

## 完全グラフによる異ジャンル間の嗜好傾向表現と 協調フィルタリングの発見性向上への応用

杉本 拓弥<sup>†</sup> 豊田 哲也<sup>††</sup> 延原 肇<sup>††</sup>

<sup>†</sup> 筑波大学 工学システム学類 <sup>††</sup> 筑波大学 大学院システム情報工学研究科

### 1 はじめに

近年、利便性や品揃えの豊富さなどを強みとする、インターネットショッピングが広く普及している。特に、世界6カ国でサイトを運営するAmazonの利用者数は、2008年から一年間で約25%増加している[1]。これらのサイトでは、ユーザが膨大な商品情報の中から目的の商品を発見しやすくするための推薦システムを実装している。推薦情報作成の代表的な手法として協調フィルタリングが用いられるが、推薦結果が同一ジャンルに偏ってしまうという問題点がある。

本研究では、従来の推薦情報に加えてそれらを補助するような形で、ユーザが新しい嗜好を発見できるような推薦情報を提示するシステムを提案する。提案システムは、ジャンルをノードとする完全グラフを用いた嗜好傾向表現によって、大局的な視点からの発見性を向上させた推薦情報を作成し、ユーザに提示する。

### 2 推薦の多様性に関する関連研究

小川らのAmazonの共購買情報を用いた研究[2]では、推薦の多様性を向上させることを目的として、正規化を伴わない予測評価値計算とアイテムの共購買情報を用いた推薦の多様化度評価手法を提案している。また、清水らの研究[3]では、ユーザの嗜好に関するプロフィールに加えて、アイテムに対する既知/未知のプロフィールを用いることで、ユーザの知らない商品を推薦するアルゴリズムが複数提案されている。これらの研究ではユーザプロフィールを用いているため、ユーザに対する負担が大きくなってしまふ。

本研究ではAmazon.co.jpの商品情報をAmazonが提供するAPIを用いて取得し、それらをクラスタリングすることで得られるジャンルをノードとする完全グラフをあらかじめ作成しておくことで、ユーザの負担をできるだけ削減する。また、ジャンルという大きな括りでの推薦を考慮することで、ユーザは商品単位のレベルではなく、より大きな視点において新たな嗜

好を発見することができる。

### 3 ジャンルネットワークレコメンデーションシステム

本研究で用いるデータは、Amazon Web Servicesで取得したAmazon.co.jp内の約8000アイテムの商品情報とする。本研究では、メディアは異なるが同じ作品に関連するもの(映画作品のサウンドトラックCDなど)を同一のジャンルとして扱うため、Amazonが商品に付与するジャンル名を識別する個別の番号であるBrowse Node IDを用いて商品のジャンルを定義する。各商品のBrowse Node IDの保有情報に基づいて、 $k$ -means法の応用である $k$ -medoids法[4]を用いてクラスタリングを行う。 $k$ -medoids法では各データ点の座標情報ではなく、任意の2データ間の非類似度を使用する。商品 $i, j$ の非類似度 $dissim_{ij}$ は、 $I_x$ を商品 $x$ が保有するBrowse Node IDを要素とした集合とみなし(1)式で求められる。

$$dissim_{ij} = 1 - \frac{|I_i \cap I_j|}{|I_i \cup I_j|} (\in [0, 1]) \quad (1)$$

$k$ -medoids法で設定した $k$ 個のクラスタをそれぞれジャンルとして定義する。本研究では、これら $k$ 個のクラスタを重み付き完全グラフのノードに対応させ、Amazon.co.jpにおいて評価を行っているユーザーの嗜好傾向を表現する。グラフの作成手順を以下に示す。

1. 商品情報(ASIN<sup>†</sup>, 評価者のID, 評価値, Browse Node ID)を取得
2. 各商品のBrowse Node IDを参照し、 $k$ -medoids法によりクラスタリングを行い新ジャンルを作成
3. 各ジャンル中の高評価ユーザに基づき、重み付き完全グラフを作成
  - (a) ジャンルをノードとする。
  - (b) 任意の2つのジャンルにまたがっている高評価ユーザ数を算出
  - (c) 一致数を重みとして、該当ノードにエッジを引き、重み付き完全グラフを作成

ユーザが選択した商品を含むノードと、より関係性の強いノードの商品を完全グラフから求め、そのノードに含まれる商品を推薦候補とする。以下に提案システムによる推薦情報作成の概略を示す。

<sup>†</sup> Amazonが各商品に与えている個別の番号。

Diffelent Genres of Trend Indication Based on Complete Graph and Application for Advancement Discoverability of Collaborative Filtering

<sup>†</sup> Takuya SUGIMOTO(sugimoto@cmu.iit.tsukuba.ac.jp)

<sup>††</sup> Tetsuya TOYOTA(toyota@cmu.iit.tsukuba.ac.jp)

<sup>††</sup> Hajime NOBUHARA(nobuhara@iit.tsukuba.ac.jp)

University of Tsukuba, College of Engineering Systems (<sup>†</sup>)  
University of Tsukuba, Graduate School of Systems and Information Engineering (<sup>††</sup>)

1. ユーザによる商品検索
2. 商品名, ASIN 取得
3. Browse Node ID, 評価ユーザ ID, 評価値取得
4. 取得情報からジャンル(完全グラフの該当ノード)を決定
5. 該当ジャンルのランキング作成
6. ランキング上位のジャンルの商品を推薦

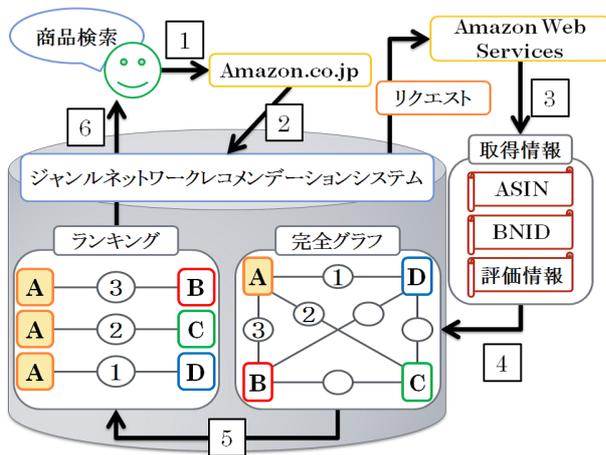


図 1: 提案システムによる推薦手順の概要

最後に先の完全グラフ中の当該ノードとそれ以外のノードとのエッジの重みを降順にソートし、リストアップする。このリストの順位に従って、ノードからいくつかの商品を選び出して推薦リストとし、ユーザに提示する。

#### 4 評価実験

本研究で作成する推薦情報はそれ単体でユーザの要求を全て満たす物ではなく、従来の推薦を補助し、ユーザに新たな嗜好を発見してもらうことを目的としている。そこで、より総合的な視点から評価するため、ユーザ満足度を指標として評価を行う。この場合のユーザ満足度は、未知の商品を発見できたかということと、推薦された商品にどの程度興味を持つことができたか、という点で判断する。被験者は大学生 23 名で、1 人当たり 5 つの自分の嗜好に合う商品を検索してもらい、Amazon と提案システムの推薦情報を提示する。同時に以下の質問項目に回答してもらう。

- Q.1 Amazon が推薦した商品の中に既知の商品がいくつ含まれているか (0~9 個)。
- Q.2 システムが推薦した商品を知っているか (0:知っている ~ 2:知らない)。
- Q.3 システムが推薦した商品に興味を持ったか (0:はい, 1:いいえ)。
- Q.4 システムが推薦した商品を欲しいと思ったか (0:欲しい ~ 4:いらぬ)。

表 1: 評価実験結果

質問		平均値	標準偏差
Amazon の推薦について	Q.1	0.554	0.185
	Q.2	0.884	0.095
システムの推薦について	Q.3	0.624	0.176
	Q.4	0.693	0.541

Q.1 と Q.2 を比較すると Amazon の推薦よりも提案システムのほうが未知の商品を多く推薦できていることが確認できた。一方、実際に欲しいという評価が得られた商品は少なく、今後改良すべき点として検討する必要がある。

#### 5 おわりに

ジャンルという大きな視点から発見性の高い推薦を作成するため、*k*-medoids 法および完全グラフを用いた異ジャンル間の嗜好傾向表現を利用したシステムを提案した。評価実験の結果から Amazon の推薦がユーザにとって既知のものが多いこと、及び提案システムがユーザにとって未知の商品を推薦するのに有効であることが確認された。また、選択商品に関わらず、ユーザの嗜好にあっていると認められるような商品が推薦される例も多く観測された。ただし、嗜好傾向の抽出には十分な結果が得られなかった。原因として、システムが推薦する商品の母集団にジャンルの偏りがあることが考えられる。今後は推薦作成時の母集団の偏りを考慮し、より多くのユーザに対応可能なシステム構築に取り組む予定である。

#### 参考文献

- [1] “compete releases top 25 retail web sites for july 200”. [http://www.competeinc.com/news\\_events/pressReleases/238/](http://www.competeinc.com/news_events/pressReleases/238/).
- [2] 小川祐樹, 諏訪博彦, 太田敏澄, 山本仁志, 岡田勇. “共購買情報を用いた多様性のある推薦アルゴリズムの提案 amazon. co. jp の関連商品ネットワークを用いて”. 日本社会情報学会第 22 回全国大会研究発表論文集, 2007.
- [3] 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に関する複数方式の検討”. 電子情報通信学会 データ工学ワークショップ, 2007.
- [4] 神鷹敏弘. “データマイニング分野のクラスタリング手法 (2)”. 人工知能学会誌, 2003.