様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　６　月　１９　日現在

機関番号：　８２６２６
研究種目：　基盤研究(C)（一般）
研究期間：　2015～2017
課題番号：　１５Ｋ００３２７

研究課題名（和文）データマイニングにおける中立・公正性に配慮するデータ変換技術

研究課題名（英文）Data Transformation that awares Fairness or Neutrality in Data Mining

研究代表者
神嶌　敏弘（Kamishima, Toshihiro）

国立研究開発法人産業技術総合研究所・情報・人間工学領域・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）　　3,600,000円

研究成果の概要（和文）：データマイニング技術の社会への適用が広がり，性別や人種といった社会的にセンシティブな情報に対する分析結果の公平性が問われる問題が見られるようになった．この問題に対処すべく研究が進められているのが公平性配慮型データマイニングである．
本研究ではデータ変換技術に対して公平性の強化を行う手法を開発し，これを推薦の各種タスクに応用してその性能を実験的に検証した．主な成果としては次の二つが挙げられる．(1) 確率的行列分解法に関して相互情報量とBhattacharyya距離を用いた二つの方法を開発した．(2) トピックモデルを用いる推薦で公平性を強化できるようにした．

研究成果の概要（英文）：Due to the wide spread of data mining technologies, the problem of unfair decisions with respect to social sensitive information, such as gender or race, have been arising. To alleviate the problem, the techniques of fairness-aware data mining have been studied.
In this project, we develop methods of data transformation while maintaining fairness, and these methods are applied to recommendation tasks. Our main contributions are as follows: (1) We developed fairness-aware probabilistic matrix factorization model with a regularization terms, adopting Gaussian mutual information and Bhattacharyya distance. (2) We developed a topic model for collaborative filtering while maintaining independence from sensitive information.

研究分野：　計算機科学

キーワード：　公平性　推薦システム　データマイニング　行列分解

１．研究開始当初の背景

(1) 膨大な個人データが集積され，またデータマイニング技術が容易に利用できるようになったため，その利用範囲が拡大している．これに伴って，データ分析技術が個人や社会に与える影響が拡大し，この分析結果の中立性や公平性が問題となる事例が指摘されはじめた．

事例 1 : Pariser は個人化技術の社会に与える影響の問題について論じ，これをフィルタバブル問題と呼んだ [Pariser]．具体例として，ソーシャルネット内で互いに情報交換をする相手である『友人』候補の推薦の例を挙げた．利用開始時には，保守派・進歩派の双方の候補が提示されていたが，進歩派の候補を友人として選ぶことが多かったため，個人化の機能により保守派が推薦されなくなった．このような場合には利用者の要求に応じた中立的な観点からの推薦が必要になる．

事例 2 : Sweeney は検索と共に表示される Web 広告において社会的な不公平が疑われる事例について報告した [Sweeney 13]．これは，個人の逮捕歴などを検索するサイトの広告で，アフリカ系の名前で検索した場合に，ヨーロッパ系の名前で検索したときより頻繁に逮捕歴を示唆するメッセージと共に広告が表示された事例である．この現象は，作為的な操作によるものではなく，広告のクリック率を向上させるようにしたことの副次的な影響によるものであった．このような場合には差別的でない社会的に公平な結果が得られるような機構が必要になる．

この事例に見られるように，単純に予測精度を向上させる目的で分析した結果を使って社会サービスに適用すると，中立性や公平性に欠けた状況が生じる可能性がある．

(2) 以上のような社会状況をふまえ，中立・公平性が保証されるような分析手法が必要となり，公平性配慮型データマイニングと呼ばれる研究が始まった．最初の研究 [Pedreschi 08] では，データベース中にある不公平な判断を検出する不公平発見タスクが提唱された．その後，[Calders 10] にて，クラス分類などの予測の結果が公平になるように配慮しつつ予測を行う不公平回避タスクが提唱された．現在では，両方のタスクに対して研究開発が進んでいる．我々もこの後者のタスクに取り組み始め，公平性を強化するための制約項である prejudice remover 正則化項を用いた手法を開発した [Kamishima 12]．この制約を導入した最適化問題への定式化はその後広くこの分野で利用されている．

２．研究の目的

公平性に配慮した分析ができるように，公平性の観点から配慮が必要なセンシティブ情報を削除するようなデータ変換，具体的には次元削減とトピックモデルの手法を開発し，その効果を検証することが本研究の目標である．以下，公平配慮型データマイニングの概要と，そのタスクの一つである不公平防止について述べたあと，本研究のアプローチによる不公平防止の枠組みについて述べる．

公平配慮型データマイニングは，大きく分析結果から不公平な部分を見つける不公平検出と，分析結果が公平となるような分析手法を開発する不公平防止の二つのタスクがあるが，本研究では後者を対象とする．分析での不公平を防止するには，公平性が問題となる属性をデータから除外するだけでは不十分である．研究の背景で述べた Sweeney の例では，どの人種かという直接的にセンシティブな情報は全く使われていないが，人種と関連する様々な他の情報の間接的な影響により，結果として不公平な状況が生じている．よって，直接的にセンシティブな情報だけでなく，これと関連した情報も削除する必要がある点が，不公平防止にとって困難な課題である．

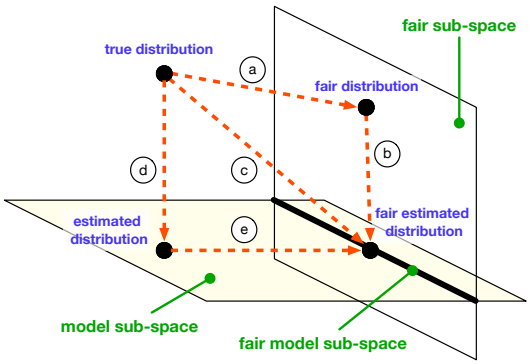

図 1 不公平防止タスクの実現手法

不公平防止では，分類問題を扱う手法に関する研究が最も進んでいる．Calders らによる研究 [Calders 10] を端緒とし，我々の [Kamishima 12] を含めいくつかの研究成果が報告されている．これらの公平性配慮型分類のアプローチは，図 1 のようにいくつかの方法がある．一つ目は前処理型で，センシティブ情報を削除するデータ変換(a)を行ってから，通常の分類器を適用する(b)．二つ目は(c)のように直接目的の公平性に配慮した分類器を作る中処理型．そして最後の三つ目は，先に通常の分類器を適用してから(d)，その分類器を公平になるように変換する(e)の後処理型である．

本研究では，このうち中処理型の枠組みに沿った新たな分析手法を開発する．従来の研究では，予測タスクとしてクラス分類問題がほとんどであり，一部に目的関数の定義域が実数となる回帰問題が対象となっていた．本研究では，新たなタスクとして代表的なデータ変換技術である次元削減やトピックモデルなどを公平性配慮型にした手法を開発する．どちらのモデルにおいても，これらの手法を推薦システムに応用する．

次元削減の手法としては，確率的行列分解法(probabilistic matrix factorization) を対象とする．本タスクに関しては，予備的な研究を行っていた．[Kamsihima 12b] では公平性を強化はできたものの，その効率は悪く，1 万件程度のデータしか処理できなかった．公平性の条件を分布の平均の一致と大きく緩和することで効率の向上をめざしたのが [Kamishima 13] であったが，公平性の改善にに関しては性能は良くなかった．そこで，本研究では新たな公平性を強化するための正則化項を開発し，計算の効率性と，公平性の強化の二つを両立するようにした．

もう一つのトピックモデルでは，推薦に用いられる Hoffmann の潜在変数モデルを利用した．このモデルにセンシティブ変数と予測評価値とが独立になるような形で，新たにセンシティブ変数を組み込むことにより公平性を達成する．

３．研究の方法

研究の方法としては順次アルゴリズムを開発し，ベンチマークデータにて検証を行う．開発にあたっては，確率的行列分解の改良，新規トピックモデルの開発，そして，確立的行列分解の新たな種類への適用を行う．詳細については次の研究成果の章で述べる．

４．研究成果

(1) 確率的行列分解法の改良

[Koren 09] の確率的行列分解モデルに，[Kamishima 12] の正則化アプローチを適用して公平性を強化するアルゴリズムを開発した．上記のように予備研究においては，計算効率性と公平性強化を同時に達成することはできていなかったが，これらを同時に達成できるようなアルゴリズムを開発できた．

これには独立性を強化するための正則化項を改良する必要があるが，これらを 2 種類開発した，一つは，推薦システムで予測するアイテムに対する評価値の分布がガウス分布に従うと仮定したのち，その分布に対するセンシティブ情報と評価値の間の相互情報量を計算するものである．しかし，評価値に対する分布は混合分布になってしまうが，これを単一のガウス分布で近似することにより，計算効率性を向上させた．もう一つの方は Bhattacharyya 距離というものにより，センシティブ特徴の値が 0 のときと，1 のときの評価値の分布を近づけるという制約項である．前者の相互情報量を用いる手法では分布に近似を用いているが，センシティブ特徴の値が 3 種類になっても適用できる利点がある．後者の Bhattacharyya 距離をモチル方法はセンシティブ特徴の値が 2 種類でなければならないが，代わりに厳密な値を計算できる．

どちらの手法においても，実験により大規模なデータを処理する計算効率性を実証できた．また，分布の平均，すなわち 1 次モーメントだけではなく，分散という 2 次モーメントまで考慮できるようになったことで，評価値の分布をより厳密に一致させることができ，より高度な公平性も達成することができた．

以上の結果は，機械学習に関する公平性に関する国際会議 FAT*2018 に採録され，〔雑誌論文〕 (4)の成果として発表した．

二つ目のデータ変換技術として [Hofmann 99] に，センシティブ特徴を導入して，公平性を達成できるようにするモデルを開発した．

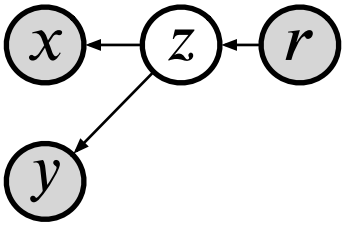

図 2 Hofmann の潜在クラスモデル

この図 2 の Hofmann のモデルでは，利用者 x がアイテム y に対して評価値 r を与えたことを表す生成モデルである．単純にこれらの同時分布を計算するとパラメータ数は非常に多数になるため，たかだか K 個の値をとる潜在クラス z を導入し，予測で汎化をできるようにしたものである．

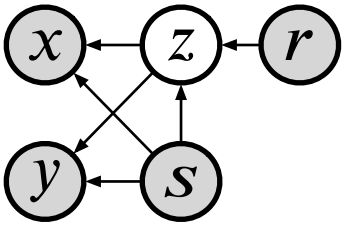

図 3 公平性配慮型潜在クラスモデル

図 3 は図 2 のモデルにセンシティブ特徴を扱う変数 s を新たに導入したものである．公平性に配慮するため，評価値 r とセンシティブ情報 s とは統計的に独立でならねばならない．一方で，s と他の変数とはできるだけ密に結合して，不要な独立性仮定を排除しなければならない．これら二つの条件をみたすように設計したのがこのモデルである．

実験により，公平性を強化したことにより予測精度の若干の低下はみられたが，図 2 のモデルと比べて公平性を図 3 のモデルでは改良できることを示した．

この成果はい CDM2016 の併設ワークショップにて採録された（〔雑誌論文〕(3)）

(3) 公平性の達成における確定的決定則の影響

本研究は直接敵にはデータ変換ではなく，クラス分類タスクを対象としたものである．しかし，この成果によって明らかになった知見はデータ変換にも当てはまるものである．

この研究では，クラス分類において，従来の公平性のモデルがクラスがクラスの事後分布に応じて確率的に生成されることを仮定している問題点を論じた．この仮定に反して，実際のクラスラベルを予測するときにはベイズ決定則と呼ばれる確定的な決定則により，クラスの事後確率が最も高いクラスに分類されるように予測され，この仮定は一般には成立しない．

そこで，クラスが確定的な決定則の結果決まる分布を使った独立性を考え，従来のモデルベース独立性に対し，実独立性の概念を提唱した．モデルベースの独立性の代わりに実独立性を達成することで，クラス分類においてより高度な公平性を達成できるだけでなく，公平性の達成に伴う予測精度の低下をも軽減できることが分かった．

この成果は Data Mining and Knowledge Discovery 誌に採録された（〔雑誌論文〕(4)）

ここでクラス分類では確定的な決定則の影響を受けていた．データ変換においても同様の現象が見られる．成果(1)の確率的行列分解の正則化アプローチでは，評価値を求めるときに分布の期待値の分布を考えていた．それに対し，成果(2)のモデルでは評価値が事後分布から確率的に評価値が生成されることを過程している．このため，確率的行列分解法の方がトピックモデルによる方法より，より公平性を達成できているという知見が得られた．

(4) 確率的行列分解のアイテム発見タスクへの適用

これは次の段階に向けたよび研究である．推薦システムのタスクには，利用者がアイテムについて付けようとしている評価値をよそくしようとする評価値予測タスクと，利用者が何か一つ自身の規準に合致したアイテムを見つけようとするアイテム発見タスクとがある．以上のうち，(1)と(2)のモデルは前者の評価値予測タスクのためのものであった．

それに対し，公平性の強化の可能性をもう一つのアイテム発見タスクにて検証した．アイテム発見タスクでは，利用者がアイテムを好むであろう度合いを予測する嗜好スコアの予測段階と，この求めた嗜好スコアに応じて好むであろう順にアイテムを並べる整列段階とに分けられる．

そして，評価値予測タスク用のモデルをそのまま，アイテム発見タスクに適用し，その有効性を検証する予備実験を行った．ここでは，評価値に対して公平性を強化するかわりに嗜好スコアに対して公平性を強化した．

実験の結果，嗜好スコアに対して公平性を強化できることが確認できた．しかし，嗜好スコアで整列した推薦リストについては公平性を強化できないことが分かった．もう少し詳細を述べると，推薦リストの上位 k 番目以上にあるかどうかということが，センシティブ情報とどくりになるかを検証したところ，この独立性の達成度は k の値によって大きく変動し，とくに k が小さな重要な部分で公平性が達成できないことが分かった．

この結果は，推薦における公平性のワークショップ FATREC2017 にて採録された（〔雑誌論文〕(2)）

＜参考文献＞

[Pariser] The Filter Bubble http://www.thefilterbubble.com/
[Sweeney 13] L.Sweeney ″Discrimination in Online Ad Delivery″ Communications of ACM (2013)
[Pedreschi 08] D.Pedreschi et al.“Discrimination-aware Data Mining”KDD2008
[Calders 10] T.Calders and S.Verwer, Three naive Bayes Approaches for Discrimination-free Classification, DMKD Journal (2010)
[Kamishima 12] T.Kamishima et al., Fairness-aware Classifier with Prejudice Remover Regularizer, ECMLPKDD (2012)
[Kamsihima 12b] T.Kamishima et al., Enhancement of the Neutrality in Recommendation, The 2nd Workshop on Human Decision Making in Recommender Systems

(2012)
[Kamishima 13] T.Kamishima et al., Efficiency Improvement of Neutrality-enhanced Recommendation, The 3rd Workshop on Human Decision Making in Recommender Systems (2013)
[Hofmann 99] T.Hofmann and J.Puzicha, Latent Class Models for Collaborative Filtering, IJCAI (1999)
[Koren 09] Y.Koren, Collaborative Filtering with Temporal Dynamics, KDD (2009)

５．主な発表論文等

〔雑誌論文〕(計 4 件)

(1) T. Kamishima, S. Akaho, H. Asoh, and I. Sato “Model-Based Approaches for Independence-Enhanced Recommendation”, Proceedings of the IEEE 16th International Conference on Data Mining Workshops, pp.860-867 (2016)

(2) T. Kamishima and S. Akaho “ Considerations on Recommendation Independence for a Find-Good-Items Task”, Proceedings of the FATREC Workshop on Responsible Recommendation (2017)

(3) T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma “Recommendation Independence”, Proceedings of the Conferene on Fairness, Accountability and Transparency, PMLR vol.81, pp.187-201 (2018)

(4) T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma “ Model-based and Actual Independence for Fairness-aware Classification”,Data Mining and Knowledge Discovery, vol.32 (2018)

〔学会発表〕(計 1 件)

(1) T. Kamishima and K. Fukuchi “Future Directions of Fairness-aware Data Mining: Recommendation, Causality, and Theoretical Aspects”, ICML2015 Workshop: Fairness, Accountability, and Transparency in Machine Learning (2015)

〔図書〕(計 0 件)

〔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/

６．研究組織

(1)研究代表者
神嶌 敏弘（KAMISHIMA, Toshihiro）
国立研究開発法人産業技術総合研究所・人間情報研究部門・主任研究員
研究者番号：50356820

(2)研究分担者
赤穂 昭太郎（AKAHO, Shotaro）
国立研究開発法人産業技術総合研究所・人間情報研究部門・研究グループ長
研究者番号： 40356340

(3)連携研究者
なし

(4)研究協力者
なし