

ジャムセッションシステムに向けた演奏特徴量の統計解析*

保利 武志[†] 中村 和幸[†] 嵯峨山 茂樹[†]

(明治大学先端数理科学研究科)[†]

1 はじめに

本研究の目的は、人間の演奏者と計算機とがインタラクティブに即興演奏を行うジャムセッションシステムを実現することである。本稿ではその前段階として、ジャズピアノトリオによる MIDI 演奏データから抽出できる物理的特徴量(以下、特徴量)を用いて、演奏意図に対応した特徴量を持つ演奏を推定するための手法について述べる(図 1)。

ジャズセッションでは即興演奏を基本とした演奏者同士のインタープレイが行われる。このとき、1 人の演奏者は他の演奏者の音を聞き、その演奏意図を解釈して適切な演奏を返すことにより協調演奏が成立する。特に即興を主体とするセッションでは、リード譜やコード、直前までの演奏に応じて遷移する演奏意図によって様々な演奏が展開される。

従来のジャムセッションシステムでは、パラメータやルール群をヒューリスティックに、あるいは特定の被験者に対する心理実験結果を用いて設定することで、演奏意図を推定し演奏を予測、生成しているが、パラメータの設計に被験者の主観が混入してしまうため、その主観評価を排除することは課題の 1 つであった^{[1][2]}。

そこで本稿では、演奏の特徴量のみから演奏の遷移を追跡することで特徴量の予測を行った。ただし、演奏意図と特徴量是一对一対応ではなく、ある種の自由度を持ち、またこれらの遷移は離散的なものである。従って、演奏はシンボリックな遷移によって表現されるものと仮定し、拍単位、小節単位それぞれについて抽出した特徴量に対して、 k -means, GMM, NMF を用いたクラスタリングを行い、クラスラベルを付与して、その遷移を追跡した。なお、クラスの遷移確率は N -gram を用いて求め、交差検定法による予測精度の比較評価を行った。

2 演奏の物理的特徴量

2.1 抽出範囲

人間が演奏を決定する際、直前の音からスム

Statistical analysis of music rendering features toward jam session systems

[†]Takeshi HORI, Kazuyuki NAKAMURA, Shigeki SAGAYAMA (Meiji University)

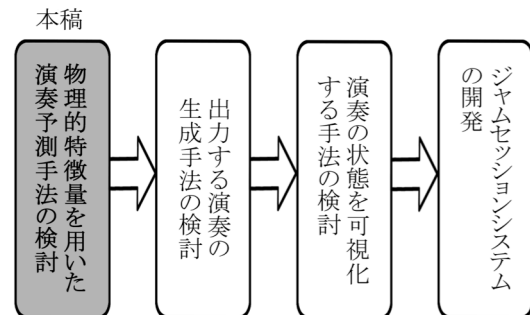


図 1 ジャムセッションシステムに向けた研究手順

ーズに次へつなげるための局所的な演奏と、より広い範囲でフレーズを構成するための大局的な演奏が同時に意識される。本稿ではこれらの局所的意図と大局的意図を追跡するために、分析に用いた曲に対して、それぞれ 1 小節単位と 1 拍単位について特徴量を抽出した。

2.2 抽出した特徴量

特徴量には Piano, Bass, Drums の各楽器について、それぞれ音高やベロシティ、テンションノート数等を抽出し、また前拍や前小節、あるいは総数に対する比率も用いた。特徴量の数は、小節単位、拍単位でそれぞれ 68 個である。これらの特徴量について、Yamaha MusicDatashop のピアノトリオ 15 曲分から抽出を行った。

3 クラスタリング

3.1 k -means

拍または小節単位で並べた特徴量ベクトルに対して、それぞれ各クラスタの重心ベクトルとユークリッド距離において最も近いクラスタを採用した。

3.2 ガウス混合モデル(Gaussian Mixture Model, GMM)

GMM は複数の正規分布の重ね合わせにより表されるモデルであり、EM アルゴリズムによってフィッティングが行われる。正規化した特徴量ベクトルについて、対数尤度関数の変化量が閾値 0.001 以下になるまで繰り返し処理を行

い、最も事後確率の高いクラスタを採用した。

3.3 非負値行列因子分解 (Nonnegative Matrix Factorization, NMF)

NMF は非負値行列 X を、同様に非負値行列である H と U の行列積によって近似する手法であり、特徴量ベクトルを並べた行列 X に対し、以下の式で表される。[3]

$$X_{i,j} \approx \sum_k H_{i,k} U_{k,j} \quad (1)$$

ただし、 k は基底に対するインデックスである。 H は演奏の特徴を表す基底行列であり、 U はそのアクティビティを与える行列である。 H と U は非負値性に基づく乗法更新により計算される。なお、基底のスケール任意性を除去するため、各基底の L1 ノルムを 1 にそろえた。また、対応するクラスラベルは行列 U に対し、

$$\arg \max_k U_{k,j}$$

であるものを付与した。

図 2 は小節単位で特徴量を抽出し、クラスタ数 25 として NMF を行ったときの基底行列のヒートマップである。他と比べて少数の特徴量のみが特に色濃く現れるクラスを見ると、例えば音域幅やテンションノート数、ブルーノート数やその変化量等がその特徴量として現れている。これらは実際のジャズセッションでも、曲調の変化に応じて顕著に変化が見られる特徴である。

4 評価実験

k -means GMM, NMF でそれぞれ得られたクラスタに対し、拍クラスタ遷移、小節クラスタ遷移、また、各小節クラスタにおける拍クラスタ遷移を追跡した小節・拍クラスタ混合遷移の 3 パターンについて、tri-gram による遷移確率を求め、交差検定法による評価実験を行った。

図 3 にクラスタ数を変化させた時の、それぞれのクラスタリング手法による正答率を示す。 k -means と GMM を用いた手法では有意な差は得られなかったが、NMF を用いた手法では他の手法に比べて比較的高い精度を維持しており、特に小節・拍クラスタ混合遷移による評価においては高いスコアを示した。

5 おわりに

本稿では実際の演奏データから特徴量を抽出し、クラスタリングを用いて予測する手法を提案した。 k -means, GMM, NMF の 3 つの手法につ

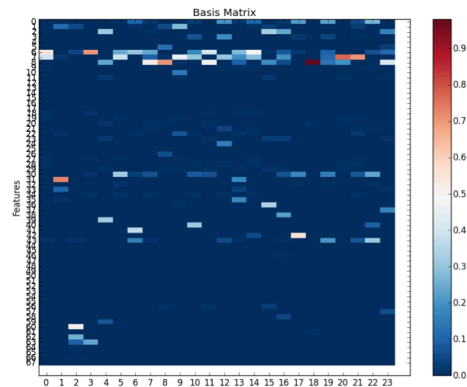


図 2 基底行列のヒートマップ (横軸:クラスラベル, 縦軸:特徴量ラベル)

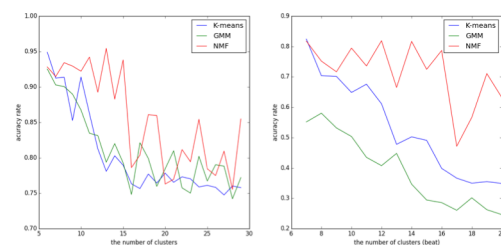


図 3 各クラスタリング手法による正答率 (左図:拍クラスタ遷移の正答率, 右図:小節クラスタ数 9 固定時の拍クラスタ遷移正答率)

いて比較検討を行った結果、NMF によって特に高精度な予測結果が得られ、また、その基底行列から実際に演奏を特徴付ける特徴量が得られることを示した。

今後は個々の楽器による演奏が他に与える影響や、より広いスパンで見た演奏の変化を反映することによって、さらなる予測精度の向上を図るとともに、得られた各クラスの特徴量から、適切な演奏を生成する手法について検討する。

参考文献

[1] 後藤真孝, ほか: 仮想ジャズセッションシステム: VirJa Session, 情報処理学会論文誌, Vol.40, No.4, pp.1910-1921, 1999.

[2] 浜中雅俊, ほか: Guitarist Simulator: 演奏者の振舞いを統計的に学習するジャムセッションシステム, 情報処理学会論文, Vol.45, No.3, pp.698-709, 2004.

[3] D. D. Lee, et al.: Algorithms for non-negative matrix factorization, *Advances in neural information processing systems*, Vol.13, pp.556-562, 2001.