様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月　 ３ 日現在

機関番号：１２１０２
研究種目：挑戦的萌芽研究
研究期間：2013～2014
課題番号：２５５４００９４

研究課題名（和文）機械学習における自己情報コントロール機構の構築

研究課題名（英文）Establishment of Self-information Control Mechanism for Machine Learning

研究代表者
佐久間　淳（Sakuma, Jun）

筑波大学・システム情報系・准教授

研究者番号：９０３７６９６３

交付決定額（研究期間全体）：（直接経費）　　2,900,000円

研究成果の概要（和文）：機械学習における中立化は、差別、不公平、偏見の要因となるような属性である視点と、学習の結果得られる分類器の出力が相関しないようにすることによって達成される。この研究では、 分類学習や推薦において、視点と予測の相関をコントロールする中立化の枠組みを主に研究した。特に分類学習において、未知の事例に対する汎化的な中立性について理論的な解析を行ったところ、 未知事例に対する汎化的な中立性が $O(1/\sqrt{n})$ で確率的にバウンドできることが導かれたことが最大の成果である。

研究成果の概要（英文）：The neutrality of predictions made by machine learning is measured by dependency between values of a specified attribute and predictions. In this research, we developed a framework that controls dependency between prediction and a specified attribute values to achieve fairness, privacy protection, and prevention of discrimination of predictions. One of the significant results of our study is the generalization analysis of neutrality, in which we proved that generalization neutrality can be probabilistically upper-bounded by $O(1/\sqrt{n})$ for unseen examples.

研究分野：プライバシー

キーワード：プライバシー　機械学習　データマイニング　セキュリティ

１．研究開始当初の背景
実世界から取得した種々の情報を入力とした、機械学習による予測モデルの構築が、様々な社会サービスの高度化に寄与することが期待されている。しかし個人から詳細な情報を収集した場合には、 プライバシ保護の観点から取扱に注意が必要であることは近年強く認識されるようになり、多くのプライバシ保護手法が提案されている。これらの多くは暗号化、匿名化やランダム化によって、データ全体が一定のプライバシ保護基準を満たすことを目指すものである。

この研究では、プライバシ情報の利用において利用者が指定した属性について、予測結果が強く依存しないことを保証することで、プライバシ保護や公正性の確保を実現することを目指す。

２．研究の目的
データ提供者はデータ利用者にプライバシ情報を提供し、データ利用者は提供データから、機械学習により予測モデルを得る状況を考える。データ利用者の立場からは、 あらかじめ指定した属性について、予測結果が強く依存しないような予測モデル学習を行うことで、プライバシを侵害するような予測結果や、差別的な予測結果について被害を受けることを防ぐことができる。このような、指定属性と予測値の間の依存性のコントロールを実現しつつ、モデルの予測精度の劣化を最小限に抑える学習法の構築と、その予測精度の劣化の上界の評価を目指す。

３．研究の方法
k 匿名性によるデータ匿名化によるプライバシ保護手法や、あらかじめ指定した属性値と予測が強く関連を持たないように学習する中立化の手法において、プライバシ・中立化と、予測精度のトレードオフ関係を理論的に解析するとともに、実際のデータによりその性能を実験的に評価する。

４．研究成果
（H25 年度の成果）技術やサービスの発展に伴い、大量のデータが蓄積され利用されるようになった。 例えば、病院では患者の疾患、治療の記録が保存され、ショッピングサイトでは購買履歴や商品の閲覧履歴などを収集している。 これらのデータを解析することで、将来の疾患リスクの予測や商品の推薦など、多くの有用なサービスへの利用が可能となる。また、これらのデータやサービスを繋ぎ合わせ連携させることによって得られる価値は大きい。

しかし、病院やサービス提供者が持つデータを解析機関に提供したり、連携のために別のサービス提供者へデータを受け渡す場合、疾患や治療の記録、住所や購買履歴など、 他人に知られたくない情報が含まれるためプライバシの問題を解決する必要がある。そのために、これらのデータが一体誰のものなのかを一意に特定できないようにする必要がある。そのための技術がデータ匿名化である。

匿名化では、定義した指標に基づき元のオリジナルデータに対して個人が識別できないようにデータを加工する。データの有用性を元データからの歪みと解釈するならば、この加工によって有用性は一般に低下する。極端な例を言えば、匿名化していないオリジナルデータは、最大の有用性を持ち、他の誰とも識別できないように匿名化したデータは匿名性は最大であるが有用性は極めて低い。この場合、匿名性と有用性の間にはトレードオフがあると考えられ、匿名性と有用性を両立させることは一般的に困難である。

一方で、有用性を匿名化のためのデータの加工が、データ解析の出力に与える変化量と解釈した場合、この匿名性と有用性の間のトレードオフは、実験的に評価はされてきた。実験的評価はデータやデータ解析手法に依存する。匿名化がデータ解析に与える影響をより一般的に議論するためには、プライバシと有用性についての理論的解析が必要である。

我々は、匿名化データを用いた協調フィルタリングに関する研究において、プライバシと有用性が単純なトレードオフの関係ではないことを過去の研究では実験的に示した。この研究では対象をより単純な線形回帰に絞り、プライバシと有用性の関係を統一的に記述した。さらに、この関係が理論的にも単純なトレードオフの関係にないことを示した。具体的には、以下の貢献があった:

① 匿名化のためのデータ変更量をロバスト最適化における摂動と解釈し、これに対しロバスト最適化の枠組みで匿名化データから学習する方法を導入した。
② 有用性とプライバシをカバリングナンバーにより統一的に記述し、関係付けた。さらに、ロバスト線形回帰について、この有用性とプライバシの関係が必ずしもトレードオフとならないことをカバリングナンバーによる記述に基づき、理論的に示した。

(H26 年度の成果) 機械学習のアルゴリズムを実用化するためには、機械学習によって行われる分類や予測から差別、不公平、偏見を排除しなければならない。中立化は、差別、不公平、偏見の要因となるような属性である視点と、学習の結果得られる分類器の出力が相関しないようにすることによってこの問題の解決する。この研究では、 経験損失最小化の目的

関数に対して、視点と分類器の出力が相関していることに対する罰則項である中立性リスクを加えることによって中立化を行う neutralized empirical risk minimization (NERM) という枠組みについて議論した。

具体的には中立性リスクとして、視点と分類器の出力の共分散を基にして定義される共分散中立性リスクを提案した。さらに共分散中立性リスクによる NERM において、未知の事例に対する汎化的な中立性について理論的な解析を行ったところ、未知事例に対する汎化的な中立性が $O(1/n^{1/2})$ で確率的にバウンドできることが導かれた。以下、具体的に研究成果を説明する。

Empirical Risk Minimization (ERM) は、入力 x と目標 y の集合に対する経験損失が最小となる仮説 f を獲得することで教師付き学習を行う枠組みである。この研究では、ERM に対して新たに視点仮説 g を導入し、視点仮説に対する中立化について述べる。

仮説 f は入力 x に対する目標の予測 $y = f(x)$ を与える関数であり、視点仮説 g は与えられた入力 x に対する視点の予測 $v = g(x)$ を与える関数である。f と g を区別するため、f を目標仮説と呼ぶ。目標仮説が視点仮説に対して中立であるとは目標 $f(x)$ と視点 $g(x)$ との間の相関が小さい状態のことを指し、中立化とは教師あり学習において予測の精度を保持したまま、与えられた g に対して中立な目標仮説 f を得るための学習法である。この研究では、ERM を基にした中立化の新しい枠組みである Neutralized ERM (NERM) を提案した。

中立化が解決する問題の一つとして、filter bubble [Pariser 11] があげられる。例えば、ユーザの興味に応じた記事配信システムを考える。このとき、入力 x としてアクセスログなどを収集し、目標 y である記事がユーザの好みかどうかを目標仮説 $y = f(x)$ によって予測する。あるユーザは世論を 2 分するような政策に偏った意見を持っており、視点仮説 $v = g(x)$ によってどちらの意見なのか予測できたとする。もし、$f(x)$ と $g(x)$ が強く相関していると、片方の意見に関する記事ばかり推薦していることになり、政策に偏見を与えかねない。偏見を排除するためには、目標仮説と視点仮説の出力 $f(x), g(x)$ が互いに相関しないようにする必要がある。

記事配信システムは、過去のユーザの記事の好みに関するデータを用いて学習を行うが、推薦はまだ読まれていない記事に対してする必要がある。従って、目標仮説から偏見を排除するためには、未知の記事に対して中立である必要がある。このように、教師あり学習における中立性は、未知の入力 x に対応する目標 $f(x)$ と視点 $g(x)$ の相関性によって測られる。これは、教師あり学習において分類器の精度を測る汎化損失と似たような基準であることから、未知の事例に対する中立性の性能を汎化中立性と呼ぶ。

教師あり学習における中立化の目的は、汎化中立性が保証されると同時に汎化損失を最小となる目標仮説を獲得することである。汎化中立性と汎化損失の間のよいトレードオフを得ることが、中立化における課題となる。

この研究では、分類問題において中立化を行う学習アルゴリズムの枠組みとして NERM を提案した。NERM は、出力 y が 2 値である ERM において、2 値である視点 v に対して中立化を行うアルゴリズムを構築できる。NERM は、ERM に対して中立性が低いことに対する罰則項を加えた最適化問題として定式化される。最適化問題の目的関数は、パラメータによって分類と中立性の性能のトレードオフを制御することができる。

NERM の最適化問題は凸となるため、NERM によって構築された学習手法は大域的最適解を保証できる。

中立化を行う方法としてヒューリスティックな方法 [Calders 10] や最適化を基にした手法 [Kamishima 12], [Zemel 13], [Fukuchi 13] があり、どの手法も与えられた事例について経験的に計算される中立性の評価量を基にして中立化を行っているため、未知事例に対する中立性の保証は無い。しかし、汎化中立性を保証するための、理論的な解析は行われていなかった。

この研究では、NERM の枠組みにおける汎化中立性の確率的バウンドに関する理論的解析を行った。NERM の枠組みを用いた適用例として、2 値の線形分類器として高い性能が示されている Support Vector Machine [Vapnik 98] において中立化を行った。提案する中立 SVM は、双対問題を導くことによってカーネル化を行うことができる。カーネル化を行うことによって入力の非線形な特徴を用いて分類を行うことができ、分類の精度と中立性の間の高いトレードオフが実現できることが期待できる。

５．主な発表論文等
〔雑誌論文〕（計 3 件）

① Kazuto Fukuchi, Jun Sakuma: Neutralized Empirical Risk Minimization with Generalization Neutrality Bound. ECML/PKDD (1) 2014: pp. 418-433. 10.1007/978-3-662-44848-9_27,査読有

② Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, Jun Sakuma: Correcting Popularity Bias by Enhancing Recommendation Neutrality. Poster Proceedings of RecSys 2014. http://ceur-ws.org/Vol-1247/recsys14_poster10.pdf 査読有

③ Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, Jun Sakuma: Efficiency Improvement of Neutrality-Enhanced Recommendation. Decisions@RecSys 2013, pp. 1-8.
http://ceur-ws.org/Vol-1050/paper1.pdf
査読有

〔学会発表〕（計 1 件）
小林星平，佐久間 淳,匿名化データからのロバスト線形回帰とその汎化誤差解，2014年暗号と情報セキュリティシンポジウム(SCIS2014)，3D4-3 2014 年 1 月 23 日，城山観光ホテル(鹿児島)

６．研究組織
(1)研究代表者
佐久間 淳（SAKUMA, Jun）
筑波大学・システム情報系・准教授
研究者番号：90376967

(2)研究分担者
神嶌 敏弘（Kamishima, Toshihiro）
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・研究員
研究者番号：50356820