

ユーザの好みに基づく音楽アーティスト推奨システムの開発

著者	岡田 吉史, 澤井 政宏, 楠 芳之, 長島 知正
雑誌名	サテライト・ベンチャー・ビジネス・ラボラトリー年報
巻	5
ページ	36-38
発行年	2003
URL	http://hdl.handle.net/10258/368

ユーザの好みに基づく音楽アーティスト推奨システムの開発

著者	岡田 吉史, 澤井 政宏, 楠 芳之, 長島 知正
雑誌名	サテライト・ベンチャー・ビジネス・ラボラトリー 年報
巻	5
ページ	36-38
発行年	2003
URL	http://hdl.handle.net/10258/368

ユーザの好みに基づく音楽アーティスト推奨システムの開発

岡田 吉史¹⁾, 澤井 政宏²⁾, 楠 芳之¹⁾, 長島 知正^{1,2)}

1) 室蘭工業大学 サテライト・ベンチャー・ビジネス・ラボラトリー

2) 室蘭工業大学 情報工学科

1. はじめに

近年、Web 上のオンラインショップの増加に伴って、個人の趣味嗜好に合った商品や情報(以下、アイテムと呼ぶ)を推奨するシステムが提案されている[1]。そのようなシステムは、ユーザ毎の各アイテムに対する評価点(例えば5段階評価など)のプロファイルや購買履歴などのログ情報を蓄積したデータベースを持ち、そこから推奨要求ユーザと興味や嗜好が似た他のユーザを探し、その人たちが好んでいるアイテムを推奨する。しかし、ユーザ間の嗜好の類似性は、ユーザのアイテムの選好履歴(どのアイテムをどの程度好んできたか)から計算され、個々のユーザがアイテムのどのような特徴や性質に着目して好んでいるかは考慮されてこなかった。一般に、複数ユーザが同じアイテムを同程度に好む場合でも、それらのユーザの“嗜好性”、すなわち嗜好理由や評価基準は多かれ少なかれ異なると考えられる。従って、アイテムの持つ様々な特徴に対する嗜好理由の個人差(すなわち“こだわり”)を考慮することで、より個人の感性に合ったアイテム推奨が可能になると期待される。

本研究では音楽アーティスト(以下、アーティスト)推奨を例に、ユーザの“好きなアーティスト”と“その嗜好理由”に基づいて、ユーザの好みに合ったアーティストを推奨する方法を提案し、その実装システムの開発を行った。本稿では、既存システムとの推奨精度に関する比較評価をととして、本システムの有用性について述べる。

2. 音楽アーティスト推奨システム

2.1 実行例

図1は本システムの実行画面のスクリーンショットである。本システムへの入力クエリとして、好きなアーティスト(図1 a)と、その嗜好理由(図1b)を選択すると、ユーザが指定したアーティストを同様の理由で好んでいる人(つまり、共通のアーティストを好むだけでなく、そのアーティストに対するこだわりも似た人)が、他にどのようなアーティストを好んでいるかを提示する(図1c)。この時、ユーザと嗜好理由が似ている人の推奨アーティストほど上位にランキングされる。ここでは、ユーザの好きなアーティストとして“宇多田ヒカル”、嗜好理由として“声質”、“音域”、“歌い方”

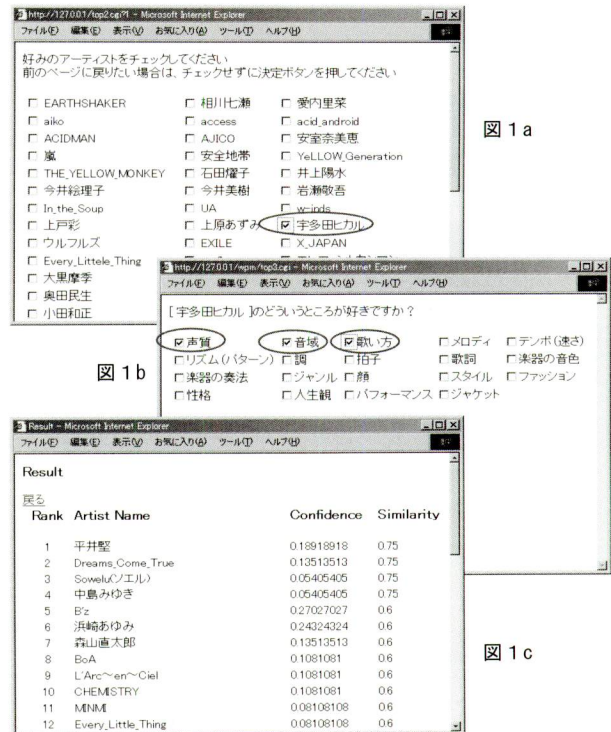


図1: 本システムの実行画面

を選択した場合の例を示している。

2.2 システム構成

本システムは、1) 相関ルールで表されたアーティストの嗜好パターン(アーティスト相関ルール)を格納したデータベースと、そこから2) ユーザの好みに合う候補アーティストを検索して提示する推奨モジュールから構成される。相関ルールとは、共起関係に着目して膨大なデータに内在する規則性を表現するための記述形式であり、スーパーマーケットの POS データから同時に売れる商品のパターンを分析する場合などに用いられる[2]。例えば、「商品 A を好む顧客の 90% は商品 B を好み、全顧客の 30% が商品 A と商品 B を好む」とき、相関ルールは“商品 A ⇒ 商品 B: 支持度=30%, 確信度=90%”のように記述される。ここで商品 A を条件部、商品 B を結論部と呼ぶ。本研究では、人々から同時に好まれるアーティストの関係を相関ルールで表し、さらに、それらの相関ルールにアーティストに対する嗜好理由を表す情報を導入する(3章参照)。

表 1 : 数値データに変換したアンケート調査結果

アンケート	アーティスト ID	Taste Vector
1	176	1000001001000000000
	222	0011000010000000101
	446	0010000010011000011
2	75	1010011010011110011
	176	0110110000001000011
⋮	⋮	⋮

3. システム構築方法

3.1 アーティスト嗜好調査

295 人の男女に対して”好きなアーティスト”と”その嗜好理由”を調査するアンケート調査を行った。ここで、好きなアーティストは自由記載とし、各アーティストに対する嗜好理由は、事前アンケートにより予め選定した 19 個の理由(図1b)からの選択形式とした。アンケート調査結果のうち、238 件をアーティスト関連ルールデータベースを作成するためのトレーニングセットとして、残りの 57 件をシステム評価のためのテストセットとして利用した。

3.2 アーティスト関連ルールデータベース

前節のアンケート調査から、各アーティストを一意的番号(アーティストID)に置き換え、各アーティストに対する嗜好理由を19次元のベクトルに変換し、表1のようなテーブルを作成した。ベクトルの各次元は各嗜好理由に対応し、各次元の要素はアンケート回答者が嗜好理由を選択したか否かにより1または0を持つ。こうして構成されたベクトルを *Taste Vector* と呼ぶ。次に、このテーブルから回答者によって同時に好まれるアーティストの関連ルールを抽出した。アーティスト関連ルールは、例えば“{Artist1 & Artist2}⇒{Artist3}; 支持度=0.05, 確信度=0.7”のように表される。これは、“Artist1, Artist2, Artist3 の3人のアーティストを好む人は全体の 5%であり、さらに Artist1 と Artist2 を好む人の 70%は Artist3 を好んでいる”ことを意味する。もし、ユーザが Artist1 と Artist2 を好むならば、Artist3 も好む可能性が高いと考え、Artist3 を推奨アーティストとしてユーザに提示する。さらに、関連ルールを構成するアーティストが回答者にどのような理由で好まれているかを表現するため、関連ルールの条件部および結論部に含まれるアーティストに対して各々の *Taste Vector* を付加し、最終的に以下のような形式で関連ルールを記述する:

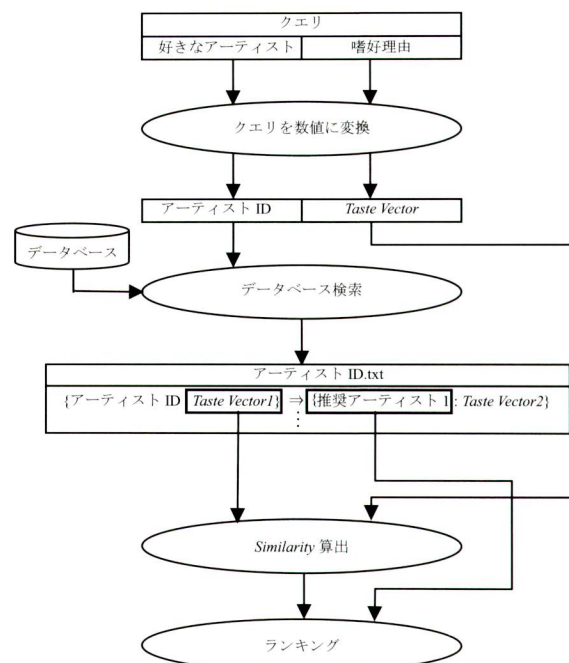


図 2 : 推奨モジュールの処理の概要図

{Artist1: Taste Vector1 & Artist2: Taste Vector2}
⇒ {Artist3: Taste Vector3}

以上の手順で、アーティスト関連ルールを抽出し、それを格納するデータベース作成した。

3.3 推奨モジュール

前節のデータベースからユーザの好みに適合するアーティストを検索して提示する、以下の推奨モジュールを開発した。図2は、推奨モジュールが行う処理の概要図である。推奨モジュールは、ユーザからのクエリとして“好きなアーティスト”と“嗜好理由”を受け取ると、それらを“アーティストID”と“Taste Vector”に変換する。次に、ユーザの指定したアーティストIDを条件部に持つ関連ルールを検索する。これにより、ユーザと同じアーティストを好む人から得られたルールのみが検索される。以下、検索されたファイルに格納される関連ルール群を検索ルールと呼ぶ。最後に、ユーザの指定した *Taste Vector* と検索ルールの条件部が持つ *Taste Vector* を用いて、嗜好理由の類似度 (*Similarity*) を次式により計算する:

$$Similarity = \frac{Count_{both}}{Count_{either\ or\ both}}$$

ここで、 $Count_{both}$ はユーザの指定した *Taste Vector* と検索ルールの条件部の *Taste Vector* の各次元において、両方の値が1である個数、 $Count_{either\ or\ both}$ は一方または両方の値が1である個数を表す。高い *Similarity* を持つ検索ルールは、ユーザと嗜好理由の似た人から得られた関連ルールであり、そのような人が好むアーティストはユーザ

にも好まれる可能性が高いと考えられる。そこで、本研究では *Similarity* の高い順に検索ルールの結論部に含まれるアーティストをランク付けする。

4. 推奨精度の評価

4.1 評価方法

本システムの性能評価として、W. Linらによって提案された推奨システムAR-CRS[3]との推奨精度の比較を行った。AR-CRSは相関ルールを用いた推奨システムであり、ユーザ相関、アイテム相関、ユーザ相関とアイテム相関の一体型という3通りの手法が提案されている。本研究では、最も高い推奨精度を示すユーザ相関に基づく手法を比較対象として選択した。

3章で行ったアンケートで回答数が多かった上位10名のアーティストをテストクエリに用いる。そこで、テスト用の57件のアンケートから各テストクエリを含むアンケートを1枚ずつ無作為に選びテストアンケートとした。各テストアンケートに含まれるテストクエリ以外のアーティストをターゲットと呼ぶ。システムにテストクエリを与えた時にターゲットが上位にランキングされるならば、実際のシステムの利用においても、ユーザにとって未知の好みのアーティストが上位にランキングされる可能性が高いと考えられる。評価は、全てのテストクエリに対する適合率および再現率を平均し、再現率-適合率グラフ[4]を作成することにより行う。再現率-適合率グラフとは、横軸に再現率を0.1刻みの11個の点(0.0, 0.1, ..., 1.0)でとり、それに対する適合率を縦軸にプロットしたグラフである。このグラフが右上に位置するほど性能(推奨精度)が高いと見なす。

4.2 結果と考察

図3に本システムとAR-CRSの再現率-適合率グラフを示す。ここで、再現率がいずれも1.0に達していないのは、ターゲットとなるアーティストが相関ルールベースにない(トレーニングセットにない)場合があったためである。図3に示されるように、本システムは再現率、適合率グラフにおいてAR-CRSを上回る推奨精度を示した。

ところで、AR-CRSは相関ルールのマイニング処理をオンラインで行うため、データベース内のユーザ数が増えるに従い推奨に要する計算量も増大する。一方、本システムはアーティスト相関ルールをオフラインで生成するため、推奨に要する時間はデータベース中の人数によらずほぼ一定である。このように、本システムは推奨精度とオンライン計算時間においてAR-CRSを上回っている。

5. まとめ

本稿では、ユーザが好むアーティストとその嗜好理由から、ユーザの好みに合うアーティストを提示する新しい

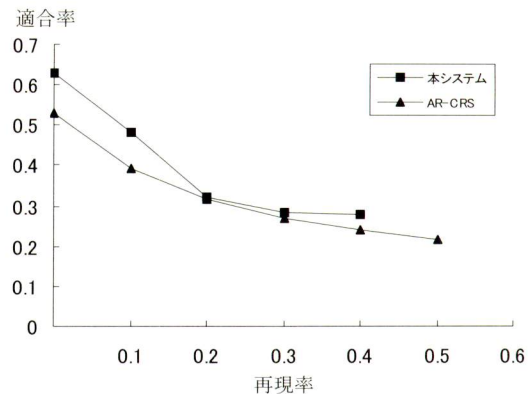


図3：再現率-適合率グラフ

推奨システムを提案した。評価実験により、本システムは従来手法の1つであるAR-CRSより、推奨精度とオンラインにおける計算時間において優れていることを示した。

一方、本研究で使用した19個の嗜好理由は事前アンケートに基づいて我々が手作業で決定したもので、それらが実際に個人の嗜好を特徴付ける要因として妥当かどうかの検証は行っていない。今後は、アーティスト嗜好に関わる特徴量を心理学的あるいは統計的分析をとおしてさらに吟味し、推奨技術に取り入れていきたい。

アーティストに限らず香水や絵画など、様々なコンテンツに対する嗜好性には、個人の感性が大きく関与していると考えられる。近年、オンラインショッピングの普及に伴い、インターネット上で購入できる商品の種類や数が急激な増加を見せている。そのような膨大かつ多種多様な商品から、本稿で述べたような、ユーザの感性的な嗜好性に基づいて商品推奨を行う技術の開発は、今後ますます重要になると思われる。

参考文献

- 1) Resnick, P., Iacovou, N., Suchak, M., Bergstorm, P. and Riedl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews, Proc. ACM Conf. on computer Supported Cooperative Work, Chapel Hill, North Carolina, 175-186, 1994
- 2) R. Agrawal and R. Srikant, Fast Algorithms for Mining Association Rules in Large Databases, Proc. 20th Int. Conf. Very Large Data Bases, Santiago, Chile, 478-499 (1994).
- 3) W. Lin, S. A. Alvarez, C. Ruiz: Collaborative recommendation via adaptive association rule mining, Proc. Web Mining for E-Commerce Workshop, Boston, (2000).
- 4) Harman, D. K., editor: Overview of the Third Text Retrieval Conference (TREC-3), National Institute of Standards and Technology, Gaithersburg, MD, (1995).