

商品に対するユーザの Community 内嗜好度を考慮した Serendipity 指向情報推薦の実験的検証

山崎 隼也[†] 中島 伸介[†]

[†] 京都産業大学大学院先端情報学研究科先端情報学専攻 〒603-8555 京都市北区上賀茂本山

E-mail: †{i1558183,nakajima}@cse.kyoto-su.ac.jp

あらまし 近年, 単に精度の高い推薦を行うだけでなく, Serendipity の高い情報推薦技術の開発が注目されている. そこで我々は, 対象ユーザの興味領域における認知度は高くないもの, かつ, この対象ユーザよりも興味領域に対して詳しいユーザグループ内では認知度が高いような商品を検出し, これを対象ユーザに推薦することを特徴とする, Serendipity 指向情報推薦方式を提案してきた. 本稿では提案手法によって有効性の検証実験を行いその有効性を確認する.

キーワード 情報推薦, Serendipity, 有用性

1. はじめに

近年, インターネットの発達よりネットを介してユーザが取得可能な情報が膨大になっている. したがって, 膨大な情報からユーザが好みそうな情報を判別して, ユーザに提示するような情報推薦技術の重要性が高まっている. 情報推薦において, 単純に精度のみを追求すると, 毎回同じような商品が推薦されることとなり, 推薦商品の目新しさ, Novelty や推薦商品の偶発力, Serendipity が著しく低下することになる. つまり, 興味はあるが既に知っている, または, 既に持っている, という状況に陥りやすくなる. すなわち, 膨大な情報からユーザが好みそうな情報を判別して, ユーザに提示するような, 真に役に立つ情報推薦を実現するためにも, Novelty や Serendipity の高い推薦方式の実現に向けた研究開発が注目されている. なお, このセレンディピティ型推薦の有効性については Sawaizumi [1] らによって既に検証されている. また, 未知の商品推定には Hijikata ら [2] のようにユーザの類似度など使用する手法がある. そこで本論文では未知の商品推定を行うために商品に対するユーザの Community 内嗜好度を考慮した Serendipity 指向情報推薦方式の提案および提案手法の実験的検証を行う.

具体的には, ある推薦対象ユーザ A の閲覧及び購入ログから興味のあるトピックを推定し, そのトピックの関係語に対してユーザ A より詳しいユーザと詳しくないユーザの閲覧及び購入ログを用いて「詳しい」ユーザに利用されているが, 「詳しくない」ユーザに利用されていない商品を推薦することで, 商品の意外性と有用性を共に確保しようとする手法である. 手法の詳細な説明については第三章で順次行っていく. また, 本手法の実験的検証については第四章で言及する. 以降の構成は, 第二章で関連研究について, 第三章では提案手法について, 第四章で検証実験, 第五章で今後の課題と展開, 第六章でまとめを述べる.

2. 関連研究

情報推薦において意外性の高い推薦手法に関する研究として, 土方ら [3] は情報推薦の研究の歴史的な発展の経緯から評価方

法, 代表的な課題について述べている. また, Serendipity の重要性, 今後の方向性などの研究が報告されている.

村上ら [4] は利用者が好むコンテンツの中で, 習慣的にアクセスしないコンテンツに意外性を感じ, 利用者の満足度向上に結びつくという仮定のもと, 推薦システムを提案している.

システム実現のために, 利用者が定期的にアクセスしているコンテンツの習慣モデルと, 利用者の好むコンテンツかそうでないかを判断する予測モデルを嗜好モデルの二つのモデルの予測結果の差異により利用者の好む商品の中で Serendipity のあるコンテンツの推薦を提案している. 推薦商品の有効性は確認できたが, アクセスの習慣がないコンテンツに意外性の低い商品も存在する事もあると報告されている.

秋山ら [5] はユーザにとって有用な情報を, ユーザが意識している情報, ユーザが意識していない情報の二つに分けユーザが意識していない有用な情報を Serendipity と感じるコンテンツとし Serendipity のある推薦システムの提案を提案している. ユーザの好みであるコンテンツのテキスト情報を形態素解析し, 生成された単語を利用している. 形態素解析されたユーザの好みのコンテンツに出現する単語と, 他コンテンツに出てくる単語との違いを距離として表現している. ユーザの認識しているコンテンツと興味のないコンテンツの間を Serendipity のあるコンテンツとしていた.

そして, ユーザの認識しているコンテンツと Serendipity のあるコンテンツの間に興味のないコンテンツがあることを明らかにした. その上で各コンテンツに出てくる単語の組み合わせから, 組み合わせられにくい組み合わせのコンテンツに絞る事によって Serendipity 指向の推薦を提案している.

他にも, 奥ら [6] は任意の 2 つの商品の特徴を混ぜることで, 偶然を人工的に生み出すシステムを提案している. ユーザが 2 つの商品を納得いくまで混ぜる事ができ, 検索する手間を手間と思わせないシステムの提案している.

3. 商品に対するユーザの Community 内嗜好度を考慮した Serendipity 指向情報推薦

本節では、商品に対するユーザの Community 内嗜好度を考慮した Serendipity 指向情報推薦について説明する。

3.1 提案手法の概要

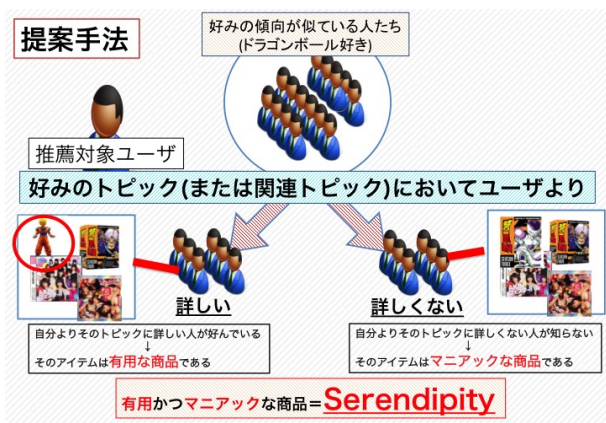


図 1 商品に対するユーザの Community 内嗜好度を考慮した Serendipity 指向情報推薦概念図

図 1 に、商品に対する Community 内認知度とユーザ嗜好度を考慮した Serendipity 指向情報推薦方式に関する概念図を示す。まず、この研究において、推薦されるべき商品とは、ユーザにとって意外かつ有用である商品であり、さらに自己発見可能性が高くない商品である。図 1 は推薦対象ユーザがドラゴンボールに興味があるとシステムに判定された際の例である。本研究において推薦されるべき商品をどのように選定していくかを説明している図となっている。推薦対象ユーザ A (以後ユーザ A) と同じ好みのトピックをもつユーザ群をその好みのトピック (または関連トピック) についてユーザ A より詳しいユーザ群が閲覧及び購入している商品はそのトピックに詳しいユーザ群が閲覧及び購入しているのでユーザ A にとって有用な商品である可能性が高いといえる。また、ユーザ A より詳しくないユーザ群が閲覧及び購入している商品はそのトピックに対してユーザ A より詳しくないユーザ群が閲覧及び購入しているのでユーザ A にとってありきたりな商品である可能性が高いといえる。つまりこのユーザ群が閲覧及び購入していない商品はユーザ A にとってありきたりでない、意外性があり、自己発見可能性が高くない商品である可能性が高いといえる。この 2つのユーザ群の閲覧及び購入ログを用いて、ユーザ A の好みのトピックにユーザ A より詳しいユーザ群が閲覧及び購入している商品かつユーザ A の好みのトピックにユーザ A より詳しくないユーザ群が閲覧及び購入していない商品を推薦することでユーザにとって意外かつ有用である商品であり、さらに自己発見可能性が高くない商品を推薦できる。関連トピックとあるが、これは好みのトピックの関係語のことであり、主にこのトピックの共起語である。関連トピックを用いて推薦を行うことにより、より意外性が高く、自己発見性が低い商品を推薦できる。なお、ユーザ A より詳しいユーザ、詳しくないユーザを探すときに、そのトピッ

クに興味のあるユーザグループ内で探す理由は、例えばマラソンというトピックの関係語として出てきたランニングと、ダイエットというトピックの関係語として出てきたランニングは単語としては同じだが、意味が違う。このような問題を避けるためユーザグループ内で探している。

この手法では、興味のあるトピックに絞ることでコンテンツベース型推薦の側面も出てくるが、コンテンツベースとは単に似ているユーザのログから推薦するものではないため、区別することができる。なお、Serendipity を考慮したコンテンツベース型推薦に関しては Iaquinta ら [7] のようなものが挙げられる。

3.2 提案手法の処理手順

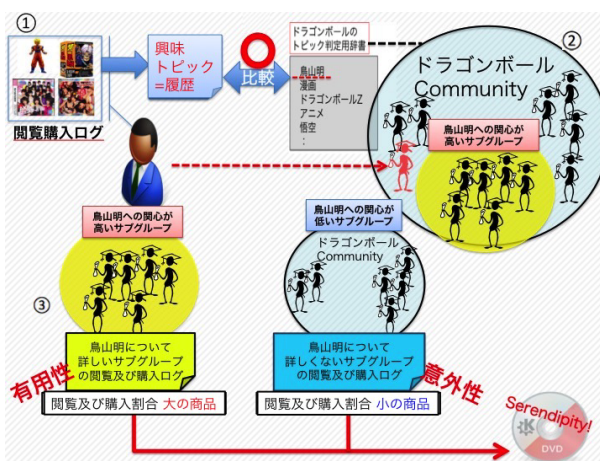


図 2 商品に対するユーザの Community 内嗜好度を考慮した Serendipity 指向情報推薦処理手順図

図 2 は処理手順であり、先ほどの概要と同じく、ユーザ A が閲覧及び購入ログからドラゴンボールというトピックに興味があると推定された場合の例を示して。この本節では提案手法の流れを順序立てて説明していく。この推薦手法では大きく分けて 3つの処理手順があり

- (1) 対象ユーザの Community 分類
 - (2) ユーザの嗜好度判定による Community 内のサブグループ判定
 - (3) 各 Community 関連商品の Serendipity Score の算出および推薦
- となる。それぞれの説明は以下で行う。

3.2.1 対象ユーザの Community 分類

図 3 では対象ユーザの対象ユーザの履歴の取得方法について説明する。まずこの研究で使われる商品にはすべて関連トピックというその商品の特徴を表すトピックが付いている。商品への関連トピックの付与過程に関しては図 4 の通り商品関連トピック自動付与システムにより自動で付与されている。ここで用いられているトピック判定用辞書という辞書は、同研究室の研究で作成された brogram 辞書というブログ分析によって得られたあるトピックに対する共起語の辞書があり、それを元に名詞以外の語句を除くなどの処理を行った辞書である。自動付与の方法としては商品の商品名と商品紹介文の txt ファイルとトピック判定用辞書との類似度を判定して、高い類似度の辞書



図 3 対象ユーザの履歴の取得方法

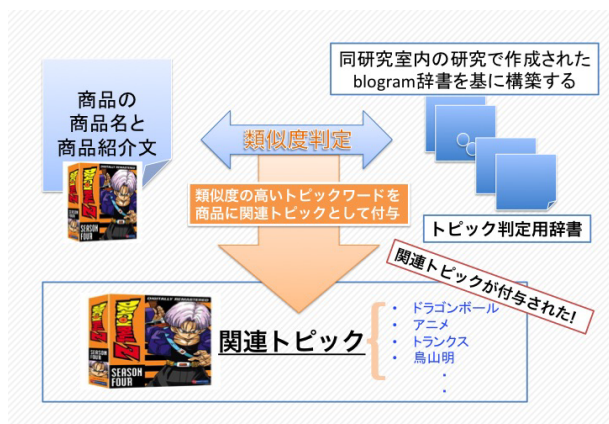


図 4 商品関連トピック自動付与方法

のトピックワードを判定されている商品に対して関連トピックとして付与している。ユーザはした商品に付与されている関連トピックの集合を履歴として持ち、その履歴を元に推薦を行う。

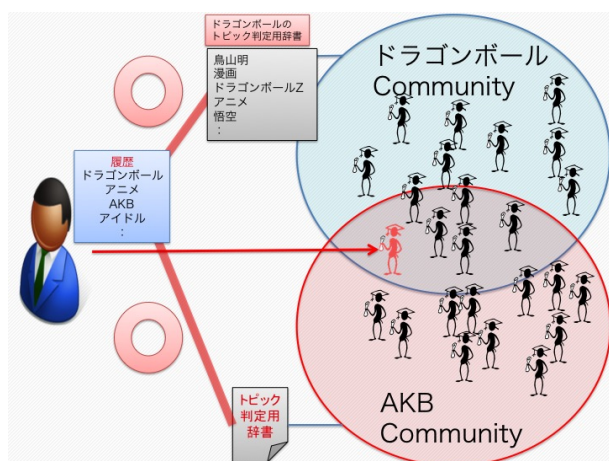


図 5 対象ユーザの Community 分類

図 5 では図 3 で作成した履歴を用いて、対象ユーザの Community 分類を行う。推薦対象ユーザは Community という、あるトピックに対して興味があるユーザグループに履歴を元に推薦対象ユーザの嗜好を推定して分類される。この Community は Community のトピックワードに対してのトピック判定用辞書

をもつ。その Community に興味があるかどうかは、Community のトピック判定用辞書と図 3 で作成したユーザの履歴との類似度をみて判定する。このとき、推薦対象ユーザは履歴から、興味があると判断されることで複数の Community に分類される。

3.2.2 ユーザの嗜好度判定による Community 内のサブグループ判定

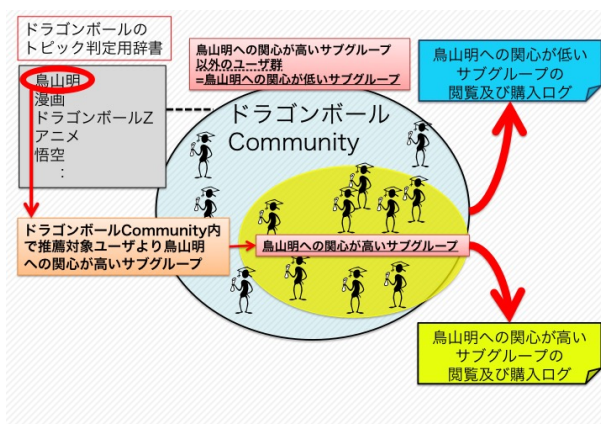


図 6 ユーザの嗜好度判定による Community 内のサブグループ判定

図 6 にて、ユーザの嗜好度判定による Community 内のサブグループ判定について説明する。先ほど説明したトピック判定用辞書のワードを使って、嗜好度により、サブグループ判定を行っていく。まず取り出したワードについて、推薦対象ユーザより、関係語に対する関心の高いユーザを Community の中からそのワードに対して詳しいグループとしてくくりだす。この時の関心の高さは取り出した関係語が関連トピックとして付与されている商品の閲覧及び購入回数で測る。そしてそのグループをそのワードに詳しいサブグループ、Community 内のそれ以外のユーザグループをそのワードに詳しくないサブグループとしてそれぞれのサブグループ内での閲覧及び購入ログを作成する。

3.2.3 各 Community 関連商品の SerendipityScore の算出および推薦



図 7 各 Community 関連商品の SerendipityScore の算出および推薦

図 7 にて、各 Community 関連商品の Serendipity Score の算出および推薦について説明する。図 6 で作成したある関連語ワー

ドに詳しいサブグループと詳しくないサブグループ、この2つのサブグループの閲覧及び購入ログから Serendipity Score を算出し、スコアの高いものから推薦していく。Serendipity Score は以下の式で算出される。

$$\text{SerendipityScore} = \frac{\text{詳しいサブグループの閲覧購入頻度}}{\text{詳しくないサブグループの閲覧購入頻度}}$$

この Serendipity Score の算出処理は詳しいサブグループの閲覧及び購入ログにある商品で、かつ、詳しくないサブグループの閲覧及び購入ログにない商品を推薦商品として決定するという処理である。この処理をトピック判定用辞書の関係語ごとに行い、推薦商品のリストを作成する。

この操作によって推薦対象ユーザの好むトピックに詳しいユーザにのみ知られているマニアックな商品を集めることができる。さらにこの商品は推薦対象ユーザの興味のあるトピックに関係のある商品の可能性が高いので有用性も高いと考えられる。この提案手法では商品を商品に関するトピックを見て取得しているのでカテゴリを問わず商品を推薦する事ができる。

4. 検証実験

4.1 検証実験の概要

ここでは本提案手法の検証実験について述べる。本実験では、ドラゴンボールに関する商品にレビューをつけている Amazon のレビュアー 15 名を被験者と同じ好みを持つユーザとして、そのレビュアーたちがレビューをつけた商品 145 件を擬似的な購入履歴として実験を行った。本実験の目的は、提案手法の有効性の確認を実験において行うことである。使用したデータは

- (1) ドラゴンボール関係の商品に頻繁にレビューを行っている Amazon レビュアー 15 名のレビュー履歴
- (2) レビューされた商品 145 件 (この商品はレビュー履歴の全商品ではなく、適切な商品を抜き出している)
- (3) トピック判定用辞書 1162 件 (blogram 辞書を元に作成したトピック判定用辞書から今回の実験用に抜き出した辞書)

であり、20 代男性 1 名に対して実験を行った。実験の手順としては

- (1) 事前に集めた 145 件の商品の中から被験者に購入したと思う商品を複数選択してもらい、擬似的な利用履歴を作成する
 - (2) 作成した利用履歴と 15 名の Amazon レビュアーのレビュー履歴をドラゴンボールの関連トピックごとに比較し、被験者より関心が低いレビュアー集団と高いレビュアー集団に分類する
 - (3) 最後に被験者に関心が高いレビュアー集団でレビューされていてかつ関心が低いレビュアー集団にレビューされていない商品を推薦し、比較手法と比較し評価してもらう
- の3つの手順で実験を行った。比較手法については協調フィルタリングを用いた。評価については提案手法と従来手法から5商品ずつ推薦し、各商品について5段階の評定尺度法を用いた3つの問いに答えてもらった。問いはそれぞれ

- (1) 商品を知っていたかどうか

- (2) 商品が欲しいと思ったかどうか

- (3) 商品を自力で見つけられたと思ったかどうか

である。(1)と(3)は値が低いほど良く、(2)は値が高いほど良い。

4.2 結果とその考察

図8の4つのグラフのうち、上部は提案手法、下部は比較手法であり、左の図は商品それぞれに対しての評価、右の図は商品の評価を正規化して15点満点で表したものである。結果として、提案手法では右図の平均が10.4点、最大が13点だったのに対し、比較手法では平均が8.6点、最大が10点と平均、最大ともに比較手法を上回ったので、提案手法の有効性を示せたと言える。三問それぞれの評価を比べてみると(2)の評価でもっとも差がでていた。比較手法の商品は認知度が均等だったのに対し、提案手法は認知度にばらつきが出たことから、知られていなかった商品に(2)の評価を高く付け、差がでたものと思われる。意外な結果として、(3)の問いではあまり差がつかなかった。なぜこうなったかは被験者数と試行回数が現状では少ないので、共に増やしていくことにより違いが顕著に出て、見えてくるだろうと考える。さらに、推薦する過程において、比較手法と同じ履歴を与えた際に提案手法の方が推薦の候補となる商品が多かったことから、提案手法はコールドスタートに強いのではないかと推測される。

5. 今後の展望及び課題

本節では、提案手法の展望と課題について説明する。

我々の提案手法は、コールドスタート問題に対して有効である可能性があるがさらに推薦されるユーザの検索サイト上での検索ワードやwebサイトの閲覧ログから好みのカテゴリを推測することや、Community 毎の流行語分析手法[8]などによって、ブログやSNSの分析により新規ユーザ用の推薦リストを作っておくなどの処理で改善するのではないかと考える。しかし、この改善方法の有効性については検証が必要であると考えられる。

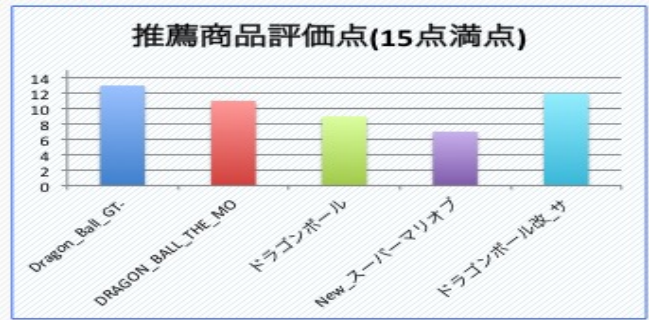
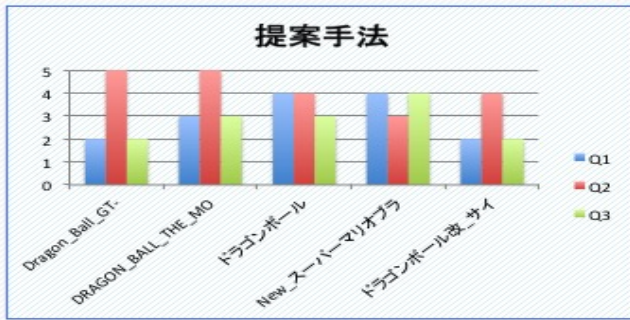
また、Community のトピック取得に使用するデータや取得方法についても考える必要があり、同時に関係語の取得の難しさも考慮する必要がある。関係語に関しては Community のトピックの共起語から商品と関連しそうなワードを関係語とする予定であるが単に共起語以外に、例えば共起は低い流行している商品が多いトピックのワードなどを考慮することで商品の範囲が広がり、より Serendipity な商品を推薦することができる。また、今後の方向性としては、より実装に近い環境での検証実験を行い、セレンディピティの評価に関しては Murakami ら [9] の評価方法を、システムに関しては Herlocker ら [10] の評価方法を参考に評価を行う予定である。

楽天データ公開 [11] におけるデータを含めて、検証実験に利用可能な公開データには利用上の制限も大きいといえる。したがってその他公開されているデータも含めて、検討を行い、適切なデータの選定と利用方法について検討するつもりである。

6. まとめ

本稿において、商品に対する Community 内認知度とユーザ嗜好度を考慮した Serendipity 指向情報推薦方式についての提

提案手法



比較手法

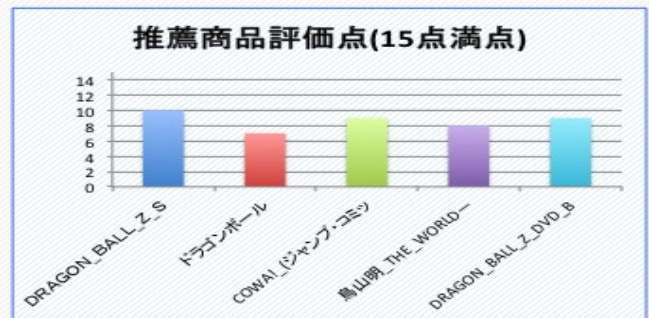
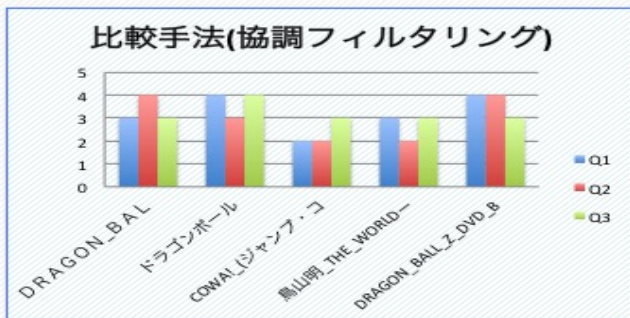


図 8 検証実験 結果

案と提案手法の検証実験を行い、有効性を示した。今後は、実装に近い状態での検証実験を行い、今回浮かび上がった課題も含め、さらなる検証と有効性及び妥当性の確認を行う。

謝 辞

本研究を遂行するにあたり、楽天データ公開において提供された楽天市場の商品データおよびレビューデータを分析対象データとして使用した。ここに記して謝意を表します。

文 献

- [1] Shigekazu Sawaizumi, Osamu Katai, Hiroshi Kawakami, and Takayuki Shiose. Use of Serendipity Power for Discoveries and Inventions. Intelligent and Evolutionary Systems, Studies in Computational Intelligence, Vol. 187/2009, pp. 163?169, 2009.
- [2] Yoshinori Hijikata, Takuya Shimizu, and Shogo Nishida. Discovery-oriented collaborative filtering for improving user satisfaction. In IUI '09: Proceedings of the 14th international conference on Intelligent user interfaces, p. 67, New York, New York, USA, 2009. ACM Press.
- [3] 土方嘉徳,『嗜好抽出と情報推薦技術』, 情報処理, Vol.48, No.9, pp963-964, 2007.
- [4] 村上知子, 森紘一郎, 折原良平, 推薦の意外性向上のための手法とその評価, 人工知能学会論文誌 24 巻 5 号 G, pp428-436, 2009.
- [5] 秋山高行, 小原清弘, 谷崎正明『ユーザの選択履歴に依存しない指標を利用した Serendipity のある推薦方式の提案と評価』, FIT2010 (第 9 回情報科学技術フォーラム), RO-007, 第 4 分冊 p157-164, 2010.
- [6] 奥健太, 服部文夫, セレンディビティ指向情報推薦のためのフュージョンベースアプローチのユーザ評価, DEIM Forum, 2012, A1-3, 2012.
- [7] Leo Iaquinta, Marco De Gemmis, Pasquale Lops, Giovanni Semeraro, Michele Filannino, and Piero Molino. Introduc-

ing Serendipity in a Content-Based Recommender System. In 2008 Eighth International Conference on Hybrid Intelligent Systems, pp. 168?173. Ieee, September 2008.

- [8] 中島伸介, 張 建偉, 稲垣陽一, 中本レン, (WebDB11 推薦論文) 大規模なブログ記事時系列分析に基づく流行語候補の早期発見手法, 情報処理学会論文誌データベース (TOD56), 6(1), 1-15 (2013-01-23), 1882-7799, 2013.
- [9] Tomoko Murakami, Koichiro Mori, and Ryohei Orihara. Metrics for evaluating the serendipity of recommendation lists. New Frontiers In Artificial Intelligence, Lecture Notes in Computer Science, Vol. 4914/2008, pp. 40?46, 2008.
- [10] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS), Vol. 22, No. 1, pp. 5?53, 2004.
- [11] 楽天データ公開, <http://rit.rakuten.co.jp/rdr/>