

ECサイトにおけるユーザの嗜好の変化の検出

Detection of Preference Shift on E-Commerce Sites

伊藤 冬子^{*1*2}

Fuyuko Ito

廣安 知之^{*3}

Tomoyuki Hiroyasu

三木 光範^{*4}

Mitsunori Miki

横内 久猛^{*4}

Hisatake Yokouchi

^{*1}同志社大学 大学院 工学研究科

Graduate School of Engineering, Doshisha University

^{*2}日本学術振興会

Japan Society for the Promotion of Science

^{*3}同志社大学 生命医科学部

Department of Life and Medical Sciences, Doshisha University

^{*4}同志社大学 理工学部

Department of Science and Engineering, Doshisha University

The purpose of this study was to increase sales opportunities by detection of the preference shift on e-commerce sites and its triggers. In this paper, a method that applies time-series clustering on preferred products was proposed to detect the preference shift timing. The features of the cluster and constraint of past clustering results were also discussed using four sets of artificial test data to determine when the clustering result had changed. As an experimental result, the preference shift timing could be detected according to the time-series variations of distance between the centroids and the difference between the spaces of two states of the same cluster. Moreover, it was confirmed that application of constraint of past clustering results helped to detect the preference shift timing correctly.

1. はじめに

近年、嗜好情報に基づいて個々のユーザにより適した情報を提示するシステムが増加している。例えば、Amazon.com^{*1}では購買履歴に基づいて、ユーザの嗜好に合う確率の高い商品をユーザに提示することで、商品販売の機会の増加を図っている。本研究ではこのようなECサイトにおける商品の推薦に着目する。これらの推薦手法はユーザの嗜好に合う商品を提示する一方で、過去の嗜好に基づいて情報を提示することから、意外性のある推薦が行えないなどの問題がある。実際には、システムを利用する時点でユーザが持っている嗜好は、過去にシステムを利用した時点やアクセス開始時点から変化し得る可能性があるが、既存の推薦手法はこれらの嗜好の変化を考慮していない。また、商品推薦を行う最大の目的は、ユーザの嗜好に合う商品を効率的に提示することであるが、サイトでの滞在時間を延ばすことも商品の販売機会の増加に繋がると考えられる。滞在時間を延ばすためには、ユーザの嗜好や興味を変化させ続けてサイトへの興味を持続させる必要がある。ECサイトのデザインなどにより購入する商品が変化したという報告などから[5, 6]、サイトの商品提示方法により嗜好の変化を促せる可能性がある。本研究ではこのような嗜好の変化とその要因の検出を行い利用することで、ユーザの嗜好の変化に対応した情報推薦、または嗜好を変化させ続けるユーザが飽きない情報推薦の実現を目指す。本稿では、ECサイトにおけるユーザの嗜好の変化を検出するために時系列クラスタリングを用いることを検討する。具体的には、ユーザが嗜好に合うと評価した商品にクラスタリングを適用し、そのクラスタリング結果の時系列変化を嗜好の変化として検出する。また、ECサイトにおけるユーザの嗜好の変化のモデルを実装したエージェントを用いて、クラスタリング結果の変化の検出について検討を行う。

2. 商品推薦におけるユーザの嗜好のモデル

2.1 一般的な嗜好のモデル

推薦手法は一般に協調フィルタリング (collaborative filtering)[4, 10] とコンテンツに基づいたフィルタリング (contents-based filtering), そしてそれらを組み合わせたハイブリッドなアプローチの3つに大別できる[1]。コンテンツに基づいたフィルタリングでは、行動履歴から獲得した嗜好情報をユーザプロファイルとしてモデル化する。まず、対象となる情報がその複数の特徴量を要素とするベクトルで表現される場合、その空間で嗜好をモデル化するアプローチの一つとして、ユーザの選好情報から特徴空間において嗜好に合う領域を特定する方法がある。この方法ではユーザの選好情報から対象となる特徴空間において嗜好に合う領域を特定したり、未知の商品の情報と選好情報の類似度を余弦で表現することでユーザの嗜好を表現している。本研究では、商品を探している瞬間に理想とする商品の持つ傾向を嗜好として定義する。このため、選好情報に基づいて特徴空間において嗜好に合う領域を対話的に特定することで、ユーザの嗜好を把握する。

2.2 特徴空間におけるクラスタリングによる嗜好の把握

本研究では、1回のアクセスにおいてユーザが興味を持ってクリックした商品の集合をユーザの嗜好とする。また、商品は特徴空間におけるベクトルで表すものとする。例えば、対象商品が衣服である場合、生地の色や袖の長さなどが商品の特徴となり、その商品のベクトルに各特徴の値を設定する。ただし、ユーザの嗜好は常に1つの傾向を示しているとは限らないため、選好商品の集合に対して特徴空間上でクラスタリングを行うことで、ユーザが同時に持つ複数の嗜好を把握する[3]。

3. 商品推薦におけるユーザの嗜好の変化

本研究では、商品推薦におけるユーザの嗜好の変化をECサイトの閲覧中の理想とする商品の傾向の変化と定義する。例えば、あるユーザがドレスを探している場合を考える。そのユーザは最初は黒いドレスばかりを探していたが、閲覧しているう

連絡先: 伊藤 冬子, 同志社大学 大学院工学研究科 知識工学
専攻/日本学術振興会, 京都府京田辺市多々羅都谷 1-3,
0774-65-6130, fuyuko@mikilab.doshisha.ac.jp

*1 <http://amazon.com/>

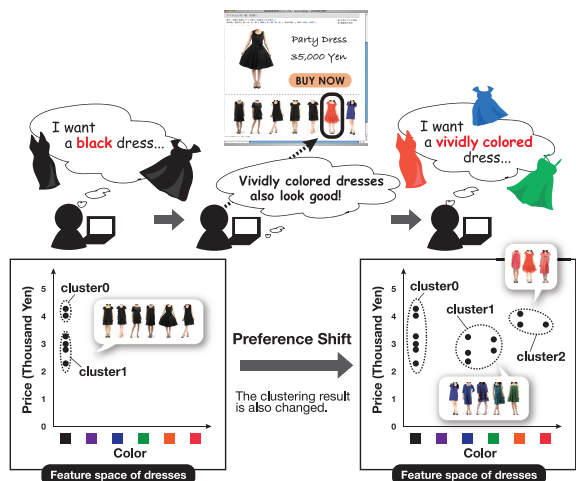


図 1: ユーザの嗜好に合う領域のクラスタリングによる特定

ちに鮮やかな色のドレスも魅力的だと気づき、ピンクやオレンジなどのドレスを探すようになる状態を嗜好の変化とする。この EC サイトにおいてドレスが色と価格の 2 次元の特徴ベクトルで表現できる場合、閲覧したドレスの履歴は図 1 のように特徴空間にプロットすることが出来る。

この嗜好の変化はユーザが商品をクリックするごとに行われるクラスタリング結果の変化として表現される。そのため、例で示した閲覧したドレスのクラスタリング結果も、図 1 のように変化する。ただし、時系列データのクラスタリングを行うためには、次の 2 点を考慮する必要がある。

- クラスタリングを適用するデータの選択
- クラスタリング結果の劇的な変化の抑制

まず、最も単純な時系列クラスタリングの方法としては、ユーザが好みの商品をクリックするごとに、その時点までに蓄積されているクリックされた商品のデータに対してクラスタリングを行うことが考えられる。しかし、長期間にわたって大量のデータが蓄積されている場合、多くのデータによって形成されているクラスタ群とは全く異なる性質を持つ新しいデータが少しずつ追加されても、クラスタリング結果はあまり変化せず、嗜好の変化を捉えられない可能性がある。そのため、クラスタリングに適用するデータの選択が必要となる。本研究では、新しい順から一定数のサンプルデータを用いるスライディングウィンドウ方式をとる。

また、商品をクリックするごとに独立に蓄積してきたデータにクラスタリングを行うと、これまで得られていたクラスタ構造から全体的にクラスタリング結果が急激に変わってしまう場合が想定される。そのため、前の時点でのクラスタリング結果を重みとして現時点のクラスタリングに加味することで [2, 9], クラスタリング結果の劇的な変化を抑制する。

本稿ではまず、クラスタリング結果の変化を検出するために、それを顕著に反映するクラスタの特徴量について検討を行う。また、蓄積されていくデータのクラスタリングに過去のクラスタリング結果を重みとした制約がどのように影響するかを検討する。

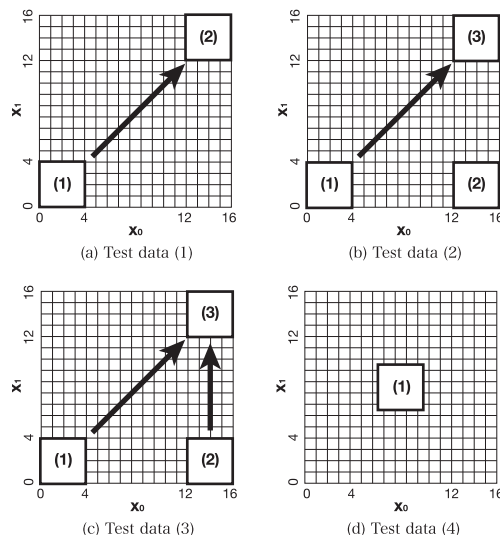


図 2: エージェントに組込まれた嗜好

4. クラスタリング結果の変化の検出に用いる特徴量の検討

4.1 実験概要

本実験ではインクリメンタルに増加する商品の時系列データを人工的に生成してクラスタリングを行い、前の時点との比較でクラスタリング結果の変化を把握するために用いるクラスタの特徴量について検討を行う。対象とするデータ、使用するクラスタリング手法、異なる時間におけるクラスタ間の関係の同定、検討するクラスタの特徴量は下記のとおりである。

4.1.1 対象データ

商品の特徴空間を 2 次元とし、商品 x は $x = (x_0, x_1)$ と表現できる。ただし、 $0 \leq x_0, x_1 \leq 16$ とし、 x_0 および x_1 は実数とする。本実験では、EC サイトにおいて想定されるユーザの嗜好の変化のモデルとして下記の各嗜好を持つエージェントにそれぞれ 24 個のデータを生成させて 4 種類のテストデータを用意した。なお、単位時間は 1 つのデータの追加とするため、 $0 \leq t \leq 24$ となる。

- (1) 単一の嗜好が変化する場合
探索の前半はユーザの嗜好に合う範囲を図 2(a) に示す範囲 (1)、後半から範囲 (2) に変化すると定義する。実際には、まず図 2(a) の範囲 (1) に 12 個のデータを一樣乱数に従いランダムに生成し、次に範囲 (2) において同様に 12 個のデータをランダムに生成する。これより、このテストデータにおいて嗜好が変化した時刻は $t = 13$ となる。
- (2) 複数の嗜好のうち一方が変化する場合
探索の最初はユーザ嗜好に合う範囲が図 2(b) に示す範囲 (1) と範囲 (2) であるとし、残りの期間は範囲 (1) の嗜好が範囲 (3) に変化すると定義する。なお、範囲 (2) は一貫して嗜好に合う範囲であると定義する。このテストデータにおいて嗜好が変化した時刻は $t = 14$ となる。
- (3) 複数の嗜好が別の単一の嗜好に変化する場合
探索の最初はユーザ嗜好に合う範囲が図 2(c) に示す範囲 (1) と範囲 (2) であるとし、残りの期間は範囲 (1) の嗜好が範囲 (3) に変化すると定義する。このテストデータにおいて嗜好が変化した時刻は $t = 17$ である。
- (4) 嗜好が変化しない場合
一貫して図 2(d) に示す範囲 (1) に 24 個のデータを一樣

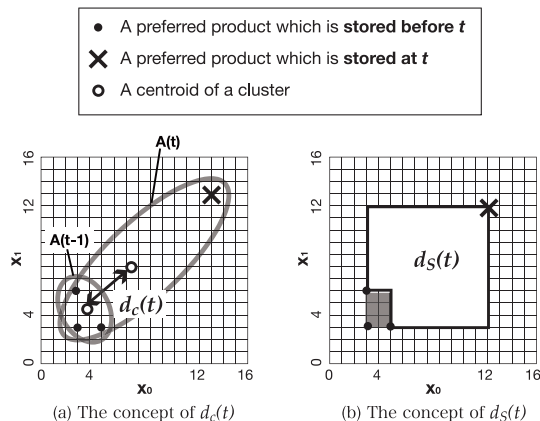


図 3: 2 つのクラスタの重心間の距離 $d_c(t)$ 及びデータが特徴空間に占める面積の差 $d_S(t)$

乱数に従いランダムに生成する．これより，このテストデータにおいては嗜好は変化しない．

4.1.2 クラスタリング手法

クラスタリング手法としては，ネットワークからコミュニティを抽出する手法として用いられている Newman 法 [7] を，重み付きネットワークに適用できるように改良した手法を用いる．なお，Newman 法においては，クラスタ数は自動的に決定される．

本研究では各商品のデータ間の類似度をエッジの重みとするネットワークを対象とする．そのため，特徴空間の各データの座標から，2 つのデータ間の距離の逆数を類似度として用いる．また，クラスタリングの対象とするデータはスライディングウィンドウ方式により最新の n 個のデータを用いるものとする．本実験では $n = 9$ とした．なお，本実験では前の時点でのクラスタリング結果の制約はかけていない．

4.1.3 異なる時刻におけるクラスタの関係の同定

あるクラスタの時系列変化を検証するためには，時刻 $t - 1$ におけるあるクラスタ $A_i(t - 1)$ が，時刻 t におけるどのクラスタ $A_j(t)$ と対応関係にあるかを同定する必要がある．本研究では，2 つのクラスタ間の相関を自動的に算出するために以下の auto-correlation function [8] を用いる．なお，下式の子分子はクラスタ間で共通しているノードの数であり，分母は和集合のノード数となる．各時刻におけるクラスタ，前の時点のクラスタとの間で $C_{A_{ij}}(t)$ を用いて相関を算出し，相関が高い組合せから同一のクラスタであるとする．

$$C_{A_{ij}}(t) \equiv \frac{|A_i(t-1) \cap A_j(t)|}{|A_i(t-1) \cup A_j(t)|} \quad (1)$$

4.1.4 クラスタの特徴量

本実験でクラスタ間の変化を検討するクラスタの特徴量を以下に示す． $d_c(t)$, $d_S(t)$ は時刻 $t-1$ と時刻 t におけるクラスタ間の特徴量であり，その概念は図 3 に示した．

- 2 つのクラスタの重心間の距離 $d_c(t)$
- 2 つのクラスタが特徴空間に占める面積の差 $d_S(t)$

4.2 実験結果及び考察

まず，単一の嗜好が変化する場合のテストデータ (1) についてクラスタの各特徴量がどのように変化しているかを検証する．図 4 にテストデータ (1) における時刻 t における全ての

クラスタの $d_c(t)$, $d_S(t)$ の合計値の推移を示す．図 4 の左側の縦軸は $d_c(t)$ の合計値を，右側の縦軸は $d_S(t)$ の合計値を表している．横軸は時刻 t を表している．なお，クラスタリングはノード数がウィンドウのサンプル数 $n = 9$ と等しくなった時点から適用を開始している．

図 4 から分かるように， $t = 13$ において， $d_c(t)$ と $d_S(t)$ が急激に増加しているのが分かる．テストデータ (1) では嗜好の変化が $t = 13$ に起こることから， $d_c(t)$ と $d_S(t)$ によりクラスタリング結果の変化を捉えられる可能性が示唆された．一方で， $t = 20$ においても， $d_c(t)$ と $d_S(t)$ が増加しているが，これはスライディングウィンドウにより，初期の嗜好に適合したデータがウィンドウに含まれなくなったためである．このことから，ウィンドウのサンプル数 n についても今後検討を行う必要がある．

次に，複数の嗜好のうち一方が別の嗜好に変化する場合のテストデータ (2) におけるクラスタの $d_c(t)$ と $d_S(t)$ の合計値の変化を図 5 に示す．このテストデータでは嗜好の変化が起こる $t = 14$ において， $d_c(t)$, $d_S(t)$ 共に増加しているが，これは $t = 17$ と比較すると増加幅が小さい．これは， $t = 13$ において範囲 (2) の 2 つのデータを含んでいたクラスタに， $t = 14$ で範囲 (3) のデータが 1 つ加わったことにより重心が移動したが，クラスタ内のデータが範囲 (2) に極端に近かったために重心の移動距離が小さく，面積もあまり変わらなかったと考えられる．これより，クラスタ内のデータの分布，共分散についても今後検討する必要がある．

複数の嗜好が別の単一の嗜好に変化する場合のテストデータ (3) ではテストデータ (1) と同様に，嗜好の変化が起こる

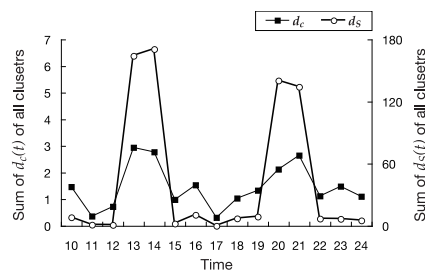


図 4: テストデータ (1) における $d_c(t)$ と $d_S(t)$ の推移

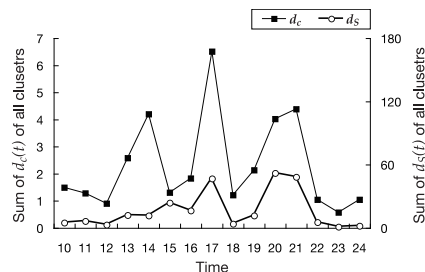


図 5: テストデータ (2) における $d_c(t)$ と $d_S(t)$ の推移

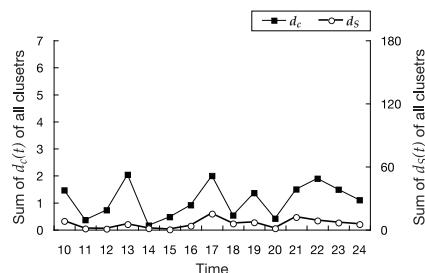


図 6: テストデータ (4) における $d_c(t)$ と $d_S(t)$ の推移

$t = 17$ において、 $d_c(t)$ と $d_s(t)$ が急激に増加した。ただし、 $t = 20, 24$ においても $d_c(t)$ と $d_s(t)$ の大きな変化が見られる。これはスライディングウィンドウにより、1つだけ残された初期の嗜好に適合したデータが、Newman法によって他のクラスタに吸収されるために起こる現象であると考えられるため、今後検討を行う。

最後に、嗜好が変化しない場合のテストデータ(4)におけるクラスタの $d_c(t)$ と $d_s(t)$ の合計値の変化を図6に示す。テストデータ(4)では嗜好の変化が起こらないため、テストデータ(1)~(3)のように $d_c(t)$ と $d_s(t)$ が変化しないかを確認する必要がある。図6から分かるように、図4や図5と比較すると、 $d_c(t)$ や $d_s(t)$ の変化が殆ど見られないことが分かる。

以上の結果から、クラスタの重心間の距離 $d_c(t)$ とクラスタ間の面積の差 $d_s(t)$ を用いることで、嗜好の変化を検出できる可能性が示唆された。また、嗜好が変化しない場合には、双方ともに変化が見られないことも確認できた。

5. 過去のクラスタリング結果による制約に関する検討

5.1 実験概要

4章の実験結果から、クラスタの重心間の距離 $d_c(t)$ とクラスタ間の面積の差 $d_s(t)$ により、嗜好商品のクラスタリング結果の変化を検出できる可能性が示唆された。一方で、実際には嗜好を変化させていない時点でも誤検出してしまう場合が確認された。そこで、本実験では4章の実験と同様のテストデータを用いて、現在の時刻 t において前の時点 $t-1$ のクラスタリング結果を制約とした制約付きクラスタリングを行い、同様の特徴量の変化を検証した。制約の強さ r は0.01, 0.03, 0.1の4つの値を用いて比較した。なお、制約なしの場合は4章の実験結果を参照するものとする。

5.2 実験結果

ここではテストデータ(2)における $r = 0.03$ 及び0.1の場合におけるクラスタの重心間の距離 $d_c(t)$ の合計値とクラスタ間の面積の差 $d_s(t)$ の合計値の推移を、それぞれ図7と図8に示した。なお、制約なしの実験結果は4章の図5となる。

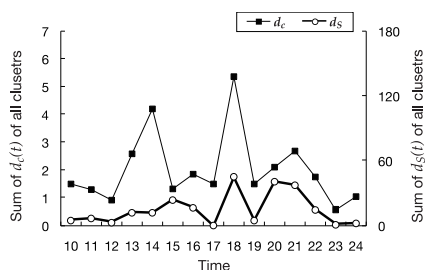


図7: $d_c(t)$ と $d_s(t)$ の推移(テストデータ(2), $r = 0.03$)

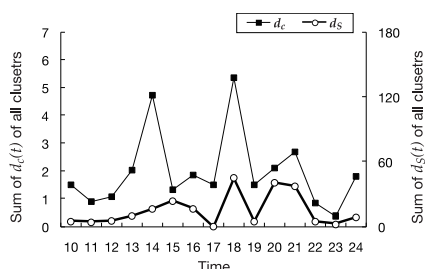


図8: $d_c(t)$ と $d_s(t)$ の推移(テストデータ(2), $r = 0.1$)

テストデータ(2)は $t = 14$ で嗜好を変化させているが、図4から分かるように制約をかけない場合、 $t = 17$ や $t = 20$ で $d_c(t)$ と $d_s(t)$ が増大しており、双方共に変化を設定した時点とは異なる時刻で嗜好の変化を検出していた。一方、制約 $r = 0.03$ をかけた場合、図7から分かるように、 $t = 20$ における $d_c(t)$ と $d_s(t)$ の増大が抑えられている。また、さらに強い制約 $r = 0.1$ を与えた場合、図8から分かるように、 $d_c(t)$ が $t = 17$ で低減し、 $t = 14$ において増大する傾向が見られた。ただし、他のテストデータにおいては制約をかけた場合でも制約をかけない場合とほぼ変わらない結果となったため、制約の強さによって検出性能にどのような影響があるかについて今後検討する。

6. 結論

本研究では、ECサイトにおけるユーザの嗜好の変化とその要因の検出を行い利用することで、販売機会の増加を目指す。本稿では、基礎的な検討として、ユーザが嗜好に合うと評価した商品に時系列クラスタリングを用いることで、嗜好の変化を検出可能か検討した。人工テストデータによる実験結果から、異なる時刻の2つのクラスタについて、その重心間の距離と、面積の差の変化から、嗜好が変化したタイミングを検出できる可能性が示唆された。また、クラスタリング結果の劇的な変化を防ぐために過去のクラスタリングによる制約を加えた場合には、変化の誤検出を抑える挙動が一部のテストデータで見られたため、引き続き制約の強さが変化の検出に与える影響についても検討を行う。今後の課題としては、ウィンドウのサンプル数を変化させた場合、前の時点でのクラスタリング結果による制約をかけた場合の検討や、被験者実験による実際の人間の嗜好変化の検出の検証が挙げられる。また、異常値検出などの他手法の利用や比較も検討する必要がある。

参考文献

- [1] Adomavicius, G., Tuzhilin, E.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, Vol.17, No.6, pp.734-749 (2005)
- [2] 福井健一, 斎藤和己, 木村昌弘, 沼尾正行: 自己組織化ネットワークによる動的クラスタの可視化編纂 人工知能学会論文誌, Vol.23, No.5, pp.319-329 (2008)
- [3] 伊藤冬子, 廣安知之, 三木光範, 横内久猛: 対話型遺伝的アルゴリズムにおける嗜好の多峰性に対応可能な個体生成方法. 人工知能学会論文誌, Vol.24, No.1, pp.127-135 (2009)
- [4] Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., Riedl, J.: GroupLens: applying collaborative filtering to usenet news. *Communications of the ACM*, Vol.40, No.3, pp.77-87 (1997)
- [5] Koufaris, M., Kambil, A., LaBarbera, P.A.: Consumer Behavior in Web-Based Commerce: An Empirical Study. *International Journal of Electronic Commerce*, Vol.6, No.2, pp.115-138 (2002)
- [6] Mandel, N., Johnson, E.J.: When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices. *Journal of Consumer Research*, Vol.29, No.2, pp.235-245 (2002)
- [7] Newman, M.E.J.: Fast algorithm for detecting community structure in networks. *Physics Review E*, Vol.69, Issue 6, 066133 (2004)
- [8] Palla, G., Barabasi, A.L., Vicsek, T.: Quantifying social group evolution. *Nature*, Vol.446, No.7136, pp.664-667 (2007)
- [9] 榎剛史, 松尾豊, 石塚満: 制約付きクラスタリングを用いた論文分類. 人工知能学会第20回全国大会講演論文集, Vol.20, 1A1-1 (2006)
- [10] Sarwar, B., Karypis, G., Konstan, J., Reidl, J.: Item-based collaborative filtering recommendation algorithms. In: *Proceedings of the 10th international conference on World Wide Web*, pp.285-295 (2001)