

HMMに基づくジャズセッションシステムにおける ベース・ドラムスの演奏に関する検討*

保利 武志[†] 中村 和幸[†] 嗟峨山 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、統計的に学習可能な枠組みにおいて、人間の演奏者と計算機とがインタラクティブにセッションできるシステムの実現である。本稿ではその数理モデルの実証を目的として、ジャズスタンダードを対象とし、入力としてピアノの MIDI 演奏データを与えると適切なベース・ドラムスの演奏が編集・生成・合成されたピアノトリオの MIDI データを出力するシステムの実現である。

従来のジャズセッションシステム[1-2]では演奏意図を判別するためのパラメータの閾値が経験則や心理実験等を用いてヒューリスティックに決められていたが、我々はこれまで演奏データのみから統計的に学習可能な枠組みで実現するために、演奏をあらゆる音楽に関する事象を内包する空間（音楽事象空間）におけるトラジェクトリで表現できるという仮説に基づき、セッションをそのトラジェクトリの絡みとしてモデル化した。そのモデルに基づき、ジャズセッションの演奏追跡に有効な特徴量選別による次元削減とクラスタリングによる空間量子化・軌跡離散化を用いて、各奏者トラジェクトリの時系列特性を *trigram* によるクラス遷移、トラジェクトリ間の絡みをクラス間の共起関係とした、事例データを利用したジャズピアノトリオ編集合成システムを提案した[3]。本稿ではこれをさらに発展させ、量子化及び離散化を連続型 HMM によって行い、トラジェクトリの時系列特性及び各奏者のトラジェクトリ間の共起関係を DNN でモデル化したジャズピアノトリオ編集合成システムを提案する。

2 HMMに基づく数理モデル

2.1 演奏の数理モデル

我々はこれまで演奏を音楽事象空間におけるトラジェクトリとしてモデル化してきたが、学習データ量に起因するスパースネス問題に対応するため、特徴量選別 (Style parameter) による次

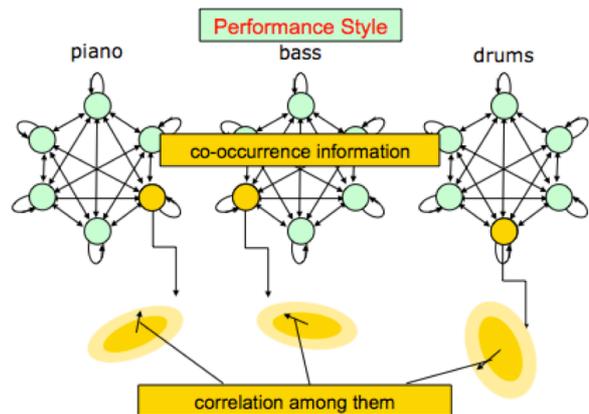


図1 HMMに基づくジャズセッションの数理モデル

元削減及びクラスタリングによる空間量子化・軌跡離散化を提案してきた。これまでの研究では 1 小節単位で Style parameter に対する Non-negative matrix factorization (NMF) によるクラスタリングを用いたトラジェクトリの離散化により、与えられたクラス遷移に対し *trigram* で時系列特性を近似し、クラス間の共起確率を用いてインタラク션을表現した。これに対し、本稿で提案するシステムは、クラスタリング時に時系列特性を同時に考慮できるように、連続混合 Hidden Markov Model (HMM) を用いた。これにより最適なトラジェクトリは Viterbi 経路として近似される。また、1 小節ごとの演奏状態を表す HMM の各状態を我々は 'Performance style' として定義した (図 1)。

各奏者のトラジェクトリは 1 小節単位で、HMM 及び Viterbi アルゴリズムによる状態が割り当てられるが、トラジェクトリの表現力は近似された HMM の状態数に依存する。しかし、状態数を増やすと十分な学習データが得られない場合スパース性が問題となるため、本システムではこれまでの研究で得られた結果から、NMF によるクラスタリングにおいて各楽器の *trigram* による演奏推定精度 80%程度を保持したクラス数 15 の場合において、全クラス中 80%以上を占めた 4 状態に限定し、GMM 混合数を 4 とし学習を行った。

Jazz session system based on HMM: Rendering of bass and drums

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

2.2 LSTM によるベース・ドラムスの状態推定

これまで trigram と共起確率で個別に考えてきた時系列特性と共起関係を同時にモデル化するために、これらを同時に学習するための手法として、本システムでは Long Short-Term Memory (LSTM) を用いた。LSTM は音声や映像データなど系列データに対する再帰的なニューラルネットワーク(RNN)としてよく知られており、長期の文脈を学習可能な NN として近年高い精度の研究結果が報告されている[4]。本稿では入力をピアノの時系列 Style parameter, 出力をベース及びドラムスの状態とした識別器として学習を行った。

2.3 Deep Belief Network (DBN) によるベース・ドラムスの特徴量の偏移推定

本システムでは少ない状態数設定による HMM の表現力の低下を補完するために、各状態間における重心からの偏移と入力特徴量との相関を DBN でモデル化した。DBN は各層ごとに Restricted Boltzmann machine (RBM) で学習した重みを初期値として、識別層を追加した fine-tuning による高精度な DNN としてよく知られている。これまでの研究結果から、クラスタリングされた局所空間で関係性を見る有用性が示されたため、本システムではベース及びドラムスの各状態における重心からの偏移を GMM でクラスタリングして偏移クラスを求め、ピアノの入力特徴量に対するベース・ドラムスの識別学習を行った。GMM のクラス数は後述のリズムパターンの分離状況を見て 8 とした。

3 ベース・ドラムスの演奏生成

ジャズセッションを行う際、多くの場合は定番のパターンを用いるか、調とコードからスケールを推定し、記憶しているリズムパターンと合成することで新たな演奏を生成し、多様な即興演奏を可能とする。本システムでもその生成過程に基づき、ベースの音高と、ベース・ドラムスのリズムパターンを個別にモデル化し、合成することによる演奏の生成を行なった。ベースの音高は多くの場合大きく変化することはなく、ルート音を中心にコードによる制約に従って選択される。本システムでは時刻 $t-1$ の音を中心とした正規分布に従い、アヴェイラブルノートスケール内の音を選択する。ベース及びドラムスのリズムパターンは、十六分音符単位で量子化した学習データから、決定木によるパターンの生成確率を学習データから計算し生成した。

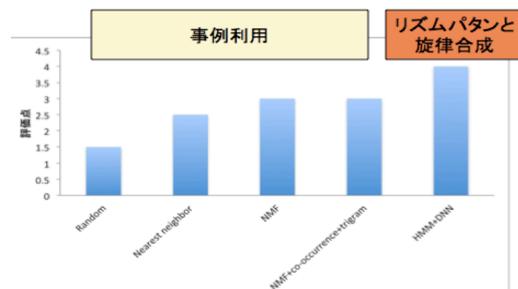


図2 本稿で提案するシステムによる予備聴取実験結果 ('Autumn Leaves')

4 評価実験

本システムを用いて生成した演奏データ ('Autumn Leaves') に対する予備的な聴取実験を行った。その結果、これまでの事例を用いたシステムよりも自然な演奏であったとの評価が得られた (図 2)。

5 おわりに

本稿では HMM に基づくジャズセッションの数理モデルを示し、その実現に向けて LSTM 及び DBN を用いた作曲生成型のジャズピアノトリオ編集合成システムを提案した。予備聴取評価実験の結果、従来のシステムに比べてより自然な楽曲が生成されていたとの評価を得た。

今後は HMM や LSTM, DBN のパラメータと生成される楽曲の依存性を調査し、より適したパラメータについて検討し、その上でよりヒューリスティックを排除するために、ノンパラメトリックなモデル化も目指す。また、MIDI による入出力ではなく音響信号を用いた実装を検討し、リアルタイムな入力に対応するジャズセッションシステムの実現を目指す。

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] 保利武志, ほか, “Bass と Drums の時系列を考慮した事例ベース MIDI 編集合成によるジャズピアノトリオの生成” *日本音響学会 2016 年秋季研究発表会*, 2016.
- [4] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, Vol.9, No.8, pp.1735-1780, 1997.