

会 議 報 告

ALT '95 (The 6th Workshop on Algorithmic Learning Theory)の報告

1. はじめに

第6回計算論的学習理論国際ワークショップ(ALT '95)が、福岡県福岡市にあるリーセントホテルにおいて、人工知能学会、九州工業大学、東京大学医科学研究所ヒトゲノム解析センターの主催により、平成7年10月18日~20日の3日間にわたり開催された。本会議は、オートマトン・言語理論や計算量の理論、アルゴリズム論などの計算機科学における理論的基礎に基づいて、機械学習の本質を捉えるとともに、遺伝子情報処理やデータベースからの知識の発見など、関連する諸分野との関わり合いを探ることを目的としている。

今回は日本も含めて12か国から94名の参加があった。海外からは20名の参加者があり、その内訳は、米国(6)、ドイツ(3)、スペイン(2)、オーストラリア(2)、フランス(1)、イタリア(1)、ベルギー(1)、ラトビア(1)、カナダ(1)、シンガポール(1)、インド(1)であった。採択率は投稿論文46件中、21件であった。うち、日本からの発表者は、全体の約1/3であり、本会議が国際会議として世界的に定着してきていることが感じられた。

2. ALT '95 会議概況

会議は、初日(18日)の午前9時に、大会プログラム委員長である九州大学のThoman Zeugmann博士のオープニングの挨拶により始まった。以下、会議の概況を報告者のわかる範囲で紹介したい。

SESSION 1 (招待講演) (座長: Thomas Zeugmann)

Yasubumi Sakakibara: Grammatical inference: An old and new paradigm.

本講演では、榊原博士および富士通研究所のグループによる研究成果を中心に、文法推論(Grammatical Inference)における従来の成果および最近の動向が非常に明快な論旨展開のもとでまとめられた。一般に、学習モデルにおいて、学習者に提示する情報(正提示、完全提示)、学習者が利用できる質問、提示情報におけ

る誤りの有無、仮説の表現形式、などを変化させればその学習問題はさまざまな難しさを呈し得る。具体的に、正則言語族や文脈自由言語族およびその部分族を対象として、上記の尺度をさまざまに変化させながら学習可能性に関するPositiveな結果とNegativeな成果がわかりやすく説明された。特に、Simple recurrent networkによる時系列学習能力を文法推論に応用する新しい話題に興味を引かれた。

SESSOIN 2 (座長: Klaus P. Jantke)

- (1) Amr F. Fahmy and Robert Roos: Efficient learning of real time one counter automata.
- (2) Takeshi Koshiba, Erkki Makinen and Yuji Takada: Learning strongly deterministic even linear languages from positive examples.
- (3) Hiroshi Sakamoto: Language learning with characteristic examples and membership queries.
- (4) Hiroki Arimura, Hiroki Ishizaka and Takeshi Shinohara: Learning unions of tree patterns using queries.

(1)は、実時間決定性 one-counter オートマトンを等価性質問と所属性質問を用いて効率的に学習するアルゴリズムを提案している。学習アルゴリズムは、AngluinのDFAのMAT学習アルゴリズムを利用して、学習対象のBehavior Graphを獲得した後、それを、制御構造とデータ構造(one-counter)に分割して目的のone-counter オートマトンを獲得する。オートマトンの分割アルゴリズムを用いるアイディアは、データ構造を持つ一般のオートマトンの学習に対する見通しの良い視野を与えてくれる点で興味深かった。

(2)は、強決定性 even-線形言語族 (Strongly Deterministic Even Linear Languages) と呼ばれる even-線形言語族の部分族を導入し、この言語族が、更新時間の意味で、正例から非常に効率良く(ほぼ線形時間で)学習可能であることを示した。この言語は、そのSzilard言語のあるhomomorphic imageが k -reversible言語となるようなeven-線形言語として特徴づけられ、その学習には、Angluinによる k -reversible言語族の学習アルゴリズムと、高田によるeven-線形言語族の学習手法が応用されている。

(3)は、Characteristic Example (以下、CE) という

新しい概念を導入し、CE と所属性質問を利用した括弧つき言語族の多項式時間学習アルゴリズムを提案した。基本的なアイデアは、榊原による構文木の構造情報を利用した文脈自由文法の MAT 学習手法と関係が深い。直観的に述べると、括弧つき言語は、その文字列に構文木の構造情報が含まれているという事実から、構造情報を使わずに MAT 学習が可能となる。さらに CE という、すべての生成規則を網羅的に適用して得られる例を利用することにより、等価性質問を使わない効率的な学習が可能となっている。

(4)は、たかだか k 個の木パターンの和により表現される言語族 TP^k の多項式時間 MAT 学習アルゴリズムを示した。まず、等価性質問と部分性質問を利用した TP^k の多項式時間学習アルゴリズムが示され、次に、定数アルファベットのサイズが k より大きいという条件のもとで、部分性質問を所属性質問で効率良く置き換える手法が示された。論理プログラムの効率的な学習に向けて着実に進歩しているようすがうかがえ、興味深い発表であった。

SESSION 3 (座長: Makoto Haraguchi)

- (1) Luc De Raedt and Wim Van Lear: Inductive constraint logic.
- (2) M. R. K. Krishna Rao: Incremental learning of logic programs.
- (3) Akira Miyashiro, Eiji Takimoto, Yoshifumi Sakai and Akira Maruoka: Learning orthogonal F-Horn formulas.

(1)は、正・負の例から一階述語論理プログラムを学習する新しいアプローチを示した。そのポイントは、与えられる正・負の例を、目標とする論理プログラムのエルプラン解釈とみなす点にあると主張されており、提案する手法が実験的にも優れていることが示されている。

(2)は、論理プログラムの漸増的学習アルゴリズムを提案している。背景知識を関数記号として項に組み込む山本らのアプローチを拡張することにより、より広いクラスの論理プログラムを多項式時間で学習できることを示した。

DNF 論理式あるいは CNF 論理式の PAC 学習問題が k -quasi ホーン論理式の MAT 学習問題に帰着されるという点で、 k -quasi ホーン論理式の質問を用いた学習に関する話題は興味深い。(3)は、F-ホーン論理式と呼ばれる k -quasi 論理式の拡張概念を導入し、その質問を用いた学習を考察している。そして、直交 F-ホーン論理式(各節のボディが互いに他を包含しないもの)という部分クラスを導入し、その等価性、部分性、

所属性質問を用いた学習アルゴリズムが示され、さらに、その部分クラスである直交 F-ホーン論理式は、MAT 学習可能であることが示された。また、直交性をはずした場合には、この MAT 学習問題に CNF 論理式の所属性質問を用いた PAC 学習問題が帰着されてしまうことも述べられている。

SESSION 4 (座長: Phil Long)

- (1) Peter Auer: Learning of nested differences in the presence of malicious noise.
- (2) Atsuyoshi Nakamura and Shinji Miura: Learning sparse linear combinations of basis functions over a finite domain.

(1)は、ノイズを許した 2 種類の学習モデルのもとで、集合積に関して閉じた (intersection-closed な) 概念族の概念を、入れ子状に差をとることによって表現できる概念 (nested differences) を学習する問題を考察している。on-line 学習モデルでは、ノイズの存在下で、誤予測の回数が最適になる効率の良い学習アルゴリズムを提案した。また、Malicious PAC 学習モデルでは、従来の効率の良い学習アルゴリズムで知られているノイズの許容値の下限を更新する結果が示された。

(2)は、有限集合上で定義された $\{0,1\}$ の値をとる関数族および実数値関数族の質問を用いた学習問題において、同定に必要な質問の回数に関する解析を行った。特に、このような関数が基底の線形結合で表現されることに注目し、基底の疎な線形結合で表現される関数族の同定に必要な質問回数の解析結果が示された。たかだか k 個の基底要素の線形結合で表現される関数族の質問回数に関する複雑度と、線形符号との関係が述べられている点が興味深かった。そこでは、BCH 符号から得られる基底により張られる関数族が、質問回数の意味で(たかだか定数倍を除いて)最適な関数族であることが示されている。

SESSION 5 (招待講演) (座長: Takashi Yokomori)

Ming Li: Infer a DNA sequence from erroneous copies (joint work with John Kececioğlu and John Tromp).

Ming Li 博士による招待講演では、博士の得意とするコルモゴロフ複雑さの理論に基づいたアルゴリズムの解析手法を、遺伝子の核酸配列の決定問題へ応用する話題がわかりやすく紹介された。そこでは、核酸配列決定の際に、挿入、削除、代入といった誤りが生じた場合に、真の核酸配列を決定する効率の良いアルゴリズムが紹介された。それは、三つの配列から順次繰り返し求めていく非常にシンプルなアルゴリズムであ

ったが、1)核酸配列がコルモゴロフ複雑さの意味でランダムであり、2)上記の3種類の誤りが均等に生ずるとする仮定のもとで、たかだか $\log^{1+\delta} n$ ($0 < \delta < 1$)個の核酸配列(誤りを含む)から、高い確率で効率良く真の核酸配列を求めてくれることが示された(ただし、 n は核酸配列の長さ)。1)の仮定は少々異論の残るところではあるが、問題自体は非常に一般性のあるものであり、研究の質の高さが感じられる講演であった。

SESSION 6 (座長: Hiroki Ishizaka)

- (1) John Case, Sanjay Jain and Arun Sharma: Machine induction without revolutionary paradigm shifts.
- (2) Lea Meyer: Probabilistic language learning under monotonicity constraints.
- (3) Frank Stephan: Noisy inference and oracles.
- (4) Bala Kalyanasundaram, Mahendran Velauthapillai: Simulating teams with many conjectures.

(1)は、パラダイムの大きな変更を許さない極限同定の学習モデルを提案し、その学習モデルの能力を詳しく分析した。ここでは、パラダイム変更の大きさは、学習者が出力する推測のサイズ変更の大きさとして捉えられ、従来の関連する学習モデルとの能力比較も示された。

(2)は、単調に概念を大きくしながら学習する単調極限学習の枠組みに、確率学習 (probabilistic learning) の規範を採り入れて、その学習能力を分析した。従来の確率学習では、離散的な学習能力の階層が得られることが知られているが、この学習モデルでは、その階層が密 (dense) になることが示され興味深かった。

(3)は、正しい例は無限回提示され、間違っただけは、たかだか有限回しか提示されないというノイズモデルのもとで、極限同定問題を考察した。そして、その主要な結果として、停止問題に関するオラクルを使用可能な有限学習モデルが、提案されたノイズを許した学習モデルと学習能力の意味で等価であることが示された。

チーム学習モデルでは、そのチームに属する少なくとも一つの学習者が学習できたとき、学習に成功したとみなされる。このモデルでは、二つのチーム学習モデルが、チームの構成員数、許される推測変更の回数、同定した関数に含まれてもよい誤りの数などのパラメータ (6パラメータ)として与えられたとき、どちらのモデルがより強い学習能力を持っているかどうかを判定する問題が、長年の未解決問題となっている。(4)は、同定結果に誤りを許さない学習 (4パラメータ)の場合において、この問題を完全に解決した。

SESSION 7 (座長: Akira Maruoka)

- (1) Charles C. Pinter: Complexity of network training for classes of neural networks.
- (2) Ricard Gavaldà and David Guisjarro: Learning ordered binary decision diagrams.
- (3) Jorge Castro and Jose L. Balcazar: Simple PAC learning of simple decision lists.

ニューラルネットの学習の分野においては、いくつかのネットワーク構造のもとでその学習問題の NP-完全性が証明されている。(1)は、これらの結果をより一般化し、任意の固定された多層ニューラルネット構造に対して、その学習問題が NP-完全であることを示すことを試みている。そして、各層のノード数が2以上の任意のネットワーク構造の学習問題が NP-完全になることを示した。

(2)は、ブール関数の表現族として知られる順序つき2進決定ダイアグラム (ordered binary decision diagram, 以下 obdd) で、変数の順序が固定されている場合の MAT 学習アルゴリズムを示した。 n 変数の obdd は、1を割り当てる代入を n 次元ベクトルと考えれば、 $\{0,1\}^n$ の部分集合を表す言語とみなせることに注目し、決定性有限オートマトンの MAT 学習問題に帰着させることにより、効率の良い学習アルゴリズムを与えている。

Li および Vitányi により提案された Simple PAC 学習モデルでは、Valiant の PAC 学習モデルとは異なり、例は universal 確率分布に基づいて生成され、さらに、学習の成功基準は“simple な確率分布のもとで”近似的な概念を高い確率で出力することに緩められている。(3)は、Li らの論文で未解決問題とされていた $\log n$ -決定木の Simple PAC 学習可能性の問題を肯定的に解決した。

SESSION 8 (座長: Taisuke Sato)

Carlos Domingo and John Shawe-Taylor: The complexity of learning minor closed graph classes.

Domingo らは、グラフマイナーに関して閉じたグラフの集合の族を等価性問題と所属性質問を用いて多項式時間で同定するアルゴリズムを示した。また、渡辺の提案した質問学習における複雑度解析の枠組みを利用して、単調 DNF の等価性質問による学習を、このグラフ集合族の等価性質問による学習に帰着させることにより、このグラフ集合族の等価性質問のみからの効率の良い学習が不可能であることを示した。

SESSION 9 (招待講演) (座長: Hiroshi Motoda)

Yves Kodratoff: Technical and scientific issues of knowledge discovery in databases.

本講演では、近年注目を集めているデータベースからの知識の発見 (Knowledge Discovery in Databases, 以下 KDD) の分野で中心的な役割を果たしている、Kodoratoff 博士による KDD の定義およびその理論に関する解説が行われた。KDD と従来の機械学習 (Machine Learning, 以下 ML) の違いが、例を交えて説明されたが、報告者がつかみ得た限りでは、1) KDD は ML, 統計, データベースという 3 分野の統合科学であり, 2) そのデータ処理が, ユーザ指向かつ目的指向であり, 3) 得られた知識の評価において, ユーザとのインタラクションが非常に重要になる, という点が強調されていたように思う。講演後, 「KDD にできて ML にできないことは何か」などの本質的な質問を含む率直な意見交換・議論がされていたことが印象的であった。

SESSION 10 (座長: Arun Sharma)

- (1) Ken Sadohara and Makoto Haraguchi: On rejecting inappropriate similarities in analogical logic program synthesis systems.
- (2) Klaus P. Jantke: Reflecting and self-confident inductive inference machines.
- (3) Satoshi Kobayashi and Takashi Yokomori: On approximately identifying concept classes in the limit.
- (4) Andris Ambainis: Application of Kolmogorov complexity to inductive inference with limited memory.

(1)は、目標となる論理プログラムを、それに類似したプログラムから正・負の例を用いて、類推に基づいて学習する問題を取り扱っている。この類推に基づく学習では、効率的な学習をするためには、不適切あるいは不要な類似性を棄却することが重要になるが、本発表では、そのような能力を持つ効率の良い類推学習アルゴリズムが示された。

(2)は、学習者が自分に学習可能かどうかを内省できる能力を持つような学習モデルを極限同定モデルに基づいて提案し、その学習能力の分析を行った。特に、学習できない場合に直ちに内省可能なモデルと、内省まである程度の時間を許すモデルが考察され、直観に反して、両者の学習能力が等しいことが示された。

(3)は、学習において、学習対象となる概念 L が仮説空間に含まれない場合における対処法として、仮説空間において L を含む最小の概念(上近似と呼ばれる)を同定する学習問題を提案し、その特徴づけを行った。また、そのような近似学習が可能となる仮説空間の位相的特徴づけも示された。

短期記憶と長期記憶を持つ学習モデルにおいては、線形オーダの長期記憶があれば、任意の同定可能な関数は、短期記憶なしで同定可能なことが知られているが、長期記憶が sublinear に制限されたときに、どの程度の短期記憶が必要になるかは未解決問題であった。(4)は、この問題に対する完全な解答を与えた。

また、2日目の最後(Session8の終了後)には、Impromptu Talks のセッションが設けられ、以下の発表が行われた。

Impromptu Talks (座長: Frank Stephan)

- (1) Ming Li: On the Learnability of Decision Lists and Decision Trees.
- (2) Taisuke Sato: Distribution Semantics: A Step Toward the Integration of Programming & Learning.
- (3) Oksana Arnold and Klaus P. Jantke: Program Synthesis Algorithms for Therapy Plan Generation.
- (4) Andris Ambainis: Probabilistic DFIN Identification General Results.

それぞれが、口頭発表の内容に優るとも劣らない発表であり、質の高さがうかがえた。報告者には、決定リスト・決定木の学習の困難さを示した(1)の発表、論理プログラムの不動点意味論に確率を採り入れた(2)の発表が印象に残った。

3. おわりに

ALT は今回で6回目を数えるが、回を重ねるにつれ、その国際色の豊かさと発表の質が向上していくように感じられる。各口頭発表に関する詳細は、会議録としてまとめられ、Springer-Verlag 社から、Lecture Notes in Artificial Intelligence 997「Algorithmic Learning Theory」として出版されているので参照されたい。特に、これらのなかから厳選されたいくつかの論文が、Theoretical Computer Science の特集として掲載される予定である。

また、今回の会議参加者は、Technical Session 以外においても、レセプション、バンケット、福岡観光ツアーなどの番外セッション(?)で積極的に交流を深めていたようである。報告者には、実行委員会による手作りながらもきめの細かいサービスが印象的であった。ちなみに、次回の ALT '96 は 10 月下旬にオーストラリアのシドニーにて行われることが決定している。

〔小林 聡(電気通信大学)〕