

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年 5月29日現在

機関番号：11301

研究種目：基盤研究（C）

研究期間：2009～2011

課題番号：21540105

研究課題名（和文） 離散幾何対象の機械学習の新しい効率評価理論の

研究課題名（英文） Efficiency for machine to learn discrete geometric objects

研究代表者

赤間 陽二（AKAMA YOHJI）

東北大学・大学院理学研究科・准教授

研究者番号：30272454

研究成果の概要（和文）：多様体の連結成分の個数の上限の評価法を研究し、主成分分析を統計的学習と見なした時の VC 次元を評価した。集合族の帰納推論の効率を表す為に新しい順序型を導入し、整礎擬順序が我々の順序型を持つ適当な集合族で表現可能で、我々の意味での順序型を持つ集合族から Cantor 連続な単調関数による像として得られる集合族も我々の意味での順序型を持つ事を証明した。整礎擬順序理論の展開である Better Quasi-Order 理論に対応するように我々の理論も展開した。

研究成果の概要（英文）：By studying evaluation methods of the number of connected components of manifolds, we evaluated the VC dimensions of principal component analysis, as statistical learning. Next, we introduce a new order type of set systems to measure the difficulty to learn. We proved (1) any well quasi-ordering can be represented by a set system having our order type; and (2) if a set system has our order type, then a continuous image of it by a Cantor monotone function does so. Finally, we developed the theory of our order type of set systems to the theory of better quasi-orderings.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,100,000	330,000	1,430,000
2010年度	700,000	210,000	910,000
2011年度	700,000	210,000	910,000
年度			
年度			
総計	2,500,000	750,000	3,250,000

研究分野：数物系科学

科研費の分科・細目：数学一般(含確率論・統計数学)

キーワード：VC次元、連結成分、順序型、整列擬順序、連続変形

1. 研究開始当初の背景

（離散）幾何手学的対象の学習が応用上重要であるが、計算論的学習の広範な広がりとともに、学習効率をまず数学の観点から定式化して評価することが重要と考えた。情報学に

においても学習の困難さの指標として Vapnik-Chervonenkis 次元 (VC 次元) が頻繁に言及されるが、ニューラルネットの学習の VC 次元を詳細に求める程度で、その他の学習システムの VC 次元は求められていなくて、代

表的な統計推論の VC 次元すら求められていない。

2. 研究の目的

組み合わせ論・確率論（大数の法則）に由来する VC 次元を数学的に洞察し、離散幾何・組合せ対象の統計的・計算論的学習理論を研究し、また、学習理論から組合せ論、確率論のテーマを開拓する。

3. 研究の方法

学習対象全体(概念クラス)を多様体と見なす一方で、学習可能な対象全体に集合族の組合せ論(Bollobas)を適用した。

前者に関しては、主成分分析と呼ばれる統計推論の効率評価のために、Vapnik、Dudley の経験過程理論と Talagrand らの測度集中不等式を適用した。

前者を適用するため、経験過程理論という一般理論における Vapnik-Chervonenkis 次元をこの場合に求めた。その概念は組み合わせ的に定義されているが、多様体の連結成分の個数を求めればよい。そこで微分幾何的手法(Warren の定理)とホモロジー論的手法(Basu, S.)について研究した。

4. 研究成果

主成分分析とは、高次元 d のデータをより低次元 $k < d$ のアフィン空間で近似する統計的推論である。主成分分析が統計的学習として誘導する概念クラスは、 k 次元アフィン空間 H から一定距離 $r > 0$ より近い点全体の集合 $B(H, r)$ を全て集めたクラス C である。

$C(d, k)$ の VC 次元を主成分分析の VC 次元と呼ぶ。主成分分析の VC 次元を求めるために、多様体の連結成分に関する上限に関して考察し、主成分分析の VC 次元の漸近的に厳密な評価として

$$\Theta((d-k+1)(k+1))$$

を与えた。

主成分分析の VC 次元の上界を求める意義は、VC 次元に関わる経験過程理論(大数の法則に関する一般理論[Dudley, van der Geer])を用いることにより、Wishart 行列の固有値の和の一致性の片側を示せることである。もう片方の一致性は、幾何学における測度集中理論(Talagrand, ...)を用いて示せることを上野 康隆氏とともに注意した。この手法は、Z. D. Bai や Johnstone らによる、乱雑行列の固有値の一致性を求める手法と異なる点で意義がある。

VC 次元の上界を求めるための多様体の連結成分に関する Warren の定理を、Milnor-Thom の定理を用いて簡単に証明し、主成分分析の VC 次元の下界のための簡潔な漸化式を与えた(入江慶氏による)。なお、我々は主成分分析の VC 次元の下界を与え点集合として、正単体と、原点および基本基底との直積を与えた。

あわせて、測度集中不等式を適用するために、Latala-Oleszkiewicz の tensorization property、対数 Sobolev 不等式、情報エントロピーと測度集中不等式をサーベイした。さらに、高次元ユークリッド空間における凸体からの一様サンプリングの効率評価との関連を見た。

また、結晶格子を正例から無矛盾な方法で学

習する困難性を表す概念である有限の弾力性 (Wright, Motoki-Shinohara-Wright) を定量化することを提案し、結晶格子と似た構造が数理言語学に現れることを注意し、組み合わせ的・代数的な手法で研究した。

その組合せ手法に着想を得て、集合族の学習過程を教師と学習者との間のゲームとみて、そのゲーム木の順序型を集合族の新しい順序型として定義し、集合族の帰納推論の効率を表す新しい順序型を導入した。我々は

- (i) 整礎擬順序が我々の順序型を持つ適当な集合族で表現可能である。
- (ii) 我々の意味での順序型を持つ集合族から Cantor 連続な単調関数による像として得られる集合族も我々の意味での順序型を持つ

事を証明した。

また、擬順序とその間の有限分岐模倣関係が作る圏から、集合族とその間の Cantor 連続な Girard 線形な単調関数全体からなる圏への、injective-on-objects, 忠実充満な関手で、擬順序の maximal order type (de Jongh-Parikh) を集合族の我々の順序型に移すものを与えた。また、その関手と左随伴関手についての、counit が恒等関手の恒等的自然変換であることを示した。擬順序とその間の有限分岐模倣関係が作る圏は、理論計算機科学における concurrency theory に有用と思われるので、集合族とその間の Cantor 連続な Girard 線形な単調関数全体からなる圏は、応用が期待される。

Nash-Williams が整礎擬順序の理論を Better Quasi-ordering 理論へと展開したように、我々の理論を展開し Better quasi-ordering に対応する集合族の理論を構築した。Better Quasi-ordering に対応する集合族が

勝手な単調関数の像で閉じていることを示した。

更に、勝手な集合族 L に対して、その有限個の要素たちの合併全体を考えたが、我々は整礎擬順序と Better quasi-ordering の理論の観点から、後者の新しい集合族が順序型を持つような、 L の特徴づけを与えた。その特徴づけは、 L の包含関係による順序構造の、通常の順序型とは関係しないことが判明した。

このように学習理論に動機づけられて、集合族の組合せ論が展開できたが、Better quasi-ordering は Petri net の性質の決定可能性を証明するのに使われているので、Better quasi-ordering の拡張概念である我々の集合族も、Better quasi-ordering と同様の応用があると期待する。

最後に、我々の集合族の連続変形が計算論における Jockusch-Owings の (weakly) semirecursive set への正帰着に近いことを指摘した。

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 7 件)

1. Yohji Akama. A new order theory of set systems and better quasi-orderings” Progress in Informatics, Vol. 9, pp. 9 - 18, (2012) (査読あり)

2. Yohji Akama. Set systems: order types, continuous nondeterministic deformations, and quasi-orders, Theoretical Computer Science, Vol. 412, pp. 6235 - 6251 ,

(2011). (査読あり)

3 . Yohji Akama, Kei Irie, Akitoshi Kawamura, Yasutaka Uwano. VC Dimensions of Principal Component Analysis” Discrete and Computational Geometry. Vol.44, No.3, pp.589-598, (2010). (査読あり)

4 . Yohji Akama and Yasutaka Uwano. VC Theory and a Concentration Inequality for Sums of Eigenvalues of Wishart Matrix. (社)人工知能学会 データマイニングと統計数理研究会, pp. 214-219, (2010). (査読あり)

5 . Yohji Akama. Commutative Regular Shuffle Closed Languages, Noetherian Property, and Learning Theory, Lecture Notes in Computer Science, Vol. 5457. pp. 93-104 (2009). (査読あり)

6 . 赤間 陽二. 可換シャッフルクローズド正規言語, ネーター性, および極限学習. 人工知能学会研究会資料 SIG-FPAI-A901. (2009). (査読なし)

7 . Yohji Akama. Gaussian Mixture models and VC dimensions. 人工知能学会研究会資料 SIG-FPAI-A801, (2009). (査読なし)

[学会発表] (計 8件)

1 . Yohji Akama. A new order theory of set systems, WQO AND BQO, SigTheory/Algs Talk, Department of Computer and Information Sciences, University of Delaware, USA, 2012年3月7日

2 . Yohji Akama. A new order theory of set

systems, WQO and BQO. Verification Seminar, 北陸先端科学技術大学院大学, 金沢, 2012年1月31日

3 . Yohji Akama. Set systems: Order types, continuous nondeterministic deformations, and quasi-orders. 第4回九州大学組合せ数学セミナー, 福岡, 2011年11月26日

4 . Yohji Akama and Yasutaka Uwano. Consistency of eigenvalues for principal component analysis Statistics and Operations Research Seminars, Department of Economics, Pompeu Fabra University, Spain, 2010年5月6日

5 . 赤間 陽二, 上野 康隆. 主成分分析の固有値の一致性について, 第12回データマイニングと統計数理研究会. 統計数理研究所, 東京, 2010年3月30日.

6 . 赤間 陽二. 可換シャッフルクローズド正規言語, ネーター性, および極限学習, 人工知能学会第74回FPAI研究会, 広島市立大学, 広島, 2009年9月14日. (招待講演)

7 . Yohji Akama. Gaussian Mixture Models and VC Dimensions, The Fourth International Workshop on Data-Mining and Statistical Science, Kyoto, 8 July 2009.

8 . Yohji Akama. Commutative Regular Shuffle Closed Languages, Noetherian Property, and Learning Theory, The 3rd international conference on Language and automata theory and application, Tarragona, Universitat Rovira i Virgili, Spain, 3 April 2009.

[図書] (計 0件)

〔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

6. 研究組織

(1) 研究代表者

赤間 陽二 (AKAMA YOHJI)
東北大学・大学院理学研究科・准教授
研究者番号：30272454

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：