6R-8

音響特徴量を用いた楽曲印象分布の推定

絵本 詩織 † 糸山 克寿 ‡ 奥乃 博 ‡

† 京都大学 工学部情報学科 ‡ 京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

近年，歌手，楽曲名，声質などの様々なクエリにより楽曲を検索するシステムが開発されている [1]．たとえば YouTube では，一つの動画を検索した時に，楽曲名や歌手名などの情報を基に関連楽曲を推薦するシステムが利用されている．しかし，楽曲名や歌手名を知らない場合，楽曲を聴いた人間が最初に知覚するのは，音やリズムによる楽曲の雰囲気 (本稿では，印象と表現する) であるため，楽曲の印象に基づく類似楽曲検索の実現が望ましい．

そこで本稿では，楽曲の印象を音響特徴量から推定する手法について述べる．さらに，本手法を応用することで，印象に基づく楽曲検索システムを実装した．本システムが実現することで，ユーザーがそのときの気分によって聴きたい曲を検索することや，ユーザーの聴いた楽曲の印象に類似する楽曲を選出して提示することなどが可能となる．

既存の楽曲印象推定の研究は，音響特徴量を用いたものと歌詞を用いたものに大別される．前者として，Support Vector Machine [2]，混合ガウスモデル [3]，判別分析法 [4] などのクラス判別機によって印象別に楽曲を分類する研究や，回帰分析を使用して楽曲の音響特徴量から楽曲印象を推定する研究 [5] などがあげられる．後者として，歌詞による特徴量ベクトルと楽曲印象ラベルを利用した k 近傍法により楽曲を分類する研究 [6]，印象表現語を利用して歌詞内容から印象を推定する研究 [7] などがあげられる．さらに，西川らは音響特徴量と歌詞の両方を用いて楽曲印象推定を行った [8]．西川らの研究では，音響特徴量と歌詞の印象曲線を別々に推定している．

上記のように，楽曲印象推定に関して多くの研究がなされているが，上記であげた手法はどれも一つの楽曲に対して一つの印象のみを推定する手法である．しかし，楽曲から受ける印象は個人によってばらつきがあり，一意には定まらないものであると考えられるため，楽曲印象を分布として推定することで，個人が楽曲から受ける印象のばらつきを表現する．本研究では，音響特徴量を用いた楽曲の印象分布推定と，本手法の類似楽曲検索への応用を目的とする．

本稿の楽曲印象分布の推定は，楽曲のフレーズごとに行われる．従来研究では，楽曲ごとに一つの印象があるとして楽曲を印象別に分類する研究 [2]- [4] が存在したが，実際の楽曲印象は明らかに楽曲内で時間推移と共に変化している．本研究では，フレーズごとの印象分布の推移を考えることで，楽曲印象の時間推移を表現する．また，本稿では，楽曲に対する印象はフレーズ内では一定であり，V-A(Valence(快-不快)-Arousal(興奮-弛緩) 軸) 座標平面上にマッピングされるものとし [8]，音響特徴量のばらつきが個人が楽曲から受ける印象のばらつきに関連すると仮定している．

Musical Mood Distribution Estimation by Using Acoustic Features: *Shiori Emoto, Katsutoshi Itoyama, and Hiroshi G. Okuno (Kyoto Univ.)*

2. 楽曲印象分布の推定及び類似楽曲検索への応用

2.1 音響特徴量に基づく印象推定モデル

本研究では，音響特徴量とあらかじめ収集した楽曲の印象を用いて重回帰モデルを学習することで，未知の楽曲の音響特徴量を入力とし，V-A 座標平面上の座標（以下，印象座標と表現する）を推定する．

楽曲から短区間フレームごとに音響特徴量ベクトルを抽出する．抽出する音響特徴量ベクトルは各短区間フレームの MFCC・クロマベクトル・スペクトルフラックス・スペクトル重心・ゼロクロッシングの平均と分散の合計 110 次元である．

学習データとして楽曲の各フレーズごとに V-A 座標平面上にマッピングされた印象データを取得する．この印象データは，被験者実験によって，実際に被験者が楽曲を聴き，その印象を V-A 座標平面上にマッピングすることで取得される．ここでは，各楽曲に対して複数人の印象データを取得した．取得された印象データを用いて，重回帰モデル

$$P = aX + b$$

（P：取得された V-A 座標平面上の座標，X：音響特徴量ベクトル）のパラメータ a, b を推定する．

推定されたパラメータ a, b を用いて，未知の楽曲に対する印象座標を推定する．まず，各時間フレームにおける音響特徴量の値を X に代入することで、フレームごとの印象座標 P を推定する．その後，得られたフレームごとの印象座標 P をフレーズごとにまとめることで，印象を分布として表現する．本分布を 2 次元ガウス分布

$$\frac{1}{2\pi\sqrt{|S|}} \exp(-\frac{1}{2}(x-\mu)^T S^{-1}(x-\mu))$$

(S: 分散共分散行列，μ: 平均) にフィッティングする．このフレーズごとの 2 次元ガウス分布の時系列を，音響特徴量に基づいて推定されたその楽曲全体の時間推移する印象として用いる．

2.2 類似楽曲検索

上記の楽曲印象推定法を，類似楽曲検索システムに応用する．本システムは，ユーザーの付した楽曲印象を検索キーとして用いることで，検索キーと類似する印象の楽曲を選出するものである．

まず，楽曲に対してユーザーが付した印象座標の時系列データを，入力としてシステムに与える．システムは，楽曲の各フレーズの印象に対する本入力の尤度を計算し，この尤度を類似度として用いてマッチングを行う．この類似度の計算は，長さの異なる時系列データの類似度が計算可能な Dynamic Time Warping(DTW) を用いて行う．以上の手法により，検索キーとして用いた本入力と近い印象分布の時系列データを持つ楽曲が求められる．

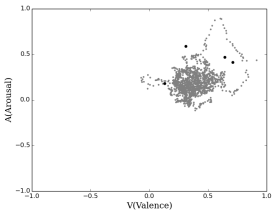

図 1: RM-P007.wav の 2 フレーズ目の印象座標の推定結果

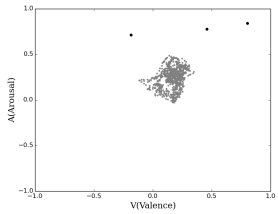

図 2: RM-P041.wav の 16 フレーズ目の印象座標の推定結果

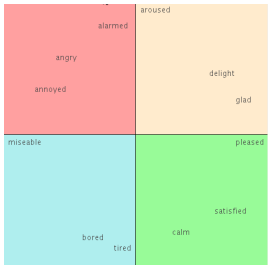

図 3: 印象座標入力用 GUI

2.3 上記手法の実装

本稿では，学習データとして用いる各楽曲の印象データは，被験者実験により取得される．これは，被験者が，プルダウンボックスの中から楽曲を選んで視聴し，V-A 座標平面を表現したウィンドウ上にマウスでクリックすることで印象をマッピングすることができて，マッピングされた印象が csv ファイルとして保存される GUI によって行った (図 3)．さらに，被験者実験によって取得した印象データを用いて，上記手法によって，フレームごとの印象座標 P を求めた．その後，印象座標をフレーズごとにまとめることで，各楽曲のフレーズにおける印象座標の分布を求める．さらに，本分布を二次元ガウス分布にフィッティングすることで定式化する．2 次元ガウス分布の平均，分散共分散行列をフレーズごとの印象の分布の時系列データとして保存することで，類似楽曲検索システムへの応用を可能にする．

上記の実験により推定された楽曲印象の二次元ガウス分布を用いて，類似楽曲検索システムを実装した．楽曲に対してユーザーが付した V-A 座標平面上の座標の時系列データを入力としてシステムに与える．システム内部では，DTW を用いて，15 曲の各楽曲に対して，各フレーズの 2 次元ガウス分布によって表された印象に対する本入力の尤度が計算される．計算された各フレーズの印象に対する本入力の尤度を類似度として用いて，各フレーズと本入力の類似度の楽曲全体での合計を求めることで，各楽曲に対するマッチングを求める．以上の手法により，15 曲の中から，楽曲に対するマッチングの類似度の合計の上位 3 位以内となる楽曲を，検索結果として提示する．

3. 評価実験

本実験では，学習データとして，RWC 研究用音楽データベース : ポピュラー音楽 [10] の中の 15 曲と，被験者実験により取得したこれらの楽曲の印象データを用いた．この 15 曲の楽曲は，ジャンル，歌手の男女の別などが重複しないように選出したものである．音響特徴量ベクトルは MARSYAS [9] を用いて抽出する．また，評価には，学習データに用いた楽曲の印象データを再び用いた．

提案した類似楽曲検索システムの性能評価を行う．性能比較のため，既存の類似楽曲検索システムを実装し，提案手法と比較した．

既存手法では類似度として，2 次元ガウス分布に対する V-A 平面座標の尤度の代わりに，分布の平均と V-A 平面座標とのユークリッド距離を用いた．本ユークリッド距離の楽曲全体での合計が下位 3 位以内となる楽曲を，検索結果とした．提案手法と既存手法の，類似楽曲検索性能を比較する．学習データとして用いた楽曲 15 曲の印象データを，再度，入力としてシステムに与える．入力として与えた印象データの該当楽曲が上位 3 位以内として検索結果に提示された場合を正解であるとした．

各手法の正解率は，提案手法が 50.0 %，既存手法が 66.0 %であった．楽曲別では，15 曲のうち，提案手法の正解率が高い曲が 2 曲，既存手法の正解率が高い曲が 6 曲，正解率が同じである曲が 7 曲であった．ここで，楽曲の印象座標の推定結果のうち，提案手法で被験者のマッピングしたデータとマッチしたものを図 1 に，マッチしなかったものを図 2 に示した．図 1，図 2 の点については，黒色が各被験者がマッピングしたデータ，灰色が提案手法による推定結果である．提案法による検索は，既存手法による検索よりも推定精度が落ちた．これは，音響特徴量のばらつきだけではなく，他の要素を考慮する必要があることを示唆している．

4. おわりに

本稿では，音響特徴量に基づく楽曲印象分布の推定法について述べ，本手法を応用した類似楽曲検索システムの実装を行った．評価実験の結果，提案法による類似楽曲検索は既存手法による類似楽曲検索よりも推定精度が落ちる結果となった．今後は，その原因を探ると共に精度の向上に努めたい．なお，本研究の一部は科研費 No.24220006 (S) の支援を受けた．

参考文献

[1] 藤原弘将, 後藤真孝 : VocalFinder:声質の類似度に基づく楽曲検索システム, *SIGMUS, 2007(81)*,pp. 27-32(2007)

[2] T. Li and M. Pgihara : Detecting Emotion in Music, *Proc. ISMIR*,pp. 239-240(2003)

[3] L. Lu, D. Liu and H. Zhang : Automatic Mood Detection and Tracking of Music Audio Signals, *IEEE Transaction un Audio, Speech, and Lang. Process*, Vol. 14, No. 1, pp. 5-18.(2006)

[4] J. Skowronek, M. McKinney, and S. van de Par : A Demonstrator for Automatic Music Mood Estimation, *Proc. ISMIR*, pp. 345-346(2007)

[5] T. Eerola, O. Lartillot, and P. Toiviainen : Prediction of Multidimensional Emotional Ratings in Music from Audio using Muitivariate Regression Models, *Proc. ISMIR*, pp. 621-626(2009)

[6] M. V. Zaanen and P. Kanters : Automatic Mood Classification Using TF*IDF based on Lylics, *Proc. ISMIR*, pp. 75-80(2010)

[7] Y. Hu, X. Chen and D.Yang : Lyric-based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method, *Proc. ISMIR*, pp. 123-128(2009)

[8] 西川直毅, 糸山克寿, 藤原弘将, 後藤真孝, 尾形哲也, 奥乃博 : 歌詞と音響特徴量を用いた楽曲印象軌跡推定法の設計と評価, *SIGMUS*, Vol. 2011-MUS-91, No. 7, pp. 1-8(2011)

[9] Tzanetakis, G. and Cook, P.: MARSYAS: A Framework for Audio Analysis, *Organised Sound*, Vol. 4, No. 3, pp. 169-175 (2000)

[10] 後藤真孝, 橘口博樹, 西村拓一, 岡隆一: ”RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース”, 情報処理学会論文誌, Vol.45, No.3, pp.728-738 (2004)