

平成 21 年 5 月 25 日現在

研究種目：若手研究（B）  
研究期間：2007～2008  
課題番号：19700223  
研究課題名（和文） 判別問題に対する実践的かつ高精度な機械学習アルゴリズムの確立  
研究課題名（英文） Development of a Practical and Accurate Machine Learning Algorithm for Classification Problems  
研究代表者  
中田 洋平（NAKADA, Yohei）  
早稲田大学・理工学術院・助手  
研究者番号：70434298

## 研究成果の概要：

本研究の目的は、「判別問題に対する実践的かつ高精度な機械学習アルゴリズムの開発」、および、「その具体的アプリケーションの確立」である。より具体的には、(1)基本アルゴリズムの開発、(2)アプリケーションの確立、(3)数理的考察をこの中で推し進めるものである。本研究はこれを遂行し、その成果は博士学位論文、国際会議、論文誌・学術誌などでの発表・掲載により公開した。

## 交付額

(金額単位：円)

	直接経費	間接経費	合計
2007 年度	2,200,000	0	2,200,000
2008 年度	500,000	150,000	650,000
年度			
年度			
年度			
総計	2,700,000	150,000	2,850,000

研究分野：総合領域

科研費の分科・細目：情報学，感性情報学・ソフトコンピューティング

キーワード：確率的情報処理，ベイズ学習，ブースティング

## 1. 研究開始当初の背景

判別問題とは、ある対象の特徴量（入力）が与えられたときに、その特徴量を基に、対象の属性（出力）を判断させるような問題である。この問題では、特徴量が与えられたとき、属性を予測する機械をつくることを最終的な目的とする。このような問題は、様々な分野で登場し、特に工学分野では、文字・音声・画像識別をはじめ、個人認証、不正検出、イベント検出など、細分化された様々な応用で見られる。

このような判別問題に対して強力なアプローチに、機械学習アプローチがあった。これは、最初に入出力関係を例題（データ）として与えて学習機を学習することで、未知の入力に対する出力を可能とするアプローチである。機械学習の中でも、特に高精度なことで注目を集めるものが、アダプストをはじめとするブースト学習アルゴリズムであった。これらのブースト学習アルゴリズムでは、単体では弱い判別能力しか保持しない学習機械、いわゆる弱学習機と呼ばれるものを、逐次的に学習・追加していくことで、高精度で柔軟性のある大規模な機械を効率良く作ることが可能となる。しかしながら、アダプストなどは、学習用データ内に潜むノイズの影響に弱いことが知られていた。そのため、様々な損失関数の改良、正則化などの観点から改良法が提案されてきたが、未だ確立された改良法はなく、より強力な改良法が求められていた。

一方、機械学習の分野では、ノイズが激しい環境などでも有効に働き、実践的なことで知られる手法として、ベイズ学習がある。ベイズ学習は、その過程にある積分計算に困難があったが、近年の計算機性能の向上とモンテカルロ手法などの数値計算アルゴリズムの発展によって実用化に至り、様々な実問題に成功裏に適用されてきていた。

## 2. 研究の目的

本研究の目的は、「判別問題に対する実践的かつ高精度な機械学習アルゴリズムの開発」、および、「その具体的なアプリケーションの確立」である。前述のブースト学習アルゴリズムとベイズ学習を統合することで、ブースト学習アルゴリズムの弱点を克服し、実践的かつ高精度な判別問題用の機械学習アルゴリズムを開発することを目的としている。また、その具体的なアプリケーションの確立も本研究の範囲である。より具体的に述べれば、本研究では、当該期間内に下記の(1)基本アルゴリズムの開発、(2)アプリケーションの確立、(3)数理的考察の3項目を実施することを目的となっている。

### (1)基本アルゴリズムの開発

実践性の高い基本アルゴリズムを開発することは、本研究の根幹である。実際の判別問題で用いるデータベースには、劣悪なノイズ（特徴量への激しいノイズ、属性のミスラベルなど）、欠損値などが混入していることが多い。このような環境を考慮した実験により、実装した幾つかのアルゴリズムのバリエーションについて、性能検証を行い、学習特性・誤差特性を明らかにする。その結果を活用し、アルゴリズムを改良する。

### (2)アプリケーションの確立

既に初期的な適用実験を実施した画像オブジェクト検出を含め、他にもどのような応用が考えられるか検討を進める。また、その応用に必要なデータベースの作成を行う。作成したデータベースには、適用実験に用いられる。適用実験で得られた結果を用いてアプリケーションに特化したアルゴリズムに改良し、その特化アルゴリズムから得られた結果を基に、アプリケーションのプロトタイプを実装する。

### (3)数理的考察

本研究は、上述の(1)、(2)に主眼を置くものであるが、同時に数理的考察も実施する。具体的には、ベイズ学習およびブースト学習アルゴリズムの数理的側面についての調査を実施する。その後、それらの結果を踏まえ、他ブースト学習アルゴリズムや他ベイズ学習との関連、判別誤差の性質などを、数理的側面から考察する。また、考察結果の妥当性を計算機実験などで検証する。

## 3. 研究の方法

前述のように、本研究では、(1)基本アルゴリズムの開発、(2)アプリケーションの確立、(3)数理的考察の3項目の遂行を目的とする。以降、申請時に立案した各項目に対する各年度での研究方法と発表・論文計画について述べる。

平成 19 年度

(1) 幾つかの基本アルゴリズムのバリエーションについて、整理・検討し、実装する。その後、Web・文献調査や公開データベースの調査を通し、様々な実データを把握する。その中で、どのような劣悪なデータが想定されるか検討する。その結果を受け、劣悪なデータ環境を考慮した複数の実験計画の策定を行う。実験計画に従い、実装した基本アルゴリズムの各バリエーションを検証する。検証結果をもとに各バリエ

ーションの学習特性・誤差特性を考察する。

(2) 実データ調査結果を基に、どのようなアプリケーションが適切か整理・検討する。その結果を受け、アプリケーション構築前の適用実験に際し、必要なデータベースを構築する。データベースが整備された段階で、適用実験を実施する。その結果をもとに、基本アルゴリズムを改良・拡張し、対象とするアプリケーションに特化した特化アルゴリズムを検討する。

(3) ベイズ学習およびブースト学習アルゴリズムの数理的側面について文献・書籍調査を実施し、それらの結果を資料化する。その結果を活用し、基本アルゴリズムと他ブースト学習アルゴリズムや他ベイズ学習との関連や、基本アルゴリズムの汎化誤差などを、数理的側面から考察する。

平成 20 年度

(1) 平成 19 年度に纏めた結果をもとに、各バリエーションの中で、その改良方法を検討する。この改良手法についても、平成 19 年度と同様の実験計画に従い、性能検証を実施する。なお、必要に応じて、実験計画は再策定する。

(2) 平成 19 年度に検討した特化アルゴリズムを実装し、適用実験を再度実施する。必要に応じて、特化アルゴリズムの見直し、改良を行い、その都度、適用実験を繰り返す。適用実験が終了した段階で、実験により得られたモデルやパラメタを内包するアプリケーションのプロトタイプを実装する。

(3) 平成 19 年度で実施した結果を受け、考察結果の正当性を数値実験などで検証する。必要があれば、再度、数理的側面からの考察を見直し、再数値実験を行い、結果をブラッシュアップする。

以上が当初の研究計画となる。なお、研究自体は後述の 4. 研究成果に示すように概ね上記の計画に沿って進めた。

#### 4. 研究成果

前述のように、本研究の目的は、「判別問題に対する実践的かつ高精度な機械学習アルゴリズムの開発」、および、「その具体的なアプリケーションの確立」である。より具体的には、(1)基本アルゴリズムの開発、(2)アプリケーションの確立、(3)数理的考察をこの中で推し進めるものである。当該期間内に上記の研究計画の基で実施

した各項目についての研究成果について述べる。

(1) 19 年度に実施した基本アルゴリズムの各バリエーションの実装をもとに、20 年度に更なる検証を進めた。またその結果を、報告者の博士学位論文：「モンテカルロ法に基づくベイズ学習の拡張に関する研究」(早稲田大学、博士論文(工学), 2009 年 2 月)の一部として書類化した。また、オンライン化に繋げるための第一歩の研究を進め、後述する国際会議：IEEE Machine Learning for Signal Processing 2008 で発表を実施した。なお、この国際会議には予稿集論文(査読あり)が存在する。

(2) 具体的なアプリケーションの候補として画像オブジェクト検出への適用を研究協力者らとともに実施し論文化した。これは、映像情報メディア学会誌、および、画像ラボに掲載された。更に、他のアプリケーションの考察を行ったが、これは現段階で未発表である。

(3) 19 年度に行った文献・書籍調査を受け、その結果を活用し、20 年度に基本アルゴリズムと他ブースト学習アルゴリズムや他ベイズ学習との関連を数理的側面から考察した。更にこの結果を前述の博士学位論文「モンテカルロ法に基づくベイズ学習の拡張に関する研究」に記載した。なお、博士論文に記載した内容については、現在、書籍化する計画もあり、博士論文の書籍化申請を早稲田大学に実施している。

#### 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 2 件)

1) 松井淳, 後藤祐, 木村彰夫, 中田洋平, 松本隆, クリッピングデルサイモン, 藤井真人, 八木伸行, "GibbsBoost 顔検出と映像監視業務への応用", 映像情報メディア学会誌, Vol.62, No.3, pp.408-413, 2008 年 3 月. 査読有

2) 松井淳, 後藤祐, 木村彰夫, 中田洋平, 松本隆, クリッピングデルサイモン, 藤井真人, 八木伸行 "GibbsBoost 顔検出と映像監視業務への応用", 画像ラボ, Vol.19, No.8, pp.16-21, 2008 年 8 月. 査読無

〔学会発表〕(計 1 件)

1) Kazue Sega, Yohei Nakada, and Takashi Matsumoto "Online Bayesian Learning for Dynamical Classification Problem Using Natural Sequential Prior", IEEE International Workshop on Machine Learning

for Signal Processing, pp.392-397, Oct.16-19,  
2008, Cancun, Mexico.

〔その他〕

中田洋平,「モンテカルロ法に基づくベイズ学習  
の拡張に関する研究」, 博士学位論文, 早稲田大  
学, 2009年2月.

## 6. 研究組織

### (1) 研究代表者

中田洋平 (NAKADA YOHEI)  
早稲田大学・理工学術院・助手  
研究者番号: 70434298