lambda^0$ の入力 $\underline{i}$ $(i \in \mathbb{N})$ に対する計算の cost $(\leq \infty)$ を表すために次の定義を行なう.

定義 2.3 $\mathbf{\Phi} \in \Lambda^0$ を

$$\mathbf{\Phi} =_{\beta\eta} \lambda m \lambda i. \, \underline{\min \ j \in \mathbb{N} \ s.t. \ \mathbf{cost} \ m \ i \ \underline{j} \ =_{\beta\eta} \ \underline{1}}.$$

すると,

事実 2.4 全ての $h \in R$ に対して適当な非負整数 i が存在して $\mathbf{E} \ \underline{i} \in \mathcal{R}$ かつ有限個を除く全ての非負整数 x に対して $\mathbf{\Phi} \ \underline{i} \ \underline{x} \ > \ \underline{h(x)}$.

実際 $\mathbf{E} \ \underline{i} \ \underline{x} \ =_{\beta\eta} \ \underline{2^{h(x)}}$ とすればよい.

3 NV: 次の値による推定

定義 3.1 (NV) $M \in \Lambda^0$ は推定機械であるとは, 勝手な $F \in \mathcal{R}$ に対して $\lambda i. \, M \ (\hat{F} \ i) \in \mathcal{R}$ とする.

[3] 入力を時刻とすればプログラムは時間発展を表すもの (例えばログ) となり自然である. また, RPC でもデータ構造は xdr でビットシーケンスに変換されて送受信される.

推定機械 M が $N \in \mathcal{R}$ を推定するとは, 有限個を除く全ての非負整数 i に対して

$$M\ (\hat{N}\ \underline{i}) =_{\beta\eta} N\ \underline{i} \tag{2}$$

と定義する.

$\mathrm{NV}(M)$ で, 推定機械 M が推定する $N \in \mathcal{R}$ 全体を表すものとする. $=_{\beta\eta}$ について閉じていることに注意する.

また, 適当な推定機械 M で $\mathcal{F} \subseteq \mathrm{NV}(M)$ となる $\mathcal{F}$ 全体を $\mathcal{NV}$ で表す.

計算可能関数の学習理論における $\mathcal{NV}$ は $\mathcal{R}$ を R に置き換えたものである.

$N \in \mathrm{NV}(M)$ ならば N の出力の履歴から次の値を推定できる. 例えば, 今までの ログ情報から次に起こりそうなことが推定できる.

Blum の抽象計算量による $\mathcal{NV}$ の特徴づけを与える.

定義 3.2 (**h 従順**) $h \in R$ とする. $N \in \mathcal{R}$ が h 従順とは, 適当な非負整数 i が存在して,

$$\mathbf{E}\ \underline{i}\ \underline{n}\ =_{\beta\eta}\ N\ \underline{n}$$

かつ有限個を除く全ての非負整数 n に対して

$$\mathbf{cost}\ \underline{i}\ \underline{n}\ \underline{h(n)}\ =_{\beta\eta}\ \underline{1}$$

$\mathcal{F} \subseteq \mathcal{R}$ は h 従順とは, $\mathcal{F}$ の全ての要素が h 従順の時を言う.

定理 3.3 (**$\mathcal{NV}$ の従順性による特徴づけ**) *1.* 全ての $h \in R$ に対して推定機械 M が存在して

$$\forall N \in \mathcal{R}. \big(N \text{ は } h \text{ 従順} \Rightarrow M \text{ は } N \text{ は推定.}\big)$$

2. 全ての 推定機械 M に対して $h \in R$ が存在して

$$\forall N \in \mathcal{R}. \big(M \text{ は } N \text{ は推定} \Rightarrow N \text{ は } h \text{ 従順.}\big)$$

この定理の帰結についてだが, 事実 2.4 より

系 3.4 $\{\mathbf{\Phi}\ \underline{i} \mid \mathbf{E}\ \underline{i} \in \mathcal{R}$ かつ $i \in \mathbb{N}\} \notin \mathcal{NV}$.

が従う. また $\mathcal{F}_i$ を h_i 従順とすると合併は $\max(h_1, h_2)$ 従順であるから, $\mathcal{NV}$ の合併定理が従う.

系 3.5 (**合併定理**) $\mathcal{F}_i \in \mathcal{NV}$ ならば $\mathcal{F}_1 \cup \mathcal{F}_2 \in \mathcal{NV}$.

定理 3.3(1) の証明は, M として手続きかた言語で書かれたものを accumulating parameter c を持つ関数型プログラムに変換すればよい. ここで $\langle \bullet, - \rangle$ を $\mathbb{N} \times \mathbb{N}$ から $\mathbb{N}$ への計算可能な全単射とする. M として $M'\ \underline{0}$ とすればよい.

$$
\begin{array}{llcl}
\multicolumn{4}{l}{M'\ c\ Ns} \\
= & & & \\
let & \langle i, k \rangle & = & c, \\
 & n & = & \underline{length}\ Ns \\
in & & & \\
\ \ if & n & = & \underline{1}\ then\ \underline{0} \\
\ \ elseif & \forall x < n. & & \Big(\mathbf{cost}\ \underline{i}\ \underline{x}\ \underline{\max(k, h(x))} \\
 & and & & \mathbf{E}\ \underline{i}\ \underline{x}\ =_{\beta\eta}\ \underline{nth}\ \underline{x}\ Ns\Big) \\
\ \ then & & & \mathbf{E}\ \underline{i}\ \underline{n} \qquad //\quad (\dagger) \\
\ \ else & & & M'\ \underline{c+1}\ Ns
\end{array}
$$

定理 3.3(2) の証明のために補題を準備する.

補題 3.6 全ての推定機械 M に対して $\sigma \in R$ が存在して $\mathrm{NV}(M) = \mathcal{S}_\sigma$.

証明. 非負整数 n に対して $\mathcal{F}_n$ で $\left\{\tilde{f} \in \Lambda^0 \mid \{i \mid \tilde{f}\ \underline{i} \text{ は正規形を持つ }\} = \{0, 1, \ldots, n\}\right\}$ を表す. するとその効果的枚挙が存在する. すなわち, 適当な $\tau \in R$ に対して (i) と (ii) が成立する. (i) 全ての非負整数 i について $f \in \mathcal{F}_n$ が存在し, $\mathbf{E}\ \underline{\tau(i)}\ =_{\beta\eta}\ f$. (ii) 全ての $f \in \mathcal{F}_n$ について非負整数 i が存在し, $\mathbf{E}\ \underline{\tau(i)}\ =_{\beta\eta}\ f$.

M を用いて $\tilde{f}_j$ を $f_i \in \mathcal{R}$ に拡張する. ここで $i = \langle j, n \rangle$ とする.

$$f_i\ \underline{x}\ =_{\beta\eta} \begin{cases} \tilde{f}_j\ \underline{x} & (0 \leq x \leq n); \\ M\ \hat{f}_i\ \underline{x} & (o.w.) \end{cases}$$

明らかに $f_i \in \mathrm{NV}(M)$. 一方 $g \in \mathrm{NV}(M)$ ならば $g = f_i$ とかける. 従って適当な $\sigma \in R$ に対して全ての非負整数 i に対して $\mathbf{E}\ \underline{\sigma(i)}\ =_{\beta\eta}\ f_i$. Q.E.D.

補題 3.7 $\sigma \in R$ が $\mathcal{S}_\sigma \subseteq \mathcal{R}$ ならば適当な $h \in R$ に対して $\mathcal{S}_\sigma$ は h 従順.

証明. $\mathcal{S}_\sigma \subseteq \mathcal{R}$ より

$$h(x) := \max\{\mathbf{\Phi}\ \underline{\sigma(i)}\ \underline{x} \text{ が表す非負整数} \mid i \leq x\}$$

は R に属し $\mathbf{E}\ \underline{\sigma(i)}$ は h 従順. Q.E.D.

上の二つの補題から定理 3.3(2) が従う.

4 EX: 説明による学習

学習理論の大きなテーマの一つである「説明による学習」の枠組を簡単にふりかえり, ラムダ計算による定式化をまず与える.

帰納的推論機械 M とは

1. 入力は適当な関数 $f \in R$ のグラフの列挙 $\langle x_0, f(x_0)\rangle, \langle x_1, f(x_1)\rangle, \ldots$ である.
2. 何も入力しない状態で $g_0 \in \mathbb{N}$ を出力する. すなわち $M\ [\,]$ は g_0. 入力から $\langle x, f(x)\rangle$ を一個を読む毎に M は高々一個の出力を行なう. M の $n{+}1$ 番目の出力を $g_n \in \mathbb{N}$ と書く. 入力列 $d_0, \ldots, d_n$ を読み込み, d_{n+1} を読み込む直前の M の出力列が $g_0, \ldots, g_m$ ならば $M\ [d_0, \ldots, d_n] = g_m$ と書く. 無限列 $d_0, d_1, \ldots,$ を読みこんだ時, M の出力列が $g_0, g_1, \ldots, g_m, g_m, \ldots$ ならば, $M\ [d_0, \ldots, d_n, \ldots] = g_m$ と書き, それ以外の時は $M\ [d_0, \ldots, d_n, \ldots] = \perp$ と書く.

$f \in R$ が帰納的推論機械 M で同定可能とは, 適当な $g \in \mathbb{N}$ に対して $\phi_g = f$ かつ $M\ \left[\langle x_0, f(x_0)\rangle, \langle x_1, f(x_1)\rangle, \ldots\right] = g$ の時である.

全ての帰納的推論機械 M は順序独立であることが知られている. すなわち, $f \in R$ のグラフのどのような順序の列挙 d に対しても $M\ d$ は不変である.

定義 4.1 (EX) 推定機械 M が $N \in \mathcal{R}$ を同定するとは, 適当な $g \in \mathbb{N}$ が存在して、有限個を除く全ての非負整数 i に対して

$$M\ (\hat{N}\ \underline{i}) =_{\beta\eta} \underline{g}, \quad \text{かつ} \quad \mathbf{E}\ \underline{g} \twoheadrightarrow_{\beta\eta} N$$

とする.

$\mathrm{EX}(M)$ で, M が推定する $N \in \mathcal{R}$ 全体を表すものとする. $=_{\beta\eta}$ について閉じていることに注意する.

また, 適当な推定機械 M に対して $\mathcal{F} \subseteq \mathrm{EX}(M)$ となる $\mathcal{F}$ 全体を $\mathcal{EX}$ で表す.

4.1 枚挙手法

定理 4.2 $\sigma \in R$ かつ次ぎの *Turing* 機械が存在すると仮定する:

- 入力: 勝手な $i, x \in \mathbb{N}$, および閉じた勝手な正規形 y.
- 出力: $\mathbf{E}\ \underline{\sigma(i)}\ \underline{x}\ =_{\beta\eta} y$ の真理値

このとき

$$\mathcal{S}_\sigma \subseteq \mathrm{EX}(M'\ \underline{1})$$

ただし

$$\begin{array}{llll}
M'\ i\ Ns & & & \\
= & & & \\
let & n & = & \underline{length}\ Ns \\
in & & & \\
\ \ if & n & = & \underline{1}\ then\ \underline{0} \\
\ \ elseif & \forall x < n. & & \mathbf{E}\ \underline{\sigma(i)}\ \underline{x}\ =_{\beta\eta}\ \underline{nth}\ \underline{x}\ Ns \\
\ \ then & & & \underline{\sigma(i)} \\
\ \ else & & & M'\ \underline{i+1}\ Ns
\end{array}$$

$i \in \mathbb{N}$ に $\mathbf{\Phi}\ \underline{i}$ のゲーデル数を対応させる関数 τ は計算可能だから, $\{\mathbf{\Phi}\ \underline{i} \mid \mathbf{E}\ \underline{i} \in \mathcal{R}$ かつ $i \in \mathbb{N}\} \subseteq \mathcal{S}_\tau$. したがって,

系 4.3 $\{\mathbf{\Phi}\ \underline{i} \mid \mathbf{E}\ \underline{i} \in \mathcal{R}$ かつ $i \in \mathbb{N}\} \in \mathcal{EX}$.

定理 4.4 $\mathrm{NV} \subsetneq \mathrm{EX}$.

証明. $\mathcal{F} \in \mathrm{NV}$ とすると適当な $h \in R$ に対して h 従順. $\mathcal{F} \subseteq \mathrm{EX}(M'\ \underline{0})$. ただし M' は定理 3.3(1) の証明 で与えた M' において行 (†) を $\underline{i}$ にしたもの.

4.2 否合併定理

定義 4.5

$$\mathcal{S}_{self} \ := \ \{N \in \mathcal{R} \mid \mathbf{E}\ (N\ \underline{0})\ =_{\beta\eta}\ N\}$$

$N \in \mathcal{R}$ で有限個を除く全ての非負整数 i に対して $N\ \underline{i}\ =_{\beta\eta}\ \underline{0}$ となるもの全体を $\mathcal{S}_0$ で表す.

定理 4.6 ($\mathcal{EX}$ の否合併定理) $\mathcal{S}_{self} \in \mathcal{EX}$ かつ $\mathcal{S}_0 \in \mathcal{NV} \subsetneq \mathcal{EX}$ であるが, $\mathcal{S}_{self} \cup \mathcal{S}_0 \notin \mathcal{EX}$.

証明. $\mathcal{S}_{self} \cup \mathcal{S}_0 \not\subseteq \mathrm{EX}(M)$ と仮定する. M を用いてできるラムダ式 F に対して "再帰定理" (1) により定義された $\mathbf{E}\ e$ が $\mathcal{S}_{self}$ に属するのに, $\mathbf{E}\ e \notin \mathrm{EX}(M)$ を示すことにより矛盾を導く.

5 今後の課題

計算可能関数の学習理論的階層は精緻なものであり, それらがラムダ計算にも対応物があることを明らかにする. たとえば, [6] の $\mathcal{EX}_b^a$ (同定までに仮説を変える回数が b 以下で, 正解と高々 a 個の引数でしか異ならないものを同定する場合の関数の集合) を考える.

また, 現実的な応用を考えるにあたっては, 確率的な妥協は必要不可欠だから PAC 学習の枠組で本稿と同様のことができるか吟味し, CFengine などの確率的学習によるコンピュータ/ネットワーク管理に応用できるかを考える.

謝辞

文献 [10] を閲覧させて下さった外山芳人 教授に感謝する.

参考文献

[1] Y. Akama, Limiting Partial Combinatory Algebras. *T.C.S.*, Vol. 311(2004), pp. 199-220.

[2] Y. Akama, S. Berardi, U. Kohlenbach, and S. Hayashi, An arithmetical hierarychy of the laow of excluded middle and related principles. Proc. of the 19th Annual IEEE Symposium on Logic in Computer Science (LICS2004), pp. 192-201.

[3] H.P. Barendregt, The Lambda Calculus, Its Syntax and Semantics. North-Holland, 2nd Edition, 1984.

[4] D. Bridges, Computability, a mathematical sketchbook. Graduate Texts in Mathematics, Springer, 1991.

[5] M. Burgess, Probabilistic anomaly detection in distributed computer networks. available via http://www.iu.hio.no/~mark/papers/anomaly.pdf.

[6] J. Case and C. Smith, Comparison of identification criteria for machine inductive inference. *T.C.S.*, Vol. 25(1983), pp. 193-220.

[7] M. Hagiya, A Typed lambda-Calculus for Priving-by-Example and Bottom-Up Generalization Procedure. *T.C.S.*, Vol. 137(1985), pp. 3-23.

[8] S. Hayashi, R. Sumitomo R., and K .Shii, Towards animation of proofs -testing proofs by examples. *T. C. S.*, Vol. 272, pp. 177–195.

[9] S.C. Kleene, λ-definability and recursiveness, *Duke Math. J.*, Vol. 2(1936), pp. 340-353.

[10] 菅原 俊治, 外山芳人, 帰納的推論の基礎理論, 1985.