ベイジアンネットによる即興音楽生成システム -少ない曲例からのモデル推定-

The Improvisational Music Generation System using Bayesian Network -Model Estimation with Small Samples-

G-022

藤本 悠 †
Yu Fujimoto

村田 昇 †
Noboru Murata

1. まえがき

本研究はジャズにおける即興演奏の生成系の確率的な構造を推定することを目指している．音楽の生成系の確率構造を記述するためには和音構成，フレーズパターン，リズムパターンなどの情報が互いにどのような依存関係にあるのかを表現する必要がある．本研究では奏者の演奏例からこれらの情報を抜き出し，離散変数として扱うことでその確率構造の表現を試みている．ここで各変数は膨大な状態数から構成されるため，確率表を実際の曲例から得るためには相応の量のサンプルが必要となる．

本稿では充分なサンプル数が得られない状況において確率表の推定を行なうための 2 つの工夫について述べる．1 つはサンプル数の調整に関するもので，少ないサンプルから確率表を推定する際に各変数状態の位相を想定することでこれを実現している．もう 1 つはモデルのパラメタ数の調整に関するもので，状態数が多いような変数群を考える際にパラメタ数の少ない周辺確率の積で表せるような混合モデルを用いることでこれを実現している．

これらの工夫を行なうことによって確率表が自然な構造を実現し得るか実験を行なったので報告する．

2. モデルの構築と問題点

2.1 ベイジアンネットの構築

物事の因果関係や推論過程を確率的に解釈，表現する際にベイジアンネットは有用である．本研究で扱うような離散変数群をベイジアンネットでモデル化する際には各変数間の条件付独立性などを論じることで因果関係を表記することが可能となる．

2.2 サンプル数の問題

ベイジアンネットを用いて離散変数間の関係を表記する際には，有向の辺によって結ばれたノード間の条件付確率を表として用意する必要がある．あるノード A における条件付確率 $\mathrm{Pr}(A|A$ の親ノード群$)$ を与えるためには各変数が取り得る状態数の積の大きさの表を用意する必要がある．

確率表をサンプルから推定するためには，場合によるが表のパラメタの数 P と比較してサンプル数 S が 5-10 倍程度必要だとされている．S/P が 5 に満たないような場合は，サンプルの生起頻度に忠実な確率表を用意しても真の確率分布からかけ離れた物になることが多いのが推定時に問題となる．

3. 提案手法

S/P が小さい場合，通常は事前知識を基に変数状態の縮約が行なわれる．これは縮約を行なうことで P を減らし，S/P を仮想的に増加させていることに相当する．

†早稲田大学理工学研究科

音楽のフレーズ群や和音群を名義尺度の離散変数として確率表を構成する場合，各変数の状態数の多さから確率表のパラメタ数は多量になる．しかし適切にフレーズ群や和音群を縮約するのは変数状態の厳密な構造が未知であるため困難である．そのため，縮約をせずに S/P を増加させる手法があれば望ましい．

この問題に対処するためには S を相対的に大きくするか，P を相対的に小さくすることが必要である．以下ではそれぞれ S の増加，P の減少を目的とした 2 種類の手法を提案する．

3.1 近傍データの利用による S の増加

機械の制御系の学習などにおいて，データ空間の位相を考慮する SOM を応用することで学習速度，汎化効果などが改善されることが報告されている [4]．この位相という考え方を確率表に導入することで S を増加させ，S/P の改善を行なうことを提案する．

確率表を推定する基になる度数表について考える．各セルの近傍を定義して近傍内の度数 n_{ij} について和をとることで，位相導入後の度数表におけるセルの値は増加する．度数表に位相を導入するイメージを表 1 に示す．

表 1: 元の度数表 (左) と位相導入後の度数表 (右)

n_{11}	n_{12}	n_{13}	n_{14}	n_{15}
n_{21}	n_{22}	n_{23}	n_{24}	n_{25}
n_{31}	n_{32}	n_{33}	n_{34}	n_{35}
n_{41}	n_{42}	n_{43}	n_{44}	n_{45}

$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\sum_{i=1}^{3}\sum_{j=1}^{3} n_{ij}$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\sum_{i=2}^{4}\sum_{j=3}^{5} n_{ij}$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

近傍にあるデータを数え上げることで各セルのデータ数が増加し S/P も増加する．近傍は度数表上のセルの隣接関係を反映する必要はないため，事前に近傍に関する知見があれば名義尺度の離散変数に対して適用することができる．近傍が適切であれば，位相導入後の度数表を基に構築した確率表は真の確率分布に近づくように改善される．

3.2 混合確率表表現による P の減少

I，J 個の状態を持つ 2 変数 X，Y の同時確率表におけるモデルのパラメタ数を減らす時，通常次のような独立モデルを考える．これは X，Y がそれぞれ状態 x_i，y_j である同時確率 $\mathrm{Pr}(x_i, y_j)$ をそれぞれの周辺確率 $\mathrm{Pr}(x_i)$，$\mathrm{Pr}(y_j)$ を用いて次のように表現することになる．

$$\mathrm{Pr}(x_i, y_j) = \mathrm{Pr}(x_i) \times \mathrm{Pr}(y_j) \tag{1}$$

しかし 2 変数間の独立性が疑わしい時には式 (1) のような縮約を行なうことはできない．このような時には変数同士が独立ではなく，かつパラメタ数が少ないモデルが望ましい．そこで式 (1) のモデルを重ね合わせることでこれを実現し，S/P の改善を行なうことを提案する．

式 (1) のような周辺確率の積で表現される確率表を K 枚重ね合わせることで，元の同時確率表は

$$\Pr(x_i, y_j) = \sum_{k=1}^{K} a_k \overbrace{\Pr{}^k(x_i)\Pr{}^k(y_j)}^{k\text{ 枚目の独立な表}} \qquad (2)$$

と表現することができる．ここで a_k は正規化のために導入したパラメタで，$\sum_{k=1}^{K} a_k = 1$，$a_k > 0$ $(k = 1, \ldots, K)$ とする．すると重ね合わせた同時確率表が，$\sum_{i,j} \Pr(x_i, y_j) = 1$ を満たすようになる．各パラメタは EM アルゴリズムで求めることができる．

K 枚の確率表を重ねることでパラメタ数 P_{mix} は $(I + J - 1)K - 1$ となる．また式 (2) において K を増やすことで真の確率分布に近づくように改善される．

ベイジアンネットではパラメタ数の少ないモデルとして naive Bayes [2] や noisy-OR [3] などが考案されている．本提案手法は，これらのモデルで仮定される制約を緩和しつつ縮約を試みた表現だと解釈することができる．

4. 実験結果

実際の曲例から 73 種類の和音と 869 種類のフレーズで構成される 22362 個のサンプルを得た．このサンプルに対して 2 つの提案手法を用いて S/P を改善しながらモデル推定を行なう実験を行なった．

4.1 近傍データの利用

和音の構成音に基づいて位相を導入し，パラメタ数 $P_{\mathrm{sat}} = 73 \times 869 - 1 = 63436$ の飽和モデルの精度を 100-fold cross validation で得た平均対数尤度で評価する実験を行なった．ここで考えている和音に対して $+1$(構成音が 1 音多い)，-1(構成音が 1 音少ない)，±1(構成音の違いが 1 音以内)，±2(構成音の違いが 2 音以内) であるような和音を位相として用意した．以上の 4 種類の位相を ±0(位相未定義) の場合と比較するのがこの実験の目的である．その際 sampling zero が存在するため，度数表の全セルに 0.5 を加えて同時確率表を作成している．結果を表 2 に示す．

表 2: S/P_{sat} と平均対数尤度の推移

位相	Samples	S/P_{sat}	平均対数尤度
±0	22,362	0.353	-8.185
$+1$	52,918	0.834	-8.174
-1	65,011	1.025	-8.263
±1	95,567	1.507	-8.160
±2	283,325	4.466	-8.393

表 2 より，位相を動かして S/P_{sat} を変化させると，近傍の適切さに応じて平均対数尤度は増減をすることが分かる．ここでは構成音が 1 音の増減で表せるような和音を近傍とすることで，元の確率表よりも尤度の高い確率表の推定が可能なことが確認できる．

4.2 混合確率表による表現

位相を設定することで S/P は多少増加させることができたが，飽和モデルをそのままモデルとして採用するには S/P が不充分である．そこで周辺確率の積で表される確率表を重ねることでパラメタ数 $P_{\mathrm{mix}} < P_{\mathrm{sat}}$ を実現し，元の確率表を近似する実験を行なった．重ねる表の数を変化させながら AIC の推移を見た結果を図 1 に示す．なお，図 1 では AIC の値のプロットと共に S/P_{mix} の概数を付記している．

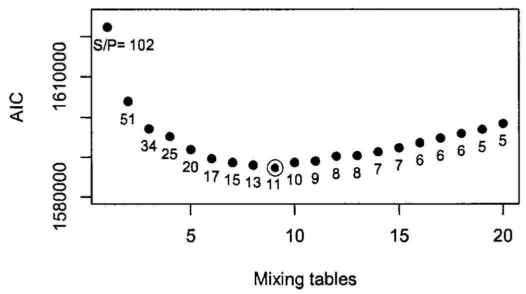

図 1: 重ね合わせる表の数と AIC の関係

図 1 から，周辺確率の積で表せるような確率表を 9 枚重ねることで最良の AIC が得られることが確認できる．なお 9 枚の表を重ねた時の S/P_{mix} は 11.286 となり，充分な S/P が得られていると言える．

5. まとめ

本稿では音楽の生成系の構造をベイジアンネットを用いてモデル化する際にデータとなる演奏例の少なさや変数の状態数の多さを問題視し，これらの影響を回避する手段として変数状態の位相の導入やパラメタの少ない混合モデルを用いた確率表の表現などを提案した．これらの工夫を行なうことでサンプル数とモデルのパラメタ数の比を操作し，より適切なモデルの取捨選択を行なえることが実験的に確認できた．

今後は限られた曲例から最大の効果を挙げられるような位相の探索を続けながら，聴取実験などを交えることで生成音楽の持つ音楽性について論じていきたいと考えている．

参考文献

[1] A. Agresti. *Categorical Data Analysis*. Wiley, 2002.

[2] N. Friedman and M. Goldszmidt. Building classifiers using bayesian networks. In *Proceedings AAAI-96 Thirteenth National Conference on Artificial Intelligence*, 1996.

[3] M. Pradhan, G. Provan, B. Middleton, and M. Henrion. Knowledge engineering for large belief networks. In *Proceedings of the 1994 Conference on Uncertainty in AI*, 1994.

[4] H. J. Ritter, T. M. Martinets, and K. J. Schulten. Topology-conserving maps for learning visuo-motor-coordination. *Neural Networks*, Vol. 2, pp. 159–168, 1989.