

早稲田大学大学院情報生産システム研究科

博士論文審査結果報告書

論 文 題 目

**Study on Nonlinear Regression with Missing Values Based
on Hybrid Models Using Quasi-Linear Kernel**

申 請 者

Huilin ZHU

情報生産システム工学専攻
ニューロコンピューティング研究

2022 年 11 月

非線形回帰は機械学習において重要な研究課題の一つであり、ニューラルネットワーク (Neural Networks: NN) は柔軟な非線形回帰モデルとして大きく注目されている。一方、実際の問題において、データの欠測がよくあり、欠損値を処理して NN モデルの学習をしなければならないことが多い。欠損値処理法として、平均値で欠測データを埋める代入法や、欠測データをモデルで予測するモデルベース手法があるが、欠測データは 100% の正確率で予測できない。そのため、欠損値の処理により、データに不確実性が増加するようになり、非線形回帰のための NN モデルの学習において、この不確実性にいかに効果的に対処するかが問題となる。一方、サポートベクター回帰 (Support Vector Regression: SVR) は、不確実性に強い線形回帰モデルとして構築できるため、近年多くの注目を集めている。また SVR では、カーネル法によりデータを高次元特徴空間に写像することによって、非線形回帰問題も特徴空間において線形回帰問題として扱うことを可能にしている。従来の SVR では、基本カーネル関数あるいは基本カーネル関数の組み合わせで合成されるカーネル関数を利用しているが、基本カーネル関数の形が固定されているので、対象問題に関する先見情報があっても利用しにくいという問題がある。それに対して先行研究 (Zhou et al., 2014, Li et al., 2016) では、分類問題において Coarse-to-Fine という Two-Step モデリング法が提案されている。この Coarse 中間モデルの構築による Two-Step モデリング法は、分類問題においてその有効性が確認されているが、回帰問題においてどのように適用するか、さらに欠損値がある問題へどのように拡張するかが課題として残されている。

そこで本研究では、Coarse-to-Fine という Two-Step モデリング法に基づいて、欠損値の処理に対処できる準線形カーネル関数を持つ SVR ハイブリッドモデルの構築を行っている。提案したハイブリッドモデルは、構造的にオートエンコーダとゲート付線形ネットワークから構成されているが、欠損値の予測機能と、ゲート制御信号生成によるパーティショニング機能およびマルチローカル線形モデリング機能という 3 つの機能を実現している。本論文では、上記の 3 つの機能を実現するため、デノイジングオートエンコーダの構築や、ウィナーテイクオール (Winner-Take-All: WTA) オートエンコーダの構築および準線形カーネル関数を持つ SVR モデルの構築などにより、欠損値に対処した高性能な非線形回帰モデルを実現している。

以下に、本論文の構成と各章の概略について述べ、評価を与える。

第 1 章では、非線形回帰、欠損値処理、および SVR モデル、準線形カーネル関数について述べ、本研究の背景・目的を明らかにしている。

第 2 章では、ガウス関数で補間されたマルチローカル線形モデリングによる準線形 SVR ハイブリッドモデルの構築について述べている。欠損値予測モデルの構築には、教師データとして信頼できる学習データがないという問題がある。本論文では、まず条件付き平均法で Affinity Propagation (AP) クラスタリングアルゴリズムを欠損値に対処できるように拡張し、欠損値ありのデータセットを自動的に複数のクラスタに分割する。これをベースにして、まず、代入

法によりクラスタの平均値で欠損値を埋めて欠損値予測モデルの構築のための学習データを生成する。そして、クラスタの中心値および大きさをマルチローカル線形モデルの構築のための補間用ガウス関数を定義する。次に、生成された学習データを用いて WTA デノイジングオートエンコーダを学習させ、欠損値予測器を構築して、欠測データの処理を行う。さらに、ガウス関数で補間されたマルチローカル線形モデリングを通して、準線形カーネル関数を持つ SVR ハイブリッドモデルを構築している。性能評価のための数値例では、提案した準線形カーネル関数を持つ SVR ハイブリッドモデルを 5 つのベンチマークデータセットに適用して、欠損値データ比率が 10%、20%、30%、40%、50%、60% の場合に、複数の従来法と比較し、Root Mean Square Error (RMSE) では、すべてのケースで提案法が従来法より優れていることを示している。例えば、欠損値データの比率が 60% の場合に、従来法の GMM-ELM (D.Sovilj et al., 2016) の RMSE 平均値は 1.6012 であるのに対して、提案法は 1.2580 である。このことから、提案した準線形カーネル関数を持つ SVR ハイブリッドモデルの有効性が明らかとなっている。

第 3 章では、ゲート付線形ネットワークに基づいた区分線形モデリングによる準線形 SVR ハイブリッドモデルの構築について述べている。第 2 章で述べたハイブリッドモデルでは、モデルの性能が AP クラスタリングの結果と強い関係があり、欠損値予測モデルの学習データの分布に大きく依存される問題がある。一方、事前学習されたエンコーダは、上位 $k\%$ の WTA 機構を導入することで、入力データをクラスタリングすることができる。そして、代入法によりクラスタの平均値で欠損値を埋めることより良い学習データが得られる。この良い学習データを利用することで、欠損値予測器の性能を向上させることができる。本論文では、まず、このようにエンコーダの更新に伴って繰り返し、徐々に更新される学習データで WTA デノイジングオートエンコーダを訓練し、高性能な欠損値の予測機能を実現している。次に、学習されたエンコーダに別の上位 $k\%$ の WTA 機構を導入することで、ゲート制御信号生成によるパーティショニング機能を実現している。さらに、ゲート制御信号に基づいたゲート付線形ネットワークが区分線形モデリングを行うことにより、高性能な準線形 SVR ハイブリッドモデルの構築を実現している。性能評価のための数値例では、提案した準線形カーネル関数を持つ SVR ハイブリッドモデルを UC アーバイン機械学習リポジトリと Tianchi の 6 個のベンチマークデータセットに適用して、欠損値比率が 10%~60% のケースで複数の従来法と比較した結果、RMSE ではすべてのケースにおいて提案法が従来法より優れていた。例えば、一酸化炭素 ‘CO’ というデータセットに対して、欠損値データ比率 20% の場合に、従来法の KSC-MVI (L.Wang et al., 2010) の RMSE 平均値は 1.2911 であるのに対して、提案法では 0.8960 である。これらのことから、提案した準線形カーネル関数を持つ SVR ハイブリッドモデルの有効性が明らかとなっている。

第 4 章では、多層ゲート付線形ネットワークに基づいた区分線形モデリングおよび敵対的学習による深層準線形 SVR ハイブリッドモデルの構築について

述べている。高性能な SVR ハイブリッドモデルを構築するために、欠損値予測器の性能向上と SVR モデルの性能向上が必要である。本論文では、前者に対して識別器を導入し、欠損値と観測値の識別を利用した敵対的学習と、第 3 章で述べた徐々に更新された教師データを利用した学習と共に WTA デノイジングオートエンコーダの訓練を行っている。後者に対して、ハイブリッドモデルを多層ゲート付線形ネットワークと深層オートエンコーダで構築している。これによって、高性能な欠損値の予測機能と多層ゲート制御信号生成によるパーティショニング機能という二つの機能を同時に実現し、さらに多層ゲート制御信号に基づいた多層ゲート付線形ネットワークが強力な区分線形モデリングを行うことによって、深層準線形カーネル関数を持つ高性能な SVR ハイブリッドモデルの構築を実現している。性能評価のための数値例では、提案した深層準線形カーネル関数を持つ SVR ハイブリッドモデルを実際の空気品質ベンチマークデータセットに適用して、複数の従来法と比較し、予測精度で提案法が従来法より優れていることを示している。例えば、敵対的学習の従来法である AT-MLP (Z.Car et al., 2020) の RMSE 平均値は 27.553 であるのに対して、提案法は 22.893 である。提案した深層準線形カーネル関数を持つ SVR ハイブリッドモデルの有効性が明らかとなっている。

第 5 章では結論として、本研究の成果について総括し、今後の研究課題を論じている。

以上を要約すると、本論文では、準線形カーネル関数という中間モデルの構築を通して、WTA デノイジングオートエンコーダとゲート付線形ネットワークによる欠損値の予測機能と、ゲート制御信号生成によるパーティショニング機能およびマルチローカル線形モデリング機能という 3 つの機能を実現し、欠損値の問題に対処できる準線形カーネル関数を持つ SVR ハイブリッドモデルの構築を行っている。また、提案した準線形カーネル関数を持つ SVR ハイブリッドモデルを複数のベンチマークデータセットや実際の空気品質データの予測問題に適用しその有効性を明らかにしている。これらの成果は機械学習、ニューロコンピューティング分野の発展に寄与するところ大である。よって、本論文は、博士（工学）の学位論文として価値あるものと認める。

2022 年 10 月 19 日

審査員

主査	早稲田大学	教授	博士(情報工学)(九州工業大学)	古月 敬之
副査	早稲田大学	教授	工学博士(早稲田大学)	吉江 修
	早稲田大学	教授	博士(工学)(早稲田大学)	藤村 茂