

## 科学研究費助成事業 研究成果報告書

平成 28 年 9 月 19 日現在

機関番号：62603

研究種目：基盤研究(B) (一般)

研究期間：2013～2015

課題番号：25280067

研究課題名(和文) 時空間ガウス過程モデルによる音データの判別予測に関する研究

研究課題名(英文) Study on predictive discrimination of sound data based on spatial-temporal Gaussian processes

研究代表者

松井 知子 (Matsui, Tomoko)

統計数理研究所・モデリング研究系・教授

研究者番号：10370090

交付決定額(研究期間全体)：(直接経費) 13,000,000円

研究成果の概要(和文)：時空間ガウス過程モデルによる分類や回帰を行うための汎用ソフトウェア、Monte Carlo Dynamic Classifier (MCDC) ツールを開発した。MCDCでは時空間ガウス過程モデルとして、状態と観測関数をガウス過程で表した状態空間モデルを考える。また時空間ガウス過程モデルによる音響空間のモデル化について、音波の位相を考慮したカーネル関数として波動方程式に基づくカーネル関数を新たに設計した。実験的に従来のガウスカーネルと比べて高いISDR値を示すことを確認した。さらに時空間ガウス過程モデルを音楽ジャンル分類、音楽ムード推定に適用し、従来法と比べて、高い性能が得られることを確認した。

研究成果の概要(英文)：We newly developed a general-purpose software “Monte Carlo Dynamic Classifier (MCDC) tool” for classification and regression with spatial-temporal Gaussian processes. In MCDC tool, state-space models are utilized and the state and observation functions are designed with Gaussian processes.

For modeling of acoustic space with spatial-temporal Gaussian processes, we designed a novel kernel function based on the wave equation in which the wave phase information was taken into account. The effectiveness of the kernel function was experimentally shown through comparison with the conventional method. Moreover, we investigated new methods for music genre classification and music emotion recognition using spatial-temporal Gaussian processes and experimentally showed that the proposed methods outperformed the conventional methods.

研究分野：統計的機械学習、音情報処理

キーワード：ガウス過程 時空間モデリング 音響信号処理 音楽情報処理

## 1. 研究開始当初の背景

(1) ビッグデータ時代を迎え、大量かつ多様なデータをいかに使いこなして、イノベーションや生産性向上に結びつけるかが重要な課題となっている。その処理の難しさは、データが多種多様であること、雑音を含むこと、様々な時空間粒度で時々刻々と観測されること、大量であることなどにある。本研究では機械学習の応用研究を展開する研究代表者の松井、音響、音楽情報処理に関して優れた研究成果を創出する研究分担者の武田、Markov がタッグを組み、それぞれのスキルを発展させる形で、音響や音楽などの音データを例として取り上げ、雑音を含み、時間的かつ空間的に変動する、複雑なデータの処理の要素技術の確立をねらう。特に、ガウス過程[1]を対象とする機械学習理論の専門家たち（海外の研究協力者の Peters と Nevat）も巻き込んで、基本的にカーネル関数を利用したガウス過程を用いて、回帰や分類の問題をモデル化することにより、データの複雑性を扱っていく。ガウス過程を用いたモデル化では、回帰や分類を表す関数の値がガウス過程に従うと仮定する。観測雑音や関数のパラメータの揺らぎを確率的に扱うので頑健な推定が可能となる。また、カーネル関数を利用することにより、データの時空間を高次元（無限次元）空間に写像して、元の時空間での複雑な局所関係を、高次元空間における線形の関係として上手く扱う。

(2) 音響データは、ハンズフリー電話や対話型ロボットなどのシステムの研究開発において対象とされる。それらのシステムではマイ

クロホンで収録される音声を利用するが、収音音声はその室内空間に依存し、時間的に変動する残響や雑音などの音響的なひずみを含むことが多い。この音響的ひずみは音声の明瞭性を低下させ、音声認識の性能を大きく劣化させるので、いかに収音音声から音響的ひずみの影響を軽減するかが重要な課題となる。

(3) 音楽データについては一般に時系列データとして扱われ、数十ミリ秒単位で短時間フーリエ変換などを用いて分析されるスペクトル特徴量ベクトル、もしくは数百ミリ秒単位で抽出されるクロマ特徴量の時系列として表される。近年、研究代表者の松井と研究分担者の Markov は、音楽データを時空間データとして扱う方法について検討し、ジャンル分類においてその有効性を確認している。この方法では音楽データを数秒単位で定 Q 変換によるスペクトル画像の時系列（～時空間データ）として表現する。数秒単位のスペクトル画像には、従来の数十、数百ミリ秒単位の分析では捉えられない、音楽の動的特徴が表現されている。

## 2. 研究の目的

(1) 本研究では時空間構造を持つビッグデータの判別予測技術の高度化に関わる基礎数理とその応用の研究を行う。音響や音楽などの音データを取り上げ、音データの回帰・分類問題をガウス過程を用いてモデル化して処理する技術の確立することを目的とする。

(2) 音響データについては、カーネル関数を利用したガウス過程を用いて、音響的ひずみの影響を室内の空間全体にわたり、より詳細に統計的にモデル化することを目指す。

(3) 音楽データについては、ガウス過程を用いて、この時空間表現された音楽データをより精密にモデル化することを目指す。

### 3. 研究の方法

(1) 本研究では基礎数理班、音響応用班、音楽応用班の三つの班を設置して実施する。音響、音楽応用班ではそれぞれに、カーネル関数を利用したガウス過程を主軸に用いた音響、音楽の事例研究を推進する。基礎数理班では、研究統括を行うとともに、海外の研究協力者との連携をはかりながら、事例研究から得られる知見に基づいて、時空間データを扱うための汎用的なガウス過程に基づく技術を洗練していく。

(2) 基礎数理班では、時空間ガウス過程モデルによる汎用的な分類・回帰手法を研究開発する。

(3) 音響応用班では、基礎数理班で検討する時空間ガウス過程モデルを、音響伝達特性の推定問題に適用し、推定精度の向上を目指す。

(3) 基礎数理班で検討する時空間ガウス過程モデルを、音楽ジャンル分類や音楽ムード推定の問題に適用し、推定精度の向上を目指す。

### 4. 研究成果

(1) 基礎数理班の主な研究成果は、時空間ガウス過程モデルによる分類や回帰を行うための汎用ソフトウェア、Monte Carlo Dynamic Classifier (MCDC) ツールである。MCDC では時空間ガウス過程モデルとして、状態と観測関数をガウス過程で表した状態空間モデルを考える。状態と観測関数のガウス過程のパラメータ推定のために、パス空間モンテカルロ法 (Path-Space Monte Carlo; PSMC)、粒子

モンテカルロ法 (Particle Markov Chain Monte Carlo; PMCMC) の二つのアルゴリズムを開発した。

(2) 音響応用班では、時空間ガウス過程モデルを用いた音響空間のモデル化について検討を行い、音波の位相を考慮したカーネル関数として波動方程式に基づくカーネル関数 (波動カーネル) を新たに設計した。定常的な音波を用いて波動カーネルとガウスカーネルによる音圧時系列の推定性能を比較する実験を行い、波動カーネルはガウスカーネルと比べて、高い SDR (signal-to-deviation ratio) 値を示し、波動カーネルの有効性を確認した。

(3) 音楽応用班では、時空間ガウス過程モデルを音楽ジャンル分類に適用し、従来の SVM (support vector machine) による方法と比べて、高い性能が得られることを確認した。また同モデルを音楽ムード推定に適用し、従来法と同等以上の性能が得られることを示した。さらに、音声や音楽データに潜在する動的な感情の認識システムの構築を行った。

(4) STM (International Workshop on Spatial-Temporal Modeling) 2013-2015 (統計数理研究所で毎年開催、研究代表者が主催) を利用して本研究成果の発表をまとめて行い、関連分野の研究者に対して本研究成果を広く周知した。

### <参考文献>

[1] C. E. Rasmussen et al., "Gaussian Processes for Machine Learning," the MIT Press, 2006.

### 5. 主な論文発表等

〔雑誌論文〕(計 8 件)

Konstantin Markov and Tomoko Matsui,  
“High level feature extraction for  
the self-taught learning algorithm,”  
EURASIP Journal and Audio, Speech and  
Music Processing, 査読有, Vol. 6,  
2013, pp. 1-12.

Tatsuya Komatsu, Gareth W. Peters,  
Tomoko Matsui, Ido Nevat, and Kazuya  
Takeda, “Modeling room impulse  
response via composite of  
spatial-temporal Gaussian processes,”  
J. Acoust. Soc. Am., 査読有, 133, 2013,  
pp. 3454.

<http://dx.doi.org/10.1121/1.4806130>

Tatsuya Komatsu, Takanori Nishino,  
Gareth W. Peters, Tomoko Matsui and  
Kazuya Takeda, “Modeling  
head-related transfer functions via  
spatial-temporal process,” IEEE  
International Conference on Acoustics,  
Speech and Signal Processing, 査読有,  
2013, pp. 301-305

DOI: 10.1109/ICASSP.2013.6637657

Konstantin Markov and Tomoko Matsui,  
“Music genre classification using  
Gaussian process models,” IEEE  
International workshop on machine  
learning for signal processing, 査読  
有, 2013, pp. 1-6.

DOI: 10.1109/MLSP.2013.6661991

Konstantin Markov, Motofumi Iwata and  
Tomoko Matsui, “Music emotion  
recognition using Gaussian processes,”

MediaEval 2013 Workshop, 査読有,  
2013.10.18-2013.10.19.

Konstantin Markov and Tomoko Matsui,  
“Music genre and emotion recognition  
using Gaussian Processes,” IEEE  
Access. 査読有, Vol 2, 2014, pp.  
688-697.

DOI:10.1109/ACCESS.2014.2333095

Konstantin Markov and Tomoko Matsui,  
“Dynamic music emotion recognition  
using state-space models,”  
MediaMil2014 Workshop, 査読有, Vol.  
1263, 2014.

Konstantin Markov, Tomoko Matsui,  
Francois Septier and Gareth W. Peters,  
“Dynamic speech emotion recognition  
with state-space models,” CEUR  
Workshop Proceedings, 査読有, Vol. 1,  
2015, pp. 2077-2081.

〔学会発表〕(計5件)

小松 達也, 西野 隆典, 松井 知子,  
武田 一哉, “ガウス過程による頭部伝  
達関数の補間,” 日本音響学会 2013 年  
秋季研究発表会, 2013.9.25-2013.9.27.

Konstantin Markov, “Gaussian  
processes based music information  
retrieval,” STM2013,  
2013.8.1-2013.8.3.

Kazuya Takeda, “Modeling  
head-related transfer functions via  
spatial-temporal Gaussian process,”  
STM2013, 2013.8.1-2013.8.3.

足立 悠輔, 西野 隆典, 武田 一哉,

“ ガウス過程を用いた楽器音の波形補間,” 日本音響学会 2014 年秋季研究発表会, 2014.9.4.

足立 悠輔, 西野 隆典, 松井 知子, 武田 一哉, “ ガウス過程を用いた楽器音の統計的雑音抑圧,” 日本音響学会 2015 年春季研究発表会, 2015.3.16.

〔図書〕(計 1 件)

Konstantin Markov and Tomoko Matsui, “ Modern Methodology and Applications in Spatial-Temporal Modeling: Chapter 3 Speech and Music Emotion Recognition Using Gaussian Processes,” Springer, 2015, 111(Chapter 3: 21).

〔産業財産権〕(計 0 件)

〔その他〕

ホームページ等

Monte Carlo Dynamic Classifier:

<http://www.ismvideo.org/NDEOUP/MCDCtool/index.html>

## 6 . 研究組織

### (1) 研究代表者

松井 知子 (MATSUI TOMOKO)

統計数理研究所・モデリング研究系・教授

研究者番号 : 10370090

### (2) 研究分担者

武田 一哉 (TAKEDA KAZUYA)

名古屋大学・情報科学研究科・教授

研究者番号 : 20273295

マルコフ コンスタンティン (MARKOV

KONSTANTIN)

会津大学・コンピュータ理工学部・准教授

研究者番号 : 80394998

### (3) 研究連携者

上野 玄太 (UENO GENTA)

統計数理研究所・モデリング研究系・准教授

研究者番号 : 40370093

### (4) 研究協力者

ピータース ギャレス W. (PETERS GARETH W.)

University College London, UK・統計学科・

講師

研究者番号 : 90763061

ネバット アイド (NEVAT IDO)

A\*STAR, Singapore・Institute for Infocom

Research (I2R)・チームリーダー