数理解析研究所講究録
1489 巻 2006 年 256-260

普遍代数における閉集合族と正データからの帰納推論
Families of Closed Sets and Inductive Inference from Positive Data

マシュー デ・ブレクト* Matthew de Brecht　小林正典** Masanori Kobayashi　徳永浩雄** Hiroo Tokunaga　山本章博* Akihiro Yamamoto

*京都大学 情報学研究科
**首都大学東京
*Graduate School of Informatics, Kyoto University
**Tokyo Metropolitan University

Abstract: Previous research has begun to show the connections between the properties of formal language classes that can be learned from positive data and algebraic structures such as the class of ideals of polynomial rings [5, 9]. In particular, it has been shown that a computable commutative ring is Noetherian if and only if the class of its ideals has finite elasticity [5]. In this paper, we generalize these results by defining languages using closure operators, and show how some sufficient conditions for learning from positive data appear naturally as properties of families of closed sets. In particular, we show that the conditions EC1, C2, and C3 as defined in [7] become equivalent within the family of closed sets of an algebraic closure operator.

1. はじめに

可換環(特に多項式環)の性質を形式言語(本稿では単に言語とよぶ)の正データからの帰納推論によって特徴付ける研究が始められている[5, 9]. 本研究は, その特徴付けを抽象化し, 再び言語の正データからの帰納推論に利用できる形にすることを目的とする.

Angluin[1]による正データからの帰納推論可能な言語族の特徴付けを出発点に, 理論上・実用上の多くの研究がなされてきた. 特に, 言語族が推論可能であるための十分条件については, 数学的にも興味深いものが4つ発見されている. 榊原ら[7]の成書ではこれらの条件を EC1, C2, C3, C4 と表している. Angluin の特徴付けは EC1 であり, この順序で後者は前者より真に強い条件である.

形式言語の要素である語を定義するための代数的な演算は文字の連結(concatenation)であり, 語全体は非可換の半群をなす. そこで, 半群よりも複雑な構造を帰納推論の対象とすることが考えられる. 実際, Angluin[1]では, 正データからの帰納推論可能条件 C4 の例として, 整数環のイデアルを取り上げている.

最近, Stephan and Ventsov[9]は, 可換環 R がネータ性を持つことは, R のイデアルを形式言語として捉えたときに, R のイデアルの全体からなる族について条件 EC1 が成立することと同値になる. 小林ら[5]は, 可換環 R のネータ性がイデアルの族について条件 C3 が成立することと同値であることを示した. このことは, 条件 EC1, C2, C3, C4 が正データからの帰納推論可能性に関する尺度として用いることができること, そして, 可換環という構造ではそのうちの EC1, C2, C3 が同値になることを示している.

本稿では, 言語族の構造を閉包演算で定義することにより, これらの同値性が形式言語の帰納推論でも成立するような条件を与える. 次節では, 正データからの帰納推論の基本定義や帰納推論可能性のための様々な条件を述べておく. 3 節で可換環と正データからの帰納推論の関係を復習する. 4 節では, 閉包演算を導入し, 主定理を述べる. 最後に 5 節で今後の展望を述べる.

2. 言語族の正データから帰納推論

最初に, 正データからの帰納推論の基本定義を述べておく. Σを記号の有限集合とし, Σ^*の部分集合を言語という. 言語の列 $L_1, L_2, L_3, \ldots$ に対して,

$$f(i, w) = \begin{cases} 1 & w \in L_i \text{ のとき} \\ 0 & \text{そうでないとき} \end{cases}$$

のような計算可能な特徴関数 f が存在するとき, $\mathcal{L} = \{ L_1, L_2, L_3, \ldots \}$を**添字付き帰納的言語族**(indexed

*京都大学情報学研究科知能情報学専攻
〒606−8501 京都市左京区吉田本町
Tel: 075-753-5636, Fax: 075-753-5628
E-mail: matthew@mbox.kudpc.kyoto-u.ac.jp
akihiro@i.kyoto-u.ac.jp
**首都大学東京 都市教養学部 理工学系 数理科学コース
〒192−0397 八王子市南大沢1−1
Tel: 0426-77-1111(代), Fax: 0426-77-2481
E-mail: masanori@math.metro-u.ac.jp
tokunaga@comp.metro-u.ac.jp

family of recursive languages)という．以下の議論では添字付き帰納的言語族しか考慮しないので，単に言語族と省略することがある．無限列 $\sigma = s_0, s_1, s_2, \ldots$ が $L=\{ s_i \mid s_i \in \sigma\}$を満たせば，$\sigma$は L の正例の提示，または**正提示**(positive presentation)であるという．

推論機械(inference machine)とは，特殊なチューリング機械で，時折入力を要求したり，仮説を出力したりする．正提示σに対して推論機械 M の出力の列を $M[\sigma]$で表す．$M[\sigma]$が無限で有限個以外の要素が全て j である場合，または$M[\sigma]$が有限で最後の要素が jである場合に，$M[\sigma]$が jに収束するという($M[\sigma]\downarrow = j$と表す)．

言語 L_i の任意の正提示σに対して$M[\sigma]\downarrow = j$ かつ $L_j = L_i$ が成り立つとき，M はL_i を正例から帰納推論するという．添字付き帰納的言語族 $\mathcal{L} =\{L_1, L_2, L_3, \ldots\}$の全ての言語を正例から帰納推論する M が存在すれば，$\mathcal{L}$ は**正例から帰納推論可能**(inferable from positive data)であるという．正例からの学習の詳細については[1, 7]などを参照されたい．

Angluin[1]は言語族の正例からの帰納推論可能性の必要十分条件を有限証拠集合を用いて証明した．

定義(Angluin[1]). $\mathcal{L} =\{ L_1, L_2, L_3, \ldots\}$を添字付き帰納的言語族とする．言語 $L_i \in \mathcal{L}$ のある有限部分集合 T_i は，全ての $j \neq i$ に対して $T_i \subseteq L_j$ ならば L_j は L_i の真部分集合ではないという条件が成り立てば，T_iを L_i の**有限証拠集合**(finite tell-tale set)といい，L_i **が $\mathcal{L}$ において有限証拠集合を持つ**という．

添字付き帰納的言語族$\mathcal{L}=\{ L_1, L_2, L_3, \ldots\}$の全ての言語が有限証拠集合を持てば，文献[7]にならって $\mathcal{L}$ は条件 C1 を満たすとよぶことにする．$\mathcal{L}$が条件 C1 を満たし，かつ，任意の i に対して L_i の有限証拠集合の要素を枚挙する手続きが存在すれば，$\mathcal{L}$ は条件 EC1 を満たすという．

定理 1(Angluin[1]). 添字付き帰納的言語族 $\mathcal{L}$ が正例から帰納推論可能であるための必要十分条件は $\mathcal{L}$ が条件 EC1 を満たすことである．

EC1 より強い条件がいくつか挙げられている．

定義(Kobayashi, S. [6]). $\mathcal{L}$ を添字付き帰納的言語族とする．全ての $L' \in \mathcal{L}$ に対して，$F \subseteq L' \Rightarrow L \subseteq L'$ を満たす $L \in \mathcal{L}$ の有限部分集合 Fが存在すれば，Fを Lの**特徴例集合**(characteristic sample set)という．$\mathcal{L}$ の全ての言語が特徴例集合を持てば，$\mathcal{L}$ は条件 C2 を満たすという．

定義(Wright[10], Motoki et al. [8]). $\mathcal{L}$ を添字付き帰納的言語族とする．2つの無限列，$L_1, L_2, L_3, \ldots$と $s_0, s_1, s_2, \ldots$ (但し，$L_i \in \mathcal{L}$, $s_i \in \Sigma^*$) が存在し，しかも $\{ s_0, s_1, \ldots, s_{n-1}\} \subseteq L_n$かつ $s_n \notin L_n$が成り立つ場合，$\mathcal{L}$ は**無限の弾力性**(infinite elasticity)を持つという．$\mathcal{L}$ が無限の弾力性を持たなければ，$\mathcal{L}$ は**有限の弾力性**(finite elasticity)を持つ，または$\mathcal{L}$は条件 C3 を満たすという．

定義(Angluin[1]). $\mathcal{L}$ を添字付き帰納的言語族とする．任意の $s \in \Sigma^*$に対して，集合 $\{ L_i \in \mathcal{L} \mid s \in L_i \}$ は必ず有限個の言語しか含まない場合，$\mathcal{L}$ は**有限の厚さ**(finite thickness)を持つ，または $\mathcal{L}$ は条件 C4 を満たすという．

条件 C1 から C4 の間には，

$$\mathrm{C4} \Rightarrow \mathrm{C3} \Rightarrow \mathrm{C2} \Rightarrow \mathrm{EC1} \Rightarrow \mathrm{C1}$$

という関係が知られている．逆方向の含意は一般には成立しないことに注意する．

3. 環のイデアルと帰納推論

以下の議論では，考慮する環の要素は枚挙可能であり，環上に定義される和演算と積演算が計算可能であると仮定する．また，定義や議論を簡潔するために，環の積演算が可換であることを仮定する．

環 R の集合 Iが次の条件を満たせば，I を Rの**イデアル**(ideal)という．

- $0 \in I$.
- $f, g \in I$ ならば$f+g \in I$.
- $f \in I$ かつ $h \in R$ならば $hf \in I$.

$X \subseteq R$から生成されるイデアル$\langle X \rangle$とはXを含む最小のイデアルであり，以下のように定義される：

$$\langle X \rangle = \left\{ \sum_{i=1}^{n} h_i f_i \mid h_1, \ldots, h_n \in R, f_1, \ldots, f_n \in X, n < \infty \right\}.$$

Rのイデアルからなる集合の包含関係で真に増加する無限列が存在しなければ，Rを**ネータ環**という．イデアルの詳細は[3]を参照されたい．

ネータ環と帰納推論の間の関係は，当初は次の定理で与えられた．

定理 2 (Stephan and Ventsov [9]). 環 Rがネータ環であるための必要十分条件は Rのイデアルのクラス $\mathcal{I}$ が正データから帰納推論可能であることである．

この定理は次のように詳細化することができる．

定理3(小林, 徳永, 山本[5]). 環Rがネータ環であるための必要十分条件はRのイデアルのクラス$\mathcal{J}$が有限の弾力性を持つことである.

すなわち, 環というデータ構造においては, EC1, C2, C3が同値になる. しかし, C4については一般には同値にならない.

例1. 有理数を係数とし, xとyの変数からなる多項式全体の集合を$Q[x, y]$とする. $Q[x, y]$はネータ環であるため, $Q[x, y]$のイデアルのクラスが有限の弾力性を持つ. 全ての自然数iに対して, $x \in \langle x, y^i \rangle$であるが, $i \neq j$のとき, $\langle x, y^i \rangle \neq \langle x, y^j \rangle$が成り立つ. 従って, xは無限個のイデアルに含まれているため, $Q[x, y]$のイデアルのクラスは有限の厚さを持たない.

上の例で示したように, 一般的にイデアルのクラスは有限の厚さを持たない. しかし, 有限の厚さを持つ有用なイデアルの部分クラスが知られている. 有理数を係数とするn変数の多項式環を$Q[x_1, \ldots, x_n]$とし, $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ ($\alpha_1, \ldots, \alpha_n \in N$)のような変数の積を単項式と定義する.

補題1 (de Brecht et al [4]). 1つの単項式から生成されるイデアルのクラス$\{ \langle x^\alpha \rangle \mid \alpha \in N^n \}$は有限の厚さを持つ.

証明. 多項式の指数上の整礎な完全順序を適切に決めて, 任意のfは$f = g\,x^\alpha$, 但しgは定数か単項式でない多項式, と一意的にかけるので, $f \in \langle x^\beta \rangle \Leftrightarrow h\,x^\beta = f$ ($h \in Q[x_1, \ldots, x_n]$) $\Leftrightarrow \beta \leqq \alpha$が成り立つ. そして, αより小さい指数は有限個しか存在しないため, fは有限個のイデアルにしか含まれていない. 多項式fは任意であるため, 1つの単項式から生成されるイデアルのクラスは有限の厚さを持つ.

4. 閉包演算と正データからの帰納推論

多項式環で成立した性質が, 言語族の正データからの学習に対しても成立するための条件を考察する.

定義. Uを集合とする. $C: 2^U \to 2^U$が次の条件を満たすとき, Cを**閉包演算**とよぶ.

- $X \subseteq C(X)$.
- $C(C(X)) = C(X)$.
- $X \subseteq Y$ならば$C(X) \subseteq C(Y)$.

さらに$X = C(X)$であれば, **XをCに対する閉集合**とよぶ. 曖昧さが生じなければ, 単に閉集合とよぶ.

以下の議論では, $U = \Sigma^*$とし, ある閉包演算Cが与えられたとき, $\mathcal{L}_C = \{X \mid X$は$C$に対する閉集合$\}$と定義する. このように定義された閉集合族と学習可能性の各条件の関係を議論するには, コンパクト性という概念が有用である.

定義. 閉集合Xと任意の閉集合族Y_i($i \in I$)に対して, $X \subseteq C(\bigcup_{i \in I} Y_i)$であれば$X \subseteq C(\bigcup_{i \in I'} Y_i)$が成立するような$I$の有限部分集合$I'$が存在するとき, **$X$はコンパクト**であるという.

閉集合族の集合のコンパクト性と有限証拠集合は次の関係を持つ.

定理4. 任意の閉包演算Cの閉集合族を$\mathcal{L}_C$とする. ある閉集合$L \in \mathcal{L}_C$が枚挙可能で$\mathcal{L}_C$においてコンパクトならばLは$\mathcal{L}_C$において有限証拠集合を持つ.

証明. $L \in \mathcal{L}_C$がコンパクトで有限証拠集合を持たないと仮定する. $s_0, s_1, s_2, \ldots$というようにLの全ての要素を枚挙すると, $\{s_0\}$はLの有限証拠集合でないため$C_0 = C(\{s_0\})$はLの真部分集合となる. 同様に, $s_{i_n} \in L - C_{n-1}$となる最小のi_nをとって, $C_n = C(\{s_0, \ldots, s_{i_n}\})$は$L$の真部分集合となる. 無限に続けると, 明らかに$L \subseteq C(\bigcup_{n \in N} C_n)$が成り立つが, 作り方により$C_0 \subset C_1 \subset \cdots$であるため, 任意の有限集合$I \subseteq N$に対して$C(\bigcup_{n \in I} C_n) = C_{\max(I)} \subset L$が成り立ち, Lのコンパクト性と矛盾する.

以上の定理はコンパクト閉集合の有限証拠集合の枚挙可能性を議論しないので, 正データからの学習可能性のための十分条件にはならない. 学習可能性を保証するために次のような閉包演算を導入する.

定義. 閉包演算Cは任意の$X \subseteq U$に対して,

$$C(X) = \bigcup \{ C(Y) \mid Y \subseteq X, Y\text{は有限集合} \}$$

を満たすとき, **代数的閉包演算**(algebraic closure operator)であるという.

例2. Rは任意の環で, $X \subseteq R$に対して$I(X)$はXから生成されるイデアルとする. 明らかに$I(\cdot)$は閉包演算である. また, 任意の$a \in I(X)$は$r_1 x_1 + \cdots + r_n x_n$(但し, $r_i \in R$, $x_i \in X$)というような有限の和であるため, $a \in I(\{x_1, \ldots, x_n\})$が成り立つ. 従って, $I(\cdot)$は代数的閉包演算である. 同様に, 任意の群Gに対して, $S(X)$が$X \subseteq G$から生成される部分群ならば, $S(\cdot)$は代数的閉包演算である.

定義. 集合Xは有限集合Yに対して, $X = C(Y)$であるとき, **有限生成の閉集合**であるという.

定理5([2]参照). Cが代数的閉包演算であれば, コン

パクト閉集合と有限生成閉集合は一致する.

この事実を用いて代数的閉包演算のコンパクト性は学習可能性の条件と次の関係を持つ.

定理 6. 代数的閉包演算 Cの閉集合族を $\mathcal{L}_C$とする. $X \in \mathcal{L}_C$がコンパクト閉集合であるための必要十分条件はXが $\mathcal{L}_C$において特徴例集合を持つことである.

証明. Xがコンパクト閉集合と仮定すれば, $X=C(Y)$を満たすXの有限部分集合 Yが存在する. 任意の閉集合 X'に対して, $Y \subseteq X' \Rightarrow C(Y) \subseteq C(X') \Rightarrow X \subseteq X'$より, YはXの特徴例集合である. 逆に, Xの特徴例集合をYと仮定すれば, 特徴例集合の定義よりXはYを含む最小の閉集合であり, $X=C(Y)$が成り立つ. 従って, Xは有限生成閉集合で, コンパクトである.

定理7. 代数的閉包演算 Cの閉集合族を $\mathcal{L}_C$とする. ある閉集合 $L \in \mathcal{L}_C$が $\mathcal{L}_C$において有限証拠集合を持てば, Lは $\mathcal{L}_C$においてコンパクトである.

証明. $L \in \mathcal{L}_C$は有限証拠集合を持つと仮定する. $T \subseteq L$をLの $\mathcal{L}_C$における有限証拠集合とすれば, $C(T) \subseteq L$が成り立つが, $C(T) \in \mathcal{L}_C$であることと有限証拠集合の定義より $C(T)=L$が成立しなければならない. 従って, Lは有限生成閉集合であり, コンパクトである.

定理4, 6と7より, 代数的閉包演算の閉集合族では, 条件C1とC2が同値になる. さらに, 環のネータ性とイデアルの有限の弾力性の関係を次のように一般化できる.

定理 8. 閉包演算 Cに対して真に増加する閉集合の無限列が存在しないための必要十分条件はCに関する閉集合のクラスが有限の弾力性を持つことである.

証明. C に関する閉集合のクラスは真に増加する閉集合の無限列を含まないが, 無限の弾力性を持つと仮定する. 従って $B_{n-1}=\{s_0, s_1, \ldots, s_{n-1}\} \subseteq X_n$かつ $s_n \notin X_n$を満たす無限の要素の列 $s_0, s_1, s_2, \ldots$と無限の閉集合の列 $X_1, X_2, X_3, \ldots$ が存在する. $C(B_{n-1}) \subseteq X_n$かつ $C(B_n) \not\subseteq X_n$であるから, $C(B_{n-1}) \neq C(B_n)$が成り立つ. また, $B_{n-1} \subseteq B_n$より $C(B_{n-1}) \subseteq C(B_n)$であるので, $C(B_{n-1})$は $C(B_n)$の真部分集合である. nは任意であるから, 真に増加する無限列となり, 仮定と矛盾する.

次に, 有限の弾力性を持ち, $X_1 \subset X_2 \subset X_3 \subset \cdots$という閉集合の無限列が存在すると仮定する. s_0をX_1の任意の要素, $s_n \in X_{n+1}-X_n$とすれば, 明らかに無限列 $s_0, s_1, \ldots$と$X_1, X_2, X_3, \ldots$が有限の弾力性の仮定と矛盾する.

閉包演算として形式化すると, イデアルのクラスでEC1, C2, C3 が同値になる理由が次の定理より明らかになる.

定理 9. 代数的閉包演算 C において, 全ての閉集合がコンパクトであるための必要十分条件は真に増加する閉集合の無限列が存在しないことである.

証明. 全ての閉集合がコンパクトであるが真に増加する閉集合の無限列 $C_0 \subset C_1 \subset C_2 \subset \cdots$が存在すると仮定する. $C_\omega = C(\cup_{i \in \mathrm{N}} C_i)$とすれば, 任意の有限集合 $I \subseteq \mathrm{N}$に対して $C(\cup_{i \in I} C_i) = C_{\max(I)} \subset C_\omega$ が成り立つため, C_ω はコンパクトではない. これは矛盾である.

逆に, 真に増加する閉集合の無限列が存在しないが, 非コンパクトの閉集合が存在すると仮定する. X を非コンパクト閉集合とすると, $X \subseteq C(\cup_{i \in \mathrm{N}} Y_i)$であるが, 全ての有限集合 $I \subseteq \mathrm{N}$ に対して Xは $C(\cup_{i \in I} Y_i)$の部分集合ではないような無限個の閉集合 Y_i ($i \in \mathrm{N}$)が存在する. X は $C(Y_0)$の部分集合ではないので, $Y_k - C(Y_0) \neq \phi$を満たす Y_k が存在する (一般性を失わずにこれを Y_1とする). 従って, $C(Y_0) \subset C(Y_0 \cup Y_1)$, かつ Xは $C(Y_0 \cup Y_1)$の部分集合ではない. 無限に繰り返すと, $C(Y_0) \subset \cdots \subset C(Y_0 \cup \cdots \cup Y_n) \subset \cdots$という真に増加する無限列が存在することより矛盾である.

系 1: 代数的閉包演算によって定義された閉集合族では, 条件C1と EC1, C2, C3 は同値である.

系 2: 任意の代数的閉包演算に対する閉集合は全て計算可能であれば, 全体の閉集合族の正データからの学習可能性のための必要十分条件は真に増加する閉集合の無限列が存在しないことである.

例 3. $i \in \mathrm{N}$ に対して $L_i = \{x \in \mathrm{N} \mid x \leqq i\}$とし, 言語族 $\mathcal{L}_\mathrm{N} = \{L_i \mid i \in \mathrm{N}\}$と定義する. $\mathcal{L}_\mathrm{N}$は無限の弾力性を持つが明らかに正データから帰納推論可能である. そして, 任意の集合 $X \subseteq \mathrm{N}$ が有限集合のときに $C(X) = \{x \in \mathrm{N} \mid x \leqq \max(X)\}$, X が無限集合のときに $C(X) = \mathrm{N}$ と定義すれば, Cは代数的閉包演算となり, $i \in \mathrm{N}$ に対して $L_i = C(\{i\})$が成り立つ. しかし, $\mathcal{L}_\mathrm{N} = \mathcal{L}_C - \{\mathrm{N}\}$のため, 上の系2は $\mathcal{L}_\mathrm{N}$に対しては当てはまらない. このように, 閉集合族に関する上の定理が部分族については一般には成り立たないことに注意する必要がある.

5. おわりに

本稿では, 正データからの帰納推論と多項式環のような代数的な構造の関係の研究を進めるために, 閉包演算により定義された言語族の性質と帰納推論可能性のための十分条件の関係を示した. 特に, 環 R がネータ環であるための必要十分条件はRのイデアルのクラスが有限の弾力性を持つという事実を, 代数的閉包演算の閉集合族まで自然に一般化できることを証明した.

今後は, 本稿の結果を広く適用できるように, 形式的文法など言語族を定義するための種々の方法と閉包演算の関係を明確にしていきたい.

参考文献

[1] D. Angluin: Inductive Inference of Formal Languages from Positive Data, *Information and Control* **45**, 117-135 (1980).

[2] Burris and H. P. Sankappanavar: A Course in Universal Algebra, Springer-Verlag, (1981).

[3] D. Cox, J. Little, D. O'Shea, *Ideals, Varieties, and Algorithms, 2nd Edition*, Springer (1996).

[4] M. de Brecht, 山本章博：有限の弾力性を含む証明の分析 ー有限の弾力性の手続き化に向けてー, 第 61 回人工知能基本問題研究会資料 SIG-FPAI-A503, 93-98(2005).

[5] 小林正典, 徳永浩雄, 山本章博：多項式のイデアルと正データからの学習, 2005 年情報論的学習理論ワークショップ予稿集, 129-134 (2005).

[6] S. Kobayashi, Approximate Identification, Finite Elasticity and Lattice Structure of Hypothesis Space, Technical Report, CSIM 96-04, Dept. of Compt. Sci. and Inform. Math., Univ. of Electro- Communications, 1996.

[7] 榊原康文, 小林聡, 横森貴: 計算論的学習, 培風館, 2001.

[8] T. Motoki, T. Shinohara, and K. Wright: The Correct Definition of Finite Elasticity: Corrigendum to Identification of Unions, *Proceedings of COLT '91*, 375 (1991).

[9] F. Stephan and Y. Ventsov: Learning Algebraic Structures from Text, *Theoretical Computer Science*, **26**, 221-273 (2001).

[10] K. Wright: Identification of Unions of Languages Drawn from an Identifiable Class, *Proc. 2nd Workshop on Computational Learning Theory*, 328-333 (1989).