

研究終了報告書

「ベイズ誤差推定及び正則化手法の研究」

研究期間：2020年11月～2023年3月

研究者：石田 隆

1. 研究のねらい

本研究では、ベイズ誤差推定や正則化の技術に関する研究を行うことを目指しています。教師あり学習におけるベイズ誤差とは、与えられた問題に対する予測誤差の下限のことを指します。ベイズ誤差を推定する技術が発達すれば、扱いたいデータ分析の問題に対して、その限界性能を予め知ることができるようになり、機械学習モデルの訓練に対して定量的な目標の設定を行うことが可能となります。また、ベイズ誤差を推定することができれば、正則化のハイパーパラメータのチューニングにも活用すると言ったように、より良い正則化手法の研究に繋がる可能性もあります。本研究のねらいは、ベイズ誤差を推定するための手法の開発に取り組み、また過剰適合・正則化手法の研究にも取り組むことです。

2. 研究成果

(1) 概要

本研究では、ベイズ誤差推定の研究と正則化手法の研究を行いました。1つ目のベイズ誤差推定の研究では、インスタンスに対する事後確率のうち、最大の値になるクラスの値を直接収集することにより、ベイズ誤差を推定する手法を提案しました。最大のポイントは、事後確率が最大となるクラスの事後確率の情報さえあれば、インスタンスは必要ないという点です。インスタスがなければ、入出力関係を表す分類器が学習できなくなるという面白さがあります。実用上は、インスタスを収集しないことでインスタス収集コストを抑えることや、データ収集の制約が存在するような状況などでもベイズ誤差推定を実施できるようになる可能性があります。しかし現実的には事後確率を直接収集するという仮定は強く、この仮定を弱める研究も行いました。具体的には、特定の雑音付きのソフトラベルが収集できる場合、複数のハードラベルが収集できる場合などであっても、一致性や不偏性、もしくは漸近不偏性などの良い性質を持つことを示しました。

2つ目の研究として、正則化手法の研究を行いました。前述のベイズ誤差を推定するための手法では、ソフトラベルに相当する情報を活用しました。機械学習プロジェクトを開始する事前ステップとしてベイズ誤差を推定するようなケースが考えられますが、その後プロジェクトを開始し機械学習モデルを学習する際に、収集したソフトラベルを再活用できれば、少数のデータからでも過剰適合を回避しやすくなることが期待できます。そこで、ソフトラベルを用いた新たな学習手法を提案しました。また、ハードラベル付き訓練データしかない場合に使うことのできる、ローカルなラマツハ複雑度に基づく正則化手法も提案しました。

(2) 詳細

研究テーマ A:「ベイズ誤差推定の研究」

1つ目の研究テーマとして、ベイズ誤差推定の研究を実施しました。教師あり学習におけるベ

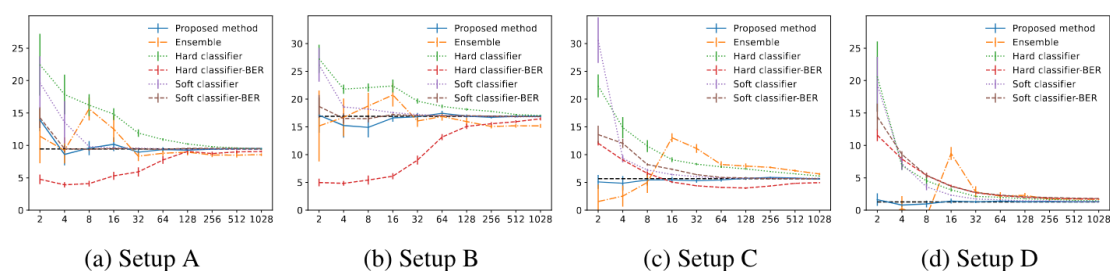
ズ誤差とは、与えられた問題に対する予測誤差の下限のことを指します。ベイズ誤差を推定する技術が発達すれば、扱いたいデータ分析の問題に対して、その限界性能を予め知ることができるようになり、機械学習モデルの訓練に対して定量的な目標の設定を行うことが可能となります。また、ベイズ誤差を推定することができれば、正則化のハイパーパラメータのチューニングにも活用すると言ったように、より良い正則化手法の研究に繋がる可能性もあります。

インスタンスに対する事後確率のうち、最大の値になるクラスの値を直接収集することにより、ベイズ誤差を推定する手法を提案しました。最大のポイントは、事後確率が最大となるクラスの事後確率の情報さえあれば、インスタンスは必要ないという点です。インスタスがなければ、入出力関係を表す分類器が学習できなくなるという面白さがあります。実应用上は、インスタスを収集しないことでインスタス収集コストを抑えることや、データ収集の制約が存在するような状況などでもベイズ誤差推定を実施できるようになる可能性があります。もちろんインスタスがすでに存在するような場合でも使える手法です。

しかし現実的には事後確率を直接収集するという仮定は強く、この仮定を弱める研究も行いました。具体的には、特定の雑音付きのソフトラベルが収集できる場合、複数のハードラベルが収集できる場合などであっても、一致性や不偏性、もしくは漸近不偏性などの良い性質を持つことを示しました。

クラス事後確率（もしくは雑音付きの事後確率）の情報を得ることができるというのは、一見強い仮定ですが、問題設定によっては自然とそのような情報を得ることができる場合もあります。例えば二値分類では通常、正と負の両方のデータが必要ですが、正信頼度分類においては正例とその信頼度だけを用いて学習します。この場合、信頼度は正のクラスの事後確率と一致します。この設定において、ベイズ誤差を正例の信頼度で表現できることを導出し、インスタスを用いずにベイズ誤差が推定できることもわかりました。

人工データを用いた実験により、ソフトラベルが与えられる場合、雑音付きのソフトラベルが与えられる場合、正信頼度が与えられる場合において、真のベイズ誤差を正確に推定できることを確認しました(図表1はソフトラベルが与えられた場合の実験結果)。



図表 1 人工データ実験の結果。データ数を増やしていくことで、提案法（青）は真のベイズ誤差（黒の点線）に近づいていくことがわかる

最後に、ベイズ誤差推定の応用として次の2種類の実データ実験を行いました。1つ目は、機械学習で頻りに用いられるベンチマークデータセットのベイズ誤差を推定することで、近年の論文で報告されているテスト誤差がベイズ誤差にどれほど近づいているかということを確認しました。CIFAR-10においてはCIFAR-10Hで提供されている各テスト画像に対する複数のハードラベ

ルからソフトラベルを構築しました。また、Fashion-MNIST でも CIFAR-10H と同様のソフトラベルを構築するために、Amazon Mechanical Turk を用いて各テスト画像に対して 60 個程度のハードラベルを収集しました。これらのソフトラベルを用いてベイズ誤差を推定しました。結果、最近の論文で用いられているニューラルネットワーク構造を用いた場合のテスト誤差は、ベイズ誤差に既に接近しており、CIFAR-10 や Fashion-MNIST は限界性能に到達済み(もしくは間近である)ということがわかりました。

2つ目の応用例として、**データセットの本質的難易度**を計測するためにベイズ誤差推定を活用する問題に取り組みました。こちらの結果は現時点で非公開のため省略します。

研究テーマ B:「正則化手法の研究」

2つ目の研究テーマとして正則化手法の研究を実施し、以下の2つのプロジェクトを進めました。1つ目では、実用上はプロジェクト開始前にベイズ誤差の推定のためにソフトラベルを収集することを想定していますが、プロジェクト開始後には、ベイズ誤差推定のために収集したソフトラベルをモデルの学習において再活用するのが望ましいと言えます。ソフトラベルを用いて分類器を学習する新たな手法の研究を現在進めています。

2 つ目のプロジェクトとして、ニューラルネットワークの正則化手法に取り組み、ローカルなラマツハ複雑度に基づく正則化手法を提案しました。

3. 今後の展開

二値分類におけるベイズ誤差を推定する手法を提案することができましたが、今後はさらに対象を広げていく予定です。例えば、多値分類におけるベイズ誤差や、マルチラベルにおける限界性能などが考えられます。また、ベイズ誤差は 01 損失に対応していますが、代理損失を用いた場合の限界性能にも広げたいと考えています。

ソフトラベルを用いた学習手法の研究については、現在研究を進めている最中ですので、引き続き研究を実施し、近々成果を論文にまとめて国際会議等に投稿する予定です。

4. 自己評価

研究目的の達成状況:大きく2つの研究テーマを掲げていましたが、1つ目のベイズ誤差推定の研究は計画通りに進みました。分類問題におけるベイズ誤差推定の手法を提案することができ、さらにはソフトラベルに雑音に乗る場合や、弱教師(性信頼度)の場合などに対するベイズ誤差推定の手法を提案することも達成しました。本研究は機械学習の国際会議 ICLR 2023 に Notable top 5%として採択されました。一方で、2つ目の正則化手法の研究ではベイズ誤差推定値を用いたモデルの学習の研究を行うことを計画していましたが、ベイズ誤差推定を行うために収集したソフトラベルを用いたモデルの学習に研究余地を見出し、現在研究を進めている最中です。また、ベイズ誤差の研究とは独立に、ハードラベルが与えられた状況での正則化手法を提案し、IEEE Transactions on Pattern Analysis and Machine Intelligence に採択されました。

研究の進め方:2020 年度～2021 年度は研究費を計画的に執行しました。2022 年度も計画的に執行できる見込みです。

研究成果の科学技術及び社会・経済への波及効果:研究期間中はベイズ誤差推定の手法や正

則化手法の開発に注力しましたが、今後は社会・経済データを活用した実応用研究も積極的に行い、本研究の社会・経済への波及効果を高めることに繋がります。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 2件

1. T. Ishida, I. Yamane, N. Charoenphakdee, G. Niu, M. Sugiyama. Is the Performance of My Deep Network Too Good to Be True? A Direct Approach to Estimating the Bayes Error in Binary Classification. *Proceedings of Eleventh International Conference on Learning Representations (ICLR2023)*.

There is a fundamental limitation in the prediction performance that a machine learning model can achieve due to the inevitable uncertainty of the prediction target. In classification problems, this can be characterized by the Bayes error, which is the best achievable error with any classifier. The Bayes error can be used as a criterion to evaluate classifiers with state-of-the-art performance and can be used to detect test set overfitting. We propose a simple and direct Bayes error estimator, where we just take the mean of the labels that show the uncertainty of the classes. Our flexible approach enables us to perform Bayes error estimation even for weakly supervised data. In contrast to others, our method is model-free and even instance-free. Moreover, it has no hyperparameters and gives a more accurate estimate of the Bayes error than classifier-based baselines. Experiments using our method suggest that a recently proposed classifier, the Vision Transformer, may have already reached the Bayes error for certain benchmark datasets.

2. Z. Lu, C. Xu, B. Du, T. Ishida, L. Zhang, & M. Sugiyama. LocalDrop: A hybrid regularization for deep neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.44, No.7, pp.3590–3601, 2022.

In neural networks, developing regularization algorithms to settle overfitting is one of the major study areas. We propose a new approach for the regularization of neural networks by the local Rademacher complexity called LocalDrop. A new regularization function for both fully-connected networks (FCNs) and convolutional neural networks (CNNs), including drop rates and weight matrices, has been developed based on the proposed upper bound of the local Rademacher complexity by the strict mathematical deduction. The analyses of dropout in FCNs and DropBlock in CNNs with keep rate matrices in different layers are also included in the complexity analyses. With the new regularization function, we establish a two-stage procedure to obtain the optimal keep rate matrix and weight matrix to realize the whole training model. Extensive experiments have been conducted to demonstrate the effectiveness of LocalDrop in different models by comparing it with several algorithms and the effects of different hyperparameters on the final performances.

(2)特許出願

研究期間全出願件数:0件

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

石田隆. 2022年度船井研究奨励賞.

T. Ishida, I. Yamane, N. Charoenphakdee, G. Niu, M. Sugiyama. Is the Performance of My Deep Network Too Good to Be True? A Direct Approach to Estimating the Bayes Error in Binary Classification. *IEICE Technical Report, IBISML2021-44*, pp.38-45. (presented at [45th IBISML, online, Mar 8-9, 2022.](#))

T. Ishida. Reliable Machine Learning from Limited Data and Supervision: A Risk Modification Approach. *PhD thesis*, 2021.

石田隆. 東京大学大学院新領域創成科学研究科長賞, 2021年3月.