

氏 名 今村 武史

学位(専攻分野) 博士(学術)

学位記番号 総研大甲第 1996 号

学位授与の日付 平成 30 年 3 月 23 日

学位授与の要件 複合科学研究科 統計科学専攻
学位規則第6条第1項該当

学位論文題目 ガウス過程を利用したピアノ演奏の自動採譜に関する研究

論文審査委員 主 査 教授 福水 健次
教授 松井 知子
准教授 南 和宏
准教授 持橋 大地
名誉教授 鹿野 清宏 奈良先端科学技術大
学院大学

論文の要旨

Summary (Abstract) of doctoral thesis contents

インターネット等を通じて大量の音楽データを扱うことが可能な時代となり、これらの音楽データから様々な特徴を引き出して楽曲の検索や推薦を行う技術が求められている。このようなニーズと、機械学習の諸技術の発展を背景とし、音楽データを科学的に扱う音楽情報処理の研究が広範に行われている。楽曲の音響信号のみを与え、これを演奏内容の表記へと変換する自動採譜は音楽情報処理の最も基本的な要素技術の一つである。自動採譜はこれまで様々な手法が提案されてきたが、現在に至るまで確定的な手法は見出されておらず、新たな手法の開発による推定精度の向上が求められている。本研究はピアノ演奏を対象にガウス過程およびその応用手法を用いた新たな自動採譜手法の提案を行い、採譜精度の向上を図ることを目的に行った。提出論文は以下の7章より成る。

第1章は研究の背景として、本研究の意義と自動採譜の研究の歴史について述べている。

第2章では、本論文で扱う音楽の音響データおよび演奏データについての説明を行う。

第3章はガウス過程による自動採譜に先立ち、識別的な手法である **Support Vector Machine** による採譜手法について記している。各鍵盤に対して識別器を割り当てて **one vs all** 識別機を構成し、それぞれの鍵盤の押鍵の有無を推定している。

第4章ではガウス過程回帰による自動採譜手法について記している。識別的な手法では、対象の鍵盤が押鍵されているのか、いないのかといった2値の結果のみが得られ、打鍵の強さは推定できない。この点は、音響データから押鍵の有無だけではなく、演奏のアーティキュレーションも抽出しようとする場合には大きな制約となる。これらを解決するために、ガウス過程回帰によって楽曲の音響データから各鍵盤の打鍵の強さ（MIDI 信号のベロシティ値）を推定している。推定結果には、非線形フィルタの一種である **rank order filter** による後処理を施して最終的な採譜結果を得ている。採譜実験においては、学習データ数や後処理手法に対する採譜性能を評価している。

第5章では **Shared Gaussian Process Latent Variable Model** による自動採譜手法について記している。従来の自動採譜の手法は、結果である演奏音から、原因である演奏内容を推定する逆問題という構図で考えられることが多かった。しかし電子楽器の演奏時のように、演奏音と共に MIDI 信号などの形で演奏データが得られる場合には、演奏データも演奏音同様に観測変数と捉えることができる。本章では演奏音の音響データと、MIDI 信号より得られる演奏データが、共通の情報源より発生した異なる形式を持つ観測変数であると捉え、これらの観測変数が共通の潜在変数を共有するモデルを **Shared Gaussian Process Latent Variable Model** により構築している。未知楽曲の音響データに対する演奏データの推定は欠損した観測データの推定として行われる。本章では、**Shared Gaussian Process Latent Variable Model** による演奏データの推定と、**rank order filter** による後処理の組み合わせにより、自動採譜アルゴリズムを構築している。第4章と同様に学習データ数、後処理手法に加え、潜在変数の次数が採譜性能の与える影響についても検証している。

第6章では **Online Gaussian Process Dynamical Model** による自動採譜について記している。音は時間と共に変化する情報であり、自動採譜においても演奏音の時間的な連続性

(別紙様式 2)
(Separate Form 2)

を考慮する必要があるとの指摘はこれまでも行われてきた。これらの指摘に対し、本章では第 5 章で導入した音響データ、演奏データに対する潜在変数に時間的な依存性を持たせることを考える。まず時間的な依存性を持たせた潜在変数を状態変数とし、音響データ、演奏データ、状態変数の関係を非線形状態空間モデルで表す。非線形状態空間モデルを構成するシステムモデル、観測モデルは **Sparse Online Gaussian Process** で実現する。この際、平均および分散をそれぞれ別の回帰モデルによって推定する異分散モデルを構成する。この非線形状態空間モデルに対して、非線形な状態推定手法である **Cubature Kalman Filter** を適用し、学習用楽曲の音響データ、演奏データを与えて状態推定を行いながら、**Sparse Online Gaussian Process** によってシステムモデル、観測モデルをオンライン学習する。また未知楽曲の音響データのみが与えられた際に、それに対応する演奏データを推定する方法についても併せて提案する。

第 7 章は全体のまとめとして、本研究の成果および今後の課題について述べる。本研究における採譜結果と、先行研究における採譜結果の比較もここで行っている。

Summary of the results of the doctoral thesis screening

2018年1月22日午後3時から約1時間40分にわたり、今村武史氏の博士論文審査委員会を開催した。申請者による1時間の公開発表による概要説明と質疑応答、さらに約40分間の審査員のみによる審査と口述試験を踏まえた上での審議の結果、以下に述べる理由により、審査委員会は申請論文が学位授与の水準に達していると全員一致で判断した。

[論文の概要]

申請論文は、ガウス過程を利用したピアノ演奏の自動採譜に関する研究をまとめたものであり、全7章97ページからなる。

1章では、研究の背景(自動採譜の研究の歴史を含む)と本研究の意義が述べられている。

2章では、本論文で扱う音楽の音響データおよび演奏データについて説明されている。

3章はガウス過程による自動採譜に先立ち、識別的な手法であるSupport Vector Machineによる採譜手法について記している。各鍵盤に対して一対他識別器を構成し、それぞれの鍵盤の押鍵の有無を推定している。

4章では、ガウス過程回帰による自動採譜手法について記している。識別的な手法では、対象の鍵盤が押鍵されているか否かの2値の結果のみが得られ、打鍵の強さは推定できない。この点は、音響データから演奏のアーティキュレーションも抽出しようとする場合に大きな制約となる。これを解決するため、ガウス過程回帰によって楽曲の音響データから各鍵盤の打鍵の強さ(MIDI信号のベロシティ値)を推定している。推定結果には、rank order filterによる後処理を施して最終的な採譜結果を得ている。採譜実験においては、学習データ数や後処理手法に対する採譜性能を評価している。

5章では、Shared Gaussian Process Latent Variable Model (SGPLVM)による自動採譜手法について記している。従来の自動採譜の手法は、結果である演奏音から、原因である演奏内容を推定する逆問題として考えられることが多かった。しかし電子楽器の演奏時のように、演奏音と共にMIDI信号などの形で演奏データが得られる場合には、演奏データも演奏音同様に観測変数と捉えることができる。本章では演奏音の音響データと、MIDI信号より得られる演奏データが、共通の情報源より発生した異なる形式を持つ観測変数であると捉え、これらの観測変数が共通の潜在変数を共有するモデルをSGPLVMにより構築している。未知楽曲の音響データに対する演奏データの推定は欠損した観測データの推定として行われる。SGPLVMによる演奏データの推定とrank order filterによる後処理の組み合わせにより、自動採譜アルゴリズムを構築し、その検証を行っている。

6章では、Online Gaussian Process Dynamical Modelによる自動採譜について記している。自動採譜において演奏音の時間的な連続性を考慮するために、5章で導入した音響データ、演奏データに対する潜在変数に時間的な依存性を持つ状態変数とし、音響データ、演奏データ、状態変数の関係を、Sparse Online Gaussian Processによる非線形状態空間モデルで実現している。この際、平均および分散をそれぞれ別の回帰モデルによって推定する異分散モデルを構成している。この非線形状態空間モデルに対して、非線形な状態推定手法であるCubature Kalman Filterを適用し、オンライン学習を行っている。また、未知

(別紙様式 3)

(Separate Form 3)

楽曲の音響データのみが与えられた際に、それに対応する演奏データを推定する方法についても提案している。

7章は全体のまとめとして、本研究の成果および今後の課題について述べている。

[論文の評価]

音楽情報処理は近年の産業応用の発展により重要度が高まっている研究分野であり、本論文の扱う自動採譜はその一つの大きな研究課題である。4章で提案されている方法は、押鍵の有無だけでなく打鍵の強さを推定している点と、rank order filterによる後処理によって信号形状が大きく改善されている点から、有効性の高い提案となっている。5章で提案している方法は、演奏音の音響データとMIDI信号より得られる演奏データの両方に共通する潜在的表現を求めており、音響データから演奏データへの写像を推定しようとしていた従来のアプローチとは異なる方向性を提案している。また従来法と比較して採譜精度のF値が向上していることも確認されており、意義の大きい方法の提案となっている。6章で提案している方法は、非線形状態空間モデルを用いて時間的な連続性を自動採譜の問題に取り入れた点で新規性が高く、時間発展を有する潜在変数モデルの一般的な推定法としても将来性のある方法と考えられる。

なお、4章の内容をまとめた論文が査読付き学術雑誌、電子情報通信学会論文誌Dに採録されている。