

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 11 日現在

機関番号：82626

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500194

研究課題名(和文) 社会的公正性に配慮したデータマイニング技術

研究課題名(英文) Data Mining Techniques that aware social fairness

研究代表者

神島 敏弘 (Kamishima, Toshihiro)

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

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：データマイニング技術の進展に伴い、採用・与信・保険などの生活上重要な決定にデータマイニング技術が利用されるようになってきている。このような決定において、社会的に決定要因として用いてはならない性別や人種などの情報に基づくことがないように、公正なデータマイニングを行う手法を開発するのが本研究の目標である。こうした配慮が必要な情報を直接的に利用しなくても、その間接的な影響を排除することまではできないというred-lining効果への対処や、統計的独立性に基づいた方法論が本研究の特色である。

研究成果の概要(英文)：As the improvement of data mining technologies, such technologies are being used for determinations that seriously affect individual' lives, such as employment applications, credit, insurance rates, and so on. Needless to say, such serious determinations must be unbiased and fair in terms of socially sensitive information, such as gender or race. For this purpose, we develop the technique to avoid the use of such sensitive information in the serious determinations. The simple elimination of sensitive information from the calculations is insufficient for avoiding inappropriate determination process, due to the indirect influence of sensitive information. This phenomena is called the read-lining effect. We treat this red-lining effect by introducing the methodologies based on the notion of the statistical independence.

研究分野：データマイニング

キーワード：データマイニング 機械学習 統計 独立性 公正性

1. 研究開始当初の背景

膨大な個人データが集積され、またデータマイニングが容易に利用できる環境が整備されるにともない、与信、保険、採用などの重要な決定にもデータマイニング技術が利用されるようになり始めている。一方で、一見すると問題を生じる個人情報を含んでいないデータでも、そのデータから個人の性別や住所などの情報を間接的に得ることができるようになってきている。例えば、ネット広告企業が、Webの閲覧履歴や広告のクリック履歴から、利用者のデモグラフィック属性を予測し、他のネットサービス企業に提供している。

こうした状況を踏まえ、たとえデータマイニング技術を用いても、社会的公正性に配慮した、すなわち、人種、性別などに基づく先入観や差別のない判断がなされるようにする必要が生じている。また、データ提供者との契約があるデータを扱う場合には、その契約に配慮が必要な場合もある。例えば、推薦システムで利用する目的で収集した顧客の個人情報のデータは他の目的には利用できない。しかし、このデータを利用した推薦システムで商品を提示すると、商品の閲覧履歴データは間接的に影響を受ける。このとき、閲覧履歴データの推薦以外の目的への利用には配慮が必要になる場合も生じる。

2. 研究の目的

社会的公正性や法令・契約遵守の観点から配慮を要する情報に、分析結果が依存しないようなデータマイニングの数理的な条件を論じ、それに基づいて社会的公正性が保たれるようなデータマイニング手法を開発するのが本研究の目的である。

詳細について述べる前に、社会的公正性の観点から、最終決定に利用してはならない情報をセンシティブ情報、判断に用いてもよい情報を非センシティブ情報と呼ぶ。例えば、採用に関する決定において、人種や性別に基づく決定は社会的に望ましくないため、これらはセンシティブ情報となる。一方で、職種に関する資格や技能については、これらを採用の決定時に考慮することは社会的に問題とはならないため、非センシティブ情報となる。この問題に対処するためのデータマイニング研究には、若干の先行研究がある。[Pedreschi 08] などによる、これらの研究では、配慮が必要な特徴を分析時に排除するだけでは、それらの特徴の間接的な影響を排除できないため、結果として差別的な決定がなされる場合があることが報告されており、この現象を red-lining 効果と呼ぶ。

本研究の目標は、この red-lining 効果をも考慮したうえで、センシティブ情報の分析結果への影響を排除することで社会的公正性に配慮したデータマイニング手法を開発する。

もう少し詳細に述べると、既存のデータマイニング手法、例えばクラス分類といったものは、将来に対する予測精度、すなわち、今後の事象においてより正確な予測ができるという点のみ注力して設計されている。それに対し公正配慮型データマイニングでは、公正性に配慮した制約を加えている。すなわち、センシティブ情報が予測結果に影響しないようにする制約の下で、最も正確な予測をめざすことになる。

[Pedreschi 08] D.Pedreschi et al. "Discrimination-aware Data Mining" KDD2008

3. 研究の方法

本研究は以下の三つの方向から進める

1. 公正性に配慮した分類アルゴリズムの開発
2. 公正配慮型データマイニングに関する理論的検討
3. 分類アルゴリズム以外への公正配慮型データマイニング手法の拡張

第1の『公正性に配慮した分類アルゴリズムの開発』とは、与えられたクラス集合を対象を分類するクラス分類問題を扱うものである。このどのクラスに分類するかという決定を、センシティブ情報の影響を受けないように行うアルゴリズムを新たに設計する。さらに、収束の安定性などの点についても考慮した改良を行う。

第2の『公正配慮型データマイニングに関する理論的検討』では、予測精度と公正性のトレードオフなど、過去の分類問題にはなかった理論的な問題について検討を行う。

第3の『分類アルゴリズム以外への公正配慮型データマイニング手法の拡張』として、分類問題以外の回帰問題や行列分解などの他の問題についても公正性を保つような手法の開発を試みる。

4. 研究成果

研究の方法で示した三つの課題それぞれについて成果を報告する。

4.1) 公正性に配慮した分類アルゴリズムの開発

既存研究では、公正配慮型の分類器としては、普通の分類器を学習した後、分類の基準を修正して公正性を確保するような後処理型や、事前に不公正な判断を見つけデータから取り除いておく前処理型が提案されていた。

それに対し、公正性の制約項を、分類器の目的関数に導入することによって、分類器の学習と公正性の達成を同時に行う手法を開発した。前処理・後処理型とは異なり、非センシティブ情報も考慮にいれつつ公正性を保つように出来る点がこのタイプの手法の特徴となる。また、理論面においては、目的関数の形

で最適化されるモデルが明示されているため解析を行いやすい利点もある。実行上の問題としては、目的関数が凸関数ではないため、多くの局所解が存在し、その最適化が難しい問題が生じた。これに対し、公正性について配慮しない分類器を予め学習しておき、このときのパラメータを初期値として公正配慮型の目的関数を最適化することで安定して最適化ができるように改良も行った。結果は国際会議 ECMLPKDD2012 などに採録されるなどの評価を得た。実験の結果を以下に示す。

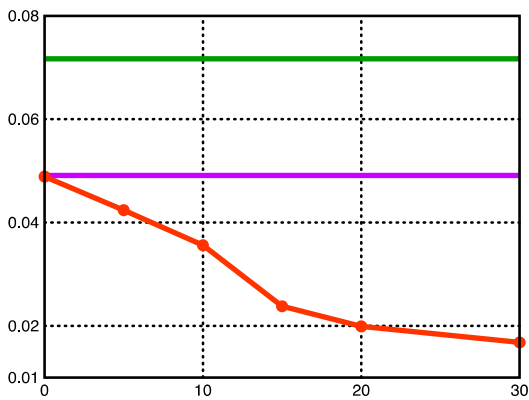


図 1(a) : 予測精度

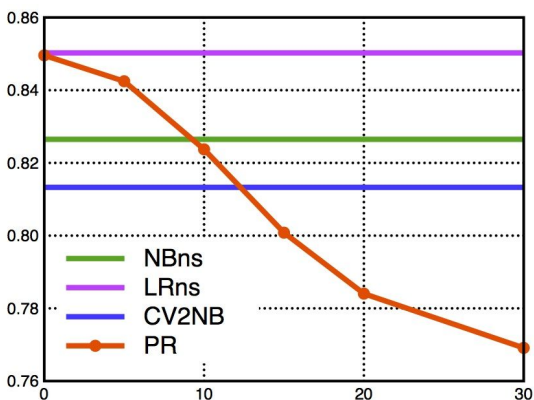


図 1(b) : 正規化相互情報量 (小さいほどより公正)

図中 NBns と LRns はセンシティブ情報を単純に削除した公正性に配慮しない手法である。図 1(b)に見られるように提案手法の PR (prejudice remover 法) はこれらの方法より高い公正性を達成できている。一方で既存手法の CV2NB 法よりは精度面では提案手法が良かったが、公正性では劣っていた。この理由については次の理論面からの検討で一つの答えを得た。

4.2)公正配慮型データマイニングに関する理論的検討

この課題としては、予測精度と公正性の関係について検討を行った。特に、提案手法 PR 法が既存手法と比べ、公正性に関して性能が

良くなかった点について調査し、その原因を理論的に解析した。

その原因として、モデルバイアスと確定的な決定則の影響があることを示した。モデルバイアスとは、実際のデータの分布と、学習用に用いるモデルが表現できる分布との乖離であり、実際のデータに対してなくすることはできない。もう一方の決定則とは、予測クラスは、そのクラスが生成される予測クラスが最大になるようなクラスに分類することで定めることである。これらのモデルバイアスや決定則の影響を無視し、PR モデルなどではモデル分布のまま、そして決定則を適用するまのクラス分布のまま統計的独立性を評価していたため、公正性において CV2NB モデルに及ぶことはなかった。

これを検証するため公正分割という技法を開発し、単純ベイズで CV2NB と同等の公正性が達成できることを示した。結果は PADM2011 ワークショップにて発表した。

4.3)分類アルゴリズム以外への公正配慮型データマイニング手法の拡張

これまでの研究では、センシティブ情報とクラスの分類結果が公正になるような公正配慮型の分類問題を扱ってきた。これに対し、推薦タスクにおいても公正性を保つ方法を提案した。

この方法は、確率的行列分解と呼ばれる方法で予測されるアイテムに対する評価値が、指定したセンシティブ情報とは独立になるようにするものである。利用が禁止されている情報の不使用の保証、情報提供者の公平な扱い、利用者が不要と考える情報の利用の排除などに応用できる。

公正性の制約項を加えることで実現したが、最初のモデルは計算量的に問題があった。そこで、公正性制約を大幅に近似することで高速化を達成するとともに、実際の性能には大きく影響の出ない手法を開発した。

この結果は、Decisions2012 と 2013 のワークショップ、および、RecSys2014 のポスターにて発表を行った。また国内では 2014 年度人工知能学会全国大会優秀賞という高い評価をえた。

実験結果を示しておく。

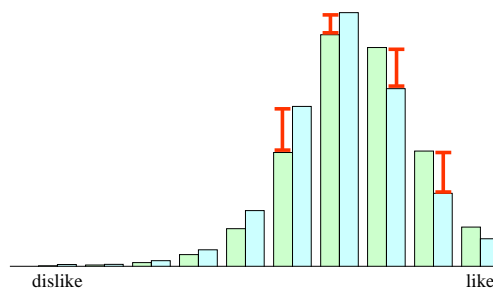


図 2(a) : 公正性に配慮しない場合

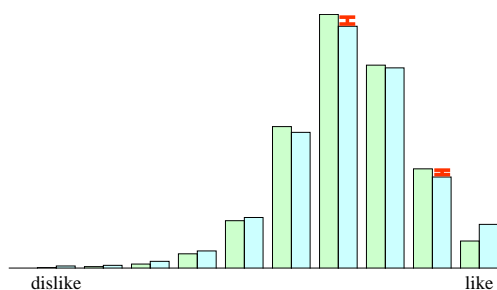


図 2(b) : 公正性に配慮した場合
これはセンシティブ情報として、利用者が不要と考えることの多いアイテムの人気という情報について独立性を保つ推薦を行った場合である。図 2(a)の公正性に配慮しない場合については赤いマークで示したように人気・不人気アイテムで評価に差がある場合が生じるが、提案手法は図 2(b)のように一般の人気には左右されず個人の好みのみを反映していることが分かる。

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 1 件)

(1) T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma “Fairness-Aware Classifier with Prejudice Remover Regularizer” Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD), Part II, pp.35-50 (2012)

〔学会発表〕(計 11 件)

(1) T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma “Correcting Popularity Bias by Enhancing Recommendation Neutrality” Poster Proceedings of the 8th ACM Conference on Recommender Systems (RecSys) (2014)

(2) 神島 敏弘, 赤穂 昭太郎, 麻生 英樹, 佐久間 淳 “公正配慮型分類器の公正性に関する分析” 電子情報通信学会技術研究報告, IBISML 2014-28 (2014)

(3) 神島 敏弘, 赤穂 昭太郎, 麻生 英樹, 佐久間 淳 “情報中立推薦システムの高速化” 人工知能学会全国大会(第 28 回)論文集, 2B5-OS-15b-5 (2014) [2014 年度人工知能学会全国大会優秀賞]

(4) T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma “Efficiency Improvement of Neutrality-Enhanced Recommendation” Proceedings of the 3rd Workshop on Human Decision Making in Recommender

Systems (Decisions@RecSys), pp.1-8 (2013)

(5) T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma “The Independence of the Fairness-aware Classifiers” Proceedings of the 4th IEEE International Workshop on Privacy Aspects of Data Mining (PADM@ICDM), pp.849-858 (2013)

(6) 神島 敏弘, 赤穂 昭太郎, 麻生 英樹, 佐久間 淳 “情報中立推薦での中立性項の改良” 電子情報通信学会技術研究報告, IBISML 2013-7 (2013)

(7) 神島 敏弘 “中立性・公正性に配慮したデータ分析” 人工知能学会全国大会(第 27 回)論文集, 3L1-OS-06a-1 (2013)

(8) T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma “Considerations on Fairness-aware Data Mining” Proceedings of the IEEE International Workshop on Discrimination and Privacy-Aware Data Mining (DPADM@ICDM), pp.378-385 (2012)

(9) T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma “Enhancement of the Neutrality in Recommendation” Proceedings of the 2nd Workshop on Human Decision Making in Recommender Systems (Decisions@RecSys), pp.8-14 (2012)

(10) 神島 敏弘 “公正・差別配慮型マイニングのサーベイ” 電子情報通信学会技術研究報告, IBISML 2012-13 (2012)

(11) 神島 敏弘, 赤穂 昭太郎, 麻生 英樹, 佐久間 淳 “情報中立推薦システム” 人工知能学会全国大会(第 26 回)論文集, 3E1-R-6-1 (2012)

〔図書〕(計 1 件)

(1) 神島 敏弘 “プライバシー保護関連データ処理技術ビッグデータ・マネジメントーデータサイエンティストのためのデータ活用技術と事例, 第 3.2.1 節, pp.229-235. エヌ・ティー・エス出版 (2014)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

ホームページ等

<http://www.kamishima.net/fadm>

<http://www.kamishima.net/inrs>

6. 研究組織

(1) 研究代表者

神鷹 敏弘 (KAMISHIMA, Toshihiro)
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

(2) 研究分担者

赤穂 昭太郎 (AKAHO, Shotaro)
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