

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

山崎 隼也[†] 井狩 憲幸[†] 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,g1044048,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, 意外性を高めつつ, 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, 世間の認知度は低いが Familiarity の高いユーザからの評価が高いアイテムを判別することで, Serendipity の高いアイテムを提示することが可能な情報推薦方式の提案を行う.

1. はじめに

インターネットの発達やネットワーク端末の普及により, 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, Novelty や Serendipity が著しく低下することになり, 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, 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.

そこで本研究では, 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を目的とする. 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. なお, アイテムの認知度を分析する際, 全ユーザ内での認知度だけでなく, あるコミュニティに限定した認知度を考慮することで, より細やかな情報推薦が可能になると考えている.

以降の構成は, 第2節で関連研究を紹介する. 第3節では, ユーザの利用履歴をもちいた推薦, 4章ではコミュニティをもちいた推薦, 5章で問題点, 6章で今後の展開, 7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 [1] 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. 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 [2] 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を習慣モデルとし, 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としている. 両モデルとも利用者の行動履歴に基づいて作成されている. 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 [3] 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, この二つに分けユーザが意識していない有用な情報を Serendipity を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. テレビ番組で実験しており, 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, 形態素解析されたユーザの好みのコンテンツにでき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注目している. 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. しかし, 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 [4] 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.

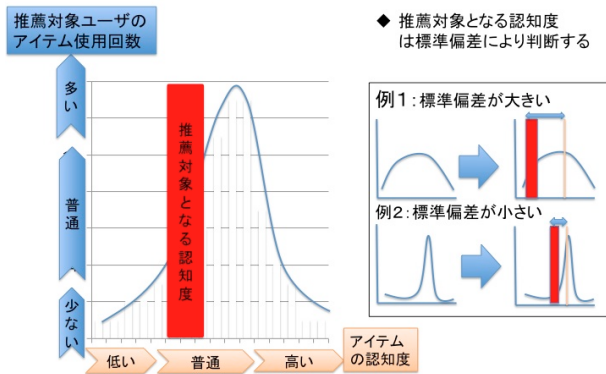


図1 アイテム認知度を考慮した Serendipity 指向情報推薦

3. 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

本節では、アイテムに対する認知度とユーザ評価を考慮した Serendipity 指向情報推薦方式について説明する。

3.1 提案手法の概要

図1に、アイテム認知度を考慮した Serendipity 指向情報推薦に関する概念図を示す。図1にて示すヒストグラムは、あるユーザが興味のあるカテゴリにおいて利用するアイテムの認知度と、その利用頻度を示している。すなわち、この例では、アイテム認知度が普通よりやや高めアイテムを利用することが多いことを示している。つまり、このユーザは普通よりやや高めの認知度のアイテムについては、良く知っているということであるため、同様の認知度のアイテムを推薦しても、既知である可能性が高い。そこで我々の手法は、ユーザの利用アイテム群の世間での認知度の平均値もしくはピークを検知し、この認知度帯よりも少し認知度が低いアイテム（恐らくこのユーザが見慣れていないアイテム）を推薦対象とすることで、このユーザにとって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。なお、図1中の例1、例2に示すように、認知度を考慮したアイテム利用頻度を分析する際には、その標準偏差（分散）についても検討する。標準偏差の大きさによって、どの程度認知度が低いアイテムを対象とするかを決定する。

ここで考えなければならないこととして、単純に認知度が低いアイテムをユーザに推薦するだけでは、ユーザにとって価値が低いアイテムが推薦されてしまう恐れがある。したがって、何らかの方法で推薦対象アイテムの有用性、つまりありふれていない、かつユーザが興味を持てるアイテムの推薦をしなければならない。そこで、“あるカテゴリに対して Familiarity が高い（詳しい）ユーザがしばしば利用するアイテムの有用性は高い”という仮説の下に、対象カテゴリに関して詳しい人が頻繁に利用するアイテムの推薦を目指す（図2参照）。

3.2 Community 粒度の考慮による Serendipity 指向情報推薦方式の拡張

前節にて説明した、Serendipity 指向情報推薦方式において、認知度を考慮する際のユーザグループの大きさを推薦対象ユーザが属する Community 粒度に基づいて変化させることにより、

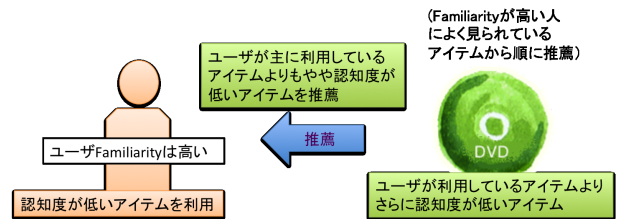


図2 ユーザ Familiarity を考慮した推薦



図3 Community 粒度の違い

いろいろな切り口による情報推薦の実現を目指している。

まずは、各ユーザが属する Community を判定する。手法としては、過去のアイテム利用履歴等进行分析することにより行う。この時、本研究で定義する Community は、ユーザが相互に友人もしくは知り合い関係である必要はなく、単に同じ興味分野を持つことを表している。

図3に、Community 粒度の違いを示した概念図を示す。このように、Community の粒度は均一なものではなく、小さな Community が合わさって、より大きな Community を形成することもある。3.1 節にて、アイテムの認知度を考慮した Serendipity 指向情報推薦方式について説明したが、考慮する Community 粒度を変化させることで、いろいろな観点からの情報推薦が可能になると考えている。図4に、Community 粒度の違いによる推薦候補アイテムの変化について示す。

ユーザ全体（社会全体）で見ると、認知度としては非常に低く、推薦対象にならない場合も、ある Community に限定してみれば、ある程度の認知度がある場合が考えられる。人間は、いろいろな興味を持ち、潜在的にいろいろな Community に属している。すなわち、このような Community 粒度の違いを考慮することにより、推薦対象アイテムに対するユーザの興味の多様性を考慮した情報推薦を実現することができると考えている。また、これにより専門的な物や知る人ぞ知るようなアイテムを推薦できるようになると考えられる。

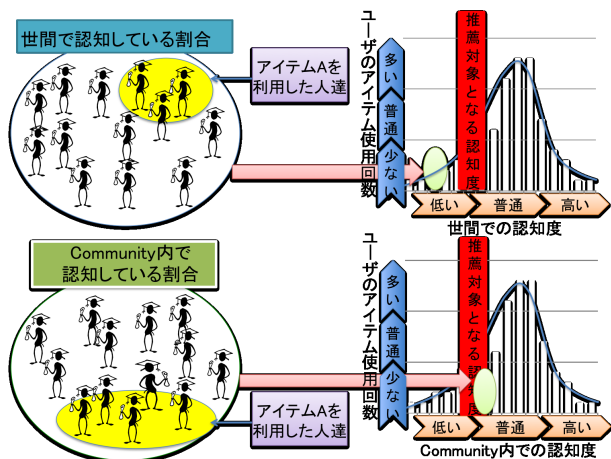


図 4 Community 粒度の違いによる推薦候補アイテムの変化

3.3 Community 毎の流行を考慮した Serendipity 指向情報推薦方式の拡張

3.2 節では、Community 粒度を考慮することによる Serendipity 指向情報推薦方式の拡張について述べた。社会全体としてはほとんど見えなかったが、Community の粒度の違いに着目することで、ある Community においては認知度が適度に高いアイテムを発見し、推薦できる可能性について述べた。本節では、Community 毎の流行を考慮した Serendipity 指向情報推薦方式の拡張について説明する。

中島らによる研究 [5] では、Community 間での流行語の広まりから、流行語を早期に発見する手法を提案している。そこで、この技術と提案手法を組みあわせることで、よりの確で Serendipity が高いアイテムを早期に推薦する手法の提案を試みる。具体的には、各 Community 内での認知度の時系列変化や、Community 間での認知度の広がり予測することで、Novelty および Serendipity が高い情報推薦を実現できる可能性があると考えている。

4. 実データに基づく評価実験方法

提案手法の有効性および妥当性を検証するための評価実験の方法について述べる。我々の提案手法では、粒度の異なる各種 Community を考慮する必要があるため、幅広いアイテムに関するデータが好ましい。したがって、本研究では楽天データ公開 [6] により公開されている商品データおよびレビューデータを利用する予定である。楽天市場の商品データには属性として、

- ・商品名
- ・商品コード
- ・商品価格
- ・商品説明
- ・販売方法別説明文
- ・商品 URL
- ・商品画像 URL
- ・レビュー件数
- ・レビュー平均
- ・店舗コード

- ・ジャンル ID
- ・登録年月日

が登録されている。商品データに登録されているアイテムの数は約 5,000 万件あり、レビュー回数は 1,600 万件登録されている。提案手法の評価実験には、商品名、商品コード、レビュー件数、レビュー平均、ジャンル ID を利用する予定である。ジャンル ID は階層構造になっており、計 32,697 ジャンルにアイテムを分類している。

提案手法では、アイテムに関する認知度を用いるが、評価実験ではレビュー件数を用いて、認知度の推定を行う。これは認知度が高いアイテムには多くのレビューが付くという仮説に基づいている。

具体的には、被験者に対して擬似的にアイテムの購入履歴を作成してもらい、被験者毎にあるカテゴリに対する利用履歴を作成する。同時に、該当するカテゴリにおいて、レビュー件数に基づいた認知度のヒストグラムを作成し、被験者が主に利用する認知度よりも低い認知度帯のアイテムのうち、レビュー評価が高いものを被験者へ推薦する。そこで、推薦されたアイテムに対する満足度を分析することで、提案手法の有効性について検証する予定である。なお、被験者の Community 分類については、アンケートにより嗜好度の高いカテゴリを抽出し、そのカテゴリを近似的にその被験者が属する Community とすることで、Community 粒度の考慮による Serendipity 指向情報推薦方式についても、その有効性の検証を試みる。

Community 毎の流行を考慮した Serendipity 指向情報推薦方式については、Community 毎の流行語分析手法 [5] との融合方法について検討を行うつもりである。

5. 提案手法の課題および方向性

本節では、提案手法がもつ課題と今後の方向性について説明する。

我々提案手法全般において、ユーザ Familiarity の値がどの分野においても小さいユーザに対しては、推薦が出来ないという問題がある。これは、普段から幅広くアイテムを利用する人に起こりうる現象であり、そもそもこの Serendipity 指向推薦ではユーザ Familiarity の値が小さいユーザは対象としていない事が原因である。

Community 粒度の考慮による Serendipity 指向情報推薦方式では、既存の Serendipity 指向情報推薦を実現するための手法と比べて、推薦アイテムがよりそのジャンルに精通した人向けになるので、そのジャンルに対しユーザ Familiarity がそこまで詳しくないユーザに対しては推薦アイテムが一般人の価値観とはかけ離れる可能性がある。これについては、実際に検証実験を通じて検証する必要があると考えている。

Community 毎の流行を考慮した Serendipity 指向情報推薦方式については、Community 毎の流行語分析手法 [5] がブログや SNS の分析による手法であるため、どのようにネットショップの利用データと融合させられるかという問題がある。ただし、ブログや SNS の利用ユーザに対する個別化されたネット広告に特化してしまうという考え方もあり、本手法の可能性は大き

いと考えている。

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にプロトタイプシステムの構築を行い、評価実験を行う。また同時に、Community 粒度の考慮による Serendipity 指向情報推薦方式や、Community 毎の流行を考慮した Serendipity 指向情報推薦方式の具体的な手法について、検討を進め、評価実験が可能なシステムの構築を目指す。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。

6. ま と め

本稿において、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、Community 粒度の考慮による Serendipity 指向情報推薦方式および、Community 毎の流行を考慮した Serendipity 指向情報推薦方式について提案した。また、楽天データ公開の商品データおよびレビューデータを用いた評価実験の方法について説明を行った。

今後は、提案手法を実現するプロトタイプシステムの構築と、これを用いた評価実験を行い、提案手法の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] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 楽天データ公開, <http://rit.rakuten.co.jp/rdr/>