

情報中立推薦での中立性項の改良

神嶋 敏弘[†] 赤穂昭太郎[†] 麻生 英樹[†] 佐久間 淳^{††}[†] 産業技術総合研究所

〒 305-8568 茨城県つくば市梅園 1-1-1 産総研つくば中央第 2

^{††} 筑波大学, 〒 305-8577 茨城県つくば市天王台 1-1-1E-mail: [†]mail@kamishima.net, ^{††}{s.akaho,h.asoh}@aist.go.jp, ^{†††}jun@cs.tsukuba.ac.jp

あらまし 情報中立推薦とは、ある与えられた視点に関して中立性を保証した推薦である。これは、個人化技術により提供される情報が偏るというフィルターバブル問題などへの対処として考案したものである。従来法では、中立性に関する制約項が解析的に微分できない問題があったため、この点を改善する。

キーワード 公正配慮型データマイニング, 推薦システム, 行列分解, 中立性, フィルターバブル

An improvement of a Neutrality Term
in an Information-neutral Recommender SystemToshihiro KAMISHIMA[†], Shotaro AKAHO[†], Hideki ASOH[†], and Jun SAKUMA^{††}[†] National Institute of Advanced Industrial Science and Technology (AIST),
AIST Tsukuba Central 2, Umezono 1-1-1, Tsukuba, Ibaraki, 305-8568 Japan^{††} University of Tsukuba, 1-1-1 Tennodai, Tsukuba, 305-8577 JapanE-mail: [†]mail@kamishima.net, ^{††}{s.akaho,h.asoh}@aist.go.jp, ^{†††}jun@cs.tsukuba.ac.jp

Abstract Information-neutral recommender systems aim to make recommendations whose neutrality from the specified viewpoint is guaranteed. Such systems are developed for dissolving a filter bubble problem, which is the bias or restriction that provided to people by the influence of personalization technologies. Our previously developed system was not scalable because efficient optimization techniques could not be applied. To address this problem, we developed a constraint term for enhancing the neutrality that can be analytically differentiable.

Key words fairness-aware data mining, recommender system, matrix factorization, neutrality, filter bubble

1. はじめに

推薦システムは、利用者が好むであろうアイテムや情報などを、利用者との対話履歴やアイテムの特徴に基づいて予測し、それらを利用者の目的に合わせて提示する [1]。90 年代中頃以降から多くの手法が研究レベルで提案され、今世紀に入ってから顧客へのサービスとして、多くの電子商取引サイトで導入されている。この推薦システムは意思決定支援システムの側面を持つため、当然ながら利用者の行動に影響を与える。こうした影響の一つとして、利用者に提供される情報の多様性が失われたり、偏りが生じるというフィルターバブル問題 (Filter Bubble) [2, 3] が指摘されている。

フィルターバブルは、Pariser により主張されたもので、推薦などの個人化技術により、利用者は、知らないうちに、自身が関心があるとされる限定された話題の情報のみしか接しな

いようになっており、まるで『バブル』の中に閉じ込められたような状態になっているとするものである。そのため、利用者が新たな話題に関心をもつ機会が失われたり、社会全体での情報や認識の共有が困難になるなどの影響があると指摘している。これに対し、2011 年の推薦システムの国際会議 RecSys 2011 では、パネル討論が開催された [3]。

このパネルでは次のようなことが述べられた。個人化の影響で話題に偏りが生じていることは確かである。同時にどのような観点からも中立な推薦は本質的に不可能であり、利用者の関心への適合を高めることと、提供する話題を多様化することはトレードオフ関係にある。そして、この問題に対処する方法として、利用者の現在の要求だけでなく長期の要求も同時に考慮すること、推薦対象を個々に決めるのではなく提示するリスト全体を最適化すること、そして利用者が視点を変更できるようにする手段を提供することなどを挙げていた。

この問題に対処するために利用者が視点を制御出来るようにした著名な手法は知る限りないので、推薦の中立性を保証する情報中立推薦システム (information-neutral recommender system) を提案した [4, 5]. あらゆる視点からの絶対的な中立性を保証することは本質的に不可能なので、このシステムでは利用者が指定した観点に対する中立性の保証をめざす。例えば、ブランドといった商品の特徴や、推薦対象者の性別といったものを観点に指定する。そして、これら指定された特徴に対して推薦結果が影響されないような推薦を行う。このシステムは、システム運用者が情報や商品提供者を公平に扱ったり、法や契約により利用が制限されている情報の推薦への影響を排除する目的でも利用可能である。

以前の研究 [4, 5] では、我々の公正配慮型データマイニングの手法 [6] を応用し、指定された観点と推薦結果の間を相互情報量に該当する制約項を加えることで情報中立推薦を実現した。しかし、この制約項を単純に実装することで情報の中立性は達成できたが、この項が解析的に微分できないため効率的に最適化できず、大規模なデータを処理できない問題があった。そこで、今回は解析的に計算できるような制約項を開発する。

2. 節ではフィルターバブル問題と推薦の中立性について論じ、情報中立推薦タスクの目的を定義する。情報中立推薦システムの提案手法を 3. 節で述べ、その実験結果を 4. 節で示す。最後の 5. 節と 6. 節は関連研究とまとめである。

2. 情報の中立性

ここではフィルターバブル問題に関する議論を紹介し、醜いアヒルの子の定理に基づいて推薦の中立性について論じ、情報中立推薦システムの目的について述べる。

2.1 フィルターバブル問題

最初に、Pariser によるフィルターバブル問題についての指摘と、RecSys 2011 のパネルでの指摘についてまとめる。フィルターバブル (Filter Bubble) [2, 3] 問題とは、推薦を含めた個人化技術によって、利用者が接する情報の話題の範囲が狭められたり、偏ったりすることが、利用者が知らないうちに行われるという懸念に対する指摘である。

この問題に関する TED Talk [7] など、Pariser は次のような例を挙げている。ソーシャルネットサービスである Facebook では、議論などを限定された利用者間で行うために『友人』関係を明示する。この友人関係の構築に役立つように、利用者に関連がありそうな、他の利用者を推薦する機能が備わっている。Pariser がサービスを利用し始めたころは、保守派の人も、革新派の人も混在して推薦されていた。ところが、Pariser 自身が革新派の人と友人関係を実際に構築することが多いため、個人化の機能により保守派の人が推薦されなくなると述べている。システムは利用者の断りもなく保守派を除外し、多様な選択の機会を奪ったと、Pariser は主張している。

また、2011 年にエジプトで政変があった時期に、Google で『エジプト』で検索してもらった結果を、多くの人に送ってもらった事例についても述べている。政変という重大な事件があったにもかかわらず、個人化の機能によって、政変に関する

情報ではなく、観光の情報が示されている場合があったと述べている。このような、多くの人が知るべき重要な情報に接する機会が失われてしまうことは問題であると主張している。

この指摘は大きく二つにまとめることができる。一つは、利用者が多様な情報に接する機会が少なくなる問題である。このため、自身の知見を広げる芸術のようなものに接する機会が失われ、安易な娯楽情報のみを消費するようになってしまうと主張している。もう一つは、各人がそれぞれ異なる限られた情報にのみ接して、互いに共有する情報が減ってしまう問題である。社会でのコンセンサスの構築には、その基盤となる情報の共有が重要であるとし、それが失われると主張している。

推薦システムの国際会議 RecSys2011 ではパネル討論 [3] で、(a) フィルターバブル問題は実在しているのか？、(b) 個人化はどれくらい問題なのか？、そして (c) この問題に対して推薦システム研究コミュニティは何をすべきか？について論じた。まず (a) について、個人化が利用者の経験範囲を狭めることは、1990 年代中頃には Resnick が指摘していた。特定の情報を選びとることは、他の情報を無視すること必然的に伴うので、利用者の関心に集中することと、多様な話題を提供することは本質的にトレードオフ関係になる。この問題は、個人化技術に特有のものではなく、一般のニュースにおいてもイスラム系の al-Jazeera と、アメリカ保守系の Fox News との報道姿勢などにもみられるものである。

(b) の点については、ある程度は問題ではあるが、何かしらのフィルタリングに人間は常に触れていて、その影響をうまく扱うすべを持っているとパネリストは指摘した。これに関して choice architecture の概念を紹介した『カフェテリアで、どんな食事の並べ方も客の食べ方に影響するので『中立な』並べ方というものはない。どんな並べ方が最高のものかというのは奇妙な質問であり、またそんなものはいったい誰が決めるのか？』ここで述べられているように、どんな情報も『中立』ではなく、人は常に何らかの『バブル』に囚われているといえる。こうした状況下にもかかわらず、人はこれらの影響をうまく扱いつつ生活できている。個人化技術には確かにある種の情報の偏りはあるが、これらの影響を今までのようにどうにか扱いつつ、この技術を使いこなしてゆく以外にはない。

最後の (c) について、フィルターバブル問題に対する技術的対応として、利用者の現在の要求だけでなく長期の要求も同時に考慮すること、推薦対象を個々に決めるのではなく提示するリスト全体を最適化すること、そして利用者が視点を変更できるようにする手段を提供することなどを挙げていた。

2.2 推薦における中立性

フィルターバブル問題への対処法として、ここでは利用者に視点を制御する手段を与えることを考える。この手段について述べる前に、醜いアヒルの子の定理をもとに中立性について論じたい。醜いアヒルの子の定理 (ugly duckling theorem) [8] とはパターン認識での古典的な定理で、分類対象のある特定の側面や特徴を、他より重視することなく『分類』というものはできないとするものである。²⁾ 匹のアヒルの子が、ちょうど n 個の二値特徴で表現され、醜い子と普通の子の 2 クラスに分類す

ることを考える．このとき，これらの特徴に基づいて，醜い子と普通の子をこれらの特徴に基づいて識別するルール数が，任意の普通の子を識別するルールの数と等しいことが容易に示せる．もし識別ルールが多いほど容易に分類できると考えるなら，これは普通の子と醜い子は同じくらい似ているということになる．全ての特徴を等価に考えるという前提のため，この直感に反する結論は，実は導き出されている．すなわち，羽根が黒いかどうか？といった特定の特徴に注目することによって，醜いアヒルの子を醜いものと見なすことになっている．より一般に，何かを分類するときには，分類対象についてのある特定の視点，側面，特徴などを必然的に重視することに伴う．推薦は推薦対象を関心のあるものとそうでないものに分類することなので，特定の特徴や観点は推薦に置いても重視せざるおえない．その結果，RecSys のパネルで指摘されたように，絶対的に中立な推薦は本質的に不可能である．

そこで絶対的に中立ではない推薦というものを考える．醜いアヒルの子の定理によれば，分類ではある特定の特徴に注目することを必然的に伴うが，このことは全ての観点ではなく，ある特定の観点であれば中立性を保つことができることを示唆している．この考えに基づき，利用者が指定した特定の観点に対して中立性を保証する情報中立推薦システム (information-neutral recommender system) を提案する．Pariser の Facebook の場合の例でいえば，推薦される友人が保守派か革新派かという特定の観点については中立性を保証するが，他の観点，例えば出身地などについては，個人の嗜好を反映した偏りを許すような推薦を行うシステムである．こうした中立性を保った情報は利用者の意思決定にとって有用であると考えられる．

この情報中立推薦システムは，利用者だけでなくシステムの管理者にも有用で，情報の提供者や，商品の販売者を推薦システム公平に扱う目的にも利用できる．自身のサービスを競合企業のそれより検索の上位に表示したとして，Google はアメリカの連邦取引委員会の調査を受けた [9]．また，オンラインモールの運営企業は，商品を推薦するとき，参加小売店を公平に扱う必要があるだろう．検索対象や，参加小売店の情報を視点として設定すれば，情報中立推薦によってこれらを公平に扱うことができるだろう．その他，情報中立推薦システムは法や契約によって制限された情報を推薦で利用しない目的にも利用できる．例えば，プライバシーポリシーで推薦への利用が制限されている情報がある場合，この情報を視点とすることで，制限された情報の利用を回避できる．

この中立性を視点を示す変数と推薦結果との統計的独立性によって形式的に定義する．すなわち，情報理論的には，視点の状態の情報は推薦結果に影響せず，それ以外の特徴から得られた情報にのみ基づいて中立性を保証した推薦を行う．Pariser の例でいえば，政治的な立場という情報は除外されることになる．なお，このため利用できる情報量は非増加となるので，推薦の予測精度はその分だけ低下することになる．

この中立性は，互いに似ていないアイテムを推薦しようとする推薦の多様性 (recommendation diversity) の概念とは明確に異なる [10–12]．中立性は推薦と単一の視点の関係として定

義されるのに対し，多様性は推薦される対象間について相対的に定義されるものである．Pariser の例では，利用者の多様性を高めた推薦では，政治的な立場以外の特徴では互いに似ていなくても，全て進歩派の人が推薦されるということはある．逆に，情報中立推薦では，進歩派か保守派かという観点については中立性が保証されていても，その他の情報に強く依存した推薦になることがある．

3. 情報中立推薦システム

この節では，情報中立推薦問題を定式化した後，この問題を解くアルゴリズムを示す．

3.1 情報中立推薦問題の形式的定義

文献 [13] によれば，推薦の問題は，利用者の関心にあったアイテムを何か見つける良いアイテムの推薦 (recommending good items)，利用者の効用を最適化する効用最適化 (optimizing utility)，そして利用者のアイテムへの評価値を予測する評価値予測 (predicting ratings) に分けられる．ここでは，アイテムの評価値予測問題を対象に，これを情報中立化する問題を扱う． $X \in \{1, \dots, n\}$ と $Y \in \{1, \dots, m\}$ はそれぞれ利用者とアイテムを示す確率変数．イベント (x, y) は，確率変数の対 (X, Y) の具現値．利用者 X によるアイテム Y への評価値を示す確率変数を R とし，その具現値を r とする．評価値の定義域は， $\{1, \dots, 5\}$ といった離散値であることが多いが，ここでは評価値の実数集合として扱う．以上の変数は，情報中立ではない評価値推定問題と共通である．

推薦での中立性を考えるために，中立性の保証の対象となる視点を表す確率変数 V を導入する．この変数が何を表すかは利用者が指定し，その値はイベントに依存して決まる．視点変数の例としては，イベント中の利用者に依存して決まる性別，アイテムに依存して決まる映画の公開年，利用者とアイテムの両方に依存する評価時刻などが挙げられる．ここでは，議論を単純化するため，視点変数は定義域が $\{0, 1\}$ である二値変数とする．訓練事例は，イベント (x, y) と，このイベントに対する視点値 v と評価値 r とで構成される．この訓練事例を N 個含むのが訓練集合 $\mathcal{D} = \{(x_i, y_i, v_i, r_i)\}$ ， $i = 1, \dots, N$ である．

ある新しいイベント (x, y) とそれに対応する視点値 v に対して，利用者 x がアイテム y に与えた評価値を予測するのが評価値予測関数 $\hat{r}(x, y, v)$ であり， $\hat{r}(x, y, v) = \mathbb{E}_{P_{\mathbf{r}}[R|x, y, v]}[R]$ を満たすようにする．この評価値予測関数は，損失関数 $\text{loss}(r^*, \hat{r})$ ，中立性項 $\text{neutral}(R, V)$ ，および正則化項 reg の三つの部分から成る目的関数を最適化することで推定する．損失関数は，真の評価値 r^* と予測評価値 \hat{r} の非類似度を表す．中立性項は，視点変数が表す視点に対する予測評価値の間の中立性を定量化する．正則化項は，過学習をさけるためのものである．与えられた訓練集合 \mathcal{D} に対し，損失関数の期待値をできるだけ小さく保つと共に，中立性項の期待値をできるだけ大きくするような評価値予測関数 $\hat{r}(x, y, v)$ を獲得することが (評価値予測問題での) 情報中立推薦の目的である．この目的を，次の目的関数を最小化するような評価値予測関数 \hat{r} を見つけることで定式化する．

$$\sum_{\mathcal{D}} \text{loss}(r, \hat{r}(x, y, v)) - \eta \text{neutral}(R, V) + \lambda \text{reg}(\Theta) \quad (1)$$

ただし, $\eta > 0$ は中立性と損失の釣り合いを調整する中立性パラメータ, $\lambda > 0$ は正則化パラメータ, そして Θ はモデルパラメータである.

3.2 確率的行列分解モデル

ここでは, 予測精度も高く, 効率的で大規模化も可能であることが知られている確率的行列分解モデル [14] を評価値予測に採用する. このモデルには細部が異なるものがあるが, 文献 [15] の式 (3) の次のものを用いる:

$$\hat{r}(x, y) = \mu + b_x + c_y + \mathbf{p}_x^\top \mathbf{q}_y \quad (2)$$

ただし, μ, b_x, c_y はそれぞれ, 大域, 利用者ごと, およびアイテムごとのバイアス項である. \mathbf{p}_x と \mathbf{q}_y はそれぞれ K 次元ベクトルのパラメータで, 利用者とアイテムの交差的な効果を表現する. 損失関数 $\text{loss}(r, \hat{r})$ として, 正則化項付きの二乗誤差を導入する:

$$\sum_{(x_i, y_i, r_i) \in \mathcal{D}} (r_i - \hat{r}(x_i, y_i))^2 + \text{reg}(\Theta) \quad (3)$$

このモデルは, 式 (2) を平均とする正規分布から真の評価値が生成されていると仮定することと等価であることが知られている. もし X と Y の定義域全ての事例があり, 正則化項も凸ならば, 目的関数は凸となり, 大域的に最適なパラメータを単純な勾配降下法により求めることができる. 実際には, 全ての事例が観測されることはないので, 損失関数 (3) の凸性は保証されず, 求まるのは局所最適解のみである. それでも, ほとんどの場合に良い解を単純な勾配降下法によって求められることが実験的に知られている [15].

この潜在変数モデルを情報中立推薦用に拡張するために, モデル (2) を視点値 v に依存するように修正する. V のそれぞれの視点値 0 と 1 に対して, パラメータ集合 $\mu^{(v)}, b_x^{(v)}, c_y^{(v)}, \mathbf{p}_x^{(v)}$, そして $\mathbf{q}_y^{(v)}$ を用意する. これらのパラメータ集合を視点値に応じて選択し, 次の関数で予測評価値を得る:

$$\hat{r}(x, y, v) = \mu^{(v)} + b_x^{(v)} + c_y^{(v)} + \mathbf{p}_x^{(v)\top} \mathbf{q}_y^{(v)} \quad (4)$$

式 (1) で, 元の確率的行列分解の場合と同じ二乗損失を採用して, 式 (4) を代入すると, 次の情報中立推薦の目的関数を得る:

$$\sum_{\mathcal{D}} (r_i - \hat{r}(x_i, y_i, v_i))^2 + \eta \text{neutral}(R, V) + \lambda \text{reg}(\Theta) \quad (5)$$

ただし, v の各値に対応するパラメータ集合の, 大域バイアス以外のパラメータの L_2 正則化の和を正則化項とする. モデルパラメータ集合 $\Theta^{(v)} = \{\mu^{(v)}, b_x^{(v)}, c_y^{(v)}, \mathbf{p}_x^{(v)}, \mathbf{q}_y^{(v)}\}$, for $v \in \{0, 1\}$ はこの目的関数を最小化することで求める. パラメータの学習後は, 式 (4) を用いて, 任意のイベントに対する予測評価値が計算できる.

3.3 中立性項

あとは, 視点変数 V に対する予測される情報中立性を定量化する中立性項を決める必要があり, 相互情報量に基づくもの

と, Calders&Vewer の差別スコアに基づくものとを提案する.

3.3.1 相互情報量

2.2 節で述べたように, 中立性を統計的独立性とみなす仮定の下で中立性を負の相互情報量で定量化する. 評価値変数 R と視点変数 V の負の相互情報量は次式:

$$\begin{aligned} -I(R; V) &= -\sum_V \int \Pr[R, V] \log \frac{\Pr[R|V]}{\Pr[R]} dR \\ &\approx -\frac{1}{N} \sum_{(x_i, y_i, v_i) \in \mathcal{D}} \log \frac{\Pr[\hat{r}_i|v_i]}{\Pr[\hat{r}_i]} \end{aligned} \quad (6)$$

ただし, $(x_i, y_i, v_i) \in \mathcal{D}$ を式 (4) に適用して \hat{r}_i は求める. なお, 第 2 行では, この R と V 上の周辺化を \mathcal{D} 上の標本平均で近似している. $\Pr[R]$ は $\sum_V \Pr[R|V] \Pr[V]$ で計算し, $\Pr[V]$ は標本分布から計算できる. $\Pr[R|X, Y, V] \Pr[X, Y]$ を X と Y について周辺化すれば $\Pr[R|V]$ を導出できるが, この周辺化を再び標本平均で近似する:

$$\Pr[r|v] \approx \frac{1}{|\mathcal{D}^{(v)}|} \sum_{(x_i, y_i) \in \mathcal{D}^{(v)}} \mathcal{N}(r; \hat{r}(x_i, y_i, v), \mathbb{V}_{\mathcal{D}^{(v)}}(R)) \quad (7)$$

ただし, $\mathcal{N}(\cdot)$ は正規分布密度関数, $\mathcal{D}^{(v)}$ はその視点値が v に等しい全ての訓練事例で構成される. $\mathbb{V}_{\mathcal{D}^{(v)}}(R)$ は標本分散 $\sum_{r_i \in \mathcal{D}^{(v)}} (r_i - \mathbb{M}_{\mathcal{D}^{(v)}}(\{\hat{r}\}))^2 / |\mathcal{D}^{(v)}|$ であり, $\mathbb{M}_{\mathcal{D}^{(v)}}(\{\hat{r}\})$ は次式:

$$\mathbb{M}_{\mathcal{D}^{(v)}}(\{\hat{r}\}) = \frac{1}{|\mathcal{D}^{(v)}|} \sum_{(x_i, y_i, v_i) \in \mathcal{D}^{(v)}} \hat{r}(x_i, y_i, v_i)$$

しかし, この式は非常に多数の要素分布を含む混合分布であり扱いにくいので, $\Pr[r|v]$ を直接的にモデル化した 2 種類のモデルを用いた.

一つ目のヒストグラムモデルは, 以前に文献 [5] で提案したものである. 実数の予測評価値を, 5 点評価尺度の場合には $(-\infty, 1.5], (1.5, 2.5], \dots, (4.5, \infty)$ のビンに離散化し, この離散化した値を多項分布でモデル化する. このヒストグラムを用いた相互情報量モデルを mi-hist で参照する. このモデルには不連続な点があるため, 連続で扱いやすいように $\Pr[\hat{r}|v]$ を単一の正規分布でモデル化するのが次のモデルである. 形式的には次式:

$$\Pr[\hat{r}|v] \approx \mathcal{N}(\hat{r}; \mathbb{M}_{\mathcal{D}^{(v)}}(\{\hat{r}\}), \mathbb{V}_{\mathcal{D}^{(v)}}(\{\hat{r}\})) \quad (8)$$

ただし, $\mathbb{V}_{\mathcal{D}^{(v)}}(\{\hat{r}\})$ は, \mathcal{D} 中の事例からの予測評価値 \hat{r}_i の標本分散である. この正規分布を用いた相互情報量モデルを mi-normal で参照する.

残念ながら, これらの中立性項の勾配は解析的には計算できない. なぜなら, mi-hist モデルでは不連続な変換である離散化をしており, mi-normal モデルでは指数分布族ではない正規混合分布を $\Pr[\hat{r}]$ で用いているためである. そのため, 勾配を計算することなく適用できる Powell 法で目的関数を最適化した. しかし, この最適化法は関数の評価階数が多く非効率なため大規模な問題には適用できないという問題があり, 実験では, これらのモデルでは 100k 規模のデータでも 2~3 日のうちには終了しなかった. 一方, 次節で提案するモデルはこの規模の計算

を数 10 分で処理できた。

3.3.2 Calders&Verwer の差別スコア

その勾配を解析的に計算できる中立性項を開発するために、公正配慮型データマイニング [16] の手法を用いる。ここでは、文献 [17] で Calders と Verwer が用いた社会的な決定の差別度の尺度（ここでは CV スコアと呼ぶ）を用いる。 $V = 0$ と $V = 1$ の間の目標変数の値の分布の差で CV スコアを定義する：

$$\Pr[R|V = 0] - \Pr[R|V = 1] \quad (9)$$

視点変数 V の評価値変数 R への影響を減らすために、CV スコアを 0 に近づけることで、二つの分布 $\Pr[R|V = 0]$ と $\Pr[R|V = 1]$ を近づけるような分類モデルを彼らは提案していた。この過程は V と R の統計的独立性を高めることであることが容易に示せる [18]。この考えに基づいて、二つの分布 $\Pr[R|V = 0]$ と $\Pr[R|V = 1]$ を近づけるような 2 種類の中立性項を設計する。

一つ目の中立性項は、二つの分布の 1 次モーメント、すなわち平均を一致させるもので、形式的には次式で定義する：

$$-(\mathbb{M}_{\mathcal{D}(0)}(\{\hat{r}\}) - \mathbb{M}_{\mathcal{D}(1)}(\{\hat{r}\}))^2 \quad (10)$$

ただし、 $\mathbb{M}(r_i^{(v)})$ は、 $v_i = v$ であるような \mathcal{D} 中の標本上の標本平均である。この平均を一致させる中立性項を m-match で参照する。二つ目の中立性項は、 x と y が同じ場合には視点値の値とは無関係に同じ予測評価値となるような制約であり、形式的には次式で定義する：

$$-\sum_{(x_i, y_i) \in \mathcal{D}} (\hat{r}(x_i, y_i, 0) - \hat{r}(x_i, y_i, 1))^2 \quad (11)$$

この評価値を一致させる中立性項を r-match で参照する。

CV スコアに基づく中立性項はどちらも単純な二次多項式なので、その導関数を解析的に求めるのは容易である。よって、Powell 法よりずっと効率的に計算できる共役勾配法を用いて目的関数を最適化した。たとえ訓練データがより大きくなったとしても、勾配を解析的に計算できるので、確率的勾配降下法といったより効率的な最適化手法が利用でき、対処が可能である。

これら CV スコアに基づく中立性項には、単純な二次式で凸ではないが比較的なめらかなので、局所解析解にあまり陥りにくいという利点もある。逆に、相互情報量の場合とは違って、視点変数が多値の離散や連続値である場合に CV スコアに基づく中立性項を拡張するのは簡単ではない。m-match と r-match を比べると、 $V = 0$ と $V = 1$ の両方の場合の予測評価値を計算する必要があるため、r-match の方が m-match の約 2 倍の計算時間が必要になる。一方、r-match は、m-match より厳密に中立性を達成しているともみなせる。m-match では、利用母集団上で平均的に中立性を強化しているため、利用者によっては中立性が強くなったり弱くなったりする場合があります。r-match では、利用者とアイテムの定義域上のほぼいたるところで一様に中立性を強化するように設計されている。m-match とは違い、r-match は現実にはない状況を想定する性質がある。例えば、利用者の性別を視点に設定すると、性別が

変化していなくても、r-match は変化した現実には生じなかった場合を想定して計算する。このことは、意味的にはあまり望ましいとはいえないだろう。

4. 実験

3.3 節の 4 種類の中立性項を用いた情報中立推薦システムを実装し、ベンチマークデータ上での実験を行った。

4.1 データ集合

実験には、映画の評価データである Movielens の 100k データ集合 [19] を用いた。3.3.1 節の二つの相互情報量に基づく方法 (mi-hist と mi-normal) は、このデータ全体を処理できるほど効率的ではない。そこで、利用者 ID が 300 以下かつアイテム ID が 200 以下のイベントのみを抽出し、データ集合を縮小した。この縮小したデータ集合は、イベント数 9,409、利用者数 200、そしてアイテム数 300 である。この縮小データでの実験の目的は、4 種類全ての中立性項の特徴を比較することである。相互情報量に基づく中立性項は、3.3.2 節の CV スコアに基づく方法より、 R と V 上の分布をより厳密にモデル化しているだけである。この縮小データで CV スコアベースの手法が相互情報量ベース手法と同様の振る舞いをするのが確認できれば、CV スコアベースのやや厳密ではないモデル化で十分であり、効率的な計算ができると結論付けるとができるだろう。

視点変数には次の二つのものを採用した。一つ目の Year は、映画の公開年が 1990 年より古いかどうかという、イベントのアイテム部分に依存した視点変数である。Koren は古い映画ほど評価が良くなる傾向を報告しており、これはおそらくは長い年月の間に傑作のみが鑑賞し続けられて残ってきたためだろう [20]。視点変数 Year を視点変数に選んだとき、情報中立推薦システムはこの『傑作バイアス』を補正することになる。二つ目の Gender は利用者の性別という、イベントの利用者部分に依存した視点変数である。映画への評価が性別によって変わることは十分に考えられ、この要因に対する中立性の確立を試みる。

4.2 実験条件

mi-hist/mi-normal と m-match/r-match の中立性項を使った目的関数 (5) をそれぞれ、SciPy パッケージ [21] 中の Powell 法と共役勾配法を用いて最適化した。訓練集合 \mathcal{D} をその視点変数の値に応じて二つに分割し、その視点値に対応する訓練集合を使って、式 (3) の情報中立ではない確率的行列分解モデルを最適化することで、それぞれの視点値に対応するパラメータの初期値を求めた。目的関数を実装するとき、損失項は訓練事例数で割り、 L_2 正則化項はパラメータ数で割ることで、大まかに各項のスケールを揃えた。四つの中立性項についても同じ超パラメータ η の値で比較できるように、大まかにそのスケールを揃えた。正則化超パラメータは $\lambda = 0.01$ とし、潜在因子の数、すなわち $p^{(v)}$ や $q^{(v)}$ の要素数は、小規模なデータなので $K = 1$ とした。元の評価値は 1, 2, ..., 5 であるので、中立性項 mi-hist では $(-\infty, 1.5], (1.5, 2.5], \dots, (4.5, \infty)$ の 5 個のピンを採用した。

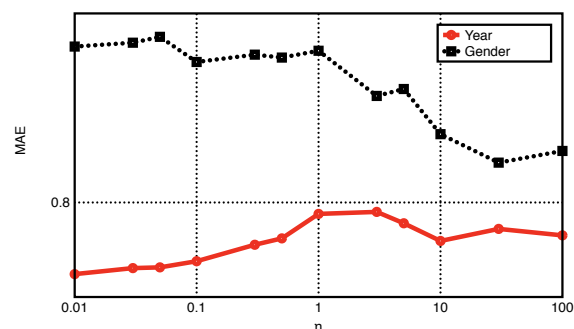
予測誤差と中立性で実験結果を評価した。予測誤差は平均絶対誤差 (mean absolute error; MAE) [13] で評価した。MAE は、テスト集合中で観測された評価値と予測評価値の差の絶対値の、テスト集合上の平均であり、小さな値ほど予測が正確であることを示す。中立性を測には、予測評価値と視点変数の間の相互情報量を用いた。この相互情報量が小さいほど、より中立性が厳密に達成されていることになる。この相互情報量は文献 [18] の幾何平均を使った方法で $[0, 1]$ の範囲に標準化した正規化相互情報量 (normalized mutual information; NMI) である。なおこの NMI の計算には $\Pr[\hat{r}|v]$ の算出が必要になるが、ここでは中立性項 mi-hist で用いたヒストグラムモデルで計算した。実験では、5 分割交差確認を実施し、予測誤差と中立性の評価尺度をテスト集合について求めた。

4.3 実験結果

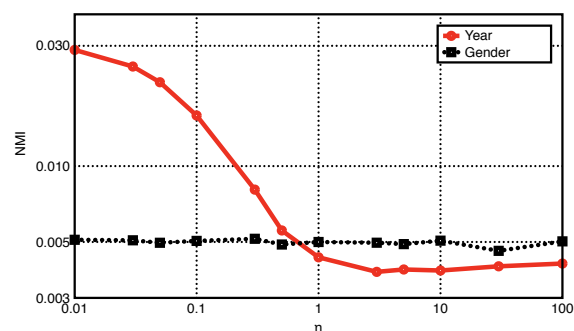
4 種類の中立性項を用いた実験結果を図 1 に示す。視点 Year に対する結果を図 1(a) と (b) に、視点 Gender に対する結果を図 1(c) と (d) に示した。図 1(a) と (c) には平均絶対誤差 (MAE) で測った予測誤差の変化を示した。図 1(b) と (d) には正規化相互情報量 (NMI) で測った中立性の変化を示した。図の水平軸は、目的関数 (1) の予測精度と中立性の釣り合いを調整する中立性パラメータ η の値である。このパラメータは中立性項がほぼ無視される 0.01 から、逆に非常に重視される 100 まで変化させた。

訓練データの評価値の標本平均である 3.74 を全員に提示したときの MAE0.90 をベースラインとする。なお、この予測誤差は、ランダムに選んだアイテムの評価値を提示した場合の誤差と近似的に等価である。このベースラインは、ランダムな選択はどんな情報をも利用していないことに対応するので、中立性を極限まで高めた状態に対応する。図 1(a) と (c) では、予測誤差はこのベースラインより良い。全般的に、中立性パラメータ η の増加に伴い MAE は増加しているが、どの中立性項でも極端に悪化はしていない。中立性項 r-match では η を増やしても誤差が減ることがあるが、これは、この項の誤差は不安定なので、統計的な変動の影響によるものと考えられる。

図 1(b) と (d) に移ると、中立性項 r-match とそれ以外とは大きく結果が異なっている。それ以外の三つの中立性項 (mi-hist, mi-normal, m-match) では、視点 Year について中立性を向上できているが、視点 Gender ではあまり出来ていない。逆に、中立性項 r-match では、視点 Gender で中立性を強化できているが、視点 Year では失敗している。この結果の差は、データでの予測評価値と視点値の独立性の差によって生じたものと考えられる。図 1(b) と (d) の左端、すなわち中立化をほとんどしていない状態に注目すると、評価値と視点値の依存性は視点 Gender より視点 Year の方がずっと大きいことが分かる。さらに、3.3.2 節で述べたように、利用者とアイテムの定義域上で一様に中立性を向上させるような設計に r-match はなっているのに対し、他の三つの項は平均的に中立性を向上させるような設計である。視点 Year では、 η の小さい状態ではあまり中立性は高くなかったため r-match 以外の項では平均的に中立性を向上できたが、r-match の制約はこの場合には強すぎたと考える。



(a) 視点 Year と Gender での MAE



(b) 視点 Year と Gender での NMI

図 3 より大きなデータ集合での予測誤差と中立性尺度の変化
Fig. 3 Changes of prediction errors and neutrality measures on a larger data set.

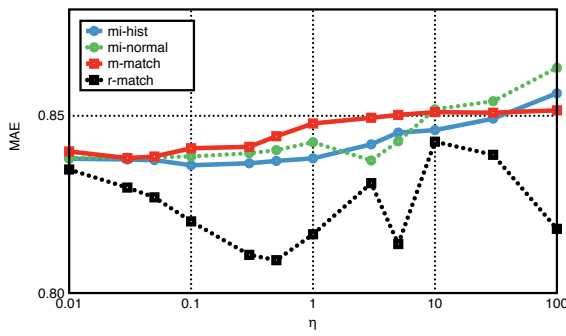
一方で、もともと中立性が高かった視点 Gender では、平均的な制約を与えるだけではそれ以上の中立性向上には失敗したが、一方で r-match の強い一様な制約では有効だったと考える。このあたりの性質についての分析は今後の課題としたい。まとめると、提案した情報中立推薦システムは、あまり予測精度を犠牲にすることなく中立性を向上させることができた。

二つの中立性項 m-match と r-match の性質をさらに調べるために、予測評価値の平均の変化を図 2 に示した。図 2 中の二つの図はどちらも、水平軸は中立性パラメータ η の値で、垂直軸は各視点値での予測評価値の平均である。図 2(a) と (b) はそれぞれ視点 Year と Gender に対する結果を示している。

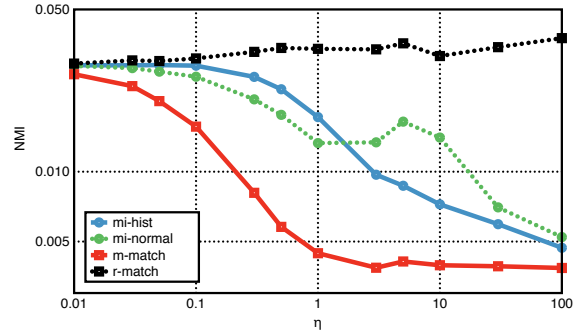
まず、以前の図 1(b) と (d) の元データの NMI の値に見られたように、この図 2(a) 左端の予測評価値の差は図 2(b) のそれよりずっと大きく、視点 Gender の方が元から中立性がより高かったことが確認できる。具体的には、視点 Year では 0.38 も差があるのに対し、視点 Gender では 0.016 である。視点 Year では、m-match では η の増加に伴い二つの平均の差が縮小できているが、中立性項 r-match は失敗している。一方、Gender では、最初から平均の差は小さいため、制約項を強めてもあまり平均の差を縮小できていない。

4.4 より大きなデータ集合での実験

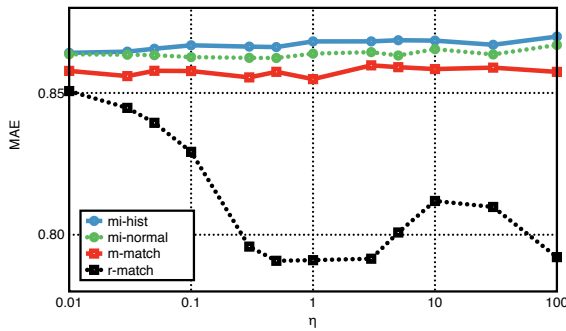
ここで提案した中立性項がより大きなデータ集合にも適用できることを示すため、前節の約 100 倍の規模の Movielens 100k データ集合全体に対して情報中立推薦アルゴリズムを適用した。以前の [5] の方法ではこの規模のデータは処理できなかった。ここでは m-match 中立性項を用い、 $K = 3$ としたこと以外は



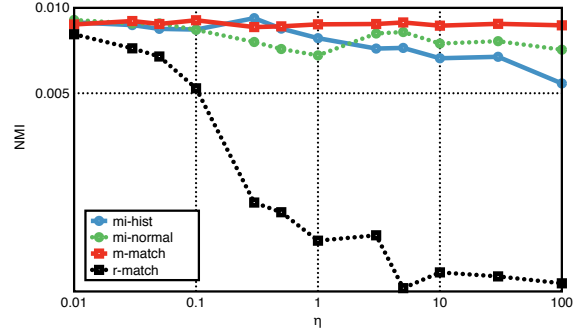
(a) Year 視点での予測誤差 (MAE)



(b) Year 視点での中立性 (NMI)



(c) Gender 視点での予測誤差 (MAE)

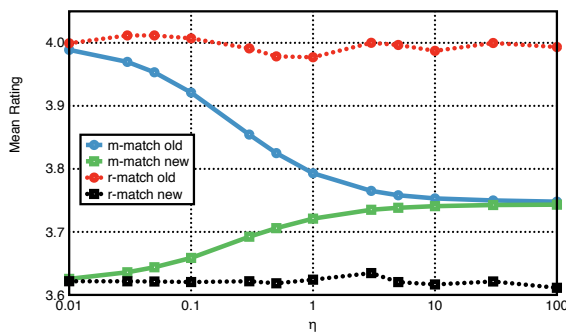


(d) Gender 視点での中立性 (NMI)

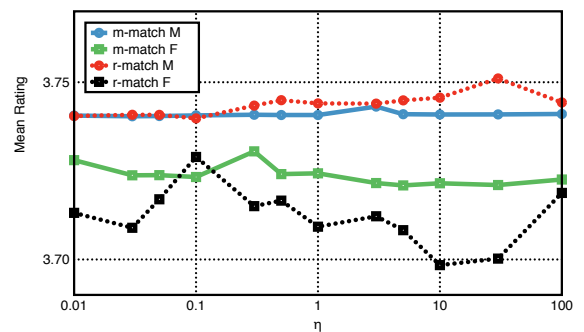
図 1 中立性パラメータの増加に伴う中立性の変化

Fig. 1 Changes of the degrees of neutrality accompanying the increase of a neutrality parameter

NOTE: 視点 Year に対する結果を図 (a) と (b) に, 視点 Gender に対する結果を図 (c) と (d) に示した. 図 (a) と (c) には平均絶対誤差 (MAE) で測った予測誤差の変化を示した. MAE は小さな値ほど, より正確な予測であることを示している. 図 (b) と (d) には正規化相互情報量 (NMI) で測った中立性の変化を示した. NMI は小さな値ほど, より厳密に中立性が達成されていることを示す. 図の水平軸は, 中立性パラメータ η の値である.



(a) 視点 Year



(b) 視点 Gender

図 2 中立性パラメータの変化に伴う平均予測評価値の変化

Fig. 2 Changes of mean predicted ratings accompanying the increase of a neutrality parameter

NOTE: どちらの図でも, 水平軸は中立性パラメータ η の値で, 垂直軸は各視点値での予測評価値の平均である. 図 (a) は視点 Year での予測評価値の平均を示し, 1990 年以前に公開された映画を “old”, 1991 年以後に公開された映画が “new” である. 図 (a) は視点 Gender での予測評価値の平均を示し, 男性と女性をそれぞれ “M” と “F” で表した.

4.2 節と同じ条件で実験した。図 3 に、 η の増加に伴う予測誤差と中立性尺度の変化を示す。この図では、以前の図 1 と同様の傾向が見られた。すなわち、中立性項 m-match では、視点 Year では予測誤差をそれほど犠牲にすることなく中立性を向上できたが、視点 Gender ではあまり中立性は向上できなかった。

最後に 4 種類の中立性項の計算時間についてふれておく。一般医、相互情報量に基づく中立性項は、CV スコアに基づくものよりも、勾配を解析的に計算できないので非常に遅い。3.3.2 節で述べたように、m-match は r-match より約 2 倍速い。実験的には、 η の増加に伴って、単純な損失項よりもより形状が複雑な中立性項の影響が大きくなり、目的関数が滑らかでなくなるため、最適化の収束は遅くなった。この η の増加に伴う影響は、中立性項 m-match よりも r-match で深刻だった。

5. 関連研究

2.2 節で述べたように、中立性は推薦の多様性と類似性がある。topic diversification は推薦リストから類似したアイテムを除外して多様性を向上させる方法である [11]。制約項を用いて多様性を向上させる手法には文献 [10] の方法がある。単一の推薦リストではなく、時間的に連続した推薦の多様性については文献 [12] で論じられている。

情報中立推薦は特定の個人情報利用を避ける目的にも利用できる。プライバシー保護データマイニング [22] とも関連がある。評価値情報に含まれる個人情報を保護するために、ダミーの評価値を加える方法が文献 [23] で提案されている。

6. まとめ

本論文では、利用者が指定した視点に対する中立性を向上させる情報中立推薦システムについて述べた。このシステムは、個人化技術が利用者の視野を狭めるという懸念であるフィルターバブル問題の緩和に有用である。そして、いくつかの中立性項を採用した情報中立推薦アルゴリズムを開発した。以前に開発した中立性項は効率的に計算できない問題があったため、この問題に対処した CV スコアに基づく中立性項を開発した。最後に、あまり予測精度を犠牲にすることなく中立性を向上できることを実験的に示した。

情報中立推薦システムに必要なとされる機能は多い。効率性を維持したままで、1 次モーメントだけでなくより厳密に視点変数と評価値変数の分布の独立性を評価できる中立性項を開発する必要がある。現状での視点変数は二値に限定されているので、これを多値離散や連続の場合にも拡張する必要がある。現状の手法は評価値予測の推薦問題を対象にしているが、良いアイテムの推薦といった問題に適用できる手法も考案する必要がある。

謝辞 本研究は JSPS 科研費 16700157, 21500154, 23240043, 24500194, 25540094 の助成を受けたものである。

文献

- [1] 神鷹敏弘, “推薦システムのアルゴリズム (1)~(3),” 人工知能学会誌, Vol. 22, No. 6 ~ Vol. 23, No. 2, 2007-2008.
- [2] E. Pariser, The Filter Bubble: What The Internet Is Hiding From You, Viking, 2011.

- [3] P. Resnick, J. Konstan, and A. Jameson, “Panel on the filter bubble,” The 5th ACM Conf. on Recommender Systems, 2011. (<http://acmrecsys.wordpress.com/2011/10/25/panel-on-the-filter-bubble/>).
- [4] 神鷹敏弘, 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.
- [5] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma, “Enhancement of the neutrality in recommendation,” Proc. of the 2nd Workshop on Human Decision Making in Recommender Systems, pp.8-14, 2012.
- [6] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma, “Fairness-aware classifier with prejudice remover regularizer,” Proc. of the ECML PKDD 2012, Part II, pp.35-50, 2012. [LNCS 7524].
- [7] E. Pariser, “The filter bubble”. (<http://www.thefilterbubble.com/>).
- [8] S. Watanabe, Knowing and Guessing – Quantitative Study of Inference and Information, John Wiley & Sons, 1969.
- [9] S. Forden, “Google said to face ultimatum from FTC in antitrust talks,” Bloomberg, Nov. 13 2012. (<http://bloomberg/PNEas>).
- [10] M. Zhang and N. Hurley, “Avoiding monotony: Improving the diversity of recommendation lists,” Proc. of the 2nd ACM Conf. on Recommender Systems, pp.123-130, 2008.
- [11] C.-N. Ziegler, S.M. McNee, J.A. Konstan, and G. Lausen, “Improving recommendation lists through topic diversification,” Proc. of the 14th Int’l Conf. on World Wide Web, pp.22-32, 2005.
- [12] N. Lathia, S. Hailes, L. Capra, and X. Amatriain, “Temporal diversity in recommender systems,” Proc. of the 33rd Annual ACM SIGIR Conf. on Research and Development in Information Retrieval, pp.210-217, 2010.
- [13] A. Gunawardana and G. Shani, “A survey of accuracy evaluation metrics of recommendation tasks,” Journal of Machine Learning Research, vol.10, pp.2935-2962, 2009.
- [14] R. Salakhutdinov and A. Mnih, “Probabilistic matrix factorization,” Advances in Neural Information Processing Systems 20, pp.1257-1264, 2008.
- [15] Y. Koren, “Factorization meets the neighborhood: A multifaceted collaborative filtering model,” Proc. of the 14th ACM SIGKDD Int’l Conf. on Knowledge Discovery and Data Mining, pp.426-434, 2008.
- [16] D. Pedreschi, S. Ruggieri, and F. Turini, “Discrimination-aware data mining,” Proc. of the 14th ACM SIGKDD Int’l Conf. on Knowledge Discovery and Data Mining, pp.560-568, 2008.
- [17] T. Calders and S. Verwer, “Three naive bayes approaches for discrimination-free classification,” Data Mining and Knowledge Discovery, vol.21, pp.277-292, 2010.
- [18] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma, “Considerations on fairness-aware data mining,” Proc. of the IEEE Int’l Workshop on Discrimination and Privacy-Aware Data Mining, pp.378-385, 2012.
- [19] “Grouplens research lab, university of minnesota”. (<http://www.grouplens.org/>).
- [20] Y. Koren, “Collaborative filtering with temporal dynamics,” Proc. of the 15th ACM SIGKDD Int’l Conf. on Knowledge Discovery and Data Mining, pp.447-455, 2009.
- [21] “Scipy.org”. (<http://www.scipy.org/>).
- [22] C.C. Aggarwal and P.S. Yu, eds., Privacy-Preserving Data Mining: Models and Algorithms, Springer, 2008.
- [23] U. Weinsberg, S. Bhagat, S. Ioannidis, and N. Taft, “Blurname: Inferring and obfuscating user gender based on ratings,” Proc. of the 6th ACM Conf. on Recommender Systems, pp.195-202, 2012.