

# 博士論文概要

## 論文題目

状態遷移モデルにおける  
ベイズ予測アルゴリズムに関する研究  
A Study on Bayes Prediction Algorithms for  
State Transition Models

申請者

須子	統太
Tota	Suko

--

2009 年 1 月

与えられたデータから次に発生するデータを予測する問題は、様々な工学分野における基本的な問題の一つである。例えば、株価予測や需要予測などの時系列データに対する予測の他にも、手書き文字認識や音声認識などのパターン認識や、信号処理における信号推定、情報理論におけるデータ圧縮など、予測問題と捉えられる問題は多岐にわたる。予測問題は、 $n$  個の時系列データが与えられたもとで  $n+1$  番目のデータを予測する問題や、 $n$  組の説明変数と目的変数の組と  $n+1$  番目の説明変数が与えられたもとで、 $n+1$  番目の目的変数を予測する問題などが含まれる。また、情報理論や制御理論などでは、 $n$  個の時系列データが与えられたもとで  $n+1$  番目のデータの出現する確率分布や、 $n$  個のデータが出現した確率分布を推定する問題が扱われている。本研究ではこれらも同様に予測問題として扱う。本研究ではこれら予測問題に対し、確率モデルを用いた予測法のひとつであるベイズ予測を適用する。

予測問題の評価基準としては、予測値と真の値との誤差を表す損失関数があげられる。損失関数はデータの関数として得られるため、与えられたデータにより値が変わる。そのため、損失関数をデータの出現確率で期待値を取ったリスク関数が評価基準として用いられることも多い。従来から様々な予測問題に対し、提案された予測法についてこれら損失関数やリスク関数により予測性能の評価が行われている。

ベイズ予測はベイズ基準のもとで損失を最小にする予測法である。近年、学習理論や情報理論において注目を集めており、その有効性が示されている。ベイズ基準のもとで最適とは、リスク関数を確率モデルにおける未知パラメータの事前分布で期待値をとったベイズリスク関数を考え、そのベイズリスク関数を任意のデータ数において最小にするという意味である。

ベイズ予測の値は、定義された損失関数により異なる。しかし、多くの損失関数に対しベイズ予測は、予測分布と呼ばれる確率分布の関数として与えられる。そのため、ベイズ予測を行うには、予測分布の計算が中心的な役割を果たす。しかし一般の確率モデルに対し、予測分布の計算は積分が含まれ解析的に解けない、和計算の計算量が膨大になる、などの問題点がある。そのため、ベイズ予測を実用化するには何らかの計算量削減アルゴリズムが必要となる。また、ベイズ予測はベイズ基準のもとで損失を最小にすることを保証するが、一般の確率モデルに対し、ベイズ基準以外の評価基準における予測性能は保証されない。そのため、ベイズ予測を実用化するにあたって、ベイズ基準以外の視点から予測性能を評価する必要がある。従来、学習理論や情報理論の分野では、損失やリスクの漸近評価により予測性能を評価する研究が多く行われている。本研究では、これらベイズ予測を実用化するにあたっての課題に着目する。

他方，予測問題に用いる確率モデルとして，マルコフモデルや隠れマルコフモデルなど，離散時間で状態遷移が起きる確率モデルが学習理論・情報理論を中心に用いられており，実用的にも有効性が示されている．本研究ではこれら離散時間で状態遷移の起きる確率モデルを，状態遷移モデルと呼ぶ．状態遷移モデルは状態が連続であるか離散であるか，状態遷移が既知であるか未知であるかで様々なモデルを分類することができる．例えば，状態が離散で状態遷移系列が既知のモデルとしては，マルコフモデルやパターン認識で用いられる決定木モデル，情報理論で用いられる **Context Tree** モデルなどがある．また，状態が離散で状態遷移系列が未知のモデルとしては，隠れマルコフモデルや混合分布モデルなどがある．さらに，状態が連続のモデルとしては，情報理論で用いられている **i.p.i.d.** (**independent piecewise-identically-distributed**) 情報源モデルなどがある．

本研究では，状態遷移モデルに含まれるいくつかのモデルに対しベイズ予測を適用する．そのもとで，ベイズ予測の実用化に伴う前述の課題に対し，計算量削減アルゴリズムの提案とリスクの漸近評価を行う．

以下，論文の構成を示す．

第 1 章では，序論として本研究の背景および目的について述べる．

第 2 章では，準備として，予測問題の分類及びベイズ予測の定式化と，確率状態空間モデルについて定義する．

第 3 章では，離散状態の状態遷移モデルで状態遷移が未知の場合に対し，計算量を削減した近似ベイズ予測アルゴリズムを提案する．この場合の状態遷移モデルにおけるベイズ予測は，取り得る全ての状態遷移のパターンについて事後確率の重み付け和を計算することで求められる．状態遷移のパターン数は状態数  $S$ ，データ数  $n$  のとき， $S^n$  個存在する．そのため，事後確率の重み付け和の計算には，データ数に対し指数オーダーの計算量がかかり計算量的に困難である．そこで提案アルゴリズムは，全ての状態遷移のパターンを重み付けるのではなく，事後確率の高い状態遷移パターンを近似的にいくつか求め，それらについての重み付け和計算を行うことで計算量を削減する．また，従来外れ値解析で用いられていた外れ値を含む回帰モデルに提案アルゴリズムを適用し，シミュレーションにより評価を行う．ここで言う外れ値とは，解析対象とする母集団以外から発生したデータのことを表す．外れ値は少量でもデータに含まれると解析結果に大きな影響を与えることが知られており，従来から外れ値の取り扱いに関する研究が数多くなされている．外れ値の発生にはモデルを仮定する場合と，仮定しない場合の研究があるが，外れ値を含む回帰モデルは前者の外れ値の発生にモデルを仮定した場合のモデルのひとつである．また，外れ値解析の目的は，外れ値の検出やパラメ

ータ推定，予測などが考えられるが，外れ値を含む回帰モデルに対し予測を目的とした研究は行われていない．そこで本研究では外れ値を含む回帰モデルに対し，予測を目的とした解析法として，計算量を削減した近似ベイズ予測アルゴリズムを適用する．シミュレーションによる評価の結果として，最適なベイズ予測の計算量がデータ数  $n$  に対し  $O(2^n)$  の計算量が必要であったのに対し，提案アルゴリズムは計算量を  $O(n^2)$  に削減しつつ，予測性能の劣化が少ないアルゴリズムであることを示す．また，一般的なベイズ予測の近似アルゴリズムである MCMC (Markov Chain Monte Carlo) 法や変分ベイズ法などとの予測性能を比較し，提案アルゴリズムは従来の近似アルゴリズムよりも予測性能が高くなることを示す．

第 4 章と第 5 章では，連続状態の状態遷移モデルである，i.p.i.d. 情報源モデルを扱う．i.p.i.d. 情報源モデルとは，情報理論におけるデータ圧縮に用いられる非定常な確率モデルである．従来データ圧縮の分野において，i.p.i.d. 情報源に対する予測アルゴリズム（圧縮アルゴリズム）が提案されているが，ベイズ予測を適用した研究はあまりなされていない．そこで本研究では i.p.i.d. 情報源モデルにベイズ予測を適用した場合の効率的なアルゴリズムの提案とリスクの漸近評価を行う．

第 4 章では，i.p.i.d. 情報源モデルに対し状態遷移が未知である場合の効率的なベイズ予測アルゴリズムを示す．状態遷移が未知の i.p.i.d. 情報源に対するベイズ予測は，状態遷移の全てのパターンについての重み付け和の計算を行うことから， $O(2^n)$  の計算量が必要であった．そこで，状態遷移がベルヌーイ分布に従って起きるという仮定を置くことで，ベイズ最適性を保持したまま計算量を  $O(n^2)$  に削減したアルゴリズムを提案する．その結果，状態遷移が未知の i.p.i.d. 情報源に対し，計算量とメモリ量の観点からベイズ予測が実用可能であることを示す．

第 5 章では，i.p.i.d. 情報源モデルに対し，ベイズ予測のリスクの漸近評価を行う．従来 i.p.i.d. 情報源モデルにおける予測アルゴリズムを漸近評価する場合には，クラスに一部制約をおいたモデルに対し評価が行われている．そこで，本研究では，状態遷移が既知であり，データ数  $n$  が無限になるとき状態遷移も無限回起きるという i.p.i.d. 情報源モデルのクラスを考える．このとき，任意の予測法に対し，リスクとベイズリスクが漸近的に一致することから，ベイズ予測が漸近的にリスクを最小にすることを示す．

最後に第 6 章では，以上の結果をまとめ結論と今後の展望について述べる．