

### 統計的予測問題におけるアンサンブル学習の有効性

数理・推論研究系 伏木 忠義

既存のデータをもとに将来のデータの値を予測する方法は、日常で広く必要とされる技術である。このような予測問題に対して、統計学や機械学習など、さまざまな分野で研究が行われてきた。予測問題は基本的で重要な問題であり、長年の研究があるが、使える計算リソースや扱う問題の大きさ、データ量といった要因は、時代と共に変化し、それらに対応した手法が提案、研究されている。

機械学習の分野では、アンサンブル学習とよばれる手法が研究されている。アンサンブル学習は、何度も学習を行い、その結果得られた多数の学習機械をうまく組み合わせて予測を行うものである。アンサンブルをとることで1つの学習機械のみで予測するよりも良い予測が実現される場合がある。特に、ブートストラップデータを用いて何度も学習を行い、それらの単純平均で予測するバグgingや学習がうまくいかなかった例題について重みをつけて次の学習を行うブースティングという手法が有名である。一方、統計学においても、Kullback-Leibler ダイバージェンスを損失関数とした統計的予測問題の文脈で、推定量をパラメータのところに代入したプラグイン予測、ブートストラップデータを用いた予測、Bayes 予測などの予測方法についてこれまで研究されている。

統計学で提案されていたブートストラップ予測は、学習理論のアンサンブル学習の立場からは、最尤推定量のプラグイン分布にバグgingを適用したものと考えることができる。Bayes 予測は、事後分布からパラメータをリサンプリングして平均をとったものと考えられるので、これもアンサンブル学習の1つとみなせる。

私のこれまでの研究では、このように2つの異

なる分野で研究されていた手法を統一的な観点でとらえ、統計的予測問題におけるアンサンブル学習の効果について調べてきた。

まず、真の分布が仮定した統計モデルに含まれている場合を考えよう。このときは、Bayes 予測が許容的となるため、Bayes 予測がある意味で最良の予測を与えているといえる。私の研究では、漸近論を用いて、ブートストラップ予測が Bayes 予測の近似になっていることを示し、さらに予測性能に関しては、ブートストラップ予測が最尤推定量のプラグイン分布よりも漸近的に良い予測を与えるということを示した。

一方、統計学においては、統計モデルに真の分布が存在すると仮定して議論する機会が多いが、現実的には、複雑な現象を扱う場合など、そのような仮定が成り立つ場合は稀であると考えられる。仮定した統計モデルの中に真の分布に十分近い分布が存在すれば問題ないが、そのような仮定が成り立つことは一般には保障できないだろう。

真の分布がモデルに含まれる場合には、ブートストラップ予測は Bayes 予測の近似と考えられたが、私の研究では、真の分布がモデルに含まれない場合は、ブートストラップ予測と Bayes 予測には違いがあるということがわかった。特に、予測性能については、ブートストラップ予測が無情報の Bayes 予測よりも漸近的には良い予測を与えるということを示した。

アンサンブル学習は、繰り返しの学習を必要とするため、大きな計算リソースを要求する。計算機性能の向上にともない、ある程度の計算コストがかかっても良い予測が欲しいという状況はあるだろう。アンサンブル学習は、そのような状況で

有効な手法であり、私の研究もその意味で、現代的な観点で意味がある研究といえる。

#### 参考文献

Fushiki, T., Komaki, F., Aihara, K. (2004). On parametric bootstrapping and Bayesian prediction. *Scandinavian Journal of Statistics*,

31, 403-416.

Fushiki, T., Komaki, F., Aihara, K. (2005). Nonparametric bootstrap prediction. *Bernoulli*, 11, 293-307.

Fushiki, T. (2005). Bootstrap prediction and Bayesian prediction under misspecified models. *Bernoulli*, 11, 747-758.