

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 8 日現在

機関番号：12612

研究種目：基盤研究(C)

研究期間：2011～2014

課題番号：23500011

研究課題名(和文)形式言語の効率的学習アルゴリズムの開発及びその応用システムの構築

研究課題名(英文)Development of efficient learning algorithms of formal languages and construction of their application systems

研究代表者

若月 光夫(WAKATSUKI, Mitsuo)

電気通信大学・大学院情報理工学研究科・助教

研究者番号：30251705

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：決定性プッシュダウン変換器のスタック記号を1種類に限定した決定性限定1カウンタ変換器について、それが最終状態受理式の場合、より一般的なε-推移を持つ場合についてもその等価性判定及び包含性判定が多項式時間で行えることを明らかにした。また、実時間最終状態受理式決定性限定1カウンタ変換器に対して、所属性質問及び等価性質問を用いた多項式時間の学習アルゴリズムを開発した。更に、正則言語の部分クラスに対する正例からの極限同定を組み込んだジュウシマツの歌構造解析ツールEUREKAを利用することによって、コンピュータ上でトランプゲームの大貧民の対戦を行うプログラムの挙動の規則性が抽出可能なことを示した。

研究成果の概要(英文)：We have developed a polynomial-time algorithm for checking the equivalence of deterministic restricted one-counter transducers (DROCT's for short), which are deterministic pushdown transducers having just one stack symbol, that accept by final state with possible epsilon-moves. By extending this technique, we have also developed a polynomial-time algorithm for checking the inclusion of these DROCT's. Then, we have presented a polynomial-time algorithm for learning real-time DROCT's which accept by final state via membership and equivalence queries. Furthermore, we have shown that the regularity of the behavior of some programs for playing games of Daihinmin, which is one of card games, on computers can be extracted by using the EUREKA system which incorporated an identification algorithm of the class of k-reversible languages, which is a subclass of regular languages, for analyzing songs of the Bengalese finch.

研究分野：総合領域

 キーワード：計算論的学習理論 質問による学習 等価性判定 包含性判定 アルゴリズム 決定性文脈自由言語
 プッシュダウン変換器 多項式時間可能性

1. 研究開始当初の背景

人工知能の実現において機械学習は最も重要な研究分野の一つであり、計算論的学習理論はその可能性を数理的で厳密に解析を行うパラダイムである。そのような研究の対象として形式言語を選択し、本研究に先立つ科学研究費基盤研究(C)「形式言語に対する例からの学習を行う効率的アルゴリズムの開発・応用」においては、形式言語の部分クラスの中で実用上重要な決定性文脈自由言語を受理する決定性プッシュダウンオートマトン(DPDA)またはそれに対応した文法、あるいは DPDA に出力機構を付与した決定性プッシュダウン変換器(DPDT)に対し、その構造に妥当な制約を課した幾つかの部分クラスを対象として、以下に述べるような研究成果を上げた。

(1) 学習アルゴリズム開発の基礎構築

DPDA や DPDT が入力を読み込むことなく動作するとき、これらは ϵ -推移を持つと言い、そうでないとき、実時間であると言う。また、DPDA、及び DPDT のスタック記号を 1 種類に限定したものをそれぞれ、決定性限定 1 カウンタオートマトン(DROCA)、決定性限定 1 カウンタ変換器(DROCT)と呼ぶ。入力を読み込むことなく出力が限りなく続くような不自然な場合以外は ϵ -推移を許した空スタック受理式 DROCT について、対象とする DROCT の持つ性質を利用することによって、その等価性判定が効率的に行えること、すなわち多項式時間で行えることを次の発表で明らかにした。

・若月光夫, 清野和司, 富田悦次, 西野哲朗: 空スタック受理式決定性限定ワンカウンタ変換器の多項式時間等価性判定アルゴリズム, 2009 年度冬の LA シンポジウム, 2010 年 2 月 2 日, 京都大学理学研究科北部構内数学教室。

上記 ϵ -受理方式の異なる最終状態受理式の実時間 DROCT に対し、その等価性判定が多項式時間で行えることを明らかにした。

(2) 学習アルゴリズムの開発

DROCA が実時間空スタック受理式で、かつ各入力記号に対応する推移規則が高々 1 つである場合、Szilard strict DROCA と呼ぶ。この Szilard strict DROCA に対する、正例からの極限同定アルゴリズムを開発し、入力記号の種類を定数とみなせば、多項式時間極限同定可能なことを次の論文で明らかにした。

・M. Wakatsuki, E. Tomita: Polynomial time identification of strict deterministic restricted one-counter automata in some class from positive data, IEICE Trans. on Inform. Syst., Vol.E91-D, No.6, pp.1704-1718 (2008)。

拡張有限オートマトンに出力機構を付与した有限状態変換器(FST)のうち、strict prefix deterministic FST と呼ぶ部分クラスに対して、正例からの極限同定が多項式時間で行えることを次の論文で発表した。

・M. Wakatsuki, E. Tomita: Polynomial time identification of strict prefix deterministic finite state transducers, Lecture Notes in Artificial Intelligence, Vol.6339, pp.313-316 (2010)。

(3) 学習アルゴリズムの応用

ジュウシマツの囀り(単に鳥の歌と呼ぶ)の解析は、バウトと呼ばれる一息ごとの連続音声データを単位として扱われる。バウトは幾つかのソングユニットから構成されており、各ソングユニットの時系列データは正則言語の真部分クラスである k -可逆言語として表現される。我々は、Angluin が提案した k -可逆言語の正例からの極限同定アルゴリズムを組み込んだ、鳥の歌構造解析ツール「EUREKA」を開発してきているが、バウトからソングユニットを自動的に分割する手法を開発した。

以上のような研究成果から、これらの手法を更に上位の言語クラスに対する学習アルゴリズムの開発や応用に適用するための手がかりを得た。

2. 研究の目的

前記のようにこれまでに蓄積されてきた実績を最大限活用して、DPDA あるいは DPDT の中の、より広範な部分クラスに対して、計算論的な手法により MAT 学習等の質問を用いた学習や極限同定による学習を行うアルゴリズムを開発する。またそれと並行して、学習アルゴリズム開発の基礎となる等価性・包含性判定アルゴリズムを開発し、時間計算量解析を行って効率化を図る。更に、開発した学習アルゴリズムを適用することによって、鳥の歌構造解析やゲノム情報処理等の実際的な問題に適用してシステムの構築・改良を行うことで、本研究手法の有効性を明確にし、前基盤研究(C)の成果を大きく発展させることを本研究の目的とする。

3. 研究の方法

(1) 学習アルゴリズム開発の基礎構築

これまでに開発してきた DPDT のある部分クラスに対する等価性判定アルゴリズムを、対象とする DPDT 対に特化して単純化することによって、多項式時間で等価性判定及び包含性判定を行うアルゴリズムを開発する。

(2) 学習アルゴリズムの開発

1. (2)で述べたような DPDA または DPDT の構造に妥当な制約を課した部分クラスに対し、所属性質問や等価性質問等の質問を用いた多項式時間学習アルゴリズムを開発する。

(3) 学習アルゴリズムの応用

1. (3)で述べた k -可逆言語の正例からの極限同定アルゴリズムを組み込んだ、鳥の歌構造解析ツール EUREKA を利用することによって、鳥の歌構造の解析を更に進める。また、EUREKA を他の実際的な問題に適用する。

4. 研究成果

(1) 学習アルゴリズム開発の基礎構築

1. (1) で述べた実時間最終状態受理式 DROCT 対に対する等価性判定アルゴリズムを拡張し、より一般的な、 λ -推移を持つ最終状態受理式 DROCT 対に対しても、その等価性判定が多項式時間で行えることを示した。また、この等価性判定アルゴリズムの手法を拡張することによって、最終状態受理式 DROCT 対に対する包含性判定が多項式時間で行えることを明らかにした。

その受理記号列集合が本質的に正則を超えるような変換器に対する等価性判定が多項式的に行えるとの結果は、上記の他には、例えば次の論文等で示されている。

・ C. Bastien, J. Czyzowicz, W. Fraczak, W. Rytter: Equivalence of simple functions, *Theoretical Computer Science*, Vol.376, pp.42-51 (2007).

この論文は、我々が対象とした DPDT の部分クラスとは比較不能な関係にある、単純決定性プッシュダウン変換器同士の等価性判定が多項式的に行えることを示しているが、thickness (各スタック記号をポップアップするのに必要とする入力記号列の最短の長さ) をパラメータの一つとして用いており、対象 DPDT の記述長の指数オーダーになり得る。これに対して我々の結果は、対象 DPDT の記述長の多項式オーダーで厳密に抑えられる点で対照的である。

また、DROCA を真に含む決定性 1 カウンタオートマトン (DOCA) に対する包含性判定問題が非可解であるという事実から、我々の包含性判定に関する結果は、DPDA または DPDT の構造にどのような制約を課せばその包含性判定が可解となるかを探究するための手がかりとなり得る。

上記の等価性判定及び包含性判定に関する成果は、対象 DPDT に対する質問による学習に利用できる。

(2) 学習アルゴリズムの開発

1. (1) で述べた実時間最終状態受理式 DROCT に対して、所属性質問及び等価性質問によって正負の例を獲得し、目標の DROCT を多項式時間で学習するアルゴリズムを提案した。

DOCA の部分クラスに対する学習問題については、次の論文が発表されている。

・ A. F. Fahmy, R. S. Roos: Efficient learning of real time one-counter automata, *Lecture Notes in Computer Science*, Vol.997, pp.25-40 (1995).

この論文では、実時間最終状態受理式で、かつ受理言語が接頭辞性質を有する DOCA を対象としており、我々が対象とした DROCT の部分クラスとは比較不能な関係にある。

(3) 学習アルゴリズムの応用

1. (3) で述べた EUREKA を利用して、親鳥の歌から幼鳥が学習する過程や、親鳥とその子が獲得した歌構造の類似性に関して解析を

行った。また、我々はコンピュータ上でトランプゲームの大貧民の対戦を行うサーバ・クライアントシステムを開発してきているが、EUREKA を利用することによって、プログラマが作成したクライアントプログラムの挙動を解析し、その規則性が抽出可能なことを示した。

(4) その他

上記に関連して、下記 5. 主な発表論文等に示したように、最大クリーク抽出アルゴリズムやプログラムのソースコードの剽窃検出、学習ゲームを用いた発達障害児向けの文字学習支援システム、遺伝的アルゴリズムを用いた大貧民の思考ルーティンの進化的学習、量子セルオートマトンを用いた画像圧縮等について、有効な研究成果を得た。特に、最大クリークの抽出については DPDA や DPDT の学習を行う際、状態の統合・分離に利用できるため、より効率的な学習アルゴリズムを開発するための手がかりとなり得る。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 10 件)

1. Hiroshi Kikuchi, Takaaki Goto, Mitsuo Wakatsuki, Tetsuro Nishino: A source code plagiarism detecting method using sequence alignment with abstract syntax tree elements, *International Journal of Software Innovation*, Vol.3, No.3, pp.41-56 (2015). 査読有
DOI:10.4018/IJSI.2015070104
2. 中西裕陽, 富田悦次, 若月光夫, 西野哲朗: 最大クリーク問題の多項式時間的可解性の拡張の改良, *電子情報通信学会論文誌 D*, Vol. J97-D, No.6, pp.1106-1121 (2014). 査読有
<http://www.ieice.org/jpn/books/transaction.html>
3. Mitsuo Wakatsuki, Etsuji Tomita, Tetsuro Nishino: A polynomial-time algorithm for checking the equivalence for real-time deterministic restricted one-counter transducers which accept by final state, *International Journal of Computer and Information Science*, Vol.14, No.2, pp.45-53 (2013). 査読有
<http://www.acisinternational.org/ijcis.html>
4. Etsuji Tomita, Yoichi Sutani, Takanori Higashi, Mitsuo Wakatsuki: A simple and faster branch-and-bound algorithm for finding a maximum clique with computational experiments, *IEICE Transactions on Information and Systems*, Vol. E96-D, No.6, pp.1286-1298 (2013). 査読有

- <http://www.ieice.org/jpn/books/transaction.html>
5. 中西裕陽, 富田悦次, 若月光夫, 西野哲朗: 最大クリーク問題の多項式時間的可解性の拡張, 電子情報通信学会論文誌 D, Vol. J95-D, No.9, pp.1716-1728 (2012). 査読有
<http://www.ieice.org/jpn/books/transaction.html>
 6. 中西裕陽, 富田悦次, 若月光夫, 西野哲朗: 最大クリーク問題の多項式時間的可解性の更なる改良結果, 電子情報通信学会論文誌 D, Vol. J94-D, No.12, pp.2037-2046 (2011). 査読有
<http://www.ieice.org/jpn/books/transaction.html>
 7. Khan Md. Mahfuzus Salam, Tetsuro Nishino, Kazutoshi Sasahara, Miki Takahashi, Kazuo Okanoya: Information-theoretic analysis for understanding the behavior of song learning by the Bengalese finch, IPSJ Transactions on Mathematical Modeling and Its Applications, Vol.4, No.3, pp.49-58 (2011). 査読有
<https://ipsj.ixsq.nii.ac.jp/ej/>
 8. 若月光夫, 清野和司, 富田悦次, 西野哲朗: 実時間最終状態受理式決定性限定 1 カウンタ変換器の多項式時間等価性判定アルゴリズム, 京都大学数理解析研究所講究録, 1744 巻, pp.1-10 (2011). 査読無
<http://www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/1744.html>
 9. 中西裕陽, 富田悦次, 若月光夫, 西野哲朗: 最大クリーク問題の多項式時間的可解性について, 京都大学数理解析研究所講究録, 1744 巻, pp.169-176 (2011). 査読無
<http://www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/1744.html>
 10. 大畑和樹, 西野哲朗, 若月光夫: 量子セルオートマトンに基づく画像圧縮のための画像変換アルゴリズム, 京都大学数理解析研究所講究録, 1744 巻, pp.181-184 (2011). 査読無
<http://www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/1744.html>

[学会発表](計 15 件)

1. Mitsuo Wakatsuki, Etsuji Tomita, Tetsuro Nishino: A polynomial-time algorithm for checking the inclusion of deterministic restricted one-counter transducers which accept by final state, Proc. 30th International Conference on Computers and Their Applications, pp.11-17, 査読有, 2015 年 3 月 9 日, Honolulu (USA).
2. Mitsuo Wakatsuki, Etsuji Tomita, Tetsuro Nishino: A polynomial-time algorithm for checking the equivalence of deterministic restricted one-counter transducers which accept by final state, 15th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 査読有, 2014 年 7 月 1 日, Las Vegas (USA).
3. Hiroshi Kikuchi, Takaaki Goto, Mitsuo Wakatsuki, Tetsuro Nishino: A source code plagiarism detecting method using alignment with abstract syntax tree elements, Proc. 15th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, pp.375-380, 査読有, 2014 年 7 月 1 日, Las Vegas (USA).
4. 中西裕陽, 富田悦次, 若月光夫, 西野哲朗: 最大クリーク問題の多項式時間的可解性の拡張の更なる改良, 電子情報通信学会コンピュータリサーチ研究会/情報処理学会アルゴリズム研究会, 2014 年 6 月 14 日 道後温泉 大和屋 (愛媛県松山市).
5. 若月光夫, 富田悦次, 西野哲朗: 実時間決定性限定 1 カウンタ変換器に対する質問による多項式時間学習アルゴリズム, 2013 年度冬の LA シンポジウム, 2014 年 1 月 30 日, 京都大学数理解析研究所 (京都府京都市).
6. 菊池紘, 西野哲朗, 後藤隆彰, 若月光夫: 類似パターンの省略によりソースコードの可読性を向上させる新技法, 電子情報通信学会知能ソフトウェア工学研究会, 2013 年 9 月 12 日, 電気通信大学 東 3 号館 (東京都調布市).
7. Mitsuo Wakatsuki, Etsuji Tomita, Tetsuro Nishino: A polynomial-time algorithm for checking the equivalence for real-time deterministic restricted one-counter transducers which accept by final state, Proc. 14th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, pp.459-465, 査読有, 2013 年 7 月 2 日, Honolulu (USA). DOI:10.1109/SNPD.2013.18
8. 綾部孝樹, 大久保誠也, 西野哲朗: 大貧民プログラムの n-gram 統計による特徴量抽出とクラスタ分析, 情報処理学会数理モデル化と問題解決研究会, 2013 年 5 月 23 日, 北海道大学 百年記念会館 (北海道札幌市).
9. 金山貴泰, 浅野久美子, 西野哲朗, 若月光夫: 学習ゲームを用いた発達障害児向

け文字学習支援システム, 情報処理学会
数理解析モデル化と問題解決研究会, 2013 年
5 月 23 日, 北海道大学 百年記念会館(北
海道札幌市).

10. 若月光夫, 富田悦次, 西野哲朗: 決定性
限定 1 カウンタ変換器のある部分クラス
に対する質問による多項式時間学習アル
ゴリズム, 2012 年度冬の LA シンポジウ
ム, 2013 年 1 月 29 日, 京都大学数理解
析研究所 (京都府京都市).
11. 中西裕陽, 富田悦次, 若月光夫, 西野哲
朗: 最大クリーク問題の多項式時間的
可解性の拡張の改良, 電子情報通信学会コ
ンピューテーション研究会, 2012 年 9 月 3
日, 法政大学小金井キャンパス 梶野町校
舎 (東京都小金井市).
12. 海川祥毅, 西野哲朗, 若月光夫: 遺伝的
アルゴリズムを用いたコンピュータ大貧
民の思考ルーティンの自動チューニング,
電気通信大学エンターテイメントと認知
科学研究ステーション 第 6 回エンター
テイメントと認知科学シンポジウム,
2012 年 3 月 19 日, 電気通信大学 西 9 号
館 (東京都調布市).
13. 若月光夫, 清野和司, 富田悦次, 西野哲
朗: 最終状態受理式決定性限定 1 カウン
タ変換器の多項式時間等価性判定アルゴ
リズム 2011 年度冬の LA シンポジウム,
2012 年 1 月 31 日, 京都大学数理解析研
究所 (京都府京都市).
14. 中西裕陽, 富田悦次, 若月光夫, 西野哲
朗: 最大クリーク問題の多項式時間的
可解性の拡張, 電子情報通信学会コンピ
ューテーション研究会, 2011 年 10 月 21 日,
東北大学青葉山キャンパス 青葉記念会
館 (宮城県仙台市).
15. 中西裕陽, 富田悦次, 若月光夫, 西野哲
朗: 最大クリーク問題の多項式時間的
可解性の更なる改良結果, 電子情報通信学
会コンピューテーション研究会, 2011 年 4
月 22 日, 京都大学吉田キャンパス ベ
ンチャー・ビジネス・ラボラトリー (京
都府京都市).

〔図書〕(計 2 件)

1. Mitsuo Wakatsuki, Etsuji Tomita,
Tetsuro Nishino: A polynomial-time
algorithm for checking the equivalence
of deterministic restricted
one-counter transducers which accept
by final state, "Software Engineering,
Artificial Intelligence, Networking
and Parallel/Distributed Computing"
in Studies in Computational
Intelligence, R. Lee (Ed.), Springer
International Publishing, Vol.569,
pp.131-144 (2015). 査読有
DOI:10.1007/978-3-319-10389-1_10
2. 西野哲朗, 若月光夫, 後藤隆彰: 応用オ
ートマトン工学, コロナ社, 174 ページ

(2012).

〔その他〕

ホームページ等

[http://kjk.office.uec.ac.jp/Profiles/5/
0000405/profile.html](http://kjk.office.uec.ac.jp/Profiles/5/0000405/profile.html)

[http://www.nishino-lab.jp/laboratory/wa
katsuki/index.html](http://www.nishino-lab.jp/laboratory/wakatsuki/index.html)

6. 研究組織

(1) 研究代表者

若月 光夫 (WAKATSUKI, Mitsuo)

電気通信大学・大学院情報理工学研究科・
助教

研究者番号: 30251705

(2) 研究分担者

富田 悦次 (TOMITA, Etsuji)

電気通信大学・名誉教授

研究者番号: 40016598

西野 哲朗 (NISHINO, Tetsuro)

電気通信大学・大学院情報理工学研究科・
教授

研究者番号: 10198484