

アイテムに対する 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

山崎 隼也[†] 中島 伸介[†]

[†] 京都産業大学コンピュータ理工学部 〒603-8555 京都市北区上賀茂本山

E-mail: †{g1145406,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. 高い Serendipity を実現するためには, 意外性を高め, かつ, ユーザにとって有用性の高いアイテムを推薦する必要がある. 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アイテムを検出し, 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する.

1. はじめに

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, 毎回同じようなアイテムが推薦されることとなり, 推薦アイテムの目新しさ, Novelty や推薦アイテムの偶発力, 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] によって既に検証されている.

そこで我々は, Serendipity の高い推薦方式の実現を目指し, アイテムに対する認知度とユーザ評価を考慮した Serendipity 指向情報推薦方式の提案してきた. 具体的には, あるユーザ A がカテゴリ X のアイテムに興味がある場合, カテゴリ X に関してこのユーザ A が普段利用するアイテム群よりも認知度が低いアイテムを推薦することで, 推薦アイテムの意外性を確保しようとするものである.

この際に, 単純に認知度が低い (マイナーな) アイテムを推薦するだけでは, ユーザに対する満足度が低下することが考えられるため, 少なくともカテゴリ X に対して Familiarity が高いユーザに支持されているアイテムに限定することで, ユーザ満足度の向上を目指すものである. Familiarity とはカテゴリに対する興味の高さを示し, 閲覧購入ログから決定する. しかしこの手法ではアイテム推定時に興味の高いカテゴリからアイテムを推定するため, 他カテゴリのアイテムの推薦ができなかった. 未知のアイテム推定には他にも Hijikata ら [2] のようにユーザの類似度など使用する手法もある. そこで本論文では未知のアイテム推定を行うためにアイテムに対する Community 内認知

度とユーザ嗜好度を考慮した Serendipity 指向情報推薦方式の提案を行う.

具体的には, あるユーザ A の閲覧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アイテムを推薦することで, アイテムの意外性と有用性を共に確保しようとする手法である. この本論文の手法ではこれまで我々が提唱してきた手法とは違い, 他カテゴリのアイテムを推薦することが可能である. 手法の詳細な説明については第三章で順次行っていく.

以降の構成は, 第 2 章で関連研究について, 第 3 章では提案手法について, 第 4 章で問題点, 第 5 章で今後の展開, 第 6 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アイテムの中で Serendipity のあるコンテンツの推薦を提案している. 推薦アイテム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アイテム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つのアイテム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つのアイテム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. アイテムに対する 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

本節では、アイテムに対する 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説明する。

3.1 提案手法の概要

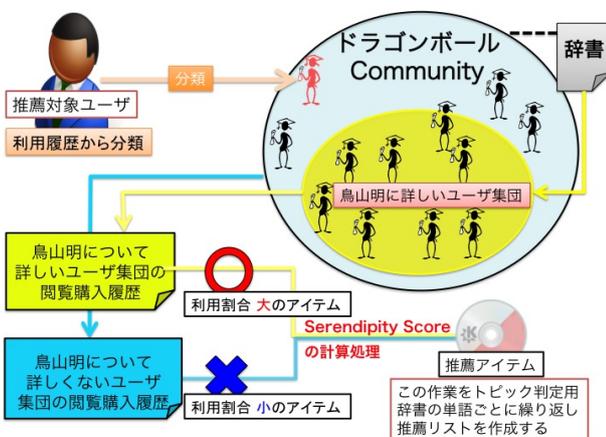


図1 アイテムに対する Community 内認知度とユーザ嗜好度を考慮した Serendipity 指向情報推薦方式概念図

図1に、アイテムに対する 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。図1について、この例では、あるユーザAが閲覧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いる。このユーザAに対してドラゴンボールに関するアイテムは有用性がある可能性は高いが、既知である可能性も高いため、意外性も低くなる。

そこで、我々の手法は、まず、ユーザAと同じく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ユーザグループを作り、このドラゴンボールという単語の関係語を集めたトピック判定用辞書を作成し、その辞書内の単語に対して、ユーザグ

ループ内でユーザAより詳しいユーザグループの閲覧購入ログと詳しくないユーザグループの閲覧購入ログを比較し、詳しいユーザグループが利用していて、詳しくないユーザグループが利用していないアイテムを推薦することで、Community 内認知度は高くないが有用である可能性の高いアイテムを推薦し、かつ、ありふれたアイテムが推薦されないようにすることを目指すものである。

つまり、ユーザAより詳しいユーザグループに利用されているということはユーザよりその関係語に関するアイテムに詳しいユーザ、つまり、よりマニアックなユーザに利用されているかつユーザAより詳しくないユーザグループに利用されていないということは、その関係語に関するアイテムに詳しくないユーザは知らないアイテムと分かる。

この推薦によって推薦されるものはユーザの興味のあるトピックの関係語のアイテムの中で自分より詳しい人は知っている、詳しくない人は知らないアイテムである。

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 節から詳細な提案手法の説明に入る。

3.2 提案手法の流れ

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4つの処理手順があり

- (1) 対象ユーザの Community 分類
- (2) Community 内のトピック判定用辞書の構築
- (3) ユーザの嗜好度判定による Community 内のサブグループ判定
- (4) 各 Community 関連アイテムの 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

図2では対象ユーザの Community 分類について説明する。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閲覧購入ログから推薦対象ユーザの嗜好を推定して分類される。このとき、推薦対象ユーザは閲覧購入ログから、興味があると判断されることで複数の Community に分類される。図3では、Community 内のトピック判定用辞書の構築について説明する。この Community はそれぞれその Community のトピックのワードに関する関係語のリストを持つことを示す。この関係語とは Community のトピックのワードの共起語かつアイテムに関するトピックになりうるワードと考えている。このリストとは共起度順にワードを並べたも

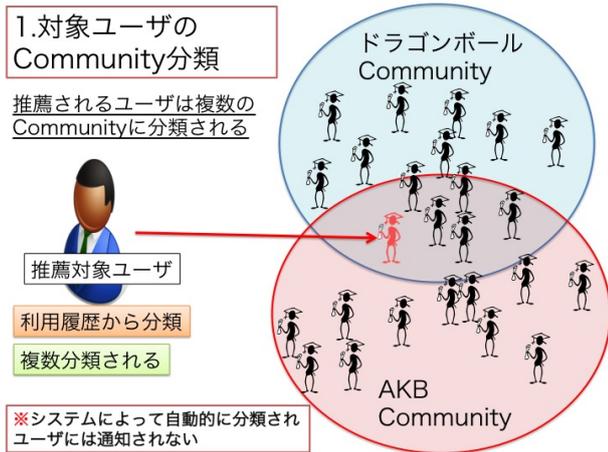


図2 対象ユーザの Community 分類

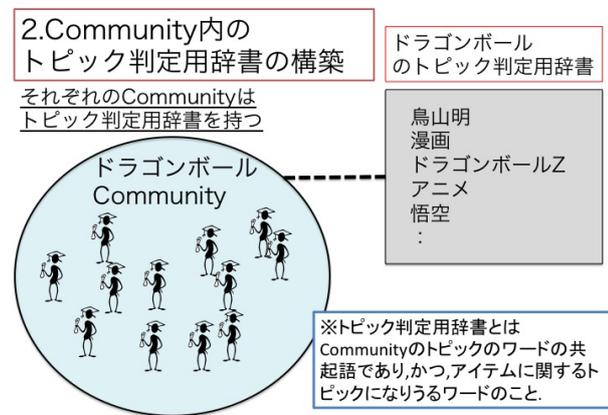


図3 Community 内のトピック判定用辞書の構築

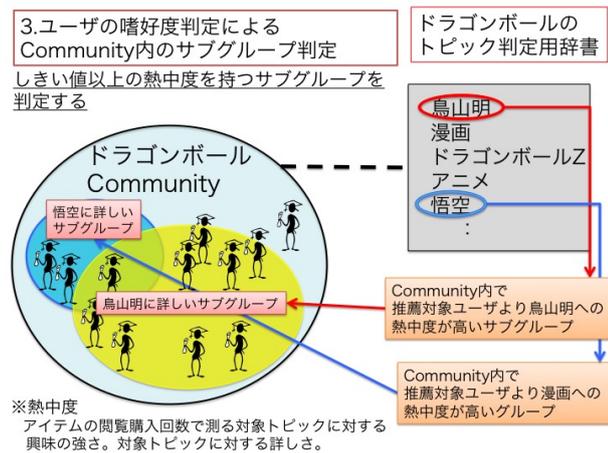


図4 ユーザの嗜好度判定による Community 内のサブグループ判定

のである。このリストをトピック判定用辞書と呼び、すべての Community で構築される。図4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、関係語に対する熱中度の高いユーザを Community の中からそのワードに対して詳しいグ

ループとしてくりだす。この時の熱中度とはアイテムの閲覧購入回数で測る対象トピックに対する興味の高さのことであり、対象トピックに対する詳しさといえる。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判定していく。図5にて、各 Community 関連アイテムの



図5 各 Community 関連アイテムの Serendipity Score の算出および推薦

Serendipity Score の算出および推薦について説明する。ユーザの嗜好度判定による Community 内のサブグループ判定によって判定され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。

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

この Serendipity Score の算出処理は熱中度の高いサブグループの閲覧購入ログにあるアイテムで、かつ、熱中度の低いサブグループの閲覧購入ログにないアイテムを推薦アイテム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、推薦アイテム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アイテムを集めることができる。さらにこのアイテムは推薦対象ユーザの興味のあるトピックに関係のあるアイテムの可能性が高いので有用性も高いと考えられる。この提案手法ではアイテムをアイテム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アイテムを推薦する事ができると考えている。

4. 提案手法の課題および方向性

本節では、提案手法の課題と今後の方向性について説明する。我々の提案手法は、ユーザの嗜好を推定して Community に所属できるまで、ある程度、購入閲覧ログが確保できていないと推薦できない、コールドスタート問題を持つ。解決方法としては、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が考えられる。しかし、この解決方法の有効性については検証が必要であると考える。

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アイテム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、例えば共起は低いが流行しているアイテムが多いトピックのワードなどを考慮することでアイテムの範囲が広がり、より Serendipity なアイテムを推薦することができると考えている。更に Community に所属する際の判定の基準も設定する必要があるが、これは閲覧購入ログとユーザの嗜好との関係を調べる実験を行い、決定する予定である。また、今後の方向性としては、我々の提案手法の有効性を確認するための予備実験を行い、推薦の有用性やセレンディピティ度の評価を行う。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。

楽天データ公開におけるデータを含めて、評価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。

5. ま と め

本稿において、アイテムに対する 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案を行った。今後は、提案手法を実現したプロトタイプシステムの構築と、これを用いたより評価実験を行い、提案手法と従来手法との比較実験により有効性および妥当性を評価する予定である。

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, 森紘一朗, 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, セレンディピ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. Introducing 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/>