

ユーザの評価と音響的特徴との確率的統合に基づく ハイブリッド型楽曲推薦システム

吉井 和佳[†] 後藤 真孝[‡] 駒谷 和範[†] 尾形 哲也[†] 奥乃 博[†]

[†] 京都大学大学院 情報学研究科 知能情報学専攻 [‡] 産業技術総合研究所
yoshii@kuis.kyoto-u.ac.jp m.goto@aist.go.jp
{komatani, ogata, okuno}@i.kyoto-u.ac.jp

本稿では、ユーザによる楽曲評価と楽曲の音響的特徴とを同時に利用するハイブリッド型楽曲推薦システムについて述べる。従来の主な楽曲推薦手法には、ユーザの楽曲評価を考慮した協調フィルタリングと音響的特徴の類似度を考慮したコンテンツベース推薦の2つがあり、それぞれ問題を抱えている。前者では、評価が与えられていない楽曲は推薦できず、推薦楽曲のアーティストのバラエティは乏しい。後者では、音響的特徴とユーザの嗜好との理論的な対応付けが不十分である。本稿ではこれらの問題を解決するため、アスペクトモデルと呼ばれるベイジアンネットワークを用いて、楽曲評価と音響的特徴とを確率的に統合する。このモデルでは、直接観測できないユーザの嗜好が隠れ変数として表現され、観測データ（楽曲評価と音響的特徴）から統計的枠組みに従って推定される。ポピュラー音楽 CD と Amazon.co.jp から収集した評価データを用いた楽曲推薦実験の結果、提案手法が上記の問題を解決し、より優れた推薦精度を示すことを確認した。

A Hybrid Music Recommender System based on Probabilistic Integration of User Ratings and Acoustic Features

KAZUYOSHI YOSHII[†], MASATAKA GOTO[‡], KAZUNORI KOMATANI[†], TETSUYA OGATA[†]
and HIROSHI G. OKUNO[†]

[†] Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University
[‡] National Institute of Advanced Industrial Science and Technology (AIST)

This paper presents a hybrid music recommender system that uses user ratings and acoustic features of musical pieces. There are two prominent recommendation methods — collaborative filtering and content-based recommendation, which have respective problems. The former cannot recommend newly-released musical pieces that have no ratings because recommendations are based on actual user ratings. Furthermore, artist variety in recommended pieces tends to be poor. The latter has unreliability in modeling of user preferences; the content similarity does not completely reflect the preferences. Our method integrates both rating and content data by using a Bayesian network called an aspect model. In this model, unobservable user preferences are directly represented by introducing latent variables, which are statistically estimated. To verify our method, we conducted experiments by using actual audio signals of Japanese songs and the corresponding rating data collected from Amazon.co.jp. The results showed that our method solves the above-mentioned problems and outperforms the two conventional methods in terms of recommendation accuracy and artist variety.

1. はじめに

近年、楽曲推薦技術の重要性はますます高まっている¹⁾。我々はインターネットを介して大規模楽曲データベースにアクセスできるようになったが、その中から好きな楽曲を見つけるのは容易ではない。現在の標準的な楽曲検索システムでは、ユーザ自身が検索クエ

リを入力しなければならないため、好きな楽曲を探すにはどのようなクエリが適切か分からず途方にくれることがある。なぜなら、ユーザが自分自身の音楽的嗜好を言明することは一般的に難しいからである。この問題に対処するには、システムがユーザの嗜好を推定し、ユーザが好むであろう楽曲を検索して提示することが必要である。これまで、主な推薦手法として、協

調フィルタリングとコンテンツベース推薦が提案されており、それぞれ相補的な性質を持っている。

協調フィルタリング手法²⁾⁻⁴⁾は、あるユーザに楽曲を推薦する際に他のユーザの楽曲評価を参考にする。例えば、楽曲 A, B を好むユーザ X に楽曲推薦を行う場合、楽曲 A, B, C を好むユーザが他に多数いれば、ユーザ X は楽曲 C を好きになってくれるはずだと予想する。このヒューリスティクスの妥当性は、オンラインショッピングサービス (Amazon.co.jp や iTunes Music Store など) で大きな成果をあげていることで実証済みである。しかし、この方法には 2 つの問題がある。1 つめの問題は、評価のなされていない楽曲は推薦できないことである*。2 つめの問題は、推薦される楽曲のアーティストのバリエーションが乏しいことである。あるアーティストのファンのユーザは、そのアーティストの多数の楽曲に高い評価をつける傾向がある。そのため、そのアーティストの楽曲が好きなユーザには、同じアーティストの他の楽曲が推薦されやすくなる。

コンテンツベースの推薦手法⁵⁾⁻⁷⁾は、音色、リズム、音響的特徴などの音楽的コンテンツの点で、ユーザが好む楽曲と類似した楽曲を推薦する。この方法では評価のなされていない楽曲も推薦可能で、アーティストのバラエティも豊かである。しかし、ユーザの嗜好と音楽的コンテンツとの関連付けに関する検討が、現実的な楽曲評価データを用いて十分に行われているとは言えない。例えば、Hoashi ら⁵⁾はユーザの嗜好を音響的特徴空間内で表現する手法を提案しているが、わずか 12 人の被験者による楽曲評価データを用いて有効性を検証している。Logan ら⁶⁾は音響的特徴の類似性に基づく楽曲推薦手法を提案しているが、各アルバム CD を 1 人のユーザが好きだと評価した楽曲群だとみなして仮想的な評価データを構築している。Celma ら⁷⁾は WEB 上の個人プロフィールを利用した楽曲推薦手法を提案しているが、定量的評価は行っていない。

我々はこれまで述べた問題を解決するため、ユーザによる楽曲評価と音楽音響信号の音響的特徴とを同時に考慮するハイブリッド型楽曲推薦手法を提案する。これにより、未評価楽曲も含めて、アーティストのバラエティが豊かな推薦を行うことができる。コンテンツベースの推薦手法では、推薦対象ユーザの楽曲評価数の不足が当該ユーザの嗜好推定の信頼性の低下に繋がっていた。一方、本手法では他のユーザの楽曲評価も参考にすることでこの問題を解消できる。

ハイブリッド型楽曲推薦を行うための一つの方法は、協調フィルタリング手法とコンテンツベース推薦の手法を並列に、あるいは連続して利用することである^{8),9)}。しかし、このような方法では、直接観測できないユーザの本質的な嗜好が依然として表現できない。観測データ (楽曲評価と音響的特徴) はユーザの嗜好を完全に反映しているわけではない。

我々の手法は、Popescul ら¹⁰⁾によって提案された 3 方向アスペクトモデル (Three-way Aspect Model) を拡張して利用する。このモデルを使えば、直接観測できないユーザの嗜好を隠れ変数として表現でき、それらの値は観測データである楽曲評価と音響的特徴から統計的に推定できる。アスペクトモデルはもともと文書推薦のために考案され、コンテンツデータとして文書中の各単語の出現頻度を利用していた。映画推薦にも利用されることもあったが、コンテンツとして出演する俳優名やレビュー文などの文字列に着目し、実際の映像から特徴量を抽出することはしていなかった。実際の音響信号から音響的特徴を抽出し、アスペクトモデルを適用したのは、本研究が初めてである。

以降、まず 2 章で楽曲推薦のタスクについて定義し、3 章で従来手法である協調フィルタリング手法とコンテンツベースの推薦手法をそれぞれ簡単に紹介する。次に、4 章で我々の提案するハイブリッド型の楽曲推薦手法を説明する。5 章で Amazon.co.jp から収集した楽曲評価データを用いた推薦実験について報告する。最後に、6 章でまとめとする。

2. 楽曲推薦システム

本章では、楽曲推薦システムに求められる 3 つの要件を述べ、楽曲推薦タスクについて定義する。

2.1 要件

本研究で我々が目指す楽曲推薦システムとは、以下の 3 つの要件を備えるものである。

推薦の正確さ 現実的には、楽曲データベース中の楽曲数に比べて、各ユーザによって評価が与えられる楽曲の数は極めて少ない。このような楽曲評価データを用いても、各ユーザの嗜好に合う楽曲を的確に推薦できなければならない。

アーティストのバラエティの豊富さ 推薦される楽曲には、さまざまなアーティストの楽曲が含まれていることが望ましい。なぜなら、ユーザが自分の好みに合う曲調の楽曲を演奏するアーティストを新発見することができるからである。

未評価楽曲を推薦できる能力 評価がなされていない楽曲 (新譜やマイナー曲) であっても、ユーザの嗜好によく合致すると推測されるならば、積極的に推薦することが望ましい。このような推薦はバラエティの豊かさにも繋がる。

1 章で述べたように、協調フィルタリング手法やコンテンツベースの推薦手法では、これらの要求を同時に満たすことはできない。我々は、協調的なデータ (ユーザの楽曲評価) とコンテンツデータ (音響的特徴) とをともに推薦結果に反映させることで、双方の手法の利点を組み合わせることができると考える。

2.2 タスク

本研究での楽曲推薦タスクとは、対象ユーザが評価を与えていない楽曲に対して、オススメ順にランク付けを行うことである。いま、全ユーザのインデックス

* この問題は cold-start problem, new-item problem, popularity problem などの名前がよく知られている。

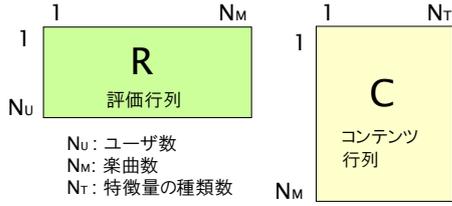


図1 観測データ：評価行列 R とコンテンツ行列 C

を $U = \{u|1, \dots, N_U\}$ とし、全楽曲のインデックスを $M = \{m|1, \dots, N_M\}$ とする。ここで、 N_U と N_M は、ユーザ数と楽曲数をそれぞれ表す。 U および M は、あらかじめシステムに登録されているものとする。その他の付加的な情報（タイトルやアーティスト名、ジャンルなど）などは楽曲推薦に利用しない。

楽曲評価データはあらかじめ登録されている必要がある。本稿では、楽曲評価は0（最低スコア：大嫌い）から4（最高スコア：大好き）の5段階評価で行うものとし、 $r_{u,m}$ ($0 \leq r_{u,m} \leq 4$) をユーザ u が楽曲 m に対して与えたスコアと定義する。ただし、ユーザ u が楽曲 m に評価を与えていない場合、 $r_{u,m}$ には「空スコア」として便宜上 ϕ を代入しておく。評価行列 R はこのようなスコアを全ユーザから集めることで得られる。

$$R = \{r_{u,m} | 1 \leq u \leq N_U, 1 \leq m \leq N_M\} \quad (1)$$

ここで注意すべきは、ユーザは M 中のごく一部の楽曲しか評価しないため、実際的なデータでは R 中のほとんどの要素が ϕ であることである。協調フィルタリング手法では、 R のみを用いて推薦を行う。

一方、コンテンツベースの推薦手法では、コンテンツデータが必要である。本研究では、 M 中の楽曲に対応する音響信号が利用可能であることを仮定している。各楽曲のコンテンツは、対応する音響信号から抽出した単一の特微量ベクトルで表現できるものとする。いま、 $T = \{t|1, \dots, N_T\}$ を特微量のインデックスとする。 N_T は特微量の種類数（特微量ベクトルの次元）を表す。さらに、 $c_{m,t}$ を楽曲 m から抽出した特微量 t の値であるとする。コンテンツ行列 C は全楽曲の特微量ベクトルを集めることで得られる。

$$C = \{c_{m,t} | 1 \leq m \leq N_M, 1 \leq t \leq N_T\} \quad (2)$$

ここで注意すべきは、対象ユーザ u に対して推薦を行う場合、 C と $\{r_{u,m} | 1 \leq m \leq N_M\}$ だけを利用すればよいことである。すなわち、 R 中の u 以外のユーザの評価スコアは参照しない。

3. 従来の楽曲推薦

本章では、協調フィルタリングとコンテンツベース推薦に関してそれぞれ典型的な手法を紹介する。これらの手法は、5章での比較実験に使用される。

3.1 協調フィルタリング

典型的な協調フィルタリング手法は、対象ユーザが評価を与えていない楽曲に対して、そのユーザが付けるであろうスコアを予測する。 $\tilde{r}_{u,m}$ をユーザ u の楽曲 m に対する評価スコアの予測値であるとする、 $\tilde{r}_{u,m}$

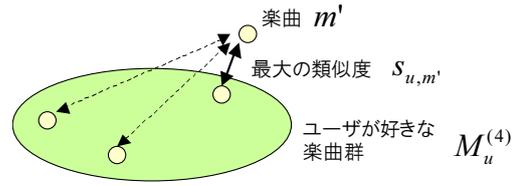


図2 コンテンツベースの推薦手法におけるユーザの嗜好と楽曲のコンテンツとの類似度の計算

は他のユーザの評価スコアを参考にすることで求まる。

$$\tilde{r}_{u,m} = \bar{r}_u + k \sum_{\{u'|u' \neq u, u' \in U\}} w_{u,u'} (r_{u',m} - \bar{r}_{u'}) \quad (3)$$

ここで、 \bar{r}_u および $\bar{r}_{u'}$ は、それぞれユーザ u, u' が付けた評価スコアの平均値である。 $w_{u,u'}$ とは、ユーザ u, u' 間の嗜好の類似度を表し、 k は $\sum_{u'} |w_{u,u'}| = 1$ とするための正規化定数である。式 (3) に従って全楽曲に対する評価スコアの予測値を求め、 $\tilde{r}_{u,m}$ の大きいものから順にランクを付けてユーザに提示する。

ユーザ間の嗜好の類似度を計算するのによく用いられる尺度の一つに、ピアソンの相関係数法がある。

$$w_{u,u'} = \frac{\sum_m (r_{u,m} - \bar{r}_u) \sum_m (r_{u',m} - \bar{r}_{u'})}{\sqrt{\sum_m (r_{u,m} - \bar{r}_u)^2 \sum_m (r_{u',m} - \bar{r}_{u'})^2}} \quad (4)$$

ここで、 \sum_m は、ユーザ u, u' がともに評価を行った楽曲についてのみ和をとることに注意する。この基本的な計算方法を用いる場合、評価行列 R が極めて疎（ほとんどの要素が ϕ ）であると推薦精度が大幅に低下することがある。

ピアソンの相関係数法に対しては種々の改良が提案されており、本稿では評価行列 R の未評価部分にはデフォルトスコア r_D を代入する手法を採用する。このとき、0 から 4 の 5 段階評価では 2 が中央値であるが、 $r_D = 2.5$ と実験的に定めた。なぜなら、3 や 4 の「好きである」という評価スコアは、0 や 1 の「嫌いである」という評価スコアよりもずっと多いため、バイアスをかける方がよい結果を生むからである。

3.2 コンテンツベース推薦

典型的なコンテンツベースの推薦手法は、ユーザの嗜好をコンテンツ空間内で表現し、それに近いコンテンツを持つ楽曲を上位にランク付けてユーザに提示する。これを行う基本的な手法として、本稿では Loganらの手法⁶⁾を紹介する。いま、楽曲 m のコンテンツベクトルを $\mathbf{c}_m = (c_{m,1}, \dots, c_{m,N_T})$ とし、ユーザ u によってスコア r ($0 \leq r \leq 4$) を与えられた楽曲の集合を $M_u^{(r)} = \{m | r_{u,m} = r\}$ とする。このとき、ユーザ u に楽曲推薦を行うアルゴリズムは以下の通りである。

(1) ユーザの嗜好ベクトルの収集

$M_u^{(4)}$ が空集合でない（ユーザ u がスコア 4 を与えた楽曲が存在する）場合、 $M_u^{(4)}$ に着目する。そうでなければ、 $M_u^{(3)}$ に着目する（両方とも空集合の場合は後述）。以降、 $M_u^{(4)}$ に着目する場合をとりあげて説明する。コンテンツベクトルの集合 $\{\mathbf{c}_m | m \in M_u^{(4)}\}$ はユーザ u の嗜好を表しており、それぞれ嗜好ベクトルと呼ぶことにする。

(2) 楽曲のコンテンツベクトルとの類似度計算

ある未評価楽曲 m' のコンテンツベクトル $\mathbf{c}_{m'}$ ($m' \in M, r_{u,m'} = \phi$) に対して, なんらかの類似度尺度を用いて, 各嗜好ベクトル \mathbf{c}_m ($m \in M_u^A$) との類似度を計算する. 計算の結果, 最大となる類似度を $s_{u,m'}$ とする. この値は, ユーザ u が楽曲 m' をどれくらい好きになりそうかを示す. $s_{u,m'}$ は未評価楽曲すべてについて計算する.

(3) 楽曲のランク付け

未評価楽曲 $\{m' | r_{u,m'} = \phi\}$ に対して, $s_{u,m'}$ が大きいものから順にランク付けを行う.

本稿では, 類似度尺度としてコサイン距離尺度を採用した. Hoashi ら⁵⁾ もコサイン距離尺度を利用している. Logan らの研究では, 各アルバム CD を 1 人のユーザが好きだと評価した楽曲群だとみなしていたので, 上記のアルゴリズムのみで問題ない.

しかし, 本研究では, ユーザ u が 0 か 1 の否定的なスコアしか与えなかった場合にも対処が必要である. このような場合, 推薦される楽曲のコンテンツベクトルと嫌いだと評価した楽曲のコンテンツベクトルとの類似度が最小化されるように上記のアルゴリズムを変形する. もし, ユーザ u が 2 という中間的なスコアしか与えていない場合, ランダムな順序で推薦する.

4. ハイブリッド型楽曲推薦

2.1 節で述べた 3 つの要件を同時に満たすため, 我々は楽曲評価データとコンテンツデータとを統合するハイブリッド型楽曲推薦手法を提案する. まず, 両方のデータを用いてユーザの嗜好をモデリングする上での課題について議論する. 次に, 確率モデルに基づく提案手法について述べる.

4.1 課題

ハイブリッド型楽曲推薦手法を実現するには, 楽曲評価データとコンテンツデータとともにユーザの嗜好のモデリングに反映させる必要がある. ここで問題となるのは, 3 章で述べたように, 協調フィルタリング手法とコンテンツベースの推薦手法ではユーザの嗜好の表現法が異なることである. 前者はユーザ u の嗜好を, N_M 次元の評価ベクトル $(r_{u,1}, \dots, r_{u,N_M})$ として表現する. 後者はユーザ u の嗜好を, N_T 次元の特徴量ベクトルの集合として表現する. これらを統合するためにアドホックな経験則を用いることはできるであろうが, 妥当性や最適性の検証が困難である.

また, 従来手法では本質的なユーザの嗜好は表現されていないため, 推薦精度に改善の余地があると考えられる. 観測可能なデータである評価スコアや音響的特徴は, ユーザの嗜好を完全には反映していないことに注意すべきである.

4.2 解決法

我々はこのような課題を解決するため, Popescul らによって提案された 3 方向アスペクトモデル¹⁰⁾ と呼ばれるベイジアンネットワークを利用する. このモデル

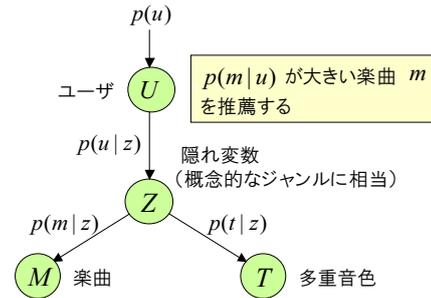


図 3 音楽的コンテンツとして多重音色を利用した 3 方向アスペクトモデルの非対称表現

は, 直接観測できない本質的なユーザの嗜好に対応する隠れ変数を備えており, この隠れ変数を介して評価データとコンテンツデータとを統合する. 隠れ変数の最尤推定値は統計的に求めることができるので, 信頼性の高い推薦に寄与すると考えられる.

しかし, Popescul らの 3 方向アスペクトモデルは文書推薦タスク用に設計されているので, 本研究での楽曲推薦タスクに直接適用できない. 彼らは, 各文書のコンテンツベクトルを言語処理分野で広く利用されている Bag-of-Words モデルに基づいて計算している. すなわち, コンテンツベクトルの各次元が辞書に登録された単語の出現頻度を表し, 総和は 1 である.

3 方向アスペクトモデルを楽曲推薦に利用するには, 各楽曲のコンテンツを, 全ての次元が等価である単一ベクトルとして表現する必要がある. 以降, これを実現する手法について述べ, 3 方向アスペクトモデルの学習方法と運用方法に関して説明する.

4.2.1 Bag-of-Timbres モデルを利用した 3 方向アスペクトモデル

上記の条件を満たすため, 我々は Bag-of-Timbres モデルを提案する. このモデルでは, 各楽曲のコンテンツベクトルが多重音色 (原語: polyphonic timbre) の重みを並べたものとして表現される. 多重音色とは個々の楽器の音色ではなく, 混合音全体の「聴こえ」を意味し, Aucouturier ら¹¹⁾ によって提案された概念である. 多重音色は楽曲の印象を特徴付ける重要な要素であるので, 我々は楽曲のコンテンツを表現する際に多重音色に着目する. また, 多重音色のモデリング法は, さまざまな音響信号に適用できる利点がある. なぜなら, 現在の技術では非常に困難である混合音中の各楽器パートの分離を必要としないからである.

本研究で用いる 3 方向アスペクトモデルを図 3 に示す. 評価スコアや楽曲のコンテンツなどの観測データは隠れ変数 $Z = \{z_1, \dots, z_{N_Z}\}$ と関連付けられる. ここで, N_Z は隠れ変数の数である. 隠れ変数の集合はユーザの嗜好を表している. すなわち, 隠れ変数それぞれが概念的にジャンルに対応し, N_Z 個のジャンルの割合が各ユーザの音楽的嗜好を反映している. あるユーザ u は自分の嗜好に従ってある確率でジャンル z を選択し, 選ばれたジャンル z が確率的に楽曲 m と多重音色 t を生起 (原語: generate) する.

上記のアスペクトモデルでは、ユーザ U 、楽曲 M 、多重音色 T の間で、隠れ変数 Z を介して条件付き独立性が成立することを仮定している。また、1人のユーザが複数のジャンルへの嗜好を持つことが許されている。これは、多くのクラスタリング手法が、各ユーザの嗜好をいずれかのジャンルクラスに割り当ててしまうのと異なる点である。

4.2.2 GMM による Bag-of-Timbres 計算

Aucouturier ら¹¹⁾ は、類似度に基づいて楽曲分類を行うため、メル周波数ケプストラム係数 (MFCC) を利用した多重音色のモデル化を考案した。MFCC は、例えばジャンル識別時に音響信号の音色をモデル化する目的でよく利用される¹²⁾。彼らはまず、各楽曲の音響信号から抽出した MFCC を混合ガウス分布 (GMM) で近似した。その後、サンプリング手法を利用して各楽曲に対応する GMM 間の距離を計算している。

本研究においても、各楽曲の Bag-of-Timbres を計算するため、対応する音響信号から抽出した MFCC ごとに GMM を構築する。形成された GMM 中の各ガウシアンが、ある多重音色の MFCC の分布を表しているとみなす。すなわち、GMM 中のガウシアン混合比が多重音色の重みに対応する。しかし、Aucouturier らの方法は本研究にそのまま適用できない。なぜなら、彼らの方法を用いると、ある楽曲の MFCC から構築された GMM と他の楽曲の GMM とでは、それらを構成するガウシアンが異なるからである。すなわち、各 GMM が別々の多重音色の組み合わせを表している。

この問題を解決するため、全楽曲の Bag-of-Timbres が同じガウシアン組み合わせを共有できる手法を提案する。まず、そのような「固定ガウシアン」の平均と分散を全楽曲から抽出した MFCC に対して GMM 推定を行うことで求めておく。この推定で求められる混合比は利用しないので捨てる。その後、各楽曲の多重音色の重みを求めるには、その楽曲の MFCC を用いて固定ガウシアン混合比のみを推定する。

具体的には、28 次元の特徴量ベクトル (13 次元 MFCC, エネルギー, それらのデルタ成分) の集合に対して GMM 推定を行った。周波数解析には、16.0 [kHz] でサンプリングされた信号に対して、窓幅 200 [ms], シフト長 100 [ms] の短時間フーリエ変換を用いた。いま、楽曲 m から抽出された 28 次元特徴量ベクトル群を $\{f_{m,i} | 1 \leq i \leq I_m\}$ とする。 I_m は特徴量ベクトルの数である。このようにして全楽曲から抽出された特徴量ベクトル群に対し、EM アルゴリズム¹³⁾ を用いて GMM 推定を行い、各ガウシアンパラメータを求める。このとき混合数を 10 とした。すなわち、 $N_T = 10$ である。

推定された t 番目のガウシアン平均と分散をそれぞれ μ_t, Σ_t とする。 $c_{m,t}$ とは楽曲 m における多重音色 t の重みであり、次式で求まる。

$$c_{m,t} = k_m \sum_{i=1}^{I_m} \frac{1}{(2\pi)^{\frac{28}{2}} |\Sigma_t|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} D^2(f_{m,i}, \mu_t)\right) \quad (5)$$

ここで、 D^2 は次式で定義されるマハラノビス距離の 2 乗を表す。

$$D^2 = (f_{m,i} - \mu_t)^T \Sigma_t^{-1} (f_{m,i} - \mu_t) \quad (6)$$

また、 k_m は $\sum_t c_{m,t} = 1$ ($m \in M$) とするための正規化定数である。

4.2.3 3 方向アスペクトモデルの定式化

本節では、3 方向アスペクトモデルの数学的定式化について説明する。 U, M, T および Z に対する同時確率 $p(u, m, t, z)$ の非対称表現 (図 3) は次式で書ける。

$$p(u, m, t, z) = p(u)p(z|u)p(m|z)p(t|z) \quad (7)$$

ここで、 $p(u)$ はユーザ u の事前確率、 $p(z|u)$ はユーザ u がジャンル z を選択する確率、 $p(m|z)$ はジャンル z から楽曲 m が生起する確率、 $p(t|z)$ はジャンル z から多重音色 t が生起する確率である。すなわち、 $p(u, m, t, z)$ とは、ユーザ u が多重音色 z を選択している状態で楽曲 m 中の多重音色 t を聴く確率である。

本研究では、 U, M および T が Z を介して独立であると仮定しているので、式 (7) は次式の対称表現に書き換えることができる。

$$p(u, m, t, z) = p(z)p(u|z)p(m|z)p(t|z) \quad (8)$$

ここで、 $p(z)$ は多重音色 z の事前確率、 $p(u|z)$ はジャンル z からユーザ u が生起する確率である。

U, M および T に対する同時確率 $p(u, m, t)$ は、式 (8) を z で周辺化することで求まる。

$$p(u, m, t) = \sum_z p(z)p(u|z)p(m|z)p(t|z) \quad (9)$$

ここで、 $p(u, m, t)$ とは、ユーザ u が楽曲 m 中の多重音色 t を聴く確率である。

本モデルにおける未知パラメータは、 $p(z), p(u|z), p(m|z)$ および $p(t|z)$ である。これらの値は、観測データである評価行列 R とコンテンツ行列 C を用いて推定しなければならない。楽曲推薦の際には、各ユーザ u に対して、 $p(m|u) \propto \sum_t p(u, m, t)$ に従って M 中の未評価楽曲をランク付けを行う。

4.2.4 EM アルゴリズムによるパラメータ推定

本節では、EM アルゴリズムを用いた未知のモデルパラメータ $p(z), p(u|z), p(m|z)$ および $p(t|z)$ の推定方法について説明する。いま、3 つ組 (u, m, t) を「ユーザ u が楽曲 m 中の多重音色 t を聴く」という事象と定義する。各事象が独立に生起するとすれば、観測データに対するパラメータの尤度 L は以下で定義できる。

$$L = \prod_{u,m,t} p(u, m, t)^{n(u,m,t)} \quad (10)$$

$$= \prod_{u,m,t} \left(\sum_z p(z)p(u|z)p(m|z)p(t|z) \right)^{n(u,m,t)} \quad (11)$$

ここで、 $n(u, m, t)$ とは、事象 (u, m, t) が生起する頻度を表す。本研究では、 $n(u, m, t)$ が次式の積で分解できると考える。

$$n(u, m, t) = r_{u,m} \times c_{m,t} \quad (12)$$

ここで、 $r_{u,m}$ と $c_{m,t}$ は以下の通り定義されている。

- $r_{u,m}$ は、ユーザ u の楽曲 m に対する評価スコアである。3.1 節で述べた通り、「空スコア」の場合はデフォルトスコアとして 2.5 が代入される。
- $c_{m,t}$ は、楽曲 m 中の多重音色 t の重みである。

式 (12) の意味は以下のように説明できる。まず、事象 (u, m, t) はユーザ u が楽曲 m を好む度合いが大きいほど頻繁に起こると考えられる。本研究では、頻度 $n(u, m, t)$ は評価スコア $r_{u,m}$ に比例すると仮定した。これには他にもいくつかの方法が考えられる。例えば、頻度 $n(u, m, t)$ により大きな傾斜を与える (例: $r_{u,m}$ の 2 乗に比例させる)、頻度 $n(u, m, t)$ が 0 にならないように評価スコアを 1 から 5 の範囲に変換しておく、これらを組み合わせる、などが考えられる。しかし、我々の実験では、上記の方法に変えて試しても性能にあまり変化が見られなかった。また、事象 (u, m, t) は楽曲 m 中の多重音色 t の重みが大きいほど頻繁に起こると考えられる。そのため、頻度 $n(u, m, t)$ は多重音色の重み $c_{m,t}$ に比例すると仮定した。

この形式の学習用データ (評価行列 R とコンテンツ行列 C) が与えられると、対数尤度 LL は次式で求まる。

$$LL = \sum_{u,m,t} n(u, m, t) \log p(u, m, t) \quad (13)$$

対数尤度 LL を最大化するパラメータを求める方法として、EM アルゴリズムが利用できる。

E ステップ

$$p(z|u, m, t) = \frac{p(z)p(u|z)p(m|z)p(t|z)}{\sum_{z'} p(z')p(u|z')p(m|z')p(t|z')} \quad (14)$$

M ステップ

$$p(u|z) \propto \sum_{m,t} n(u, m, t) p(z|u, m, t) \quad (15)$$

$$p(m|z) \propto \sum_{u,t} n(u, m, t) p(z|u, m, t) \quad (16)$$

$$p(t|z) \propto \sum_{u,m} n(u, m, t) p(z|u, m, t) \quad (17)$$

$$p(z) \propto \sum_{u,m,t} n(u, m, t) p(z|u, m, t) \quad (18)$$

E ステップと M ステップは対数尤度 LL が収束するまで繰り返す。実際には、局所解に陥りにくいように改良された EM アルゴリズム^{15),16)} を利用することが望ましい。また、本研究では $N_Z = 10$ と定めた。

5. 評価実験

提案するハイブリッド型楽曲推薦手法の有効性について検証するため、3 章で述べた協調フィルタリング手法とコンテンツベースの推薦手法との比較実験を行った。まず、楽曲評価データの収集方法と実験方法について議論する。次に、実験結果について考察する。

5.1 実験条件

信頼性の高い実験を行うには、各ユーザがある程度の数の楽曲に評価を与えている大規模楽曲データベースを用いるのが理想的である。しかし、多数の被験者

を募り、アンケートを実施して評価データを得るのは膨大な手間がかかる。また、楽曲評価は本来自発的なものであるので、請われて行われた楽曲評価は実際的な評価データと性質が異なるかもしれない。

この問題に対処する現実的な方法の一つは、評価スコアを WEB サイトから収集することである¹⁴⁾。本研究では、オンラインショッピングサイトである Amazon.co.jp から収集した。ここでは専用の API¹⁷⁾ が提供されており、5 段階の評価スコアを含む公開されているほとんど全ての情報が XML 形式で獲得できる。

実験で使用する楽曲は、2000 年 4 月から 2005 年 12 月までに、日本のシングル CD 売り上げランキングで上位 20 位にランクインしたものである。それらに与えられた評価スコアは前述の通り Amazon.co.jp から収集した。評価スコアにはユーザ ID が付与されているため、あるユーザが複数の楽曲に評価を行っていても、それらの評価は同一ユーザのものであることが分かるようになってきている。しかし、この状態のデータには、わずかな楽曲しか評価していないユーザや、ほとんど (あるいは全く) 評価するユーザがいない楽曲などが多く含まれているため、信頼性の高い実験を行う目的で適切ではない。そこで、評価数が 4 に満たないユーザと楽曲を削減することを繰り返し、評価数が 4 以上のユーザと楽曲のみを抽出することにした。最終的に、ユーザ数 $N_U = 316$ 、楽曲数 $N_M = 358$ となった。評価行列 R における評価要素 (ϕ 以外の部分) の密度は 2.19% であった。Breese らの研究³⁾ では、3 種類の推薦タスクに対する評価データが WEB サイトから収集されており、評価行列の密度がそれぞれ 1.3%, 4.7%, 2.8% であった。このことから、本研究のデータも実際的なスパースネスを持っていると言える。

こうして得られた評価データを用いて、我々が提案するアスペクトモデルに基づくハイブリッド型推薦手法 (AM と呼ぶ) を協調フィルタリング手法 (CF と呼ぶ) およびコンテンツベースの推薦手法 (CB と呼ぶ) と比較した。ただし、多重音色数 $N_T = 10$ 、隠れ変数の数 $N_Z = 10$ と定めた。

5.2 評価方法

我々が目指す楽曲推薦システムは 2.1 節で述べた通りであるので、推薦結果は推薦精度とアーティストのバラエティの 2 点で評価する。未評価楽曲も適切な推薦ができるかの検証については後述する。実験は 10 クロスバリデーションで行う。まず、図 4 に示すように、評価行列 R 中の評価要素の 10% をランダムにマスクし、学習用行列 R_l と評価用行列 R_e に分割する。AM, CF および CB の各手法は、各ユーザごとに R_l 内で未評価であるとされた楽曲をランク付けする。

推薦精度は、Hoashi ら⁵⁾ と同様に、各ユーザごとにランク付きで推薦される楽曲上位 x ($x = 1, 3, 10$) 曲、すなわち全ユーザで合計 $x \times N_U$ 曲に着目して評価する。このとき、推薦精度が高いとは、推薦ランク上位には「実際には」高い評価スコア (3 や 4) が与えら

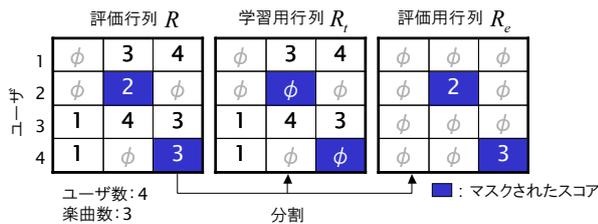


図4 評価行列 R の学習用行列 R_l と評価用行列 R_e への分割

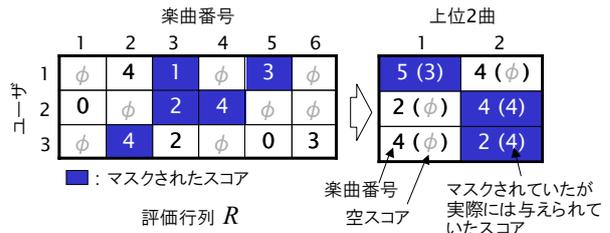
れていた楽曲の割合が高く、低い評価スコア (0 や 1) の楽曲がほとんどない状態をさす。しかし、 R_e 内にはわずかな評価要素しかない (R 内の評価要素数の 10% なので R_e の密度は 0.219%) ので、推薦される $x \times N_U$ 曲全てについて、実際に与えられていた評価スコアを参照することはできない。そこで、 $x \times N_U$ 曲の中から実際に評価スコアが与えられていた楽曲のみを抽出し、そのうち高い評価スコアが与えられている楽曲の割合を算出することにした。ここで、抽出された楽曲数を N ($N < x \times N_U$) と定義しておく。

以下に具体的な算出方法について述べる。いま、 N_r を R_e 中で実際はスコア r が与えられていたが、 R_l 中でマスクされた要素数とする。 $x = 2$ とした場合の例を図 5 に示す。 $N = \sum_r N_r$ となる。さらに、 N に対する N_r の割合を $F_r = N_r / N$ で定義する。 F_4 の値が大きいくほど、精度の高い推薦ができたことを示す。ここで注意すべきは、ランダムに楽曲を推薦した場合でも、 F_4, \dots, F_0 は 20% にはならないことである。 F_4, \dots, F_0 のチャンスレートは、 R 中の評価要素数に対する評価スコアごとの割合に等しく、それぞれ 57.9%, 19.1%, 8.6%, 4.9%, 9.5% である。すなわち、Amazon.co.jp の楽曲評価データでは、全評価スコアのうち半分以上がスコア 4 の「大好き」という評価であった。

アーティストのバラエティは各ユーザごとに評価する。各楽曲のアーティスト情報は楽曲推薦そのものには必要がなく、評価にのみ利用する。いま、対象ユーザ u に対して楽曲推薦を行う場合を考える。このとき、 v_A を推薦ランク上位 x 曲中のアーティスト数と定義する。また、 v_M を x 曲中でユーザ u によって評価を与えられたことがないアーティストの楽曲数と定義する。さらに、 V_A, V_M を全ユーザの v_A, v_M の平均値であるとする。当然 $V_A, V_M \leq x$ が成り立つ。 V_A, V_M の値が大きいくほど、バラエティ豊かな推薦ができたことを示す。

5.3 実験結果

表 1 に推薦精度を示す。 $x = 1, 10$ のときの F_4 に着目すると、AM は CF や CB よりもずっと高い値を示した。 $x = 3$ のときは、 F_4 は AM と CB とで同程度であった。ただし、CB では $x = 10$ のときの F_4 の値は、 $x = 3$ のときと比べて大幅に低下しているが、AM では同程度を保つことができていた。 F_0 や F_1 の値は、CF や CB に比べて AM の方が低い傾向が見られた。これらの結果から、AM が CF や CB よりも優れた推薦精度を達成できることが確認できた。



$$N_4 = 2 \quad N_3 = 1 \quad N_2 = N_1 = N_0 = 0 \quad N = 3$$

$$F_4 = 0.67 \quad F_3 = 0.33 \quad F_2 = F_1 = F_0 = 0$$

図5 楽曲推薦精度の算出例

ただし、図 2 に示すように、 N_4 や N の値は CF よりも AM の方がずっと小さい。これらの値は、各ユーザに推薦された上位 x 曲、全 $x \times N_U$ 曲のうち評価に利用できた楽曲数を示す。すなわち、CF では推薦された楽曲のうち、実はすでに評価スコアが与えられていた (実験のため便宜上マスクしていた) 楽曲が多いということである。この現象を説明するため、我々は多くのユーザが同じアーティストの複数の楽曲に好きだという評価をつける傾向があるため、CF ではすでに評価を与えたことのあるアーティストの楽曲が複数推薦されやすいという仮説を立てた。

この仮説を検証するには、表 3 で示されるアーティストのバラエティに着目すればよい。CF における V_A, V_M の値は、AM よりもずっと小さい。また、AM における V_A の値は CB よりも小さいが、 V_M の値は大きい。これらのことから、AM ではユーザが知らないオススメのアーティストを厳選して、そのようなアーティストの楽曲を複数推薦しやすいことが分かる。この性質は、コンテンツベース推薦と協調フィルタリングの両者の性質がうまく融合したものであり、我々の意図どおりの推薦が達成できていることを示す。

5.4 未評価楽曲を推薦する能力の検証

次に、AM が未評価楽曲を適切に推薦できるかどうかを検証するための実験を行った。この実験も 5.2 節で説明したのと同様の 10 クロスバリデーションに従うが、 M 中からランダムに 10% の楽曲を選び、それらに与えられた評価スコアを全てマスクする点が異なる。表 4 に実験結果を示す。CF では未評価楽曲を推薦することはできないため、表には載せていない。AM は音楽的コンテンツだけでなくユーザの楽曲評価も考慮するので、推薦された未評価楽曲のうち評価できた楽曲数は多くない。しかしながら、AM は CB と同様に評価がついていない楽曲でも適切に推薦できていると考えられる。

6. おわりに

本稿では、ユーザによる楽曲評価と音響信号から自動抽出した音楽的コンテンツとを同時に考慮することができるハイブリッド型楽曲推薦手法について述べた。本手法はベンジアンネットワークの一種である 3 方向アスペクトモデルに基づいたもので、直接観測できな

表1 推薦精度の評価

x	F ₄			F ₃			F ₂			F ₁			F ₀		
	CF	CB	AM	CF	CB	AM	CF	CB	AM	CF	CB	AM	CF	CB	AM
1	77.6%	85.0%	92.0%	13.8%	5.00%	4.00%	3.45%	5.00%	4.00%	0.86%	5.00%	0.00%	4.31%	0.00%	0.00%
3	77.5%	82.5%	80.3%	15.4%	12.5%	11.5%	3.08%	2.50%	6.56%	0.04%	2.50%	1.64%	3.52%	0.00%	0.00%
10	70.8%	69.4%	79.5%	18.1%	17.9%	10.6%	6.51%	4.48%	6.21%	0.10%	5.22%	1.24%	3.61%	2.99%	2.48%

表2 評価できた楽曲数

x	N ₄			N		
	CF	CB	AM	CF	CB	AM
1	90	17	23	116	20	25
3	176	33	49	227	40	61
10	294	93	128	415	134	161

表3 アーティストのパラエティの評価

x	V _A			V _M		
	CF	CB	AM	CF	CB	AM
1	1.00	1.00	1.00	0.627	0.933	0.938
3	2.49	2.93	2.80	2.09	2.76	2.78
10	7.58	9.30	8.68	8.01	9.17	9.33

表4 未評価楽曲を推薦できる能力の検証

x	F ₄		F ₃		F ₂		F ₁		F ₀		N	
	CB	AM	CB	AM								
1	89.5%	100%	5.26%	0.00%	0.00%	0.00%	5.26%	0.00%	0.00%	0.00%	19	1
3	76.9%	80.0%	12.8%	0.00%	5.12%	0.00%	5.12%	0.00%	0.00%	20.0%	39	5
10	64.1%	72.7%	19.8%	9.10%	5.34%	0.00%	6.87%	0.00%	3.81%	18.2%	131	11

い本質的なユーザの嗜好を隠れ変数の集合として表現できる。このモデルにおけるユーザ、評価およびコンテンツの間の確率的な関係は統計的に求めることができる。従来手法である協調フィルタリング手法やコンテンツベース推薦手法との比較実験の結果、提案手法が従来手法よりも推薦精度の点で優れた性能を示すことを確認した。このような高い推薦精度は、ユーザの嗜好が正確にモデル化できたおかげであると考えられる。また、提案手法による推薦の性質として、ユーザが知らないオススメのアーティストを厳選し、そのアーティストの楽曲を複数推薦しやすいことが分かった。

今後、音楽的コンテンツとして、テンポやリズムなどにも着目する予定である。これらは音響信号から自動的に解析できることが望ましい。また、事前にシステムに登録されていない楽曲やユーザの追加に対応するために、アスペクトモデルをインクリメンタルに追加学習する方法を検討していく。

本稿で述べた楽曲推薦手法の応用例としては、音楽ファン向けのソーシャルネットワークサービスが考えられる。このようなサービスで重要なのは、ユーザ推薦を行う点である。我々手法を利用すればユーザの嗜好を正確に推定することができるので、多数のユーザの中から嗜好の近いユーザを推薦し、コミュニティ形成を促進させることができると考えている。

謝辞 本研究は、科研費 (A) No.15200015, 日本学術振興会特別研究員 (DC1) 科研費の補助を受けた。

参考文献

- 1) A. Uitendbogerd and R. van Schyndel, "A review of factors affecting music recommender success," *ISMIR*, 2002.
- 2) U. Shardanand and P. Maes, "Social information filtering: Algorithms for automating "Word of Mouth"," *ACM CHI'95 Conference on Human Factors in Computing Systems*, 1995, pp. 210–217.
- 3) J. Breese, D. Heckerman, and C. Kadie, "Empirical analy-

sis of predictive algorithms for collaborative filtering," *UAI*, 1998, pp. 43–52.

- 4) W. Cohen and W. Fan, "Web-collaborative filtering: Recommending music by crawling the Web," *Computer Networks*, vol. 33, no. 1–6, pp.685–698, 2000.
- 5) K. Hoashi, K. Matsumoto, and N. Inoue, "Personalization of user profiles for content-based music retrieval based on relevance feedback," *ACM Multimedia*, 2003, pp.110–119.
- 6) B. Logan, "Music recommendation from song sets," *ISMIR*, 2004, pp. 425–428.
- 7) O. Celma, M. Ramirez, and P. Herrera, "Foafing the music: A music recommendation system based on RSS feeds and user preferences," in *ISMIR*, 2005, pp.464–457.
- 8) P. Melville, R. Mooney, and R. Nagarajan, "Content-boosted collaborative filtering," *SIGIR*, 2001.
- 9) C. Hayes, "Smart Radio: Building Community-Based Internet Music Radio." *Doctoral Thesis*, Trinity College Dublin, 2003.
- 10) A. Popescul, L. Ungar, D. Pennock, and S. Lawrence, "Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments," *UAI*, 2001, pp. 437–444.
- 11) J.-J. Aucouturier, F. Pachet, and M. Sandler, "The way it sounds": Timbre models for analysis and retrieval of music signals," *IEEE Trans. Multimedia*, vol. 7, no. 6, pp. 1028–1035, 2005.
- 12) J.-J. Aucouturier and F. Pachet, "Musical genre: A survey," *New Music Research*, vol. 32, no. 1, pp. 83–93, 2003.
- 13) A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Royal Statistical Society, B*, vol. 39, pp. 1–38, 1977.
- 14) M. Zadel and I. Fujinaga, "Web services for music information retrieval," *ISMIR*, 2004, pp.478–483.
- 15) T. Hofmann, "Probabilistic latent semantic analysis," *UAI*, 1999, pp.289–296.
- 16) N. Ueda and R. Nakano, "Deterministic annealing EM algorithm," *Neural Networks*, Vol.11, No. 2, pp.271–282, 1998.
- 17) Amazon Web Services:
http://www.amazon.com/gp/aws/landing.html.