

機関番号：17102

研究種目：基盤研究（C）

研究期間：2008～2010

課題番号：20500001

研究課題名（和文） ランダム写像に基づく非線形概念の学習

研究課題名（英文） Learning non-linear concepts based on random projection

研究代表者

瀧本 英二 (TAKIMOTO EIJI)

九州大学・大学院システム情報科学研究所・教授

研究者番号：50236395

研究成果の概要（和文）：

ランダム写像と線形分類器を組み合わせ、非線形の幾何概念の学習を効率よく行う方式の実現可能性について、多方面から検討を行った。特に、与えられた分類規則の集合から適当なものをランダムに選び、それらの線形結合を仮説とするブースティングの手法が、ある意味で本学習方式を自然に実現していることを示した。一方、本学習方式において最も自然な仮説表現である、深さが定数のしきい値回路の複雑さの下界を与え、本学習方式の理論的な限界を示した。

研究成果の概要（英文）：

We investigate new schema for learning non-linear concepts by combining random projection with linear classifiers. In particular, we show that a boosting algorithm naturally results in our learning scheme, where the algorithm randomly chooses appropriate rules in a given rule class and produces as a final hypothesis a linear combination of them. On the other hand, we give lower bounds on the complexity of constant depth threshold circuits, which serve the most natural hypothesis class for our learning schema. The bounds imply theoretical limitations for our schema.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,300,000	390,000	1,690,000
2009年度	1,300,000	390,000	1,690,000
2010年度	900,000	270,000	1,170,000
年度			
年度			
総計	3,500,000	1,050,000	4,550,000

研究分野：計算学習理論

科研費の分科・細目：情報学・情報学基礎

キーワード：計算学習理論，ブースティング，マージン最大化，ランキング学習，オンライン予測，エネルギー複雑度，論理式複雑さ

1. 研究開始当初の背景

機械学習に現れる多くの問題は、 $n$ 次元ユークリッド空間上にサンプル  $S \subseteq \mathbb{R}^n \times \{-1, 1\}$  が与えられたとき、正のラベルが付いたベクトル集合  $\{(x, y) \in S \mid y = 1\}$  と負のラベ

ルが付いたベクトル集合  $\{(x, y) \in S \mid y = -1\}$  を分離する仮説超曲面を求める問題として定式化される。このような問題を統一的に扱う1つの方法は、適当な写像  $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^N$  を用いてサンプル  $S$  を線形分離可能なサン

プル $\{(\phi(x), y) \mid (x, y) \in S\}$ に変換し、元の幾何概念の学習問題を $N$ 次元空間における線形分離関数の学習問題に還元することである。しかし、そのためには一般に $N$ を指数的に大きくする必要があるので、写像先の $N$ 次元空間で陽に線形分離関数を求めることは極めて非効率的である。広く普及しているサポートベクトルマシンは、カーネル $K(x, y) = \phi(x) \cdot \phi(y)$ を用いて仮説線形分離関数の双対表現を求めることにより、計算量を $N$ に依存しないようにする手法である。しかし、その一方で規模の大きい2次計画問題を解く必要があり、サンプルのサイズが大きいとき、その計算量を減らすことが中心的な課題の1つとなっていた。

## 2. 研究の目的

本研究は、ランダム写像の技術を用いることにより、サンプル $S$ を線形分離可能とするような低次元空間への非線形写像を構成し、その空間で高速な線形計画法等を用いて陽に分離超平面を求めるという学習方式について、その可能性と限界を明らかにすることを目的とする。

## 3. 研究の方法

(1) 写像 $\phi$ とランダム行列による写像との合成写像について検討する。ランダム行列は高い確率でサンプルの線形分離性を保存する性質があるため、合成写像の計算効率の問題さえ解決できれば、直ちに本研究の目的としている学習方式が得られる。

(2) ブースティングの技術を用いて、仮説クラス $H$ の仮説を $\phi$ の各次元として逐次的に求めていく手法、すなわち、各ラウンド $t=1, 2, 3, \dots$ において、 $\phi_t(x) = (h_1(x), h_2(x), \dots, h_t(x))$ ,  $h_t \in H$ のように写像を構成する手法について検討する。

(3) 本研究で提案する学習方式において、最も自然な仮説は、深さが3のしきい値回路としてモデル化することができる。そこで、深さ定数のしきい値回路の複雑さを明らかにし、本手法の計算論的な限界について検討する。

## 4. 研究成果

(1) 短いランダムシードを用いて構成した擬似ランダム行列が、なお高い確率でサンプルの線形分離性を保存することを示した。これは、本研究課題で考える学習方式を実現するための条件の一つを解決するものである。一方、全体の計算効率については、依然として不明であり、今後の展開が待たれるところである。

(2) データ間の順序付けを学習するランキング学習問題に対し、ブースティングの手法を用いた新しいアルゴリズム SoftRankBoost

を提案した。SoftRankBoostは、ランキング関数のクラス $H$ とサンプル $S$ が与えられたとき、 $H$ に属する関数の線形結合によって定義される関数の中で、 $S$ のソフトマージンをほぼ最大にするものを効率よく求めるものである。本手法は、理論的な性能保証が得られている手法の中で最も効率が良く、実際の性能も優れていることが実験でも明らかとなっている。また、 $H$ をランダム写像のクラスとし、カーネル手法と組み合わせることにより、本研究課題で考える学習方式が得られる可能性がある。

(3) 深さ定数のしきい値回路は、エネルギー複雑度を制限すると、多くの自然な関数を表すのに指数個のゲート数が必要となることを示した。回路のエネルギー複雑度は、回路を特徴写像 $\phi$ とみなしたとき、特徴ベクトル $\phi(x)$ のノルムの大きさの上界を与えるものである。したがって、エネルギー複雑度が小さいことは、線形分離仮説の汎化性能の観点から望ましいことであるが、本結果は、提案学習方式の効率化の限界を示唆している。また、回路の素子の機能をしきい値素子ではなく非線形な関数に拡張しても、特徴ベクトルのノルムを小さく抑えようとすると、対称性を持つ関数を学習するためには指数的に大きい素子数が必要であることを示した。この結果は、提案学習方式を拡張したとしても、その効率化に限界があることを意味している。

(4) 与えられたサンプルに対し、 $l_1$ ソフトマージンが最大となる仮説超平面を求める新しいアルゴリズム SparseLPBoost を提案した。この問題は、事例集合と特徴空間への写像の集合をそれぞれ行と列に対応させた行列によって定まる線形計画問題として定式化できるが、本手法は、この問題を小さい部分行列上の問題に帰着することによって、計算効率の飛躍的な改善を図るものである。実際、さまざまなデータに対する計算機実験によって、本手法の有効性が確かめられた。これは、疎なプロジェクションをランダム写像として用いた効率のよい非線形概念の学習方式とみなすことができる。

(5) 写像に基づく学習方式の開発の一環として、 $n$ 個の要素の順列(ランキング)を仮説空間とするオンライン学習の問題に取り組んだ。この問題は、情報検索を始め、商品推薦、倒産リスク予測、バイオインフォマティクス、自然言語処理など様々な応用分野を持つ。標準的なオンライン学習の手法では、特徴ベクトル空間の任意の点を、与えられた凸多面体に写像する過程が本質的な部分であることが知られている。本研究では、順列の特徴ベクトルを $O(n^2)$ 次元の比較ベクトルで表現した場合と、 $O(n)$ 次元の置換多面体の端点で表現した場合について、それぞれ効率

の良い写像アルゴリズムを構築するとともに、提案学習方式の理論的な精度保証を与えた。

(6) 量子質問計算量の下界を表す指標を新たに導入し、決定的計算に対する量子計算の効率化の限界を示した。また、ある制約を満たす単調積和形論理式のうち、充足割り当ての個数を最大にする、あるいは最小にするものを特定することに成功した。さらに、ある広い論理関数のクラスに対し、最小のブール式表現、もしくはその表現のサイズの精密な下界を求める効率のよい手法を与えた。多項式サイズで深さ定数のしきい値回路のクラスは、多項式サイズのブール式のクラスに含まれる ( $TC^0 \subseteq NC^1$ ) ため、この結果は、本学習方式の効率の精密な見積もりを与える可能性がある。

(7) ランキング学習の手法を用いて、将棋の局面評価関数を学習するシステムのプロトタイプを開発した。これは、実際の棋譜データを用いて、プロ棋士が打ちそうな手を、それ以外の手に比べてランクが上位になるように学習を行おうとするものである。予備実験では、わずか100局程度のデータで学習したにもかかわらず、美濃囲いや穴熊などの駒組みに高い評価値を与えるなど、特に序盤において自然な振る舞いを示すことが観察された。

## 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11件) (すべて査読有)

- ① Kei Uchizawa, Eiji Takimoto, Takao Nishizeki, Size-energy tradeoffs for unate circuits computing symmetric Boolean functions, *Theoretical Computer Science*, 412, 773-782, 2011.
- ② Kei Uchizawa, Takao Nishizeki, Eiji Takimoto, Energy and depth of threshold circuits, *Theoretical Computer Science*, 411, 3938-3946, 2010.
- ③ Hideaki Fukuhara, Eiji Takimoto, Kazuyuki Amano, NPN-representatives of a Set of Optimal Boolean Formulas, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E-93A, 6, 1008-1015, 2010.
- ④ Hideaki Fukuhara, Eiji Takimoto, Lower Bounds on Quantum Query Complexity for Read-once Formulas with XOR and MUX Operators, *IEICE Transactions on Information and Systems*, E-93D, 2, 280-289, 2010.
- ⑤ Kohei Hatano, Eiji Takimoto, Linear

Programming Boosting by Column and Row Generation, *Lecture Notes on Artificial Intelligence*, 5808, 401-408, 2009.

- ⑥ Kei Uchizawa, Takao Nishizeki, Eiji Takimoto, Energy Complexity and Depth of Threshold Circuits, *Lecture Notes on Computer Science*, 5699, 335-345, 2009.
- ⑦ Kei Uchizawa, Takao Nishizeki, Eiji Takimoto, Size and Energy of Threshold Circuits Computing Mod Functions, *Lecture Notes on Computer Science*, 5734, 724-735, 2009.
- ⑧ Hideaki Fukuhara, Eiji Takimoto, Lower bounds on quantum query complexity for read-once decision trees with parity nodes, *Theory of Computing 2009. Proc Fifteenth Computing: The Australasian Theory Symposium (CATS 2009)*. CRPIT, 94, 474-487, 2009.
- ⑨ Kei Uchizawa, Eiji Takimoto, Exponential lower bounds on the size of constant-depth threshold circuits with small energy complexity, *Theoretical Computer Science*, 407(1-3), 474-487, 2008.
- ⑩ Jun-ichi Moribe, Kohei Hatano, Eiji Takimoto, Masayuki Takeda, Smooth Boosting for Margin-Based Ranking, *Lecture Notes on Artificial Intelligence*, 5254, 227-239, 2008.
- ⑪ Takayuki Sato, Kazuyuki Amano, Eiji Takimoto, Akira Maruoka, Monotone DNF Formula that has a Minimal or Maximal Number of Satisfying Assignments, *Lecture Notes on Computer Science*, 5092, 191-203, 2008.

[学会発表] (計16件)

- ① Shota Yasutake, Online Prediction over Permutahedron, 情報処理学会アルゴリズム研究会, 2011年3月7日, 琉球大学.
- ② 奥山 洋平, 確率的評価値をもつゲーム木における最善手探索, 冬のLAシンポジウム, 2011年2月2日, 京都大学.
- ③ 末廣 大貴, カーネル法を用いたコンピュータ将棋の評価関数の学習, 第15回ゲームプログラミングワークショップ (GPW-10), 2010年11月12日, 箱根セミナーハウス.
- ④ 安武 翔太, オンラインランク統合問題, 第13回情報論的学習理論ワークショップ (IBIS 2010), 2010年11月5日, 東京大学.
- ⑤ 末廣 大貴, SVMによるバイパータイトラ

- ンキング学習を用いたコンピュータ将棋における評価関数の学習, 第13回情報論的学習理論ワークショップ(IBIS 2010), 2010年11月4日, 東京大学.
- ⑥ Shota Yasutake, Online rank aggregation, 第9回情報科学技術フォーラム(FIT 2010), 2010年9月9日, 九州大学.
- ⑦ Kei Uchizawa, Size and Energy of Unate Circuits Computing Symmetric Boolean Functions, The 13th Japan-Korea Joint Workshop on Algorithms and Computation (WAAC 2010), 2010年7月23日, 金沢市文化ホール.
- ⑧ 吉田 真一, 重みつき窓を用いた適応型オンライン予測, 人工知能学会データマイニングと統計数理研究会, 2010年3月30日, 統計数理研究所.
- ⑨ Kei Uchizawa, Size-Energy Tradeoff of Unate Circuits Computing Symmetric Functions, 電子情報通信学会コンピュータセッション研究会, 2010年1月25日, 九州大学.
- ⑩ 畑埜 晃平, 行と列の生成による線形計画ブースティング, 第12回情報論的学習理論ワークショップ(IBIS 2009), 2009年10月19日, 九州大学.
- ⑪ 福原秀明, 長方形分割数に対する緩和による論理式のサイズの下界, 夏のLAシンポジウム, 2009年7月22日, かんぼの宿松嶋.
- ⑫ Hideaki Fukuhara, Lower Bounds on Quantum Query Complexity for Read-once Formulas with XOR and MUX Operators, Annual Meeting of Asian Association for Algorithms and Computation (AAAC 2009), 2009年4月11日, 杭州(中国).
- ⑬ 福原秀明, 最簡な論理式でNPN同値類の代表のみを生成するアルゴリズム, 電子情報通信学会コンピュータセッション研究会, 2009年3月2日, 東京工業大学.
- ⑭ 内沢啓, ブール剰余関数を計算するしきい値論理回路のサイズとエネルギー複雑度のトレードオフ, 電子情報通信学会コンピュータセッション研究会, 2008年10月10日, 東北大学.
- ⑮ Hideaki Fukuhara, Lower bounds on quantum query complexity for decision trees, 電子情報通信学会コンピュータセッション研究会, 2008年10月10日, 東北大学.
- ⑯ Jun-ichi Moribe, Smooth Boosting for Margin-Based Ranking, 人工知能学会データマイニングと統計数理研究会, 2008年7月23日, 小樽市民センター.

〔図書〕(計1件)

- ① 瀧本 英二, 丸善, 現代数理科学事典(第2版) 3.4.3 節学習アルゴリズム, 2009年.

〔産業財産権〕

○出願状況(計0件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
出願年月日:  
国内外の別:

○取得状況(計0件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
取得年月日:  
国内外の別:

〔その他〕

ホームページ等

<http://www.i.kyushu-u.ac.jp/~ei ji>

## 6. 研究組織

### (1) 研究代表者

瀧本 英二 (TAKIMOTO EIJI)

九州大学・システム情報科学研究院・教授

研究者番号: 50736395

### (2) 研究分担者

( )

研究者番号:

### (3) 連携研究者

( )

研究者番号: