

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 25 年 5 月 30 日現在

機関番号：82626

研究種目：基盤研究(C)

研究期間：2010～2012

課題番号：22500142

研究課題名（和文）プライバシー保護のための情報幾何的協調フィルタリング

研究課題名（英文）PRIVACY PRESERVING COLLABORATIVE FILTERING BASED ON INFORMATION GEOMETRY

研究代表者

赤穂 昭太郎（SHOTARO AKAHO）

独立行政法人産業技術総合研究所・ヒューマンライフテクノロジー研究部門・

研究グループ長

研究者番号：40356340

研究成果の概要（和文）：

個人情報保護を保証した協調フィルタリングを情報幾何の枠組みにおける統計的推定問題として捉え、そのための評価規準であるロバスト推定の規準について理論的な最適性規準を発見した。また、情報幾何で用いられるベータダイバージェンスがロバスト制御の分野で研究されている $H\infty$ フィルタと関連していることを示し、協調フィルタリングの時系列への拡張可能性を示した。一方、協調フィルタリングを行うための次元圧縮とクラスタリングの同時最適化についてベイズ推定の枠組みから高速な推定法を開発した。また、プライバシー保護を発展させた公正配慮型学習という枠組みを考案し、そのための基礎的な検討を行った。

研究成果の概要（英文）：

In this project, the collaborative filtering preserving privacy was treated as a statistical inference in the information geometric framework. We studied the robust criterion for the privacy preservation and obtained the optimal condition. We further showed the relationship between beta divergence criterion of the information geometry and the H infinity filter in the field of robust control, which shows the possibility to apply the collaborative filtering to time series data. For the collaborative filtering, we developed the simultaneous optimization procedure of dimension reduction and clustering from Bayesian framework. Moreover, we found out a fairness-aware learning framework based on privacy preservation and investigated its fundamental property.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,200,000	360,000	1,560,000
2011年度	1,100,000	330,000	1,430,000
2012年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：知識発見とデータマイニング、情報幾何、推薦システム

1. 研究開始当初の背景

個人個人の嗜好に適応するために協調フ

ィルタリングによる推薦システムの構築が可能になってきたが、医療診断や創薬に適用

するには個人情報取扱いが実用化のボトルネックになっていた。一方、このように個人情報を保護しつつデータ解析を行う枠組みとしてプライバシー保護データマイニングという匿名化の手法が研究され始めており、これらを融合することにより、推薦システムの応用範囲を拡大できるのではないかと考えた。ちょうど研究開始当初に代表者らは、情報幾何学の枠組みにおける多変量解析法を開発していたが、それを拡張することによって協調フィルタリングの匿名化が実現できると考えて本課題を提案した。情報幾何の枠組みを導入することにより、個人情報保護を情報論的に論じることができるといふほかに、計算論的な観点からも有利であることが示唆された。すなわち、数年前から並列計算の枠組みとして分散データマイニングという研究が盛んに行われるようになってきていた。分散データマイニングによって匿名化処理を分散化して行うことができ、リスクをより軽減させることができる。匿名化の後の統合化のところの問題となるが、情報幾何の枠組みを使うと、統計量の統合をシステムティックに行うことができるという利点があり、本研究課題が開発する技術によって大規模なプライバシー保護協調フィルタリングのシステムを実現することができる考えた。

2. 研究の目的

個人の嗜好や特性に合わせた推薦などのサービスを提供する際にプライバシー保護が問題となる。本研究課題では、応募者が従来培ってきた情報幾何学的データ解析法を「全体」と「個」という二つの相反する規準が統一的に扱えるように拡張し、プライバシー保護の問題に情報理論的な視点から新たな規準を提案するとともに、その規準を満たしながら精度の高い推薦システムを構築するための手法の開発を目的とする。

3. 研究の方法

研究は大きく分けて(1)情報幾何に基づく評価規準に関する基礎的な研究、(2)協調フィルタリングを用いた推薦システムの実現に向けた実装上のアルゴリズムに関する研究、の2つに分けられる。

(1) 情報幾何に基づく評価規準については、主に個人情報保護に用いられる規準は通常の最小二乗のようなものではなく、より外れ値に鈍感なものである必要がある(そうでなければ個人情報が守れない)。そこでロバスト推定との関連性で情報幾何的な多変量解析法を見直すことによって結果がどのように変わるかを理論や数値実験を用いて調べた。

(2) 協調フィルタリングを用いた推薦シ

ステムの実現へ向けたアルゴリズムについては、情報幾何に基づいた多変量解析では情報量が重要な役割をはたすため、情報量に基づく協調フィルタリング法について研究を進めた。

4. 研究成果

(1) まず、プライバシー保護に対する情報幾何学的データ解析法について以下のような結果を得た。まず、個人情報保護のための評価基準として、ロバスト推定で用いられる k 乗ノルムと Least Median Square 規準の拡張である $L_{\alpha}PS$ 規準について考察した。特に従来凸性という制約から $k \geq 1$ のみが考慮されてきたが、我々はよりロバスト性の強い $0 < k < 1$ の場合について考慮した点、および Least Median Square 規準を拡張した点に新規性がある。我々の理論的な解析により、回帰問題においては k 乗ノルム最小化の最適解が $0 < k < 1$ の範囲では必ずサンプル点を通るものであることを示した。また、この理論的な成果に基づき、高速な近似最適化をランダムサンプリングによって行うアルゴリズムを開発した。数値実験の結果、 $L_{\alpha}PS$ 規準が k 乗ノルム規準と同等にロバストな性能をもつことを示した。

また、情報幾何において用いられるベータダイバージェンスが制御の分野で研究されている H_{∞} フィルタと関連していることを示し、協調フィルタリングの概念を時系列を含む枠組みへと拡張することができる可能性を示した。

一方、情報幾何の枠組みで協調フィルタリングを行う際には、次元圧縮とクラスタリングを同時に行い、ユーザのグループ分けと情報の圧縮を行う必要がある。その観点でベイズ的な定式化を行い、高速に解を得るための近似手法である変分ベイズ法を用いた最適化を行い、実データを用いて実装を行った。

(2) プライバシーを保護したデータマイニングについて研究を進めるうち、単なる推薦ではなく、プライバシー情報をユーザが指定し、それに対して情報量の観点から推薦などのためのデータ解析を行う公正配慮型学習という枠組みを考案した。公正配慮型学習では、学習の精度(真の分布と学習分布との間のダイバージェンス)と公正性(出力と差別属性との相互情報量)との間のトレードオフを適切に調節することによって公正性を保ちつつデータ解析を行うという目指す。そのための基礎的な検討を行い、最終的に実データを用いた数値実験により有効性を確認した。

(3) 本研究課題が今後に残した課題を以下でまとめる。まず、プライバシーを保護するための評価基準については、プライバシー保護データマイニングで研究されている規準

との関係は完全には明らかになってはいない。そもそも従来は情報幾何的なデータ解析法に基づくプライバシー保護データマイニングの手法は全く研究されて来なかったため、異なる枠組みの概念を比較する必要がある。今後はプライバシー保護に関する社会的なコンセンサスなど実用的な観点からそれらを統合するような枠組みが必要となると考えられる。今回の研究課題の中では、全体と個という当初掲げたトレードオフについては十分に検討できなかった。これについても引き続き検討の余地が残されている。次に公正配慮型学習の研究について述べる。公正配慮型学習はプライバシーの中でも差別的な属性に着目し、得られる結果が差別的にならないことを目指す研究である。これまでの研究では、公正性は差別属性と出力属性の独立性をできるだけ高く達成することが必要であるが、予測精度とこの独立性のトレードオフを適切に制御することは現状ではまだかなりの困難を伴うため、今後さらに計算アルゴリズム的な研究が必要である。また、元々のプライバシー保護の観点も同様のモデル化によって最適化することができれば統一的な枠組みになると考えられる。

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8件)

- ① T. Kamishima, S. Akaho, H. Asoh, J. Sakuma, “Fairness-Aware classifier with prejudice remover regularizer”, Proc. of ECMLPKDD (LNCS7524), pp. 35-50, 2012
DOI: 10.1007/978-3-642-33486-3_
- ② T. Kamishima, S. Akaho, H. Asoh, J. Sakuma, “Considerations on Fairness-aware Data Mining”, Proc. of IEEE ICDM Workshops, pp. 378-385, 2012
DOI: 10.1109/ICDMW.2012.101
- ③ J. Fujiki, S. Akaho, H. Hino, N. Murata, “Robust hypersurface fitting based on random sampling approximations”, Proc. of ICONIP (LNCS7665), pp. 520-527, 2012
DOI: 10.1007/978-3-642-34487-9_63
- ④ J. Fujiki, Shotaro Akaho, “Hypersurface fitting via Jacobian nonlinear PCA on Riemannian space”, Proc. of CAIP2011, pp. 236-243, 2011
DOI: 10.1007/978-3-642-23672-3_29
- ⑤ J. Fujiki, S. Akaho, H. Hino, N. Murata, “Robust Hyperplane Fitting Based on k-th Power Deviation and

alpha-Quantile”, Proc. of CAIP 2011, pp. 278-285, 2011

DOI: 10.1007/978-3-642-23672-3_34

- ⑥ T. Kamishima, S. Akaho, J. Sakuma, “Fairness-aware Learning through Regularization Approach”, Proc. of IEEE ICDM Workshops, pp. 643-650, 2011
DOI: 10.1109/ICDMW.2011.83
- ⑦ 赤穂昭太郎, 渡辺一帆, 岡田真人, “指数分布族の空間におけるデータ解析法について”, 統計数理, Vol. 58, pp. 167-183, 2010
<http://www.ism.ac.jp/editsec/toukei/toukeisuri-58j.html>
- ⑧ K. Watanabe, S. Akaho, S. Omachi, M. Okada, “Simultaneous Clustering and Dimensionality Reduction Using Variational Bayesian Mixture Model”, Classification as a Tool for Research, Springer, pp. 81-89, 2010
DOI: 10.1007/978-3-642-10745-0_8

[学会発表] (計6件)

- ① 福永修一, 赤穂昭太郎, “ベータダイバージェンスを用いた動的システムの状態推定”, 情報論的学習理論ワークショップ IBIS2011, 奈良, 2011
- ② 赤穂昭太郎, 藤木淳, “曲線あてはめにおける特徴空間の特性について”, 情報論的学習理論ワークショップ IBIS2011, 奈良, 2011
- ③ 福永修一, 赤穂昭太郎, “ H^∞ フィルタとベータダイバージェンスの関係について”, 計測自動制御学会制御理論シンポジウム, 大阪, 2011
- ④ 赤穂昭太郎, 藤木淳, “半定値計画緩和による多項式コスト関数の大域的最適化”, 情報論的学習理論ワークショップ IBIS2010, 東京, 2010
- ⑤ 藤木淳, 赤穂昭太郎, “回帰大作戦—1次元低い超平面あてはめとしての回帰—”, 情報論的学習理論ワークショップ IBISML 研究会, 大阪, 2010
- ⑥ T. Kamishima, S. Akaho, “Nantonac Collaboration Filtering - Model Based Approach”, ACM Recommender Systems, バルセロナ (スペイン), 2010

6. 研究組織

(1) 研究代表者

赤穂 昭太郎 (SHOTARO AKAHO)

独立行政法人産業技術総合研究所・ヒューマンライフテクノロジー研究部門・情報数理研究グループ長

研究者番号: 40356340

(2) 研究分担者

神寫 敏弘 (TOSHIHIRO KAMISHIMA)
独立行政法人産業技術総合研究所・ヒュー
マンライフテクノロジー研究部門・主任研
究員
研究者番号：50356820

藤木 淳 (JUN FUJIKI)
福岡大学・理学部・准教授
研究者番号：10357907
(H23 年度追加)