

Personalizability を考慮した推薦システムの提案

吉川 大弘¹ 森 貴章² 古橋 武¹

概要: 近年, インターネットの普及とそれに伴う各種メディアの発達により, ユーザがアクセス可能な情報の量は膨大なものとなったが, 一方で, ユーザにとって価値のある情報を選択することは困難となっている. そのため, ユーザのアイテム選択を支援する“レコメンデーション”に注目が集まり, 様々な推薦システムが実用化され始めている. このような推薦システムの評価には, これまで主に“精度”が用いられてきたが, 近年, ユーザ満足度の観点から, 他にも様々な側面を評価する必要性が指摘され始めている. “Serendipity”はその中の一つであり, 推薦アイテムの目新しさ, 意外性を表す. 本論文では, Serendipity 定量化の試みとして“Personalizability”を定義し, Personalizability を考慮した推薦手法を提案する. また, 提案手法をベンチマークデータに適用し, 推薦における精度と Personalizability のバランスが調節可能であること, 従来手法と比べ, Personalizability において優れていることを示す.

キーワード: 推薦システム, Serendipity, Personalizability, アソシエーションルール, 精度

A Proposal of Recommendation System considering Personalizability

Abstract: Recently, a user can access a huge amount of information with the popularization of Internet and the development of media associated with that. On the other hand, it becomes difficult for him/her to choose desirable or valuable ones. Then various recommendation systems have been studied and put into practical use to support users' selection. Though “accuracy” has been used as the evaluation of these systems so far, it is said that other evaluation factors are also needed from the aspect of the satisfaction of users. “Serendipity,” which includes novelty or unexpectedness, is one of the evaluation indexes for user satisfaction. This paper defines “Personalizability” as an approach to quantify the “Serendipity.” This paper proposes a recommendation system considering “Personalizability,” and applies the proposed method to benchmark data. It shows that this method can adjust the balance between “Accuracy” and “Personalizability” and it is superior to the conventional method in terms of “Personalizability.”

Keywords: Recommendation System, Serendipity, Personalizability, Association Rule, Accuracy

1. はじめに

近年, インターネットの普及とそれに伴う各種メディアの発達により, ユーザがアクセス可能な情報の量は膨大なものとなったが, 一方で, ユーザにとって価値のある情報を選択することは困難となっている. そのため, ユーザのアイテム選択を支援する“レコメンデーション”に注目が集まり, 様々な推薦システムが実用化され始めている [1][2][3]. 例えば, EC サイトの Amazon.com [4] では, ユーザの購入・閲覧履歴に基づき, ユーザが好むであろう

アイテムを“おすすめ”として推薦している. また, 音楽配信サイトの Pandora [5] では, ユーザの評価履歴 (Like / Don't Like) に基づき楽曲の推薦を行っている.

このような推薦システムの評価には, これまで主に“精度”が用いられてきた [1]. 精度は, 推薦アイテムのうちユーザが「Like」と答えたものの割合であり, 手法の性能を測る上で極めて重要な指標である. しかし, 近年, ユーザ満足度の観点から, 他にも様々な側面を評価する必要性が指摘され始めており, 例えば, Explainability [6] や Serendipity [1][7] などが挙げられている. Explainability は, 推薦の根拠をユーザに理解可能な形で説明できるかどうかを表す. また Serendipity は, 推薦アイテムが「Like」であることに加え, 目新しさ, 意外性があるかを表す. 本論文では, この

¹ 名古屋大学
Nagoya University
² ブラザー工業株式会社
Brother Industries, Ltd.

“Serendipity”に着目する。

ほとんどのユーザが好むアイテムのみを推薦すれば、一般的に精度は高くなるが、目新しさ、意外性は少ない [1]。よって、個人性の高いアイテムの方が Serendipity が高くなる可能性が高いと考える。そこで本論文では、Serendipity 定量化の試みとして “Personalizability” を定義する。これは推薦アイテムの個人特化性であり、推薦アイテムが全体においてどれだけ好まれていない中でその個人に特化して好まれるかを表す指標である。本論文では、推薦における Personalizability の向上が Serendipity の向上に結びつくとの仮定のもと、Personalizability を考慮した推薦手法を提案する。また、提案手法をベンチマークデータに適用し、推薦における精度と Personalizability のバランスが調節可能であること、従来手法と比べ、Personalizability において優れていることを示す。

2. 提案手法

2.1 推薦システム

本論文で着目する推薦システムを図 1 に示す。システムはユーザに 1 つのアイテムを推薦し、ユーザは推薦されたアイテムに対して評価 (Like / Don't Like) を行う。システムが持つデータベースにはユーザの評価履歴が蓄積されている。図 1 のデータベースにおいて、1 は “Like”, -1 は “Don't Like” とそれぞれ評価されたアイテムであることを示し、“0” は未評価のアイテムであることを表す。

推薦システムのアルゴリズムには様々なものが用いられるが、代表的なものとして k -Nearest Neighbor 法 [8] が挙げられる。これは、各アイテムのスコアをその近傍 k 個のアイテムに対する評価履歴に基づいて算出し、最もスコアの高いアイテムを推薦する手法である。近傍アイテムは、各アイテムの評価履歴間の類似度を基に決定される。スコア算出式の例を以下に示す。

$$score(I_i) = \sum_{j=1}^k (s_{I_i I_j} \cdot e(I_j)) \quad (1)$$

ここで、 I はアイテムを表す。 I_j は I_i の近傍アイテムであり、 $s_{I_i I_j}$ は I_i - I_j 間の類似度を、 $e(I_j)$ は I_j に対する評価履歴 (1 / -1) をそれぞれ表す。

また、最も単純な推薦アルゴリズムに Non-Personalized 法 [9] がある。これは、全ユーザに対し同じ推薦を行うものであり、本論文では、全ユーザ中における「Like」の割合が高い順にアイテムを推薦する手法を指す。

2.2 アソシエーションルールに基づく推薦

アソシエーション分析とは、データに内在する項目どうしの関連を、アソシエーションルールとして抽出する手法である [10]。ルールは $\{A \rightarrow B\}$ の形式で表現され、 A を条件部、 B を結論部と呼ぶ。ルールには、その重要性を表す

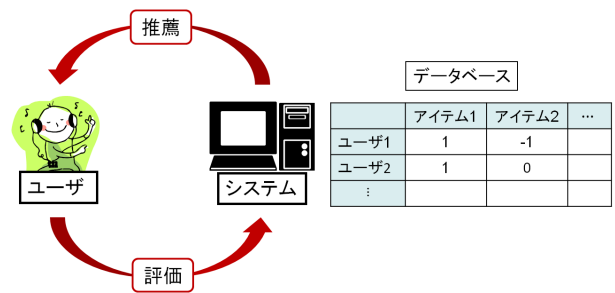


図 1 推薦システム

Fig. 1 Recommendation System

評価指標 [11] が存在し、閾値などの形で評価指標に関する条件を解析者が指定することによって、条件を満たすルールが抽出される。多く用いられる評価指標である *support*, *confidence*, *lift* を以下に示す。

$$support = P(A \cap B) = \frac{N(A \cap B)}{N_d} \quad (2)$$

$$confidence = P(B|A) = \frac{N(A \cap B)}{N(A)} \quad (3)$$

$$lift = \frac{P(B|A)}{P(B)} \quad (4)$$

ここで、 N_d は全データ数、 $N(A)$ は条件 A を満たすデータ数である。

confidence は、“同じ評価履歴を持つ人が、対象となるアイテムをどれだけ好きであるか”を表し、評価指標における“精度”の期待値に相当すると考えられる。しかし例えば、ある評価履歴を持つ人の 90% がアイテム a を好きであるが、そもそも全ユーザの 90% がアイテム a を好きである、といった場合のように、*confidence* による指標だけでは、Non-Personalized 法に近い、個人性の低いアイテムが多く推薦され得る。そこで、“ある評価履歴を持つ人があるアイテムをどれだけ好きであるか”と、“皆があるアイテムをどれだけ好きであるか”の比の値である *lift* を併せて用いる。*lift* は、*confidence* が大きく、全ユーザ中における「Like」の割合が低くなるほど大きな値となるため、2.5 で示す評価指標 Personalizability の期待値に相当すると考えられる。以下、提案手法の詳細について述べる。

2.3 Personalizability を考慮した推薦システム

提案手法では初めに、全ユーザの評価履歴データから、アソシエーション行列を作成する。アソシエーション行列の例を表 1 に示す。行頭は各アイテムに対する評価「Like」と「Don't Like」、列頭は「Like」のみである。各要素 a_{ij} は、行列の行頭 A_i 、列頭 B_j を用いて以下のように算出される。

$$a_{ij} = confidence_{(A_i \Rightarrow B_j)} \cdot lift_{(A_i \Rightarrow B_j)}^\alpha \quad (5)$$

ここで、 α は *lift* の重みに関するパラメータであり、 α を小さくするほど *lift* の影響は小さくなる。 $\alpha \neq 0$ のとき、

a_{ij} は $lift$ が 1 より大きい場合には $confidence$ より大きくなり, $lift$ が 1 より小さい場合には $confidence$ より小さい値となる. なお, アソシエーション行列の作成はオフライン (ユーザとのコネクションを行っていない状態) で行う.

表 1 アソシエーション行列
Table 1 Association Matrix

	アイテム 1 =Like	アイテム 2 =Like	アイテム 3 =Like
アイテム 1 =Like		0.88	0.60
アイテム 1 =Don't Like		0.18	0.90
アイテム 2 =Like	0.50		0.88
アイテム 2 =Don't Like	0.78		0.94

ユーザに対するアイテムの推薦は, そのユーザの評価履歴とアソシエーション行列を用いて以下のように行う.

- (a) アソシエーション行列から, 行頭がユーザの評価履歴に該当する行を抽出する. ただし, ユーザが評価済みのアイテムに関する列は除く.
- (b) (a) で抽出した行群において, 列毎に要素の値の和をとり, その値を各アイテムのスコアとする.
- (c) 最もスコアの高いアイテムを推薦する.

提案手法では, アソシエーションルールに基づく推薦を行うことで, アイテムのスコア算出に用いた評価履歴を推薦の根拠として提示可能であり, Explainability においても優れていると考えられる.

2.4 関連研究

Sarwar ら, Kim らは, 購入履歴データなどの 0/1 データを対象に, アソシエーションルールに基づく推薦手法を提案している [13][14]. 両者とも, ユーザの評価履歴をルールの条件部とみなし, その評価指標を用いたスコアづけを各アイテムに対し行う点では, 提案手法と類似している. しかし, 用いる評価指標は $confidence$ のみであり, また, アイテムのスコアを算出する際, 提案手法ではカルバック・ライブラー情報量上位 k 個の和をとるが, Sarwar らは最大値を, Kim らは総和をとるなど, 推薦における Personalizability を考慮していないところに, 提案手法との大きな違いがみられる.

2.5 推薦システムの評価指標

1章で述べた Serendipity は, 感情的な応答を伴うものであり, 定量的に評価することは難しいが, これまでいくつかの点から定量化の試みがなされている. Ziegler らは,

推薦するアイテムの多様性を重視することで Serendipity を高めようとした [15]. しかし, これにはアイテムの特徴 (メタ情報) が必要となる. Herlocker らは, グループ, もしくは全体向けの (Non-Personalized な) 推薦には現れないが, 個人向けの推薦には現れるアイテムを Serendipity が高いとみなしている [7]. またこの Non-Personalized な推薦との違いを, Chandrashekar ら, Schickel らはそれぞれ "Novelty" と呼んでいる [16][17].

本論文では, 推薦アイテムが全体においてどれだけ好まれていない中で, その個人に特化して好まれるかを表す指標として Personalizability を定義し, その定量化を試みる. 具体的には, 推薦したアイテムが「Like」であることに加え, 推薦アイテムの全ユーザ中における「Like」の割合の低さを情報量 [12] により表す.

以下に, 精度, Novelty, Personalizability に対する評価指標を示す.

精度

推薦回数を N , 推薦アイテムの集合を $I = I_1, I_2, \dots, I_N, I_i$ に対する評価履歴を $e(I_i) = 1 / -1$ とすると, 以下の式で表される.

$$\frac{1}{N} \sum_{i=1}^N t_i \quad (6)$$

$$t_i = \begin{cases} 1 & \text{if } e(I_i) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Novelty

式 (7) における t_i を以下のものとして表される.

$$t_i = \begin{cases} 1 & \text{if } e(I_i) = 1 \text{ and } I_i \notin I_{NP} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

I_{NP} は Non-Personalized 法における推薦アイテムの集合を表す. これは, 推薦アイテムが「Like」であり, かつ Non-Personalized な推薦には現れない割合である [16][17].

Personalizability

式 (6) における t_i を以下のものとして表される.

$$t_i = \begin{cases} \log_2 \frac{1}{P(e(I_i)=1)} & \text{if } e(I_i) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

式 (9) の上段は, 推薦アイテムの全ユーザ中における「Like」の割合の低さを情報量 [12] により表したものである. ここでは対数の底として "2" を選んでいるため, 「Like」の割合が 1/2 より小さい, すなわちマイノリティに好まれるアイテムにおいて, 情報量は 1 より大きくなる.

3. 実験

3.1 使用データ

本実験では, ベンチマークデータである MovieLens デー

タ [18] と Jester Joke データ [19] を用いた。MovieLens データは、映画に対する評価履歴データであり、ユーザ数は 943、アイテム数は 422 である。また、1 ユーザ当りの平均評価アイテム数は 76.6 である。評価は 1~5 の 5 段階でつけられているが、Like / Don't Like のバランスを考慮し、1~3 の評点を「Don't Like」、4,5 を「Like」に変換して実験を行った。また、Jester Joke データは、ジョークに対する評価履歴データであり、ユーザ数は 24,983、アイテム数は 100、平均評価アイテム数は 72.5 である。評価は -10.00 ~ 10.00 であるが、MovieLens と同様に、2.00 未満の評点を「Don't Like」、2.00 以上を「Like」に変換して実験を行った。

3.2 実験方法

実験手順は次の通りである。初めに、3.1 で述べた評価履歴データにおいて、「Like」と 21 回以上答え、「Don't Like」と 20 回以上答えたユーザを実験の対象とした（以降、単にユーザと呼ぶ）。ユーザ数は、MovieLens データにおいて 416、Jester Joke データにおいて 13,801 であった。次に、ユーザ群に対する 10-fold cross-validation [20] により、2.5 で示した評価指標に基づき評価を行った。すなわち、ユーザ群をランダムに 10 分割し、その 9/10 でトレーニング（アソシエーション行列の作成）、1/10 でテスト（評価）を行うことを、分割されたすべてのグループについて繰り返した。

実験の際、テストユーザは新規ユーザ（全アイテムに対し未評価）として扱うが、初期の評価履歴として、「Like」と答えたアイテムの中からランダムに選ばれたものを 1 つ用いることとした。これは、新規ユーザには、最初に好きなアイテムを 1 つ入力してもらうことを想定している。評価履歴データで「未評価」となっているアイテムが推薦された場合は、カウントせず次のアイテム推薦（次にスコアの高いアイテム）を行い、評価済みアイテムが 10 個推薦されるまで推薦を行った。

3.3 実験 1

提案手法において、式 (5) に示す重みパラメータ α に対する性能の変化を調べた。MovieLens データ、Jester Joke データにおける実験結果をそれぞれ図 2、図 3 に示す。なお、提案手法における k の値は、 $\alpha = 0$ の場合において最も精度が高くなるものを選び、MovieLens データ、Jester Joke データにおいてそれぞれ $k = 5$ 、 $k = 4$ となっている。

図から、 α を大きくするほど、精度は低くなるものの、Novelty、Personalizability は高くなっていくことがわかる。これより、重みパラメータ α により、推薦における“精度”と“Personalizability”のバランスが調節可能であると考えられる。実際の運用の上では、初期のうちは $\alpha = 0$ として精度を重視した推薦を行い、徐々に α を大きくして

Personalizability 重視の推薦を行うことや、 α を変えたときの「Like」/「Don't Like」の評価状況の変化に応じて α を調整するなどの運用が考えられる。

なお、図 2 と図 3 の α の値から見て取れる通り、重みパラメータ α に対する性能の変化はデータによって異なる。式 (5) における *lift* は、アイテムの評価履歴間の関連性（あるアイテムを Like / Don't Like であることと、別のあるアイテムを Like であること）の強さに関係しているため、全体的に関連が高いデータの方が α の変化に対して敏感になると考えられる。そこで、カルバック・ライブラー情報量の期待値である相互情報量 [12] を用いて、適用したデータにおける評価履歴間の関連性を調べた。その結果、MovieLens データ、Jester Joke データにおけるアイテム間の相互情報量の平均値は、それぞれ 0.0385、0.0080 であった。このことから、MovieLens データの方が、評価履歴間の関連性が高く、それにより α の変化に対して各評価指標の変化が敏感になっていたと考えられる。

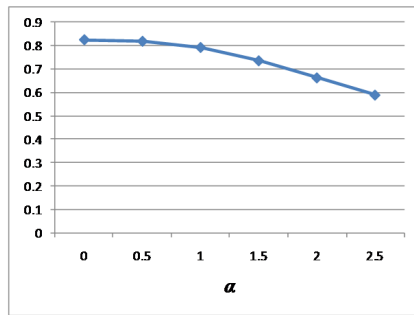
3.4 実験 2

提案手法 (AR (Association Rule))、Non-Personalized 法 (NP)、 k -Nearest Neighbor 法 (kNN) における性能比較を行った。MovieLens データ、Jester Joke データにおける結果をそれぞれ図 4、図 5 に示す。ただし、図 4(b) 及び、図 5(b) において、その定義上、Non-Personalized 法による Novelty の値は 0 となる。なお、kNN におけるアイテム間類似度にはコサイン類似度 [8] を用いた。また、kNN における k の値は最も精度が高くなるものを選び、MovieLens データ、Jester Joke データにおいてともに $k = 10$ とした。

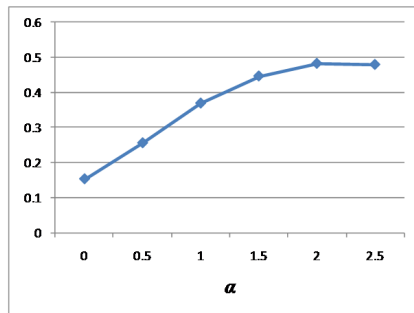
図から、提案手法において Personalizability を重視した場合、Novelty、Personalizability の基準においてともに最も高くなることがわかる。また図 4(b)、図 5(b) から、提案手法 (Personalizability を重視) では Novelty のあるアイテムを、MovieLens データにおいて 48.1%、Jester Joke データにおいて 43.7% の割合で推薦していることがわかる。さらに図 4(c)、図 5(c) において、提案手法による推薦において受け取る情報量は、Non-Personalized 法と比べ、MovieLens データにおいて 2.24 倍、Jester Joke データにおいて 1.67 倍であった。精度については、提案手法において精度を重視 ($\alpha = 0$) した場合、その性能を Non-Personalized 法と同程度にまで高くすることが可能であり、これは、 k -Nearest Neighbor 法に勝るものであることがわかる。

4. まとめ

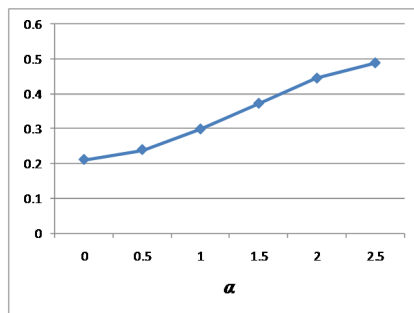
本論文では、推薦システムにおける Serendipity に着目し、その定量化の試みとして“Personalizability”を定義した。推薦における Personalizability の向上が Serendipity の向上に結びつくとの仮定のもと、Personalizability を考



(a) 精度



(b) Novelty



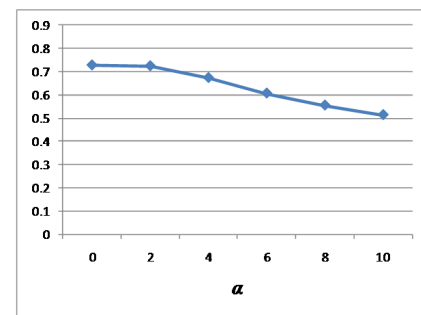
(c) Personalizability

図 2 重みパラメータ α に対する性能の変化
(MovieLens データ)

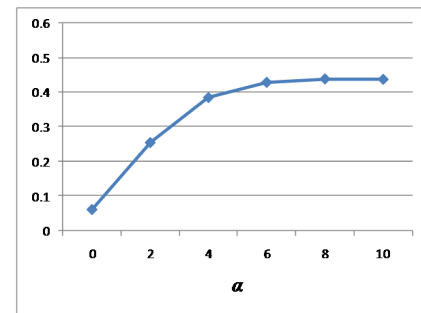
Fig. 2 Performance for Difference of α (MovieLens Data)

慮したアソシエーションルールに基づく推薦手法を提案した。提案手法では、ルールの条件部にユーザの評価履歴、結論部にアイテムに対する評価「Like」を当てはめ、その評価指標 *confidence* と *lift* を用いたスコアづけを未評価のアイテムに対し行い、最もスコアの高いアイテムを推薦する。さらにカルバック・ライブラー情報量に基づき、用いる評価履歴の選択を行う。提案手法を映画及びジョークの評価に関するベンチマークデータに適用し、*lift* に課した重みパラメータにより、推薦における“精度”と、“Personalizability”とのバランスが調節可能であること、また、従来手法と比べ、Personalizability において優れていることを示した。

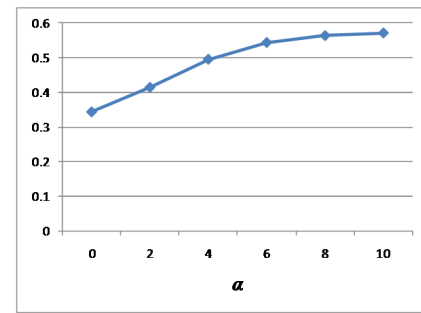
今後の課題として、より大規模なデータに適用し、多くの推薦回数を設定した実験による、提案手法の有効性の検証を行うことが挙げられる。また合わせて、提案手法による推薦アイテムについて、Serendipity としての妥当性を、



(a) 精度



(b) Novelty



(c) Personalizability

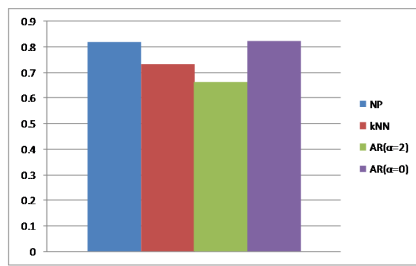
図 3 重みパラメータ α に対する性能の変化
(Jester Joke データ)

Fig. 3 Performance for Difference of α (Jester Joke Data)

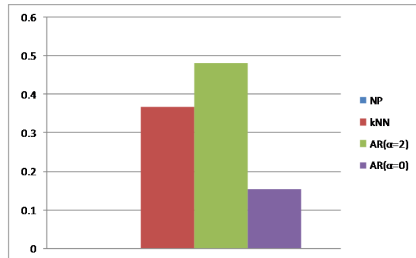
ユーザ評価により検証していく必要がある。

参考文献

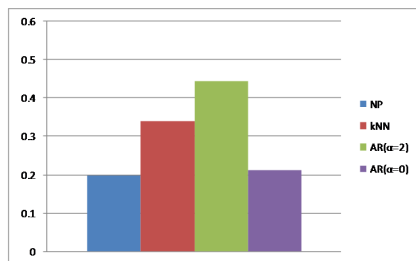
- [1] 神島敏弘：推薦システムのアルゴリズム (1)，人工知能学会誌，Vol. 22，No. 6，pp. 826-837，(2007).
- [2] 神島敏弘：推薦システムのアルゴリズム (2)，人工知能学会誌，Vol. 23，No. 1，pp. 89-103，(2008).
- [3] 神島敏弘：推薦システムのアルゴリズム (3)，人工知能学会誌，Vol. 23，No. 2，pp. 248-263，(2008).
- [4] <http://www.amazon.com>
- [5] <http://www.pandora.com>
- [6] Y. Koren: Tutorial on recent progress in collaborative filtering, Proc. of the 2008 ACM conference on Recommender Systems, pp. 333-334, (2008).
- [7] J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J.T. Riedl: Evaluating collaborative filtering recommender systems, ACM Transactions on Information Systems, Vol. 22, (2004).
- [8] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl: Item-based collaborative filtering recommendation algorithms,



(a) 精度



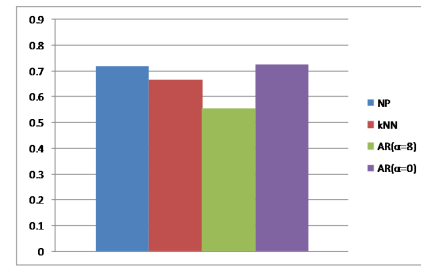
(b) Novelty



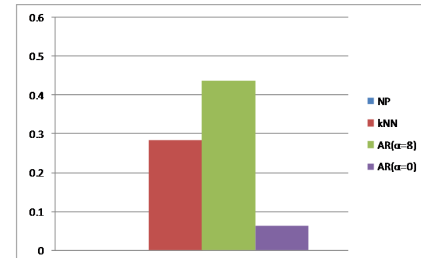
(c) Personalizability

図 4 従来手法との性能比較
(MovieLens データ)

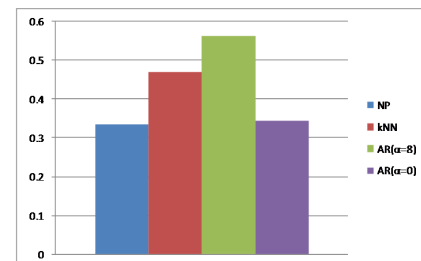
Fig. 4 Comparison of Performance with
Conventional Method (MovieLens Data)



(a) 精度



(b) Novelty



(c) Personalizability

図 5 従来手法との性能比較
(Jester Joke データ)

Fig. 5 Comparison of Performance with
Conventional Method (Jester Joke Data)

- Proc. of the 10th international conference on World Wide Web, pp.285-295, (2001).
- [9] W.S. Lee: Collaborative learning for recommender systems, Proc. of the 18th international conference on Machine Learning, pp. 314-321, (2001).
- [10] R. Agrawal, and R. Srikant: Fast algorithms for mining association rules, Proc. of the 20th international conference on Very Large Data Bases, pp. 487-499, (1994).
- [11] L. Geng, and H.J. Hamilton: Interestingness measures for data mining, A survey, ACM Computing Surveys, Vol. 38, (2006).
- [12] S. Kullback: Information theory and statistics, Dover Publications, Mineola, N.Y., (1997).
- [13] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl: Analysis of recommendation algorithms for e-commerce, Proc. of the 2nd ACM conference on Electronic Commerce, pp. 158-167, (2000).
- [14] C. Kim, and J. Kim: A recommendation algorithm using multi-level association rules, Proc. of the 2003 IEEE/WIC international conference on Web Intelligence, pp. 524-527, (2003).
- [15] C.N. Ziegler, S.M. McNee, J.A. Konstan, and G. Lausen: Improving recommendation lists through topic diversification, Proc. of the 14th international conference on World Wide Web, pp. 22-32, (2005).
- [16] H. Chandrashekar, and B. Bhasker: Personalized rec-

- ommender system using entropy based collaborative filtering technique, Journal of Electronic Commerce Research, Vol. 12, No. 3, pp. 214-237, (2011).
- [17] V. Schickel-Zuber, and B. Faltings: Inferring user's preferences using ontologies, Proc. of the 21st national conference on Artificial Intelligence, pp. 1413-1418, (2006).
- [18] B.N. Miller, I. Albert, S.K. Lam, J.A. Konstan, and J. Riedl: MovieLens unplugged: experiences with an occasionally connected recommender system, Proc. of the 8th international conference on Intelligent User Interfaces, pp. 263-266, (2003).
- [19] D. Gupta, M. Digiovanni, H. Narita, and K. Goldberg: Jester 2.0 (poster abstract): evaluation of a new linear time collaborative filtering algorithm, Proc. of the 22nd annual international ACM SIGIR conference on Research and Development in Information Retrieval, pp. 291-292, (1999).
- [20] R. Kohavi: A study of cross-validation and bootstrap for accuracy estimation and model selection, Proc. of International Joint Conference on Artificial Intelligence 1995, pp. 1137-1143, (1995).