

機関番号：32601

研究種目：若手研究（B）

研究期間：2009～2010

課題番号：21700262

研究課題名（和文） ベイズ推論の導入による機械学習アルゴリズムの実践化と高精度化

研究課題名（英文） Improvement of machine learning algorithms via Bayesian inference

研究代表者

中田 洋平（NAKADA YOHEI）

青山学院大学・理工学部・助手

研究者番号：70434298

研究成果の概要（和文）：

本研究は、報告者が継続的に成してきている「ベイズ推論の導入による機械学習アルゴリズムの実践化と高精度化」に関する一連の研究の中に属するものである。当該期間内には、主として、(1) オンライン型ベイズ学習用の自然な事前分布に対する研究課題、(2) ベイズ学習の実応用に関する研究課題に関する研究成果が得られた。また、これらの研究課題に関連して、著書 1 件の発刊し、共同研究者らとともに計 8 件の学会発表を実施した。

研究成果の概要（英文）：

This study is included in a series of studies of “Improvement of machine learning algorithms via Bayesian inference” by the reporter. In the period of this study, the results are obtained, which are mainly related to (i) natural prior for online Bayesian machine learning and (ii) applications of real data with Bayesian machine learning with real data. In addition, one book related to this study was published, and the results related to this study were presented in 8 conferences/workshops with the co-researchers.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009 年度	2,300,000	690,000	2,990,000
2010 年度	500,000	150,000	650,000
年度			
年度			
年度			
総計	2,800,000	840,000	3,640,000

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

1. 研究開始当初の背景

研究開始当初から現在に至るまで、情報ネットワーク網の整備、高性能計算機の普及、データベースの拡充など社会全体を覆う情報インフラの急速な発展の中において

いる。

このような中で、個人、企業、教育機関、研究機関などあらゆる主体が、膨大な量の多様な形式のデータを蓄積し、また、その間に膨大な量のデータが情報ネットワーク網を介して電子的に流通するようになってきて

いる。このような状況の変化に伴い、これまで得られてきた情報科学や統計科学、その周辺分野の様々な研究成果を活用し、膨大かつ多様な形式のデータから有用な法則、モデルなどを自動的に獲得するようなデータ分析に関連する技術群も急速な発展を見せてきている。

特に、データを用いて予測機能などを持つ複雑な学習機械モデルを構築するような機械学習技術はその発展が著しく、情報化社会の高度化の要請により、現在も更なる発展も望まれている。

このような中、疎なデータや雑音の大きいデータに対する頑健性を保持し、先験知識を推論体系に取り込む柔軟性のあるベイズ推論を機械学習技術に取り入れることは、機械学習技術の実践性や予測精度の向上の観点から非常に有望な方策と考えられている。研究開始当初はもとより、現在でも活発化し続けている研究分野である。

2. 研究の目的

本研究は、上記のような研究背景の下で、報告者が継続的に成してきている「ベイズ推論の導入による機械学習アルゴリズムの実践化と高精度化」、および、「その方法論の確立」を最終目標とする一連の研究の中に属するものである。

当該期間においては、ベイズ推論を導入した機械学習（以降、ベイズ学習）の中にある課題の中でも、データの背後に潜む動的变化への対応などを主たる解決対象としている。これには、動的变化に対応可能なオンライン型のベイズ学習を行う際に用いる事前分布が重要であり、当該期間においては、この事前分布に関する研究課題を中核にしている。

さらに、本研究では、ベイズ学習の実践的な研究課題の推進も目的の範疇にあり、幾つかのデータに対する応用も重要な研究課題と位置付けている。

3. 研究の方法

当該期間中は、主に以下に示す項目を中心の作業とした。

(1) オンライン型ベイズ学習用の自然な事前分布に対する研究課題：

オンライン型ベイズ学習用の自然な事前分布に対する研究課題については、報告者が研究してきている事前分布の性質や他の事前分布との関連などを明らかにすべく、数理

的考察を実施した。

(2) ベイズ学習の実応用に関する研究課題：

ベイズ学習の実応用に関する研究課題については、共同研究者らとともに多様なデータでの適用可能性を示すべく、スポーツデータ、画像データなどへの実適用に関連する研究を推進した。

また、上記(1)(2)とは別に、平成22年度からは、ベイズ学習のさらなる発展のために、ベイズ学習の実装手段の高度化が必須であることなどから、ベイズ学習の実装手段に関する研究課題にも着手した。

4. 研究成果

前述のような研究方法の遂行により得られた主たる研究成果を以下に示す。

(1) オンライン型ベイズ学習用の自然な事前分布に関する研究課題：

データの背後に潜む動的变化への対応を主たる解決対象として、これまでも報告者が研究してきている自然な事前分布に関する数理的側面について研究を進めた。その結果の一部を含む「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」と題した書籍を平成21年度に発刊した。なお、同書籍には非統計学的モデルに対するベイズ学習の導入に関する研究成果なども示されている。

更に、報告者が提案してきたオンライン型の事前分布と、主として一括型のベイズ学習で用いられてきたエントロピー事前分布との関係を明らかにし、その関係を示した論文を執筆した。なお、これは本報告書作成時点において投稿済みである。

また、上記とは他の観点から、共同研究者らとともに混合分布族の自然な事前分布として考えられるディリクレ過程を用いたベイズ学習に対して、一括型、オンライン型の双方から事後確率最大化 EM アルゴリズムによる実装方法の研究課題に取り組んだ。これについては、当該期間中に

- “Maximum a posteriori estimation for Dirichlet process language models”,

- “Dirichlet process EM algorithm for semi-supervised learning”

と題した海外発表2件が関連する研究成果として得られている。

(2) ベイズ学習の実応用に関する研究：

本研究に関連するモデルやアルゴリズムを用いた実応用に関する研究については、当該期間中に

- “An MCMC algorithm for gene regulatory network prediction with Bayesian network” ,

- “ディリクレ過程事前分布を用いた一般物体認識のための確率生成モデルの拡張と推定法” ,

- “A gene regulatory network prediction algorithm using a Gaussian Bayesian network model with a Box-Cox transformation” ,

- “A non-parametric Bayesian algorithm for predicting gene regulatory networks with a Gaussian process” ,

- “A novel gene ontology prediction algorithm using infinite mixtures of hidden Markov and binary models with a Dirichlet process prior” ,

- “ベイズ隠れマルコフモデルを用いたスポーツイベント検出の高精度化” ,

と題した計6件の学会発表が研究成果として得られている。なお、この中の1件は前述のディリクレ過程を用いたベイズ学習の実装方法を用いたものでもある。

また、これらの成果は、スポーツイベントデータ、画像データなど多様なデータに対するベイズ学習の応用に関するものであり、本研究に関連するモデル、アルゴリズムの幅広い適用可能を示す成果群と位置付けることができる。

更に、当該期間中では本研究課題に関する幾つかの論文投稿や投稿した論文査読対応作業を実施している。また、前述のように平成22年度からは、ベイズ学習の新たな実装手段に関する研究なども進めており、これについては、今後、適切な発表機会にて発表していく所存である。

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)

[学会発表] (計8件)

① H. Miyachika, J. Maruyama, T. Kaburagi, Y. Nakada, T. Matsumoto, T. Kimiwada, and K. Wada, “An MCMC algorithm for gene regulatory network prediction with Bayesian network”, InCoB2010, 2010年9月28日, Tokyo, Japan.

② 能鹿島 武志、徳田 貴昭、中田 洋平、松本 隆、“ディリクレ過程事前分布を用いた一般物体認識のための確率生成モデルの拡張と推定法”、MIRU2010、2010年7月29日、釧路市観光国際交流センター。

③ H. Miyachika, T. Kimiwada, J. Maruyama, Y. Nakada, T. Kaburagi, T. Matsumoto, and K. Wada, “A gene regulatory network prediction algorithm using a Gaussian Bayesian network model with a Box-Cox transformation”, ISMB2010, 2010年7月12日, Boston, MA, USA.

④ T. Kikuchi, Y. Nakada, T. Kaburagi and T. Matsumoto, “A non-parametric Bayesian algorithm for predicting gene regulatory networks with a Gaussian process”, ISMB2010, 2010年7月12日, Boston, MA, USA.

⑤ T. Kaburagi, N. Tagoto, K. Oota, T. Tokuda, Y. Nakada and T. Matsumoto, “A novel gene ontology prediction algorithm using infinite mixtures of hidden Markov and binary models with a Dirichlet process prior”, ISMB2010, 2010年7月11日, Boston, MA, USA.

⑥ 矢崎智浩、三須俊枝、中田洋平、本井滋、小林剛、松本隆、八木伸行、“ベイズ隠れマルコフモデルを用いたスポーツイベント検出の高精度化”、パターン認識・メディア理解研究会、2010年3月16日、鹿児島大学工学部。

⑦ T. Kimura, Y. Nakada, T. Tokuda and T. Matsumoto, “Dirichlet process EM algorithm for semi-supervised learning”, 7th workshop on Bayesian nonparametrics, 2009年6月23日, Tronto, Italy.

⑧ T. Tokuda, T. Kimura, Y. Nakada and T. Matsumoto, “Maximum a posteriori estimation for Dirichlet process language models”, 7th workshop on Bayesian nonparametrics, 2009年6月23日, Tronto, Italy.

〔図書〕（計 1 件）

①中田洋平、早稲田大学出版、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、2009、306.

6. 研究組織

(1) 研究代表者

中田 洋平 (NAKADA YOHEI)
青山学院大学・理工学部・助手
研究者番号：70434298