

## 音響モデルのベイズ学習

渡 部 晋 治<sup>†1</sup>

本稿では、音響モデルの機械学習をテーマに、ベイズ学習を用いた音声認識用音響モデルの構築について説明を行う。ベイズ学習は、機械学習分野での進展もあり、音声認識用音響モデル構築のような大量データ学習においてもその効果が示されている。本稿は、それらについての解説を行うとともに、音響モデル学習分野での研究例を紹介する。

### Bayesian training of acoustic models

SHINJI WATANABE<sup>†1</sup>

This paper describes acoustic model construction by using Bayesian training according to a theme of machine learning approaches in acoustic models. Bayesian training is successfully applied to the construction of acoustic models by using large amount of training data based on the progress of the Bayesian training in the machine learning field. We briefly explain about the Bayesian training and introduce some research examples in the acoustic model field.

#### 1. はじめに

パターン認識問題を機械学習を用いて扱う場合によく使われるのがベイズ推論にもとづくベイズ学習である。ベイズ推論では

- 確率の和 ( $p(X) = \sum_Y p(X, Y)$ )・積 ( $p(X, Y) = p(X|Y)p(Y)$ ) の公式, 独立性 ( $p(X, Y) = p(X)p(Y)$ ) などの数学的な操作が存在
- ベイズの定理により逆問題が解ける

- ベイズ推論によって所望の量の確率値・期待値を求めることができるという性質を持つ。このような性質からそれぞれ
- パターン認識の諸問題の定式化を確率・統計理論の枠組みで統一的に扱うことができる
- データの生成過程をモデルに取り込むことができる
- 期待値効果により、汎化性能を高める

といった特長を持っている。本稿では、このベイズ推論にもとづくモデル構築問題に注目し、特に音声認識用音響モデルパラメータの学習と、HMM 状態数や GMM<sup>\*1</sup>混合数決定などの構造決定(モデル選択)に焦点を当てる。音響モデル学習・モデル選択を説明する上で、学習データとして、音声データから特徴抽出(例えば MFCC 変換)された多次元特徴量ベクトルの時系列集合  $\mathbf{x} = \{\mathbf{x}^1, \dots, \mathbf{x}^t, \dots\}$ , モデルパラメータ  $\theta$ (HMM パラメータなど), モデル構造  $m$ (HMM, GMM の構造, 規模など)としたとき, モデル学習・選択は, データ  $\mathbf{x}$  が与えられたとき, 如何に適切にモデル変数  $m, \theta$  を求めるかという問題として捉えることができる。以下では、音響モデルにおけるベイズ学習の代表的な手法として、MAP(Maximum a posteriori) 適応法を例に説明を行う。

#### 2. モデルパラメータの事後分布と MAP 適応

ベイズ学習ではパラメータ  $\theta$  の値を学習データ  $\mathbf{x}$  から求めるのではなく、 $\mathbf{x}$  が与えられた上での  $\theta$  の確率分布  $p(\theta|\mathbf{x})$ (これを事後分布と呼ぶ)を推定対象とする。有名なベイズの定理を用いるとモデルパラメータ  $\theta$  の事後分布  $p(\theta|\mathbf{x})$  は次のように表現することができる。

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})} \quad (1)$$

ここで、 $p(\mathbf{x}|\theta)$  は尤度関数、 $p(\theta)$  及び  $p(\mathbf{x})$  はそれぞれモデルパラメータ、データの事前分布である。ベイズの定理(式(1))が示唆することは、ベイズ学習においては、1) 尤度関数で表現されるデータの生成過程(音響モデルの場合は HMM)、2) 事前分布により与える事前知識の 2 つの情報を加味した確率分布として事後分布を得ることができるという点である。

ベイズ学習はこのような分布表現をもとにした枠組みであるため、その解析的扱いや実装法は、パラメータ表現を扱う最尤学習に比べて一般に困難である。ただし、ベイズ推論の近似的実現である事後確率最大化(MAP, Maximum A Posteriori)法は、事後分布そのものではなくそれを最大化するモデルパラメータ(つまり  $\hat{\theta} = \operatorname{argmax}_{\theta} p(\theta|\mathbf{x})$ )を求める枠組みである。分布表現から離れるため、ベイズ学習の効果は限定的となるが、事前知識を利用した

<sup>†1</sup> 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories, NTT Corporation

\*1 隠れマルコフモデル (HMM), 混合ガウス分布モデル (GMM)

少量データの学習における過学習緩和問題においては十分にその威力を発揮する。MAP 法は、音声認識分野では、90 年代に単一ガウス分布 HMM の適応問題において適用され、その後混合分布学習に拡張された [1,2]。MAP 適応法は、不特定話者音響モデルを事前分布として、MAP 学習により特定話者音響モデルを構築する手法であり、その実装の容易性、最尤学習との親和性、適応タスクでの十分な改善効果から適応技術の標準的技術として今日の音声認識において広く用いられている。また、MAP 法の過学習緩和効果は、識別学習においても正則化として用いられ、効果を発揮している [3]。

### 3. モデル構造の事後分布とモデル選択

次に、モデル選択におけるベイズ推論の利用について説明を行う。ベイズ推論では、モデルパラメータ  $\theta$  以外の、モデル化に際して導入されるあらゆる変数を、確率変数として扱うことができるうえ、数学的操作によってその事後分布を解析的に表現できる。例えば、モデル構造変数  $m$  に対しての事後分布も次のようにして導出することが可能である。

$$p(m|x) = \frac{\int p(x|\theta, m)p(\theta|m)d\theta p(m)}{p(x)} \quad (2)$$

このように、学習データ  $x$  が与えられた際どのモデル構造が適切かという問題は、上記の式をもとに、尤度関数（データの生成過程） $p(x|\theta, m)$ 、モデルパラメータの事前分布  $p(\theta|m)$ 、モデル構造の事前分布  $p(m)$ 、データの事前分布  $p(x)$  の情報をもとに推定できるといえる。MAP 法の議論と同様、上式の解析的扱い及び実装は一般に困難であり、学習データが十分に多いという仮定の下に、上式を簡略化したベイズ情報量基準が一般的に広く用いられている。特に、音声認識用音響モデルの研究において、HMM の状態数やガウス分布の混合数決定問題に、ベイズ情報量基準を用いたモデル選択法が提案され [4]<sup>\*1</sup>、幅広く利用されている。

### 4. ベイズ推論の進展

90 年代後半には、機械学習の分野で期待値最大化学習や MAP 学習の拡張として変分ベイズ法が提案される [5]。変分ベイズ法は、MAP 法やベイズ情報量基準とは違い、ベイズ推論のなかで最も困難な問題であった期待値操作を変分法によって解くことによりベイズ推論

を効率よく実現するアルゴリズムを提供したものである。この変分ベイズ法を音声認識用音響モデルに適用することにより、MAP 法によって実現される事前知識の活用やモデル選択を同一の枠組みで実現することが可能になった [6]。また本手法は、事後分布の期待値操作をもとに識別を行うベイズ予測識別法も内包しており、ベイズ推論がもつ汎化性能向上の性質を十分に利用した音声認識が実現されている。

さらに近年では、変分ベイズ法に加えてサンプリング手法も大きく進展し、大規模データへの適用が活発化している。このようなベイズ推論の解析・実装面での進展により、事前分布の効果が階層的に表現される階層ベイズや、それらの複雑な依存関係をグラフで用いるグラフィカルモデル、時間発展をベイズ推論で記述する sequential Bayes、無限混合の扱いが可能なノンパラメトリックベイズ、複数モデルの効果をベイズ推論をもとに統合するベイズモデル合成法などの様々なベイズ推論が機械学習の分野で提案され、またその一部が音声認識用の音響モデル学習においても利用され始めている。また、ベイズ推論と識別学習が有機的に組み合わせることにより、汎化性能向上および識別性能向上をバランスよく実現する枠組みも近年注目されている。このように、音響モデルのベイズ学習は、機械学習分野の進展もあり、今後も大きな進展が望まれる分野である。筆者としては、本稿の議論をきっかけに、ベイズ学習のみならず識別学習やアンサンブル学習などの最先端機械学習手法の音響モデルへの適用に大きな注目が集まれば幸いである。

### 参 考 文 献

- 1) Lee, C.-H., Lin, C.H. and Juang, B.-H.: A study on speaker adaptation of the parameters of continuous density hidden Markov models, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol.39, pp.806–814 (1991).
- 2) Gauvain, J.-L. and Lee, C.-H.: Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains, *IEEE Transactions on Speech and Audio Processing*, Vol.2, pp.291–298 (1994).
- 3) Povey, D.: Discriminative training for large vocabulary speech recognition, PhD Thesis, Cambridge University (2003).
- 4) Shinoda, K. and Watanabe, T.: Acoustic modeling based on the MDL criterion for speech recognition, *Proc. Eurospeech1997*, Vol.1, pp.99–102 (1997).
- 5) Atias, H.: Inferring Parameters and Structure of Latent Variable Models by Variational Bayes, *Proc. Uncertainty in Artificial Intelligence (UAI) 15* (1999).
- 6) Watanabe, S., Minami, Y., Nakamura, A. and Ueda, N.: Variational Bayesian Estimation and Clustering for Speech Recognition, *IEEE Transactions on Speech and Audio Processing*, Vol.12, pp.365–381 (2004).

\*1 [4] では、情報量基準として記述長最小化基準を用いているが、ガウス分布を対象にしたモデル選択問題において、両基準は等価な評価関数となるため、[4] はベイズ情報量基準を音声認識用音響モデルの選択に先駆けて適用した研究だといえる