

決定木に基づく楽曲の印象推定モデルの分類とその評価

Classification of Estimation Model of Music Emotion Based on Decision Tree And Its Evaluation

岩月 靖典[†]
Yasunori Iwatsuki

西尾 圭一郎[†]
Keiichiro Nishio

酒向 慎司[†]
Shinji Sako

北村 正[†]
Tadashi Kitamura

1. はじめに

近年、インターネットを介した楽曲の配信が発達し、音楽販売の主流となりつつある。多くの楽曲配信サービスでは、古い曲から最新の曲まで膨大な数の曲が保管されているため、利用者が目的の曲を瞬時に検索できるよう、楽曲を分類する技術が期待されている。

楽曲の分類、検索のアプローチとしては、アーティスト名やアルバム名といった楽曲情報を利用したキーワード検索が一般的である。その他にもメロディやリズムを入力としたコンテンツベースの方法も研究され、実用化されているものもある。しかし、これらの検索方法を利用するには、ユーザが所望の楽曲の断片的な情報をあらかじめ得ている必要があり、検索範囲が限定的になる。

一方で、より直感的な検索の手法として、人が楽曲を聴いたときに感じる印象を手がかりとした検索手法も有望であり、研究が進められている。先行研究 [1] では、感性に個人差があることを考慮し、個人に対してのみ印象推定を行い、有効性が確認されている。しかしながら、個人別に学習し、楽曲の印象を得る方法では、未知の人物に対応することが困難である。そこで、被験者の付加情報（プロフィール）に着目し、類似した印象傾向を持つグループに分類することで、未知の利用者にも適応可能な印象推定モデルを提案する。

本研究では、感性の個人性と共通性の双方を考慮するために、被験者のプロフィールにより印象推定モデルを木構造的に自動分類し、その有効性を評価する。

2. 感性モデル

2.1 GMM に基づいた感性モデル

人間が抱く心理的特徴と音響信号から得られる物理的特徴には何らかの相関が存在し、この相関関係を利用して、音響信号から楽曲の印象に変換することを考える。これらの特徴の複雑な対応関係を学習的に獲得する手法として、高精度な声質変換の手法の一つである混合正規分布 (Gaussian Mixture Model: GMM) に基づいた最尤特徴量変換 [2] を用いる。この変換は、特徴量間の変換関数を確率モデルとして記述できる。この変換方法の概略を図 1 に示す。学習楽曲から得られた物理量 X と心理量 Y を結合した Z から GMM のパラメータ λ を ML 基準で推定する。そして、入力楽曲の物理量 X' と学習した GMM のパラメータ λ より、入力楽曲の心理量 Y' を推定する。

2.2 特徴量抽出

人間が感じる楽曲の印象と音響信号の特徴を表す特徴量として、物理量と心理量を抽出する。

物理量としては、強弱、リズム、調性、音色を抽出し、各フレームの平均と分散を特徴量として一曲あたり 60 次元のベクトルで表現する。

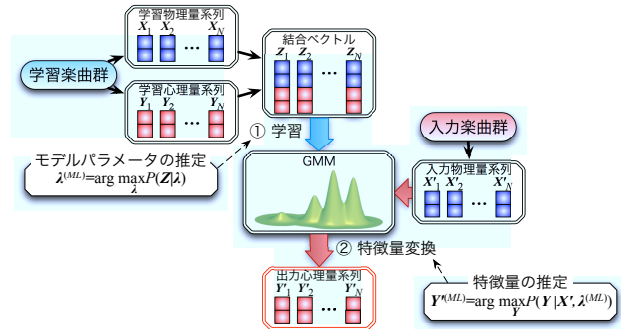


図 1: GMM に基づいたパラメータ変換の概略図

表 1: 使用した感性語対

| | | | | | |
|-------|---|---------|-------|---|------|
| 軽い | - | 重い | 明るい | - | 暗い |
| しみりした | - | うきうきした | 迫力のある | - | 静かな |
| 穏やかな | - | 激しい | 陰気な | - | 陽気な |
| のびやかな | - | 抑えたような | 速い | - | 遅い |
| 優雅な | - | 荒々しい | 寂しい | - | 賑やかな |
| 慌しい | - | のんびりとした | 重厚な | - | 軽快な |
| 安らぐ | - | 緊張した | 華やかな | - | 素朴な |

心理量としては、表 1 に示した感性語対を SD 法を用いた聴取実験を行い、14 対の印象を 7 段階の尺度で数値化した。したがって、心理量は一曲あたり 14 次元のベクトルで表現する。

2.3 感性の個人性と共通性

感性には個人差があり、楽曲聴取時に感じる印象も個人により異なる。2 節で示した手法に基づいて個人毎にモデルを作成した場合、モデルの数によるデータ量の肥大化・学習データ数の確保、未知利用者への対応という三つの問題点がある。

また、同様の手法で、複数人物の心理量を用いることにより、平均的な印象推定が可能なモデル (平均モデル) となる。この場合は、共通性を考慮する反面、個人性を考慮することができないという問題点がある。

本研究では、類似した印象傾向を持つ人物で分類することにより、個人性と共通性の双方を考慮したモデルを考える。過去に聴いた音楽や、音楽に対する知識が、音楽の認知に影響を与えているという知見 [3] に基づいて、音楽に関する知識や楽器の経験等の個人のプロフィールを利用することで、類似した印象傾向を分類する手法を提案する。

2.4 木構造による分類

プロフィールによる分類を考えた場合、どのプロフィールが印象傾向の違いに影響しているのかが不明である上に、プロフィールの要素数に応じて組み合わせの数が増大する。この問題を解決するために、決定木に基づいたクラスタリング [4] を用いて適切なモデルを選択できるモデル生成法を試みる。決定木のルートノードは、個人性を考慮しない全被験者から学習された GMM (平均モデル) とし、プロフィール分類の質問によりクラス

[†]名古屋工業大学, Nagoya Institute of Technology

表 2: 被験者のプロフィール

| 性別 | 男性: 10 名 | 女性: 10 名 |
|------------|----------|----------|
| 楽器の経験 | 有り: 7 名 | 無し: 13 名 |
| クラシック音楽の知識 | 有り: 10 名 | 無し: 10 名 |

タを分割する。質問は二者択一のものとするため、木の構造は二分木となる。分類の基準には対数尤度を用いて、最も尤度が高くなる分類で分割する。すなわち、 q をある分類を行うための質問とし、 S を分類前の状態のデータ、 $S_y(q)$ と $S_n(q)$ を q で Yes と No に分けたときの状態のデータ、 $L(S)$ をモデルに対する S の対数尤度としたとき、式 (1) の ΔL_q が最大となる q を基準にクラスタの分割を作成する。 ΔL_q が閾値を下回ったときにクラスタの分割の作成を終了する。

$$\Delta L_q = L(S_y(q)) + L(S_n(q)) - L(S) \quad (1)$$

3. 実験

3.1 実験条件

心理量は、20～60 代の男女 20 名の被験者に対し、一定の印象を持つ 10 秒間のクラシック音楽 200 曲を評価させ、14 次元の特徴量を取得した。被験者のプロフィールとその内訳を表 2 に示す。

物理量は前述した内容を、音響信号から各種特徴量が抽出可能な MIRToolbox を用いて同じ 200 曲から 60 次元の特徴量を抽出した。分析条件は文献 [5] と同じである。これら 200 曲のうち 180 曲を学習に使用し、残りの 20 曲の未知曲を評価に用いた。GMM の混合数は 8 とした。

比較対象としては、個人モデル、平均モデル、プロフィールを用いたモデルの結果の比較を行う。プロフィールを用いたモデルでは、三つのプロフィールの分類を単純に全てを用いた場合のモデルと木構造で分類されたモデルを作成した。また、得られた特徴量を全て学習データに用いたモデルと、個人モデルに合わせて学習データを 180 曲に統一したモデルを作成した。詳細な条件は下記の通りである。

- 平均モデル (Average): 被験者 20 名の平均モデル
- プロフィールモデル (Profile): 三つのプロフィールを用いたときに 3 名以上となるモデル
- 木構造で分類されたモデル (Tree): 20 名の被験者から木を作成したモデル
- 個人モデル (Personal): 被験者 20 名の各人のモデル

結果の評価法として、一つの推定値と全未知データとのユークリッド距離を算出し、その距離が近い順に順位付けを行い、本来の楽曲が 5 位以内評価された楽曲数の割合を調査した。

3.2 印象推定実験

プロフィール分類の有効性を確認するために、それぞれのモデルにおける正解率を比較した。結果を図 2 に示す。何れの結果も、三つのプロフィールを用いたときに 3 名以上となるモデルに分類される 11 名の結果の平均であり、エラーバーはその標準偏差である。木構造で分類されたモデルにおいて、使用したモデルの階層の平均は、クローズでは 3.46、オープンでは 3.25 となった。

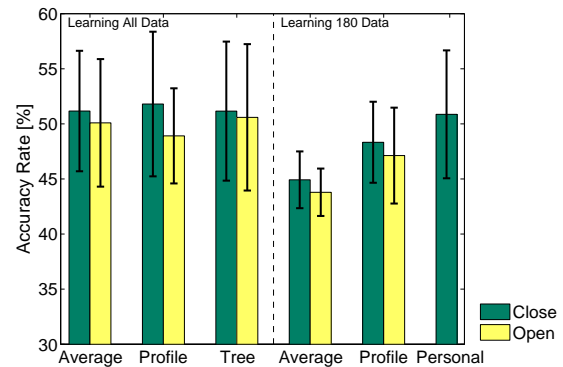


図 2: 印象推定評価実験の結果

全ての特徴量を学習に用いた未知の被験者に対する結果は、木構造で分類されたモデルが最も正解率が高いが、平均モデルもほぼ同等となり、プロフィールモデルが少し低い正解率となった。しかし、学習データ量を揃えた場合にはプロフィールモデルは平均モデルよりも高い正解率となっている。

これら結果から、学習データ量が正解率に大きく影響していることが分かる。したがって、学習データ量が十分な量となれば、プロフィールを用いたモデルは平均モデルより正解率が向上するものと予測できる。また、木構造で分類されたモデルは、学習データ量が平均モデルの半分に満たないにもかかわらず同等以上の正解率が得られ、11 名の被験者のうち 7 名で平均モデルの正解率を上回る結果となった。このことから、木構造により適切なモデルが選択されたといえる。

4. むすび

本研究では、未知利用者に適応可能な楽曲の印象推定を可能にするために、印象推定モデルを個人性を表したプロフィールによって木構造的に分類する手法を提案した。被験者 20 名の評価実験で、木構造で分類したモデルは未知被験者に対して最も高い正解率となり、個人モデルと同等の精度となった。現状では、学習データ量を増加させるほど正解率が高くなる傾向があるが、木構造で分類されたモデルの正解率が、すべてのデータを用いた平均モデルと同等以上となったことから、木構造を用いることで少ないデータから効率的に学習できる可能性を示した。今後の課題としては、より詳細な比較を行うために、被験者とプロフィールの増加があげられる。

参考文献

- [1] 平江遼, 西隆司. 感性に基づくクラシック音楽の分類. 日本音響学会誌, Vol. 64, No. 10, pp. 607–615, 2008.
- [2] 戸田智基. 最尤特徴量間変換法とその応用. 電子情報通信学会技術研究報告, Vol. 105, No. 571, pp. 49–54, 2006.
- [3] R.E Radocy. Psychological Foundations of Musical Behavior. Charles C Thomas Publisher, 1979.
- [4] S. J. Young, J. J. Odell, and P. C. Woodland. Tree-Based State Tying for High Accuracy Acoustic Modeling. *Proceedings of the workshop on Human Language Technology*, pp. 307–312, 1994.
- [5] 西尾圭一郎, 酒向慎司, 北村正. 個人性を考慮した GMM に基づいた音楽の印象推定モデルに関する研究. HCG シンポジウム 2010 論文集, pp. 281–286, 2010.