カテゴリにおける所属度と非典型度に基づく希少なWebページの推薦

多田 亮平† 湯本 高行††,a 新居 学††,b 佐藤 邦弘††,c

† 兵庫県立大学大学院工学研究科（研究当時） †† 兵庫県立大学大学院工学研究科

a) *yumoto@eng.u-hyogo.ac.jp* b) *nii@eng.u-hyogo.ac.jp* c) *ksato@eng.u-hyogo.ac.jp*

概要 本研究では，認知度が低いが有益な情報の発見のために，ユーザが指定したカテゴリにおける Web ページの希少度を提案する．希少な Web ページをユーザが指定したカテゴリに所属し，かつそのカテゴリ内で典型的でないとみなし，その確率を希少度と定義する．指定したカテゴリに Web ページが所属する確率を所属度，指定したカテゴリに所属している場合に典型的でない確率を非典型度とすると，希少度はこれらの積である．各確率は，ソーシャルブックマークに登録されているページにおけるタグの使用状況とページ内の語の出現状況から算出する．所属度をカテゴリの分類精度によって評価したところ，被タグ付け回数による場合と同等であった．所属度の算出には対象のページがブックマークされている必要がなく，適用範囲が広いという利点がある．また，希少度によるランキング結果を所属度と非典型度のそれぞれのみを用いた場合と DCG で比較を行い，有用性を確認した．

キーワード 典型性，確率モデル，ソーシャルブックマーク, ソーシャルタグ

1 はじめに

近年，Web をインフラとして膨大な量の情報が利用できるようになり，情報検索や情報推薦の技術の発展も相まって，ユーザが欲する情報は得やすくなっていると言える．一方，中村らの調査によると，Web 検索においては上位の 5 位以内，1〜3 ページを閲覧するユーザが多いと報告されている [1]．このような状況を考慮すると，ユーザが知らない情報，特に多くのページに書かれないような情報については，依然として取得は難しいと言える．これに対して，推薦結果や検索結果を多様化する [2, 3]，ユーザにとって未知の情報を優先して推薦する [4] などの方向性で研究が行われている．

我々は，認知度は低いが有益な情報を優先的に推薦する手法について研究を行っている．本研究では，特に典型的ではない情報に注目する．検索や推薦を行う場合はユーザの意図に合致していることが必須であり，単に典型的でないだけではユーザの意図に無関係なものが多数になることが予想される．そこで，Web ページが，ユーザの指定したカテゴリに所属し，かつ典型的でないことを表す指標として希少度を定義する．希少度は指定したカテゴリに所属する確率を表す所属度と指定したカテゴリ内で典型的でない確率を表す非典型度の積である．

所属度および非典型度の算出にはソーシャルブックマーク (以下，SBM) のデータを用いる．SBM サービスは Web ページに対するブックマークをオンラインで管理するサービスである．Web ページごとの SBM の件数は，Web ページの人気を表す指標としても用いられる [5]．ユーザは Web ページをブックマークする際にタグを付与することができる．タグはユーザごとに自由に決めることができ，様々な使い方がされている．Golder と Huberman は，Delicious のタグをその機能から 7 種類に分類した [6]．そのうち，“Identifying What (or Who) it is About”に分類されるタグは，Web ページの内容を表している．本研究では，このようなタグがカテゴリを表しているとみなす．なお，本研究では，カテゴリはユーザが指定することを想定しているため，カテゴリを表さないタグは指定されないことを仮定している．タグ付けされた Web ページ中に含まれる語を用いて，ある語が出現する場合にカテゴリに所属する確率やカテゴリに所属する場合にある語が出現する確率を用いて，所属度および非典型度を算出する．

2 関連研究

情報推薦の分野では主に推薦精度を向上を目指し，協調フィルタリング [7, 8] をはじめとして，さまざまな手法が開発されてきた．これに対して，Herlocker らは，精度以外の指標の必要性を指摘し，その例として，目新しさやセレンディピティについて言及している [9]．これらの観点から性能の向上を目指す研究も存在する．

推薦リストの多様化によってこれを実現しようとした研究としては，小川ら [2] や Ziegler ら [3] の研究がある．小川らの手法では，推薦リストを作成する際に異なるトピックを選択することで推薦リストの多様化を行っている．また，Ziegler らの手法では，推薦リストを構成するアイテム間での類似度が小さくなるように推薦リストを構成することで推薦リストを多様化を図っている．

意外なアイテム推薦することを目的とする研究としては，加藤ら [10] や奥ら [11] らの研究がある．加藤らの

手法では，推薦の正確性を保ちながら，意外性を高めるために，閲覧中の商品が所属するカテゴリから商品を推薦する順マッチングとこれ以外のカテゴリから商品を推薦する交差マッチングを併用している．順マッチングと交差マッチングのバランスは過去の選択履歴を考慮して動的に決定される．奥らは，2 つの推薦アイテムの特徴を混ぜ合わせることによって新たなアイテムを推薦するフュージョンベース推薦を提案している．

既知でないことを重視した研究としては，清水らの研究がある [4]．この手法では，既知のアイテムかどうかの評価値を導入し，既知でないアイテムが重視されるように嗜好の評価値と組み合わせて推薦を行う．また，本研究は，特定のユーザにとって未知かどうかは考慮していないが，多くのユーザにとって未知であると考えられる非典型的な Web ページを推薦対象にしている．

また，佃らは典型性を考慮したオブジェクト検索を提案している [12]．この手法では，典型的なオブジェクトを探すことを目的としているが，本研究では，これとは逆に典型的でないものが発見することが目的である．

3 希少な Web ページの推薦手法

本研究では，ユーザが指定したカテゴリにおける，Web ページの希少度を算出する．対象の Web ページは RSS による新着記事や SBM サービスで提供される Web ページ集合，検索結果などを想定しており，これらから希少度の高いものを順に推薦する．希少度は所属度と非典型度の積として算出するが，算出には SBM でタグ付けされた Web ページを学習データとして用いる．なお，希少度の算出の対象となる Web ページは学習データに含まれている必要はなく，任意の Web ページで希少度が算出できる．

3.1 希少な Web ページ

本研究では，ユーザが指定したカテゴリに所属し，かつ典型的でない情報を記載した Web ページを希少な Web ページと定義する．ここで，ある Web ページ d がカテゴリ c に所属するという事象を C，d が c において典型的な内容でないという事象を $\bar{T}$ とする．このとき，$\bar{T}$ は c と関係がない情報を記載している場合も含む．希少な Web ページとは C と T の事象を同時に満たしていると考え，d の c での希少性を表す希少度 $Rarity$ を d が C と T を同時に満たす確率 $P(C \cap \bar{T})$ として定義する．このとき，同時確率 $P(C \cap \bar{T})$ を条件付き確率 $P(\bar{T}|C)$ を用いて式 (1) のように表すことができる．

$$Rarity(c,d) = P(C \cap \bar{T}) = P(C)P(\bar{T}|C) \qquad (1)$$

$P(C)$ は d が c に所属する確率を示し，$P(\bar{T}|C)$ は d が c に所属する場合，c 内で典型的でない確率である．前者を所属度と定義し，$Belong(c,d) = P(C)$ で表す．後者を非典型度と定義し，$Atypicality(c,d) = P(\bar{T}|C)$ と表す．具体的な算出方法については後述する．

3.2 学習データの構築

Web ページ d のカテゴリ c への所属度および d の c 内での非典型度の算出には，$|BM^c|$，$|BM_w|$，$|BM^c \cap BM_w|$ の値が必要である．$|BM^c|$ は SBM サービス内でタグ c を使用しているブックマークの数で，$|BM_w|$ は SBM サービス内で本文中に名詞 w が記載された Web ページへのブックマーク数である．また，$|BM^c \cap BM_w|$ はタグ c を用いて，本文中に名詞 w を含む Web ページをブックマークしている数である．

これらの値は，SBM サービスに登録されている Web ページに付与されているタグの情報とこれらの Web ページの本文に含まれる語の情報を格納したデータベースを構築し，これを学習データとして用いることで取得する．具体的には，まず SBM サービスに登録されている Web ページに対して，タグの情報を収集して，Web ページとタグの対応関係をデータベースに格納する．続いて，これらの Web ページをダウンロードして本文抽出を行う．本文抽出ができたページに対して形態素解析を行い，名詞のみを抽出して Web ページと名詞の対応関係をデータベースに格納する．

3.3 所属度の算出手法

所属度は $P(C)$ と定義するが，d が c に所属しない確率 $P(\bar{C})$ との間に $P(C) = 1 - P(\bar{C})$ の関係が成り立つため，$P(\bar{C})$ の算出方法について述べる．ここで，d の本文中において主要な語のすべてが c に関係しない場合，d は c に所属しないと考える．ある語 w が c に関係しない確率を $P(\bar{c}|w)$ とし，d の本文中のすべての語は独立に出現すると仮定すると，$P(\bar{C})$ は式 (2) のように定義できる．

$$P(\bar{C}) = \prod_{i=1}^{n} P(\bar{c}|w_i) \qquad (2)$$

w_i は d において i 番目に主要な語であり，n は主要な語の数である．また，w が c に関係する確率を $P(c|w)$ とすると，式 (3) が成り立つ．

$$P(\bar{c}|w) = 1 - P(c|w) \qquad (3)$$

次に，$P(c|w)$ を SBM サービスのブックマーク情報から算出する．$P(c|w)$ は，w が本文中に出現する Web ページに対するブックマークのうち，タグ c を使用しているものの割合とし，式 (4) のように定義する．

$$P(c|w) = \frac{|BM_w \cap BM^c|}{|BM_w|} \qquad (4)$$

式 (4) でブックマーク数の代わりにページ数を使うことも考えられるが，その場合は，少数のユーザが間違えて

タグ付けをした影響を受けやすいと考えられる．これに対して，ブックマーク数を用いた場合は多くのユーザがタグ付けしたページの影響が強くなるため，相対的に誤ったタグ付けの影響は小さくなる．このような理由からブックマーク数を用いている．

以上の式 (2), (3), (4) から $P(\bar{C})$ は以下のようになる．

$$P(\bar{C}) = \prod_{i=1}^{n} \{1 - P(c|w_i)\} \quad (5)$$

したがって，所属度 $Belong(c,d)$ は以下のように定義できる．

$$Belong(c,d) = P(C) = 1 - \prod_{i=1}^{n} \{1 - P(c|w_i)\} \quad (6)$$

また，主要な語は，式 (7) で定義される TFIDF 値の値の上位 n 語とする．

$$TFIDF(w,d) = \frac{TF(w,d)}{DF(w,D)} \quad (7)$$

$TF(w,d)$ は文書 d の本文中での名詞 w の頻度である．また，$DF(w,D)$ は文書集合 D 内で w を本文に含む文書の数であり，D は SBM データベースに登録されている Web ページの集合とする．

3.4 非典型度の算出手法

d の主要な語が c 内でほとんど出現しない語である場合に，d は c 内で典型的でないと考え，d の本文中において主要な語のすべてが c 内では出現しない確率を d は c 内で典型的ではない確率であると考える．このとき，d 中の n 個の主要な語について，各語 w_i が c 内で出現しない確率 $P(\bar{w}_i|c)$ とする．d の本文中の n 個の主要な語は独立に発生すると仮定すると，$P(\bar{T}|C)$ は式 (8) として表すことができる．

$$P(\bar{T}|C) = \prod_{i=1}^{n} P(\bar{w}_i|c) \quad (8)$$

また，c 内で w が出現する確率を $P(w|c)$ とすると，式 (9) が成り立つ．

$$P(\bar{w}|c) = 1 - P(w|c) \quad (9)$$

次に，$P(w|c)$ を，タグ c を使用しているブックマークのうち，w が出現するページに付与されているブックマークの割合とし，式 (10) のように定義する．

$$P(w|c) = \frac{|BM_w \cap BM^c|}{|BM^c|} \quad (10)$$

式 (10) は式 (4) と同様の理由でブックマーク数を用いている．

式 (8), (9), (10) から非典型度を式 (11) のように定義する．

$$Atypicality(c,d) = P(\bar{T}|C) = \prod_{i=1}^{n} \{1 - P(w_i|c)\} \quad (11)$$

4 評価実験

4.1 使用したデータ

本研究では，はてなブックマークから取得したデータを学習データとして用いた．具体的には，2011 年 4 月 14 日から 2011 年 10 月 27 日までの期間に，表 1 に示すホットエントリーの最新の URL 一覧を定期的に取得し，これを対象ページとした．また，これらのページをブックマークしているユーザの一覧を取得し，これらのユーザがブックマークしているページの一覧を対象に加えた．これらの対象ページをダウンロードし，本文抽出ができた Web ページの情報のみをデータベースに格納し，実験対象および学習データとして用いた．表 2 に収集したデータの規模を示す．なお，本文抽出には ExtractContent[1] を，形態素解析には Mecab[2] を用いた．

表 1 基点とするホットエントリーのトピック

政治・経済	社会	スポーツ・芸能・音楽
生活	科学・学問	コンピュータ・IT
ゲーム・アニメ	おもしろ	

表 2 収集したデータの規模

総ブックマーク数	1149475
総タグ種類数	28996
ユニーク URL 数	11873
取得した名詞の種類数	188227

評価実験では，SBM サービス内で使用回数が多く，かつカテゴリを表しているタグ 10 件をカテゴリとして用いた．表 3 に実験に使用したカテゴリの一覧を示す．

表 3 評価実験に使用したカテゴリ

政治	ビジネス	音楽	ゲーム	アニメ
テレビ	映画	google	twitter	iphone

4.2 所属度の評価

所属度についてはカテゴリ分類器としての評価を行う．所属度の評価実験用に，指定したカテゴリに所属しているかを人手で評価した Web ページ集合 D_{test} を用意した．この D_{test} に含まれる各 Web ページに対して，c への所属度がしきい値以上の Web ページは c に所属し，c への所属度がしきい値未満の Web ページは所属しないと分類するときの分類性能で評価を行う．このとき所属度の比較対象として，SBM サービス内で Web ページに対してタグ c が付けられた回数 (以下，被タグ付け

[1] http://labs.cybozu.co.jp/blog/nakatani/2007/09/web_1.html

[2] http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html

回数) がしきい値以上の Web ページは c に所属し，しきい値未満の Web ページは所属しないと分類する場合と比較する．

4.1 節で作成したデータベースの一部を D_{test} として用い，残りを学習データとして用いた．D_{test} の作成方法について述べる．まず，タグ c が付与されているページとされていないページをそれぞれ 50 件用意する．タグ c が付与されているページについては，タグが付けられた回数が偏らないように，c が付けられた回数に応じてページ集合を 10 等分し，それぞれの Web ページ集合から，無作為に 5 件ずつ取得する．このように取得した，合わせて 100 件の Web ページを実験データ D_{test} とする．この D_{test} の各ページに対して著者のうちの 1 名が c に所属するかを判定し，正解を決めた．

表 3 のカテゴリ c ごとに用意した D_{test} の各 Web ページに対して，所属度と被タグ付け回数を取得した．それぞれのしきい値を変化させ，ROC 曲線を描き，算出した AUC[13] およびそれらの平均を表 4 に示す．

表 4 被タグ付け回数，所属度を用いた分類器の AUC

カテゴリ	被タグ付け回数	所属度
google	0.99	0.94
iphone	0.99	0.88
twitter	0.98	0.98
アニメ	0.98	0.96
ゲーム	0.97	0.93
テレビ	0.95	0.84
ビジネス	0.91	0.97
映画	0.96	0.93
音楽	0.99	1.00
政治	0.97	0.96
平均	0.97	0.94

表 4 から被タグ付け回数を用いた場合の AUC の平均は 0.97 と 1.0 に非常に近く，SBM ユーザによるカテゴリ分類が正しいことを示している．また，所属度を用いて分類した場合も，被タグ付け回数を用いた場合と同様に AUC の平均は 0.94 と大きい値をとなり，分類基準としては被タグ付け回数に匹敵する性能であると言える．このことから，所属度によって Web ページがカテゴリに所属していることを表現することは妥当であると言える．さらに，所属度は SBM サービスに登録されていない Web ページに対しても算出できることから，被タグ付け回数よりも適用範囲が広いため，有用であると言える．

4.3 希少度の評価実験

希少度については希少度の降順でランキングした場合に，被験者がカテゴリ c 内でより典型的でないと感じる Web ページが上位となるかを評価した．比較対象としては，所属度，非典型度のそれぞれの降順でランキングした結果を用いる．評価指標は，DCG[14] を用いる．DCG はランキング結果について，正解の重みを考慮した指標である．よりスコアの大きい正解がランキング上位に多く，それが上位であるほど良質なランキングであるとし，DCG の値は大きくなる．ランキング結果の上位 m 件についての DCG の値である DCG_m は式 (12) で表される．

$$DCG_m = rel_1 + \sum_{j=2}^{m} \frac{rel_j}{\log_2 j} \tag{12}$$

rel_j はランキング中の j 位のアイテムのスコアである．rel_j のスコアは表 5 のような基準で被験者 5 名に決めてもらい，その平均値を用いた．なお，スコア付けの際には，カテゴリごとに希少度，非典型度，所属度のそれぞれ上位 10 件ずつを取得してこれらの重複を除いて混ぜた状態でユーザに提示した．

表 5 DCG 算出時のスコア

評価基準	スコア
典型的である．	1
どちらかといえば典型的でない．	2
明らかに典型的でない．	3
指定されたカテゴリとは関係がない．	0

このとき，被験者ごとに知識量に大きな差がある可能性がある．そのため，カテゴリ c について詳しくないユーザが，Web ページの内容が典型的かを判断するための予備知識を提供する．Web ページの典型性を評価する前に学習データ内からタグ c が付けられた Web ページを無作為に 10 件用意し，被験者に閲覧してもらい，これを予備知識としてページごとにスコアをつけてもらった．

カテゴリごとにそれぞれの指標でランキングした場合の DCG_{10} の値を表 6 に示す．太字はカテゴリごとの DCG_{10} の最大値である．

表 6 から，“ゲーム”，“テレビ”，“映画” を除く 7 件のカテゴリで希少度の DCG_{10} が最も大きくなった．過半数のカテゴリにおいて希少度を用いたランキングは，非典型度のみ，所属度のみを用いた場合よりも DCG_{10} が高いため，希少度は有効であるといえる．実際に “google” カテゴリでの希少度が上位 10 件の結果からは「Google が自動車の運転情報を収集･分析し，運転時のエネルギー

表 6 ランキング結果の DCG_{10} の比較

カテゴリ	非典型度	所属度	希少度
google	0.54	7.71	**9.41**
twitter	1.86	7.22	**8.32**
iphone	0.00	7.79	**10.4**
アニメ	0.13	7.42	**8.51**
ゲーム	1.00	**8.74**	5.98
テレビ	1.48	**10.5**	8.88
ビジネス	2.16	9.57	**9.92**
映画	0.00	**8.44**	8.17
音楽	9.19	8.58	**9.38**
政治	1.24	8.36	**8.99**

表 7 “ゲーム”カテゴリでの評価結果

順位	希少度	平均	A	B	C	D	E
1	0.98	1.4	2	2	0	1	2
2	**0.97**	**0**	**0**	**0**	**0**	**0**	**0**
3	0.93	1.8	1	2	1	2	3
4	**0.92**	**0**	**0**	**0**	**0**	**0**	**0**
5	0.90	1.2	1	2	1	1	1
6	0.89	0.8	0	0	1	1	2
7	0.84	2	2	2	1	2	3
8	0.84	2.2	2	3	2	1	3
9	0.83	2	2	2	2	1	3
10	0.82	1.8	0	2	3	1	3

効率を最適化する API を発表した」という内容の記事や,「Google が公開した無料パズルゲームが面白い」という内容の記事なども得られた．このことから，“Google”カテゴリの Web ページにおいて一般的に記載されない“自動車”や “パズル”などの名詞が有効に機能したと言える．

一方，非典型度だけを用いてランキングした場合は，指定したカテゴリと関係のない Web ページが上位に多く存在した．そのため，表 6 の “音楽”を除くカテゴリで DCG_{10} は 0 から 2 の範囲となり，他の手法に比べて非常に小さくなった．このことから，希少度を用いてランキングした場合では，所属度が有効に機能したことで，希少な Web ページが得られたことがわかる．

所属度だけを用いてランキングした場合は，ランキング精度が “ゲーム”，“テレビ”，“映画”カテゴリでは希少度よりも高くなった．これは，これらのカテゴリにおいて c に所属しない Web ページの希少度が大きくなったことが原因だと考えられる．例として “ゲーム”カテゴリにおける，希少度が上位 10 件の Web ページに対する各被験者の評価結果を表 7 に示す．表 7 には希少度が上位から 10 件の Web ページに対して，希少度の順位と値，各被験者 (A〜E) の評価とその平均値を示している．この結果から，希少度が上位 2，4 番目の Web ページは “ゲーム”と関係がないと評価されていることがわかる．このような場合は，所属度が低くなり，その結果，希少度も低くなるべきであるが，所属度が高くなったことが問題である．

また，所属度を用いたランキングの DCG_{10} は，その他のカテゴリでも希少度を用いた場合に近い値をとっている．この原因を考察するため，所属度と非典型度の関係について調査した．横軸に所属度，縦軸に非典型度として “google”カテゴリの Web ページをプロットしたグラフを図 1 に示す．所属度は “google”内でよく見かけるような名詞が多ければ大きくなるため，所属度が大きい場合は非典型度が小さくなると考えられる．また，非典型度が大きい，“google”カテゴリ内で見かけない名詞が多い Web ページは所属度が小さくなるとも考えられる．そのため，グラフは左上から右下への対角線上に多くのページが分布すると予想していた．しかし，図 1 から，所属度が大きいグラフの右側の範囲でも非典型度の値が大きい Web ページが存在することがわかる．このようなページの存在によって，所属度でランキングした場合の DCG_{10} は大きくなっていると考えられる．これも所属度が低くなるべきページの一部の所属度が高くなっていることが予想される．

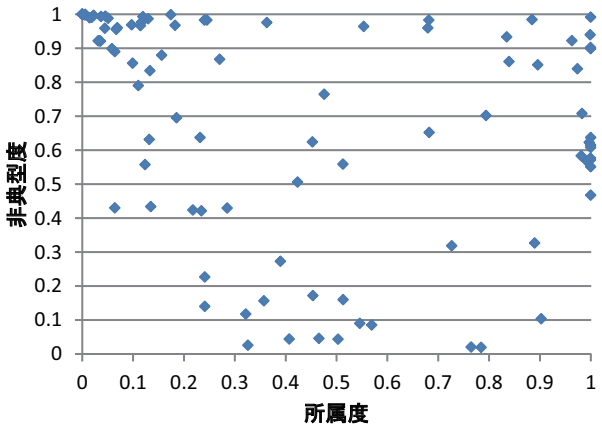

図 1 “google”に関する所属度，非典型度の分布

このように所属度が低くなるべきページで所属度が高く算出されてしまう理由であるが，DF 値が小さい語の影響であると考えている．まず，DF 値が小さい語は TFIDF 値が大きくなるため，主要な語になりやすい．このような語は $|BM_w|$ も小さくなると予想されるため，式 (10) および式 (11) より非典型度は高くなる．一方，式 (4) および式 (6) より所属度も大きくなる．しかし，

他のページにあまり含まれない語をカテゴリへの所属の根拠とした場合，一部のページでの語の使い方の影響を受けやすいため，主要な語の抽出の基準についての再検討が必要である．

5 おわりに

本研究では，認知度が低いが有益なページの推薦のため，Web ページの希少度を提案した．希少な Web ページを，ユーザが指定したカテゴリに所属し，かつそのカテゴリ内で典型的でないとみなし，その確率を希少度と定義した．ユーザが指定したカテゴリに所属する確率を所属度，ユーザが指定したカテゴリに所属している場合に典型でない確率を非典型度とすると，希少度はこれらの積である．各確率は，Web ページに付与されている SBM のタグと Web ページに出現する語の関係から算出する．評価としては，所属度による Web ページのカテゴリの判定精度と被タグ付け回数によるものを比較し，被タグ付け回数による場合と同等であることを示した．所属度の算出には対象のページがブックマークされている必要がないため，被タグ付け回数による手法よりも適用範囲が広いという利点がある．また，希少度によるランキング結果を所属度と非典型度のそれぞれのみを用いた場合と DCG の値で比較を行い，有用性を確認した．今後は，さらなる性能の向上を図るため，Web ページから主要な語の抽出する基準について検討を行う．

謝辞

本研究の一部は，平成 24, 25 年度科研費若手研究 (B)「情報の詳細関係に基づく Web ページの組織化」(課題番号 : 24700097) によるものである．ここに記して謝意を表す．

参考文献

[1] 情報検索に対する信憑性に関する調査，http://www.dl.kuis.kyoto-u.ac.jp/i-explosion/report/index.html

[2] 小川 祐樹，諏訪 博彦，山本 仁志，岡田 勇，太田 敏澄 : 動的なトピック分類に基づく Novelty を考慮した推薦アルゴリズムの提案，情報処理学会論文誌，Vol.50, No.6, pp.1636–1648, 2009.

[3] Ziegler, C., McNee, S. M., Konstan, J. A. and Lausen, G.: Improving recommendation lists through topic diversification, Proc. of the 14th international conference on World Wide Web (WWW'05), pp.22–32, 2005.

[4] 清水 拓也，土方 嘉徳， 西田 正吾: 発見性を考慮した協調フィルタリングアルゴリズム, 電子情報通信学会論文誌, Vol.J91-D, No.3, pp.538–550, 2008.

[5] 山家 雄介, 中村 聡史, Adam Jatowt, 田中 克己: ソーシャルブックマークの特性分析とそれに基づく Web 検索の再ランキング手法, 情報処理学会論文誌 データベース (TOD38), pp.88–110, 2008.

[6] Golder, S. and Huberman, B. A.: The Structure of Collaborative Tagging Systems, Journal of Information Science, 32(2), pp.198-208, 2006.

[7] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J., GroupLens: an open architecture for collaborative filtering of netnews, Proc. of Conference on Computer Supported Cooperative Work, pp.175–186, 1994.

[8] Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based collaborative filtering recommendation algorithms, Proc. of 10th International Conference on the World Wide Web (WWW'01), pp.285–295, 2001.

[9] Herlocker, J. L., Konstan, J. A., Terveen, L. G. and Riedl, J. T.: "Evaluating collaborative filtering recommender systems", ACM Transactions on Information Systems, Vol.22, Issue 1, pp.5–53, 2004.

[10] 加藤 由花，川口 賢二，箱崎 勝也 : オンラインショッピングを対象とした正確性と意外性のバランスを考慮したリコメンダシステム，情報処理学会論文誌 データベース，Vol.45, SIG 13(TOD 27), pp.53–64, 2005.

[11] 奥 健太，服部 文夫 : セレンディピティ指向情報推薦のためのフュージョンベース推薦システム，知能と情報，Vol.25, No.1, pp.524–539, 2013.

[12] 佃 洸摂，中村 聡史，山本 岳洋，田中 克己 : オブジェクトの典型度分析とその検索への応用，WebDB Forum 2011, 2G-1-1, 2011.

[13] Bradley, A. P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognition, vol.30, no.7, pp.1145–1159, 1997.

[14] Järvelin, K. and Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques, ACM Transactions on Information Systems (TOIS), Vol.20, Issue 4, pp.422–446, 2002.