

情報中立推薦システム

Information Neutral Recommender System

神嶋 敏弘*¹
Toshihiro Kamishima

赤穂 昭太郎*¹
Shotaro Akaho

麻生 英樹*¹
Hideki Asoh

佐久間 淳*²
Jun Sakuma

*¹産業技術総合研究所

National Institute of Advanced Industrial Science and Technology (AIST)

*²筑波大学 / 科学技術振興機構

University of Tsukuba / Japan Science and Technology Agency

This paper proposes an algorithm for recommendation so as to be guarantee the neutrality from specified information or viewpoint. This algorithm is useful for alleviating the filter bubble, which is the problem that information provided to people is biased and narrowed by the influence of a personalization technology. We empirically show that the independence from the specified variable is enhanced by our algorithm.

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 [神嶋 08]. 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は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に、フィルターバブル問題がある。

フィルターバブル (Filter Bubble) [Pariser 11, パリサー 12] とは、Pariser が主張する、利用者が接する情報の範囲に関する問題である。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にしか接しないようになっており、まるで『泡』の中に閉じ込められたような状態になっている。そのため、利用者がより新たな話題に関心をもつ機会が奪われたり、社会の中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 [Resnick 11].

この討論中では、全ての可能性を網羅することは不可能で、話題の広さと、関心のある情報の効率的な収集は本質的なトレードオフ問題であるとの指摘がなされた。何かを推薦する、より広く選び出すということは、ものごとの何らかの側面に注目し、それ以外の部分を無視することを伴う。その他に、現状の推薦システムは、より広い話題を求めるといった利用者の情報要求を拾い上げ、それを十分に満たすことに失敗しているといえるのではという指摘もあった。実際、推薦システムには、人気のあるアイテムをより推薦しやすいという人気バイアス [Celma 08] があるとの報告もある。

この問題への対処の一つとして、中立性や情報の偏りに配慮した情報中立推薦システムを提案する。パネルでの指摘にあったように、あらゆる観点からの中立性を保つことは不可能なので、利用者などが選んだある特定の視点・情報に対する中立性を、このシステムでは扱う。例えば、推薦される商品のブランドや、利用者の性別や年齢など、利用者などが選んだ特定の情報が推薦の結果に影響しないようにすることが目的となる。また、この枠組みは、法などにより利用を制限された情報を推薦に利用しないようにする手段としても利用できる。

この情報中立推薦システムを実現するために、著者が提案した公正配慮型マイニング [神嶋 11b, Kamishima 11a] の枠組みを利用する。これは、推薦などの目的タスクと、ある種の情報との中立性・独立性を保証するために、それらの間の相互情報量を小さくする制約を加えるものである。この枠組みの利用により、アイテムの予測評価値と、指定した情報との中立性・独立性が確保できることを、予備的な実験により示す。

2. 節はフィルターバブル問題と推薦の中立性について論じる。3. 節では情報中立推薦システムの定式化について、4. 節では予備的な実験結果、そして5. 節ではまとめを述べる。

2. 情報中立性

ここでは推薦の中立性について論じ、情報中立推薦システムの目標について述べる。

2.1 フィルターバブル問題

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) [Pariser 11, パリサー 12] 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問題である。

この問題に関する TED Talk [Pariser] などで、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。この推薦機能によって、当初は、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ったと述べている。Pariser は、多様な意見を取り入れたいため、保守派の人とも議論したいと考えているが、利用者から断りなく保守派の人を推薦リストから消したことは問題であると主張している。

また、2011年にエジプトで政変があった時期に、Googleでエジプトについて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生じてしまうことは問題であると主張している。

このフィルターバブルによって、大きく二つの問題が生じると主張している。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。

この問題について、推薦システムの国際会議 RecSys2011 では、以下のようなパネル討論 [Resnick 11] が、三つの話題を中心に開催された。第一の話題は、フィルターバブルは問題として存在するかどうかについてである。利用者の経験範囲を、個人化が狭めるというこの問題の可能性は 1990 年代から Resnick が指摘しており、何かの情報を送ることは、他の情報を隠すことに繋がり、本質的なトレードオフであるとの意見が述べられている。アルジャジーラと Fox ニュースの内容が異つるように、個人化技術と関連なく存在する問題であることや、情報を省く影響の度合いを正確に測ることの難しさなどの問題も指摘された。

第二に、この問題の重大さについて討論が行われた。まず、膨大な情報の全てを、自身で網羅することは不可能であり、また、絶対的に客観的な観点も存在しないことから、フィルタリング技術の利用が不可避であるとの意見が述べられた。既に信じていることを強化してしまうという selective exposure については研究がなされており、そのような傾向をもつ人も確かに存在するが、一方で影響されない人も確かに存在し、全体としては大きな影響にはならないとの見解が述べられた。

最後は、この問題に対して推薦システムはどのように対処すべきかという問題である。この問題は、短期的でなく、長期的な利用者の情報要求を反映させることや、個々のアイテムではなく、リスト・ポートフォリオ全体を考慮した推薦などの技術を進展させることで対処していくべきとの意見が出された。

2.2 推薦における中立性

絶対的に客観・中立な選択は存在しないことについては、上記のパネルでも議論されていたが、このことに関し、みにくいアヒルの子の定理 [Watanabe 69] に基づいて議論する。このみにくいアヒルの子の定理は、機械学習のクラス分類に関する原理的な問題を指摘する定理である。 n 匹のアヒルの子が、 $\log(n)$ 個の二値の特徴量によって表現されており、ある特徴量が 0/1 のいずれであるかを指定したルールによりアヒルの子を分類する場合を考える。すると、みにくいアヒルの子と、任意の普通のアヒルの子とを識別できないルールの数は、任意のアヒルの子の対を識別できないルールの数と等しくなり、この観点からはみにくいアヒルの子と普通のアヒルの子とは同程度にしか変わらないということになってしまう。では、なぜみにくいアヒルの子はみにくいのであろうか？ 全ての特徴を同等に扱っていた場合は識別出来なかったが、色が黒いといった特定の特徴に注目することにより、みにくいアヒルの子はみにくくなるのである。すなわち、分類するという行為は、分類対象の特定の特徴・側面・視点に注目することを必然的に伴うことが分かる。推薦とは、関心のあるものと、そうでないものとを識別する分類問題ともみなせるため、何らかの特徴や観点に注目することが推薦には不可欠となる。よって、あらゆる観点を同等に扱う、絶対的に中立な推薦は原理的に不可能である。

ここで、みにくいアヒルの子の定理に戻ると、分類には特徴や観点の重要性に差が必要になる。このことは、全ての観点ではなく、ある特定の観点に対する中立性であれば原理的

には可能なこと示唆している。そこで、利用者や他の主体が指定した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 の場合を例にとれば、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。

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定式化

文献 [Gunawardana 09] では、推薦システムのタスクを、利用者の関心に適合するアイテムの発見、利用者の効用の最適化、そしてアイテムに対する評価値の予測に分類している。それぞれについて、情報中立なものを想定できるが、ここでは最後の評価値予測タスクを扱う。 $x \in \{1, \dots, n\}$ と $y \in \{1, \dots, m\}$ はそれぞれ、利用者とアイテムを表し、特定の利用者 x と、特定のアイテム y との対をイベント (x, y) と呼ぶ。利用者 x がアイテム y に対して与えた評価値を s とする。評価値は、1~5 の 5 段階といった離散値などをとることが一般的だが、ここでは単に実数として扱う。さらに、イベントに依存してその値が決まり、中立性保証の対象となる目標変数 t を導入する。例えば、利用者 x のみに依存して決まる利用者の性別、アイテム y のみに依存して決まる映画の公開年、そして両方に依存して決まる利用者がアイテムを評価した時刻などを表す変数である。ここでは、 $t \in \{0, 1\}$ が二値変数の場合を扱うが、多値のカテゴリ変数への拡張は容易である。訓練事例集合は、これら四つの変数の組を N 個集めたものである、 $D = \{(x_i, y_i, s_i, t_i)\}, i = 1, \dots, N$ 。

ある新しいイベントと、それに付随する目標変数の値の組 (x, y, t) が与えられたとき、利用者 x が、アイテム y に与えるであろう評価値を予測して出力する関数を、予測評価値関数 $\hat{s}(x, y, t)$ と呼び、この関数を訓練事例から学習する。一方で、損失関数 $\text{loss}(s^*, \hat{s})$ と、中立性関数 $\text{neutral}(\hat{s}, t)$ は与えられる関数である。損失関数は、真の評価値 s^* と予測評価値 \hat{s} の非類似性の度合いを表す。一方、中立性関数は予測評価値 \hat{s} の、予測評価値を求めるときに用いたイベントに付随する目標変数の値 t に対する中立性の度合いを表す。訓練事例集合 D が与えられたとき、任意の (x, y, t) について、期待的に損失関数を小さく、中立性関数の値を大きく保つような予測評価値関数 $\hat{s}(x, y, t)$ を求めることが、情報中立推薦問題の目標となる。ここでは、損失と中立性を調整するパラメータ $\eta > 0$ を導入した次の目的関数を最小化するような \hat{s} を獲得することが具体的な目標である。

$$\text{loss}(s^*, \hat{s}(x, y, t)) - \eta \text{neutral}(\hat{s}(x, y, t), t) \quad (1)$$

3.2 予測モデル

この論文では、潜在因子モデルと呼ばれる、評価値予測モデルを情報中立推薦用に拡張する。この潜在因子モデルは、[Koren 09] の式 (3) で定義された行列分解を用いたモデルの一種で、次のようなものである：

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

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 $\text{loss}(s, \hat{s})$ として二乗誤差を導入し、次の目的関数を最小化することにより、パラメータを学習する。

$$\sum_{(x_i, y_i, s_i) \in \mathcal{D}} (s_i - \hat{s}(x_i, y_i))^2 + \lambda R \quad (3)$$

ただし、 R はパラメータ b_x, c_y, \mathbf{p}_x , および \mathbf{q}_y に対する L_2 正則化項をまとめたものであり、 λ は正則化パラメータである。パラメータを学習できれば、式 (2) 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。

この潜在変数モデルを情報中立推薦用に拡張する。まず、式 (2) を、目標変数 t の値に依存するように拡張する。ここでは、 $t = 0$ と 1 それぞれの場合に対してパラメータ $\mu^{(t)}, b_x^{(t)}, c_y^{(t)}, \mathbf{p}_x^{(t)}$, および $\mathbf{q}_y^{(t)}$ を用意する。そして、 t の値に応じてパラメータを使い分けてモデルを使い分ける

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

損失関数には、元と同じ二乗誤差を採用する。

次に、情報中立推薦を実現するために、中立性関数 neutral を定める。ここでは、文献 [神寫 11b, Kamishima 11a] のアイデアを利用し、中立性を統計的独立性と考え、その強さを負の相互情報量で測る。

$$\begin{aligned} -I(\hat{S}; T) &= \sum_{t \in \{0,1\}} \int \text{Pr}[\hat{s}, t] \log \frac{\text{Pr}[\hat{s}|t]}{\text{Pr}[\hat{s}]} d\hat{s} \\ &= \sum_{t \in \{0,1\}} \text{Pr}[t] \int \text{Pr}[\hat{s}|t] \log \frac{\text{Pr}[\hat{s}|t]}{\text{Pr}[\hat{s}]} d\hat{s} \end{aligned} \quad (5)$$

ここで、分布 $\text{Pr}[t]$ を掛けて、 t についての和をとる期待値計算を、訓練事例 \mathcal{D} についての和である標本平均と置き換えて次式を得る。

$$\frac{1}{N} \sum_{(t) \in \mathcal{D}} \int \text{Pr}[\hat{s}|t] \log \frac{\text{Pr}[\hat{s}|t]}{\text{Pr}[\hat{s}]} d\hat{s}$$

ここで、 $\text{Pr}[\hat{s}]$ は、 $\sum_t \text{Pr}[\hat{s}|t] \text{Pr}[t]$ であり、 $\text{Pr}[t]$ は訓練事例の経験分布から計算できる。あとは $\text{Pr}[\hat{s}|t]$ が評価できれば、この項を評価できる。

しかし、 \hat{s} は、 x, y, t により、予測評価値関数で確定的に定まる値であるので、分布 $\text{Pr}[\hat{s}|x, y, t]$ は、Dirac の $\delta(\hat{s}(x, y, t))$ のような分布となる。これを x と y について周辺化した分布である $\text{Pr}[\hat{s}|t]$ は同様の超関数となり、扱いが難しい。そこで、評価値は実際には離散であることが多いため、 \hat{s} の値域を離散化して、ヒストグラムにより表現した分布 $\tilde{\text{Pr}}[\hat{s}|t]$ を導入する。

$$\frac{1}{N} \sum_{(t) \in \mathcal{D}} \sum_{\hat{s} \in \text{Bin}} \tilde{\text{Pr}}[\hat{s}|t] \log \frac{\tilde{\text{Pr}}[\hat{s}|t]}{\tilde{\text{Pr}}[\hat{s}]}$$

確率密度関数 $\text{Pr}[\hat{s}|t]$ を確率質量関数 $\tilde{\text{Pr}}[\hat{s}|t]$ と置き換えたため、 \hat{s} に関する積分は、 \hat{s} に関するビンについての和となる。 $\tilde{\text{Pr}}[\hat{s}|t]$ は、訓練事例集合 \mathcal{D} のうち目標変数の値が t となっているものを集めた \mathcal{D}_t を作り、その中の各イベントについて式 (4) で $\hat{s}(x, y, t)$ を計算し、それらのヒストグラムを作ることによって計算できる。

式 (1) の形に、全体をまとめると、情報中立推薦システムの目的関数は次式となる。

$$\mathcal{L}(\mathcal{D}) = \sum_{(x_i, y_i, s_i, t_i) \in \mathcal{D}} (s_i - \hat{s}(x_i, y_i, t_i))^2 + \eta I(\hat{S}; T) + \lambda R \quad (6)$$

ただし、正則化項 R は、 t の各値についての全てのパラメータについての L_2 正則化項となる。この目的関数を最小化するように、モデル (4) のパラメータを計算すればよい。ただし、ヒストグラムを用いて $\text{Pr}[\hat{s}|t]$ の分布を計算したため、目的関数の勾配は解析的には計算できない。そのため、目的関数値のみで計算が可能な Powell 法によりパラメータを最適化した。

4. 実験

前節のアルゴリズムを実装し、小規模なデータに対して適用する実験を行った。実験には、MovieLens 100k データ集合 [Gro] を利用した。利用者 ID が 200 以下、アイテム ID が 300 以下のもののみを抽出することで、総評価値数 9,409 個の小さなデータ集合を作成した。前節では、勾配を利用しない最適化手法を採用したと述べたが、このような手法では目的関数の評価回数が非常に多くなり、計算量が大きくなる。そのため、大規模なデータは処理できないので、このような小さなデータ集合で実験を行った。

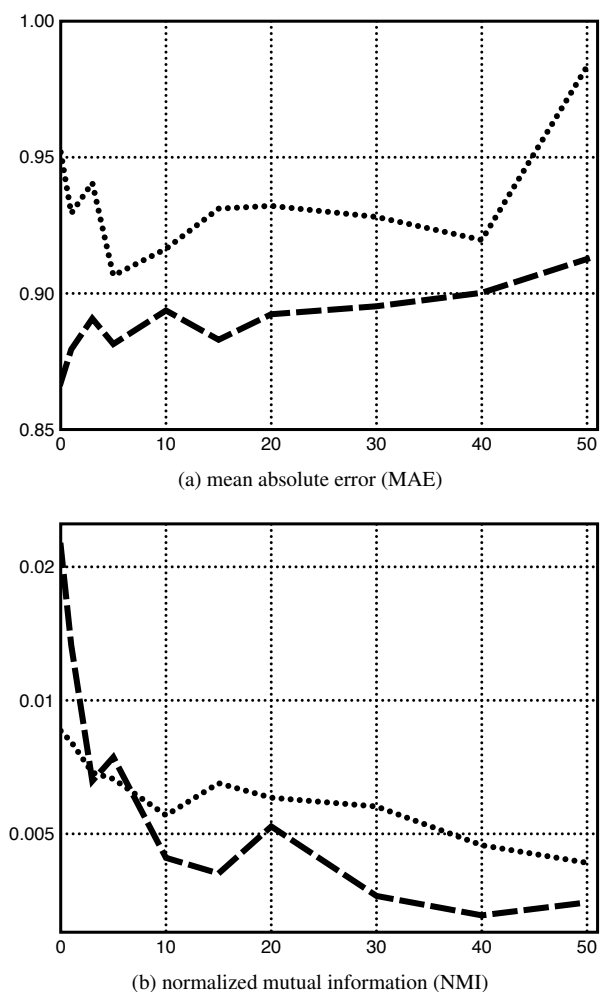
目標変数には、次の 2 種類のものを用いた。一つは、アイテムのみに依存したもので、映画の公開日が 1990 年より新しいかどうかを表す変数である。これは、映画は古いものほど、名作として鑑賞し続けられて残るため評価値が高くなる傾向がある [Koren 09] ためであり、その影響に対する中立性を強化する。もう一つは、利用者だけに依存するもので、利用者の性別を表す変数である。性別によって評価の差が生じるだろうが、その影響に対する中立性を強化する。

アルゴリズムは、前節の式 (6) を、SciPy [Sci] の Powell 法の実装を用いて、最適化した。パラメータは、訓練データ集合 \mathcal{D} を、目標変数 t の値によって二つの集合に分割し、それぞれで、情報中立用ではない、式 (3) の目的関数を、各 t ごとに最適化することで初期化した。実装上の問題から、目的関数のうち損失関数は訓練事例数で割って、 L_2 正則化項はパラメータ数で割ってスケールした。正則化パラメータ λ は 0.01 とし、 $\mathbf{p}^{(t)}$ と $\mathbf{q}^{(t)}$ 潜在因子の次元数 K は 2 とした。ヒストグラムの値は、評価値が 5 段階であるため、1~5 を中心とした 5 個のビンに分割して求めた。予測精度と中立性の評価指標は、5 分割の交差確認により求めた。

実験結果を図 1 に示す。図 (a) は予測精度を、平均絶対誤差 (Mean Absolute Error; MAE) [Gunawardana 09] で測った結果で、値が小さいほど評価値の予測精度が高いことを示す。図 (b) には、予測評価値と目標変数の値の相互情報量を、文献 [Strehl 02] の幾何平均を用いた手法で正規化した値を示した。なお、この相互情報量の計算には $\text{Pr}[\hat{s}|t]$ の評価が必要になるが、これには前節のヒストグラムモデルを用いた。横軸は、目的関数中の中立性関数の重みパラメータ η であり、中立性関数が無効となる 0 から、非常に中立性を重視する 50 まで変化させたものである。目標変数が映画の公開年である結果は破線で、利用者の性別である結果は点線で示した。相互情報量が小さいことは、高い中立性を示しているため、図 (b) の結果から、どちらの目標変数でも、設計どおりに η の増加によっても中立性が強化されている。なお、Y 軸は対数目盛であることに特に留意されたい。その一方で、図 (a) の予測誤差の増加はそれほどではなく、予測精度をあまり犠牲にすることなく中立性の強化が実現されていることが分かる。

5. まとめ

本論文では、利用者などが指定した特定の観点・情報に対して中立性を保証する情報中立推薦システムを提案した。このシ

図 1: η の変化に伴う予測精度と中立性の変化

NOTE: (a) は予測精度を, MAE で測った結果で, 小さいほど評価値の予測誤差が小さい. (b) は正規化した相互情報量で中立性を測ったもので, 小さいほど予測評価値の目標変数に対する中立性は高い. 横軸は, 中立性関数に対する強さのパラメータ η の値であり, 大きな値ほど予測精度より, 中立性を重視した推薦となる. 目標変数が映画の公開年である結果は破線で, 利用者の性別である結果は点線で示した.

システムは, 個人化技術によって利用者が多様な情報に接する機会が減るとする, フィルターバブル問題を緩和する目的で利用できる. 予測評価値と目標変数の間の相互情報量で, 中立性を測る項を, 潜在変数モデルによる推薦アルゴリズムに組み込むことで, 情報中立推薦を実現するアルゴリズムを提案した. 実験により, 提案アルゴリズムによって, 情報中立性が強化できることを示した.

情報中立推薦は, フィルターバブル問題だけでなく, プライバシーポリシーや, その他の規定により推薦に利用してはならない情報を排除する目的などにも利用でき, 広い応用を分野があると考えている. 今後の課題としては, ここで提案した手法は, 勾配を解析的に計算できないため, 大規模処理に向いていないため, この問題に対処する方法が必要である. 統計的独立性を相互情報量で測ったが, 独立成分分析で用いられている尖度などの他の指標も試したい. 今回は, 行列分解に対する拡張であったが, pLSI / LDA などの確率モデルを用いた手法などにも今後は対応してゆきたい.

謝辞

MovieLens データを公開している GroupLens プロジェクトに感謝する. 本研究は科研費 16700157, 21500154, 23240043, および 24500194 の助成を受けた.

参考文献

- [Celma 08] Celma, Ò. and Cano, P.: From Hits to Niches?: or How Popular Artists Can Bias Music Recommendation and Discovery, in *Proc. of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition* (2008)
- [Gro] GroupLens research lab, University of Minnesota (<http://www.grouplens.org/>)
- [Gunawardana 09] Gunawardana, A. and Shani, G.: A Survey of Accuracy Evaluation Metrics of Recommendation Tasks, *Journal of Machine Learning Research*, Vol. 10, pp. 2935–2962 (2009)
- [神島 08] 神島 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)
- [Kamishima 11a] Kamishima, T., Akaho, S., and Sakuma, J.: Fairness-aware Learning through Regularization Approach, in *Proc. of The 3rd IEEE Int'l Workshop on Privacy Aspects of Data Mining*, pp. 643–650 (2011)
- [神島 11b] 神島 敏弘, 赤穂 昭太郎: 公正配慮型学習 — 正則化によるアプローチ, 人工知能学会全国大会 (第 25 回) 論文集, 1D3-4 (2011)
- [Koren 09] Koren, Y.: Collaborative Filtering with Temporal Dynamics, in *Proc. of the 15th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 447–455 (2009)
- [Pariser] Pariser, E.: The Filter Bubble: (<http://www.thefilterbubble.com/>)
- [Pariser 11] Pariser, E.: *The Filter Bubble: What The Internet Is Hiding From You*, Viking (2011)
- [パリサー 12] パリサー イーライ, 井口 耕二: 閉じこもるインターネット — グーグル・パーソナライズ・民主主義, 早川書房 (2012)
- [Resnick 11] Resnick, P., Konstan, J., and Jameson, A.: Panel on The Filter Bubble, The 5th ACM conference on Recommender systems (2011), (<http://acmrecsys.wordpress.com/2011/10/25/panel-on-the-filter-bubble/>)
- [Sci] SciPy.org (<http://www.scipy.org/>)
- [Strehl 02] Strehl, A. and Ghosh, J.: Cluster Ensembles — A Knowledge Reuse Framework for Combining Multiple Partitions, *Journal of Machine Learning Research*, Vol. 3, pp. 583–617 (2002)
- [Watanabe 69] Watanabe, S.: *Knowing and Guessing — Quantitative Study of Inference and Information*, John Wiley & Sons (1969)