

多様性のある協調フィルタリングアルゴリズムの提案 Amazon.comの関連商品ネットワークを用いて

A proposal of collaborative filtering algorithm through diversification:

By using similar-items network in Amazon.com.

小川祐樹[†] 諏訪博彦[†] 山本仁志^{††} 岡田勇^{†††} 太田敏澄[†]
Yuki OGAWA[†] Hirohiko SUWA[†] Hitoshi YAMAMOTO^{††} Isamu OKADA^{†††} Toshizumi OHTA[†]

[†] 電気通信大学 大学院情報システム学研究科

^{††} 立正大学 経営学部

^{†††} 創価大学 経営学部

[†] Graduate School of Information Systems, University of Electro-Communications

^{††} Faculty of Business Administration, Rissho University

^{†††} Faculty of Business Administration, Soka University

要旨:

我々は、推薦の正確性を維持し、かつ多様性を向上させることによりユーザ満足度を向上させる推薦アルゴリズムを提案する。従来の協調フィルタリングによる推薦アルゴリズムでは、予測評価値計算において、近傍ユーザの類似度の合計値で値の正規化を行っている。このため、評価値情報から多くの人々が薦めているという本来の情報が損失してしまう。我々は、予測評価値計算において類似度の合計値で正規化しない計算式を提案した。また、推薦リストへのユーザ満足度を、適合性、納得性、意外性の3つの指標を用いて定義した。評価実験において既存のアルゴリズムとの比較を行い、実際の推薦リストに対するアンケート評価の結果、高い適合性と意外性の値を得た。

Abstract:

In order to improve satisfaction of users in recommendation lists, we will propose an alternative recommendation algorithm that increases diversity of lists keeping that of accuracy. A current recommendation algorithm, such as collaborative filtering, employs normalization by the neighboring similarity among users to calculate a prediction evaluation value. However, the algorithm with normalization can not include invaluable information concerning how many people recommend certain item. An alternative formula without normalization will be proposed, and also three indices to describe satisfaction of users; i.e. conformity, assent, and windfall. In conclusion, our algorithm can be better than the current algorithm with respect to recommendation list, in that our recommendation list can be found higher conformity and windfall than that of current algorithm in questionnaire evaluation by users.

1. はじめに

推薦システムは、Amazon.com など多くの商用サイトで用いられている。協調フィルタリングは、推薦システムを作る際に最もよく使われる手法であり、実用的な手法とされている。従来の協調フィルタリングによる推薦システムの研究では、推薦の正確性（MAE, 精度, 再現率）の向上を重視しており、推薦の多様性に関する研究は少ない。正確性を重視した推薦システムは、ユーザの嗜好を正確にとらえる反面、推薦されるアイテムが似かよったものばかりになってしまい、意外なアイテムの推薦が行われにくいという問題を抱えている。

本研究では、推薦の正確性をある程度維持し、かつ推薦の多様性を向上させることにより、ユー

ザの満足度を向上させる推薦アルゴリズムの提案を目指す。

2. 先行研究

推薦システムの多様性の問題に対し、Zieglerら[1]は従来の協調フィルタリングから求められる推薦リストを、リスト内のアイテムトピック（Amazon.com で付与されているアイテム分類）が最も多様になるアイテムの組み合わせに改善する推薦アルゴリズムを提案している。また実験により、推薦に多様性を導入することで、正確性のみを重視した推薦よりも高いユーザ満足度が得られることを示している。清水ら[2]はアイテムの既知・不既知のユーザプロファイルを用いることで、発見性を考慮した推薦アルゴリズムを提案

している．これら先行研究の問題点として，[1]では多様化の定義が十分でない点，[2]ではユーザプロフィールを作成することのユーザの労力が大きい点などがあげられる．

本研究では，正確性と多様性の両立を考えるため，既存の強調フィルタリングによる Top-N 推薦リストの計算アルゴリズムを改善し，共購買ネットワークを用いた新たな推薦アルゴリズムを提案する．

なお，今回の研究においては，協調フィルタリングによる Top-N 推薦リストの計算方法の改善アルゴリズムを提案し，提案するアルゴリズムと既存のアルゴリズムについて，推薦リストのシステム評価とユーザ満足度の評価について調査を行う．

3. 評価者数を考慮した協調フィルタリングアルゴリズム

協調フィルタリングによる推薦は，ユーザと好みの似たユーザグループが好きなアイテムをそのユーザに推薦するというものである．手順としては，類似度計算と予測評価値計算の2つの手順で行われる．

従来の予測評価値計算では，予測するアイテムの評価値を，近傍ユーザの類似度の合計値で正規化する手法をとる．しかし，個人の評価情報を正規化すると，本来，気に入ったアイテムを他者に薦めたいと考えてお気に入りアイテムを評価したユーザの情報は，正規化の段階で失われてしまう．また，多くの人が薦めているという情報も評価者数で割ることで失われてしまう．

そこで我々は，評価情報の損失を防ぎ，評価情報を保存した協調フィルタリングアルゴリズムを提案する．

3.1. 従来の協調フィルタリングによる推薦

類似度計算では，推薦を受けるユーザと似た嗜好を持っているユーザを推定するために，ユーザ間の類似度を計算する．一般的に類似度計算は，相関係数法とベクトルを用いた Cosine 類似度による計算方法とに分別される．本研究では，GroupLens[3]において用いられたピアソン相関係数と、事前実験において高い精度が得られた Adjusted Cosine Similarity による2つの類似度計算を用いた．

予測評価値計算では，類似するユーザの評価をもとに，未評価のアイテムに対して，そのアイテムの予測評価値を計算する．この値の上位 N アイテムが，Top-N 推薦リストとしてユーザに提示される．本研究では，GroupLens[3]で用いられた Adjusted Weight Sum (1)と、事前実験において高い精度が得られた Weight Sum (2)による2つの予測評価値計算を，提案する予測評価値計算式との比較対象として用いた．ユーザ u の未評価アイテム i へ対する予測評価値 $P_{u,i}$ をそれぞれ以下に示す．

$$P_{u,i}^{AWS} = \bar{r}_u + \frac{\sum_{m \in M} \text{sim}(u, m)(r_{m,i} - \bar{r}_m)}{\sum_{m \in M} \text{sim}(u, m)} \quad (1)$$

$$P_{u,i}^{WS} = \frac{\sum_{m \in M} \text{sim}(u, m) r_{m,j}}{\sum_{m \in M} \text{sim}(u, m)} \quad (2)$$

ここで， $r_{m,i}$ はユーザ m のアイテム i へ投票した評価値， \bar{r}_u はユーザ u の投票したすべての評価値の平均， \bar{r}_m はユーザ u の類似度の高い上位 M 人の近傍ユーザとする．

3.2. 提案手法

我々は，正規化による評価情報の損失を防ぐために，予測評価値計算において，値を類似度の合計値で正規化しない Simple Adjusted Summation 手法 (3) を提案する．

$$P_{u,i}^{SAS} = \bar{r}_u + \sum_{m \in M} \text{sim}(u, m)(r_{m,i} - \text{standard value}) \quad (3)$$

ここで，standard value は，評価値の基準値（例えば評価値が 1~5 の場合，基準値は 3 となる）とする．上記の式により，近傍ユーザの多くが高く評価したアイテムほど値が高くなり，推薦リストの上位に現れやすくなると考える．

4. 推薦リストの評価指標

4.1. システム評価

精度と再現率は，情報検索システムの評価指標としてよく知られており，協調フィルタリングによる推薦リストの正確さの指標として用いられている[4]．これらの値は交差検証法によって計算される．交差検証法では，各ユーザの評価情報を K 個のブロックに分割する．ここで， $K-1$ 個のブロックはトレーニングセットとされ，残りの1ブロックであるテストセットを予測するデータとされる．本研究では， $K=10$ の k -folding cross-validation によって以下の指標を計算した．

$$\text{精度} = \frac{|T_i^x \cap L_i^x|}{|L_i^x|} \quad (4)$$

$$\text{再現率} = \frac{|T_i^x \cap L_i^x|}{|T_i^x|} \quad (5)$$

ここで， T_i^x はユーザ i のテストセット中の好きなアイテム集合， $|T_i^x|$ はそのアイテム数， L_i^x はユーザ i の推薦リスト内のアイテム， $|L_i^x|$ はそのアイテム数とする．

正確さ以外の指標としては，推薦が当たり前でないことを表す Novelty という指標が提案されている[5]．Novelty における精度・再現率を以下に

示す．

$$\text{精度(Novelty)} = \frac{|C_i^x \cap L_i^x|}{|L_i^x|} \quad (6)$$

$$\text{再現性(Novelty)} = \frac{|C_i^x \cap L_i^x|}{|C_i^x|} \quad (7)$$

C_i^x は、ユーザ i のテストセット中の知らない好きなアイテムとする．

この他に、推薦リスト内に知らないアイテムがどれだけあるかを測定するための指標として、発見性という指標が提案されている[2]．

$$\text{発見性} = \frac{|D_i^x \cap L_i^x|}{|L_i^x|} \quad (8)$$

D_i^x は、ユーザ i の知らないアイテムとする．

4.2. ユーザ満足度評価

我々は、ユーザ満足度の評価において、適合性(9)、納得性(10)、意外性(11)の3つの評価方法を提案し、アンケートによってユーザ満足度の評価を行う．

$$\text{適合性} = \frac{\sum_{i \in N} -2|La_i^1| - |La_i^2| + |La_i^4| + 2|La_i^5|}{|N|} \quad (9)$$

$$\text{納得性} = \frac{\sum_{i \in N} -2|Lb_i^1| - |Lb_i^2| + |Lb_i^4| + 2|Lb_i^5|}{|N|} \quad (10)$$

$$\text{意外性} = \frac{\sum_{i \in N} -2|Lc_i^1| - |Lc_i^2| + |Lc_i^4| + 2|Lc_i^5|}{|N|} \quad (11)$$

ここで、 $|N|$ は被験者数、 $|La_i^x|$ はユーザ i の推薦アイテム中の、見たアイテムへの評価値 x のアイテム数とする．同じように、 $|Lb_i^x|$ は見てないけど知っているアイテム、 $|Lc_i^x|$ は知らないアイテムへの評価値 x のアイテム数とする．

本研究では、この指標の適合性と納得性を推薦リストの正確性、また意外性を多様性として捉え、ユーザ満足度の評価指標とする．

5. 評価実験

5.1. 実験概要

実験の手順を以下に示す．

1. 売上 Top1000 の DVD アイテムを全集合とし、各ユーザ（学生 12 人）にランダム 32 個のアイテムを提示する．
2. ユーザは提示された各アイテムに対し、アイテム知識（見た、見てないけど知っている、知らない）と、嗜好の評価値 1～5（1:最低、5:最高）を入力する．
3. これらの評価値情報をもとに、それぞれのアルゴリズムで推薦リスト（上位 Top10 アイテ

ム）を生成する．

4. 各アルゴリズムで生成した推薦リスト内のアイテムの和集合をランダムでユーザに提示し、ユーザは各アイテムに対して、アイテム知識と評価値を入力する．

手順 1 における Top1000 アイテムは、Amazon.co.jp の全 DVD アイテム数に対する各ジャンルのアイテム数の割合を考慮し、各ジャンルでの売上上位のアイテムを選定した．

手順 4 におけるアルゴリズムは、類似度計算アルゴリズムとして Adjusted Cosine Similarity (ACS)、ピアソン相関係数 (Pe) を用い、それに対応する予測評価値計算アルゴリズムとして、既存手法の Weight Sum (WS)、Adjusted Weight Sum (AWS) と、提案手法である Simple Adjusted Summation (SAS) を用いた．その他、推薦アイテムをランダムに選ぶアルゴリズム (Ra) を比較対象として用いた．

5.2. データセット

本実験では、協調フィルタリングにおいて標本となる評価値情報として、Amazon.co.jp の DVD アイテムのレビュー情報を用いた．これらの情報は、Amazon.com が提供する Amazon API によって取得し、ユーザ数 3108 人（1 ユーザあたりの平均評価アイテム数 3.2 個、平均評価値 4.2）の評価値情報を用いた．

5.3. 評価方法

本実験では、推薦リストの評価方法として、評価値情報のみから求めることができるシステム評価と、実際の推薦リストに対するユーザ満足度のアンケート評価の2つによる評価を行う．

5.4. 実験結果

表 1 は、各ユーザがランダムアイテム 32 個へ行ったアイテム評価と、評価値の内訳である．

表1 ランダムアイテム32個におけるアイテムの内訳（平均アイテム数）

		嫌いなアイテム (評価値1～3)	好きなアイテム (評価値4～5)
アイ テ ム 知 識	見た	0.4	2.1
	見てないけど 知っている	3.6	3.6
	知らない	18.0	4.3

表 2 に、推薦された Top10 アイテムにおける、アイテム知識別のアイテム数を示す．

表 3 は、3 つのシステム評価（精度・再現率、Novelty、発見性）の計算結果である．

表 4 は、アンケート評価によるユーザ満足度の結果である．

表2 推薦された Top10 アイテムにおけるアイテム知識の内訳 (平均アイテム数)

	類似度計算 & 予測評価値計算	ACS & SAS	ACS & WS	ACS & AWS	Pe & SAS	Pe & WS	Pe & AWS	Ra
アイテム知識	見た	3.6	0.6	1.6	4.2	1.2	0.8	0.6
	見てないけど知っている	2.6	2.4	2.8	3.4	2.8	3.0	2.0
	知らない	3.8	7.0	6.2	2.4	6.0	6.2	7.4

表3 システム評価 (精度・再現率, Novelty, 発見性)

類似度計算アルゴリズム & 予測評価値計算アルゴリズム	ACS & SAS	ACS & WS	ACS & AWS	Pe & SAS	Pe & WS	Pe & AWS	Ra
精度	.00017	.00100	.00033	.00092	.00167	.00118	.00041
再現率	.00105	.00875	.00208	.00833	.01243	.00868	.00234
精度 (Novelty)	.00000	.00000	.00000	.00000	.00067	.00000	.00025
再現率 (Novelty)	.00000	.00000	.00000	.00000	.00444	.00000	.00248
発見性	.00058	.00275	.00183	.00033	.00150	.00076	.00165

表4 アンケート評価 (ユーザ満足度)

類似度計算 & 予測評価値計算	ACS & SAS	ACS & WS	ACS & AWS	Pe & SAS	Pe & WS	Pe & AWS	Ra
適合性	4.6	1.5	2.3	2.6	0.5	0.3	0.6
納得性	3.8	2.3	1.9	1.0	0.4	0.6	1.0
意外性	0.8	-1.8	-3.1	0.3	-0.7	-2.3	-2.0
合計	9.1	2.1	1.1	3.9	0.3	-1.4	-0.4

5.5. 考察

表2より, ACSとPeの両方のSASにおいて, 他の予測評価計算手法よりも, 見たアイテムが多く推薦されていることが分かる。これは, SASにおいて, 近傍ユーザの多くが高く評価したアイテムが, 実際にユーザが見たアイテムである傾向が高いことを示している。

表3の精度・再現率において, SASでは他の手法よりも値が低くなった。これは, SASにおいて, 近傍ユーザの中で評価数が多いアイテム, つまり一般的にレビュー数が高いものが推薦リストの上位にきたため, ユーザごとのパーソナライズされた推薦アイテムの順位が相対的に下がったためと考えられる。また, 精度 (Novelty)・再現率 (Novelty)において, Pe & WSとRa以外の計算結果が0となったのは, 推薦するアイテム数を上位10アイテムに設定したことと, 表2より, SASは知らないアイテムが推薦される数が少ないため, このような結果になったと考えられる。発見性については, 表2の結果と対応した結果となった。

表4では, 適合性において, 提案したSASが, 高い値を示している。これは, SASが見たアイテムをより多く推薦できたことを示している。また, SASが意外性の項目において唯一プラスの値と

なっている。これは, SASでは知らないアイテムを推薦される機会は少ないが, 知らないアイテムを推薦されたときには, そのアイテムの多くは高く評価されているといえる。この理由として, 近傍ユーザの多くが高く評価しているアイテムについては, ユーザの潜在的興味が高いアイテムであることが考えられる。

6. 結論

本研究では, 予測評価値計算において近傍ユーザの類似度の合計値で正規化しない予測評価値計算式を提案した。評価実験において既存のアルゴリズムとの比較を行い, 精度・再現率においては既存の手法よりも値は低くなったが, 実際の推薦リストに対するアンケート評価の結果では, 高い適合性と意外性の値を得た。これらの結果より, 提案アルゴリズムによって, 正確性と多様性を両立したユーザ満足度の高い推薦リストを作成することができた。

7. 今後の課題

本研究では, 提案手法により適合性の高い推薦リストの作成が実現できた。しかし, 見たアイテムばかりが推薦されてしまうと, 推薦アイテムが当たり前すぎてユーザはその推薦に興味を持たなくなってしまうことが考えられる。このため, 推薦リストでの, 正確性・納得性・意外性のバランスにおいても検討する必要があると考える。また, Amazon.comの共購買ネットワークを用いた多様性の導入についても検討し, その効果についても検証する予定である。

参考文献

- [1] Ziegler, C., McNee, S.M., Konstan, J.A. and Lausen, G., "Improving Recommendation Lists Through Topic Diversification", WWW2005, pp.22-32, 2005.
- [2] 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基礎検討", 情報処理学会研究報告 データベース・システム研究会報告, vol.2006, No.59, pp.53-60, 2006.
- [3] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", In *Proc. ACM Conf. on Computer Supported Cooperative Work*, Chapel Hill, NC, pp.175-186, 1994.
- [4] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J.: "Application of dimensionality reduction in recommender system", In *ACM WebKDD Workshop*, 2000.
- [5] Herlocker, J., Konstan, J., Terveen, L., and Riedl, J.: "Evaluating collaborative filtering recommender systems", *ACM Transactions on Information Systems*, Vol.22, No.1, pp.5-53, 2004.