

グループ独自の性格を考慮したグループ向けコンテンツ推薦手法 A Content Recommendation Method for Group Recommendation using Group Character

瀬古 俊一[†]
Shunichi SEKO

八木 貴史[†]
Takashi YAGI

茂木 学[†]
Manabu MOTEGI

武藤 伸洋[†]
Shin'yo MUTO

1. はじめに

近年、ビデオホスティングサービスや e コマースの台頭によりアクセス可能なコンテンツが日々増大しており、ユーザは興味のあるコンテンツを見つけ出すことが困難になってきている。そこで、ユーザが興味のあるコンテンツを推定するレコメンデーション手法の研究が盛んに行われている[1][2]。しかしながら、多くのレコメンデーション手法は個人向けにデザインされたものである。レコメンデーションシステムは個人向けだけでなく、グループ向けにも適応可能となるべきである。例えば、映像コンテンツは友達や家族、恋人といったグループで一緒に視聴するシーンが数多く存在する。このようなグループ向けのレコメンデーションは、個人向けの場合と比較して複雑であり、グループを対象としたレコメンデーション手法が必要となる。そこで本論文では、この問題を解決するために、グループに対して適したコンテンツを推薦可能な手法を提案する。

2. グループレコメンデーションの関連研究

グループ向けの推薦(以下、グループレコメンデーション)の手法は、各グループメンバーのプロファイル(例えばコンテンツやジャンルに対する嗜好など)もしくは個々人の推薦結果を統合させるものが主流である。統合の方法は、一人でも嗜好スコアが低いコンテンツの推薦スコアを低くする・平均が高く分散が小さいコンテンツの推薦スコアを高くするなどの規則に従うルール型[3][4][5]や、TV の視聴時間帯や嗜好スコアの分散具合に基づいた加重平均型[6][7][8]などがある。しかしながら、グループの特性や傾向によって最適な統合方法は異なると考えられる。Satelo ら[9]はこれらを考慮し、グループの特性によって適応させるアルゴリズムを変えるという手法を提案しているが、2つの特性にしか分類を行っていない。グループの特性はもっと複雑であるため、より多様なグループの特性に対応可能な手法が必要である。

以上の問題に着目し、我々はグループの特性を示す Power Balance Map に基づいたレコメンデーション手法を過去に提案した。Power Balance Map とは、各メンバーの嗜好スコアを軸とした空間上にグループでの行動履歴(例えば一緒に視聴した映像コンテンツ)をプロットした分布図である。図1はユーザAとユーザBからなるグループである場合の Power Balance Map の例を示した図である。ひし形のプロットは各行動履歴(例えば映像コンテンツ)一つ一つを表しており、横軸はユーザAの嗜好スコア、縦軸はユーザBの嗜好スコアを示している。この Power Balance Map 上での高密度領域(図中の塗りつぶし領域)はグループの特性(グループメンバー間の力関係)を示していると考え、その領域の内部もしくは近くにプロットされる推薦対象のコンテ

ンツをグループに適したコンテンツと推定する。図1に示した例のように、グループでの行動によって高密度領域は異なる。このため、Power Balance Map を利用することで様々なグループの特性を考慮した推薦を行える。

しかし、以前の提案手法では最終的な推薦を行う際にコンテンツの内容を示す情報(ジャンルや出演者などのメタ情報)を考慮していなかった。そのため、まったく異なるジャンルをもつコンテンツ同士が Power Balance Map 上で同じ位置にプロットされ、グループに不適なコンテンツまで推薦してしまうという問題が生じていた。

そこで本論文では、この問題を解決し推薦精度を高めるために、Power Balance Map にグループの行動傾向を加味した手法を提案する。グループの行動傾向とは、コンテンツのメタ情報に基づくコンテンツの選択傾向を示すものである。この行動傾向を加味することで、不適なコンテンツをより多く除去可能になると考える。

3. 提案手法

図2は Power Balance Map にコンテンツのメタ情報(例えばジャンルなど)に対するグループの行動傾向を加味した提案手法の処理フロー図である。提案手法は初めに、グループでの行動履歴に基づいた行動傾向スコアの算出と、各グループメンバーの嗜好とグループでの行動履歴に基づいた Power Balance Map の作成を行う。そして、行動傾向スコア、Power Balance Map、各グループメンバーの嗜好スコアを用いて推薦対象となるコンテンツの推薦スコアを算出する。Power Balance Map に行動傾向スコアの軸を加えることによって、前章で述べた課題を解決可能と考える。図3は以前の提案手法(Power Balance Map のみを使用)と本論文で提案する新しい手法の差異を示した図である。ただし、図中の

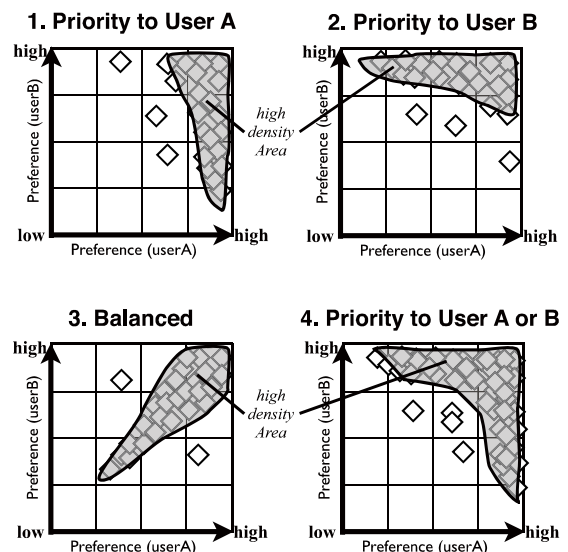


図1 Power Balance Map の例

[†] NTT サイバソリューション研究所 NTT Cyber Solutions Laboratories

すべてのコンテンツは異なるジャンル情報が紐付けられているとする。以前の提案手法では、コンテンツ A と C が同じポイントにプロットされる。これは、グループにとって不適なコンテンツを含みやすいことを意味している。一方、本論文で提案する新しい手法では、コンテンツ A と C を分離し、異なる推薦スコアを与えることが可能となる。したがって、Power Balance Map のみを使用する手法と比較して、グループに不適なコンテンツをより除去可能と考える。

以前の研究[10]では、Power Balance Map を作成し、それを用いて推薦を行う手法について述べた。しかしながら、Power Balance Map の作成法は簡潔な記述しかしていなかった。そのため、本章では初めに Power Balance Map の詳細な作成法について述べる。その後、Power Balance Map を行動傾向に基づいて拡張した推薦手法について述べる。

3.1 Power Balance Map の作成

Power Balance Map を作成するために、各グループメンバーのコンテンツに対する嗜好スコアと、グループでの行動履歴を用いる。コンテンツに対する嗜好スコアは、各グループメンバーのジャンルに対する嗜好スコアから決定される。表 1 はユーザ A とユーザ B の嗜好スコアの例を示したものである。ジャンルに対する嗜好スコアは、個人向けの嗜好推定技術[11]を用いるか、GUI などを通して各ユーザに尋ねることで取得する。もし、コンテンツが 2 つ以上のジャンルを持っていた場合、そのコンテンツに対する嗜好スコアは各ジャンルに対する嗜好スコアの平均値となる。表 2 はコンテンツの例を示しており、行動履歴も推薦対象コンテンツも所持するジャンルの数はコンテンツによって異なる。コンテンツ c に紐付けられたジャンルの集合を G 、ユーザ u のジャンル g に対する嗜好スコアを $r(u, g)$ とすると、ユーザ u のコンテンツ c に対する嗜好スコア $R(u, g)$ は以下の式(1)であらわせる。

表 1 ジャンルに対する嗜好スコアの例

Genre	User A Rating	User B Rating
News	5	3
Sports	5	1
TalkShow	3	4
Variety	2	4
Baseball	5	1
Comedy	3	3

表 2 コンテンツがもつジャンルの例

	Genre 1	Genre 2	Genre 3
History1	News	Sports	-
History2	News	TalkShow	Variety
NewContentA	Sports	Baseball	-
NewContentB	Comedy	TalkShow	-
NewContentC	Drama	Mystery	-
NewContentD	News	Documentary	Travel

$$R(u, c) = \frac{1}{|G|} \sum_{g \in G} r(u, g) \quad (1)$$

続いて、各メンバーのコンテンツに対する嗜好スコアを軸とした空間(以下、嗜好空間)にグループでの行動履歴(例えば一緒に視聴した映像コンテンツ)をプロットするために、嗜好ベクトルを作成する。嗜好ベクトルとは、各メンバーのコンテンツに対する嗜好スコアを成分とした多次元ベクトルである。グループのメンバー数を n とすると、コンテンツ c に対する嗜好ベクトル V_c は式(2)で定義される。

$$V_c = \{R(u_1, c), R(u_2, c), \dots, R(u_n, c)\} \quad (2)$$

この嗜好ベクトルを用いてグループでの行動履歴を嗜好空間にプロットする。これをすべての行動履歴に対して行うことで、Power Balance Map が完成する。

3.2 行動傾向による Power Balance Map の拡張

以前の提案手法[10]では Power Balance Map 上における全ての行動履歴と推薦対象となるコンテンツとの嗜好ベクトルの類似性を用いて推薦スコアを計算していた。本論文では、より推薦精度を向上させるために、Power Balance Map をグループの行動傾向により拡張した特長空間を用いたグループレコメンデーション手法を提案する。各メンバーの嗜好を示す次元と、グループでの行動傾向を示す次元とが統合された特徴空間を用いることで、グループにとって不適なコンテンツを除去可能と考える。以下に、新しい手法を実現するためのプロセスを述べる。

グループの行動傾向の次元軸を Power Balance Map へ加えるために、行動傾向スコアを算出する。行動傾向スコアは、行動履歴に含まれるメタ情報(例えばジャンルなど)の出現頻度を用いて求める。この値は高ければ高いほど、その行動を起こす(映像コンテンツならそれを視聴する)可能性が高いことを示す。行動履歴の集合を H 、行動履歴 h に紐付けられているジャンルの集合を h_G としたとき、ジャンル g に対する出現頻度 $F(g)$ は以下の式(3)で求める。

$$F(g) = \sum_{h \in H} \frac{exst(h, g)}{|h_G|} \quad (3)$$

$$exst(c, g) = \begin{cases} 1 & (\text{if content } c \text{ has genre } g) \\ 0 & (\text{if content } c \text{ does not have genre } g) \end{cases}$$

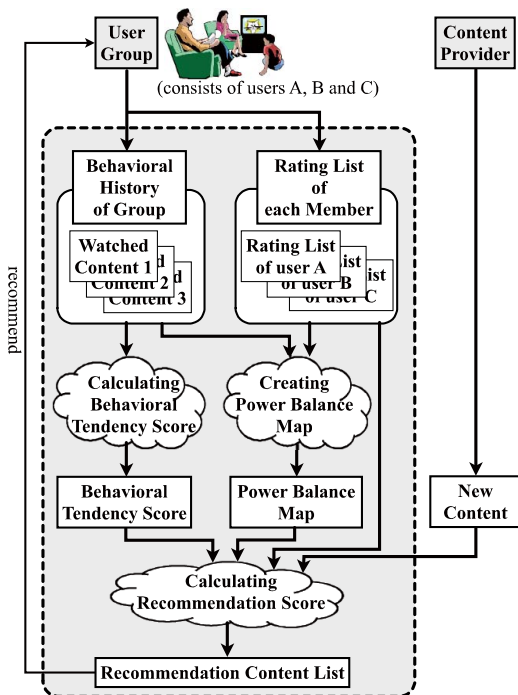


図 2 提案手法の処理フロー図

このメタ情報(ジャンル)の出現頻度を用いて各コンテンツに対するグループの行動傾向スコアを算出する. コンテンツ c に紐付けられているジャンルの集合を c_G としたとき, コンテンツ c の行動傾向スコア $BS(c)$ は以下の式(4)で定義される.

$$BS(c) = \frac{1}{|c_G|} \sum_{g \in c_G} F(g) \quad (4)$$

続いて, グループに適したコンテンツを推薦するために各推薦対象コンテンツに対する推薦スコアを計算する. まず, Power Balance Map を行動傾向で拡張するために, 嗜好ベクトルの成分に行動傾向スコアの次元を追加する. 以下, 拡張された Power Balance Map を "Groupality Space", 拡張された嗜好ベクトルを "Groupality Vector" と呼ぶ. 推薦スコアを計算するために, 推薦対象コンテンツの Groupality Vector と行動履歴の Groupality Vector との Groupality Space 上における類似度を計算する. 図4は推薦対象コンテンツと各行動履歴とをユークリッド距離に基づいた類似度計算を示した図である. この図において, 各行動履歴とのユークリッド距離が小さくなるものほど類似しているとみなす. 推薦対象コンテンツ c の Groupality Vector を V_c , 行動履歴 h の Groupality Vector を V_h とすると, 両ベクトルの類似度 $sim(V_h, V_c)$ は以下の式(5)で表せる.

$$sim(V_h, V_c) = \frac{1}{1 + \|V_c - V_h\|} \quad (5)$$

最終的な推薦スコアを計算するために, それぞれの軸(図4の例では, ユーザAの嗜好スコア, ユーザBの嗜好スコア, 行動傾向スコア)ごとに値を正規化する. 各軸で正規化を行うことで, 軸ごとに異なる最小値や最大値, 分散などの差異に依存せずにユークリッド距離を算出可能となる. 正規化は正規分布を利用した Probability Density Function [12]に基づいて行う(式(6)).

$$P(-\infty < X \leq x) = \int_{-\infty}^x \left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \right) dt \quad (6)$$

$\mu :=$ average of the score unit

$\sigma :=$ standard deviation of the score unit

例えば, ユーザAの嗜好スコアが表1に示された値である場合, ユーザAの "TalkShow" に対する正規化された嗜好スコア ($normr(userA, TalkShow)$) は以下ようになる.

$$normr(userA, TalkShow) = \int_{-\infty}^3 \frac{1}{\sqrt{2\pi}\sigma_A} \exp\left(-\frac{(x-\mu_A)^2}{2\sigma_A^2}\right) dx \approx 0.2461$$

$$(\mu_A \approx 3.8333, \sigma_A \approx 1.2134)$$

同様の処理を各ユーザの嗜好スコアおよび行動傾向スコアに行うことで, 各軸の正規化された値が算出される. これらの値を用いて正規化された Groupality Vector ($normV$) を作成し, 式(5)の類似度を計算する.

最後に, 推薦対象コンテンツに対する推薦スコアを算出する. 推薦スコアは, 正規化された Groupality Vector でありわされた推薦対象コンテンツ c と各行動履歴 h の類似度 ($sim(normV_c, normV_h)$) の総和とする. コンテンツ c に対する正規化された Groupality Vector を $normV_c$, 行動履歴 h に対する正規化された Groupality Vector を $normV_h$, 行動履歴の集合を H とすると, コンテンツ c に対する推薦スコアを $S(c)$ は以下の式(7)で表される.

$$S(c) = \sum_{h \in H} sim(normV_c, normV_h) \quad (7)$$

この推薦スコアを求める処理を全ての推薦対象コンテンツに対して行い, そのスコアの高い順に推薦対象コンテンツのランク付けを行う. そして, トップ k 位以上のコンテンツをグループに適したコンテンツと判断するなど, 算出したランクに基づいて推薦を行う.

4. 検証実験

Groupality Space を用いた提案手法が Power Balance Map のみを用いた手法よりグループに不適なコンテンツを除去可能かの検証を行った. 3組の被験者夫婦に対して TV 番組の視聴履歴(行動履歴)を約1ヶ月間収集し, そのデータを用いて評価を行った. 本実験では, 二人で一緒に視聴した番組のみを視聴履歴の対象とした. 以下に本実験の詳細

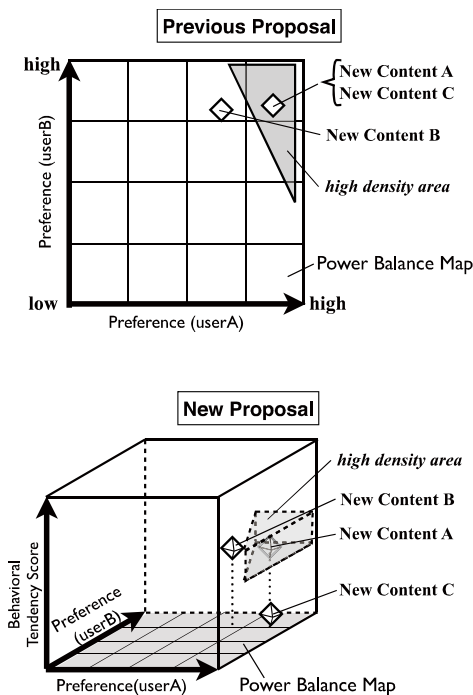


図3 旧提案手法と新提案手法の差異

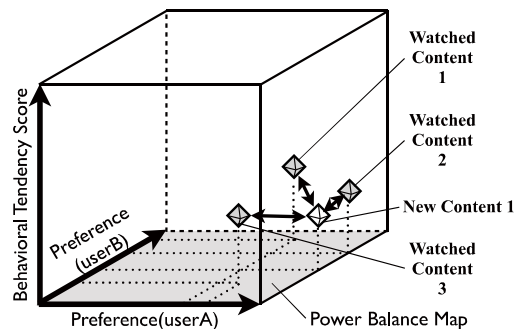


図4 Groupality Space における推薦対象コンテンツ(New Content)と各行動履歴(Watched Content)の類似度計算のイメージ図

を述べる。

4.1 データセット

各被験者夫婦の Groupality Space を作成するために、個々人の嗜好スコア、夫婦での視聴履歴、TV 番組データという3つのデータセットを収集した。

- **個々人の嗜好スコア**：個々人の嗜好スコアとは、被験者個人ごとの TV ジャンルに対する嗜好度合いを示すものである。本実験では、実験開始時に 104 種類の TV ジャンルに対する嗜好度合い(5 段階)をアンケートにて収集した。
- **夫婦での視聴履歴**：視聴履歴は夫婦一緒に見た TV 番組を示す情報の集合である。本実験では、被験者夫婦と一緒に TV 番組を視聴した際に、その番組タイトルを記録してもらうことで視聴履歴を収集した。
- **TV 番組データ**：TV 番組データはいくつかのメタ情報を含む地上波番組情報である。メタ情報として、番組タイトル、番組概要、開始・終了時刻、放送局、少なくともひとつのジャンルが紐づけられている。実験期間中のデータ数は全 12,234 番組となり、上記視聴履歴はこれらの番組データのサブセットとなる。

4.2 評価基準

新旧どちらの提案手法も、ある空間内での視聴履歴と推薦対象 TV 番組とのユークリッド距離に基づいて推薦する手法である。そこで本実験では、推薦対象 TV 番組と各視聴履歴とのユークリッド距離の平均値に基づいてランク付けを行い、被験者夫婦に適している番組が上位にどの程度含まれているかによって優位性を検証した。推薦対象 TV 番組データとして、上記 TV 番組データとは別に新たに 763 番組のデータを収集した。各被験者夫婦の組に、この 763 番組全てに対して”適”か”不適”のタグづけを行ってもらった。このアノテーションに基づいて評価を行う。

4.3 実験結果と分析

図 5 は旧提案手法(PowerBalanceMap)、新提案手法(GroupalitySpace)の評価結果を示したグラフである。横軸はユークリッド距離の平均値が低い順に上位 K 位まで推薦した場合を示す値であり、縦軸は上位 K 位以内に適した番組がいくつ含まれているかを示した値である。したがって、横軸の値が同じであれば、縦軸の値が大きいほど不適なコンテンツが含まれていなく、優れた結果であることを示している。この結果をみると、どの順位においても新提案手法のほうが上回っているといえ、上位 100 位では旧提案手法が適したコンテンツが 10 個(不適 90 個)に対し、新提案手法では 45 個(不適 55 個)存在する。つまり、新提案手法である Groupality Spaced 上でユークリッド距離を算出したほうが不適なコンテンツを除去可能といえる。よって、Power Balance Map にグループの行動傾向の軸を加えることは有効であるといえる。

5. まとめと今後

本論文では、Power Balance Map とグループの行動傾向に基づいた特徴空間(Groupality Space)を用いてグループに適したコンテンツを推薦する手法の提案を行った。提案手法の有効性を検証するために、被験者夫婦の個々人の嗜好スコアと 1 ヶ月間の TV 視聴履歴を収集した。これらのデー

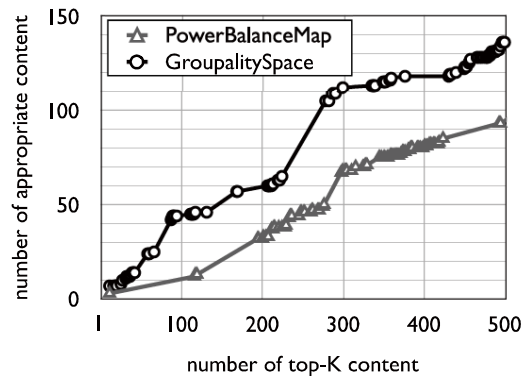


図 5 旧提案手法(PowerBalanceMap)と新提案手法(GroupalitySpace)の評価結果(上位 500 位以内)

タを用いて Groupality Space を用いる手法と Power Balance Map を用いる手法との優位性に関する比較検証を行った。その結果、Groupality Space 上でユークリッド距離を算出した方が不適なコンテンツを含まないことが示された。

今後は、レコメンドを行った際の精度を他手法も含めて比較検証を行う。また、グループのメンバ数が 3 人以上の場合においても、提案手法が有効であるかの検証を行う予定である。

参考文献

- [1] Gediminas Adomavicius and Alexander Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, vol.17, no.6, pp.734-749, June 2005.
- [2] Y. Nakamura, et al., "Personalized tv-program recommendations based on life log," International Conference on Consumer Electronics (ICCE), pp.143-144, 2010.
- [3] Judith Masthoff, "Group modeling: Selecting a sequence of television items to suit a group of viewers," User Model and User-Adapted Interaction, vol.14, pp.37-85, 2004.
- [4] M O'connor, et al., "Polylens: A recommender system for groups of users," ECSCW 2001, pp.199-218, 2001.
- [5] Anthony Jameson and Barry Smyth., "Recommendation to groups," The Adaptive Web, pp.596-627, 2007.
- [6] Dina Goren-Bar and Oded Glinansky, "Fit-recommending tv programs to family members," Computers & Graphics, vol.28, no.2, pp.149-156, 2004.
- [7] Zhiwen Yu, et al., "Tv program recommendation for multiple viewers based on user profile merging," User Modeling and User-Adapted Interaction, vol.16, no.1, pp.63-82, 2006.
- [8] Choonsung Shin and Woontack Woo, "Socially aware tv program recommender for multiple viewers," IEEE Transactions on Consumer Electronics, vol.55, no.2, pp.927-932, 2009.
- [9] Rafael Sotelo, et al., "Tv program recommendation for groups based on multidimensional tv-anytime classifications," IEEE Transactions on Consumer Electronics, vol.55, no.1, pp.248-256, 2009.
- [10] Shunichi Seko, et al., "Video content recommendation for group based on viewing history and viewer preference," IEEE International Conference on Consumer Electronics (ICCE), pp.359-360, 2011.
- [11] 手塚 博久, 他, "ライフログを活用したレストランレコメンド," NTT 技術ジャーナル, vol.22, no.7, pp.29-32, 2010.
- [12] Emanuel Parzen, "On estimation of a probability density function and mode," The Annals of Mathematical Statistics, vol.33, no.3, pp.1065-1076, 1962.