

科学研究費助成事業 研究成果報告書

令和 4 年 6 月 15 日現在

機関番号：32503
研究種目：若手研究
研究期間：2018～2021
課題番号：18K18112
研究課題名（和文）カーネル平均埋め込みに基づく集合データに対する機械学習フレームワークの構築

研究課題名（英文）Constructing Machine Learning Framework for Set Data Based on Kernel Mean Embeddings

研究代表者
吉川 友也（Yoshikawa, Yuya）
千葉工業大学・人工知能・ソフトウェア技術研究センター・主任研究員

研究者番号：30772040
交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究では、解釈可能な機械学習の問題に対して、集合データに適用可能な機械学習モデルの構築及び、ソフトウェアの開発を行った。機械学習モデルは主に2種類構築した。1つ目はガウス過程に基づく手法で、各事例に対応する局所線形モデルの係数がガウス過程に基づいて生成されるように定式化した、新たなガウス過程回帰モデルを開発した。2つ目はニューラルネットワークに基づく手法で、各事例に対応する局所線形モデルの係数を重要なものからK個生成するニューラルネットワークモデルを開発した。また、ガウス過程回帰に基づく手法については、ユーザが自身の人工知能システムの中で容易に利用できるように、ソフトウェアを開発・公開した。

研究成果の学術的意義や社会的意義

ガウス過程回帰モデルは、カーネル法における教師あり学習の代表的な手法であり、幅広い分野で応用されている。本研究の成果は、従来のガウス過程回帰モデルの予測精度を維持したまま、現在の機械学習システムにおいて必要不可欠な予測結果の解釈可能性を高めるものであり、学術的・社会的の両面で影響を与えるものである。これは、ニューラルネットワークに基づく手法についても同様である。また、ユーザが人工知能システムに本研究の成果を容易に導入できるようにソフトウェアを開発・公開しており、本邦の「人間中心のAI社会原則」で求められている「AIの透明性の確保」の一助になると考える。

研究成果の概要（英文）：We developed machine learning models and a software that can be applied to aggregated data for the problem of interpretable machine learning, which reveals which features contribute to the prediction and to what extent, along with the predictions of target variables. Two main types of machine learning models were developed: the first was based on Gaussian process regression. We developed a new Gaussian process regression model, in which the coefficients (weights) of the local linear model corresponding to each sample are formulated to be generated based on a Gaussian process. The second method is based on neural networks. We developed a neural network-based model that generates the K coefficients (weights) of the local linear model corresponding to each sample from the most important ones. For the developed Gaussian process regression-based model, a software was developed and released so that users can easily use it in their own machine learning systems.

研究分野：機械学習

キーワード：機械学習 解釈可能性 透明性 ガウス過程 ニューラルネットワーク 集合データ カーネル法 人工知能

1. 研究開始当初の背景

集合データとは、1つのデータが特徴の多重集合によって構成されるデータで、機械学習モデルの入力として頻繁に現れる。例えば、文書を単語や N グラムの多重集合で表現したものが集合データである。その他にも、生物情報学ではタンパク質を部分構造の集合で表現したり、グループに対する推薦システムでは、個人の集合をグループとして表現したりしており、集合データが入力の機械学習モデルの適用範囲は極めて広い。

集合データに対しては、カーネル法の枠組みの一つである、「カーネル平均埋め込み」の有効性が多くの研究で示されている。一方で、カーネル平均埋め込みを用いた新しいモデルを開発した場合、そのモデルのパラメータ推定法も導出しなければならず、それがユーザにとっての障壁となる。このような理由から、カーネル平均埋め込みに基づく新たな機械学習フレームワークの開発し、モデルの構築及びパラメータ推定を簡易にすることで、ユーザの負担を減らし、集合データに関連する多くのタスクの性能向上が期待できる。

2. 研究の目的

集合データに対して有効な新たな機械学習モデルを構築し、その機械学習モデルをユーザが容易に利用できるようにするフレームワーク(ソフトウェア)を開発する。

3. 研究の方法

本研究では、具体的なタスクとして、解釈可能な機械学習の問題を考える。このタスクでは、通常の機械学習モデルによる予測とともに、どの特徴が予測にどの程度貢献しているかをユーザに示す。この問題に対して、集合データに対して適用可能な機械学習モデルの構築、及びフレームワークの開発を行う。

4. 研究成果

本研究によって得られた主要な成果を以下に要約する。

(1) ガウス過程に基づく解釈可能な機械学習モデルの開発

ガウス過程回帰モデルは、カーネル法に基づく教師あり学習モデルの一つとして広く研究されている。また、カーネル平均埋め込みに基づくカーネル関数を用いることで、集合データから連続値を予測する集合データ用のガウス過程回帰モデルを自然に導出できる[引用文献]。

ガウス過程回帰モデルでは、カーネルトリックの副作用で、予測結果に対してどの特徴がどの程度貢献しているのかが明確ではない。これは AI システムの透明性の確保が求められる現在においては重大な欠点と言える。この問題に対して、研究代表者は、各事例に対応する局所線形モデルの係数(重み)がガウス過程に基づいて生成されるように定式化した、新たなガウス過程回帰モデルを開発した。表形式データ、画像データ、単語の集合で表現した文書データを用いた実験では、開発したモデルが、従来のガウス過程回帰モデルと同程度の精度を達成しながら、どの特徴が予測にどの程度貢献しているかを忠実に説明できることを実証した。

本研究の論文は、影響力のある国際論文誌(2022年3月時点のインパクトファクター: 10.451)である IEEE Transactions on Neural Networks and Learning Systems に掲載された。

(2) ニューラルネットワークに基づく解釈可能な機械学習モデルの開発

ニューラルネットワークは、機械学習モデルの中で最も高い予測精度を達成できる方法の一つであるが、ガウス過程回帰モデル同様に、予測結果に対してどの特徴がどの程度貢献しているのかがユーザから見えないという問題がある。この問題に対して、研究代表者は、各事例に対応する局所線形モデルの係数(重み)を重要なものから K 個生成するニューラルネットワークを開発した。表形式データ、画像データ、単語の集合で表現した文書データを用いた実験では、本二

ニューラルネットワークで生成される局所線形モデルが、高い予測精度を達成しながら、重要な特徴を適切に K 個選択して説明できることを実証した。

研究代表者は、これまでの研究において、カーネル平均埋め込みをニューラルネットワークで利用することで予測精度が向上することを確認している[引用文献]。研究期間内では実行できなかったが、本技術と組み合わせることで、カーネル平均埋め込みを利用したニューラルネットワークに基づく解釈可能な機械学習モデルも実現可能であると考えます。

本研究の論文は、影響力のある国際論文誌 (2022 年 3 月時点のインパクトファクター: 12.975) である Information Fusion に掲載された。

(3) ソフトウェアの開発と公開

ガウス過程に基づく解釈可能な機械学習モデルについて、ユーザが容易に利用できるようなソフトウェアを開発した。本ソフトウェアは、機械学習コミュニティにおいて標準的に用いられている機械学習ライブラリ「Scikit-learn」に準拠する形で実装しているため、Scikit-learn を利用している多くの機械学習システムで、本ソフトウェアを容易に利用できる。また、本ソフトウェアは深層学習フレームワーク PyTorch を利用して実装されているため、ユーザが自身のモデルの一部としてこれを利用することができる。また、モデルパラメータの推定も PyTorch が提供する自動微分の機能により、ユーザは追加の実装が必要なく行える。本ソフトウェアは <https://github.com/yuyay/gpx> で公開されている。

< 引用文献 >

Yuya Yoshikawa, Tomoharu Iwata, Hiroshi Sawada, “Non-linear Regression for Bag-of-Words Data via Gaussian Process Latent Variable Set Model,” The 28th AAAI Conference on Artificial Intelligence, Austin, Texas, USA, Jan. 2015.

Yuya Yoshikawa, Tomoharu Iwata, “Randomized Kernel Mean Networks for Bag-of-Words Data,” IPSJ Transactions on Mathematical Modeling and Its Applications (TOM), Vol.10, No.3, pp.32-38, Dec. 2017.

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 2件）

1. 著者名 Yoshikawa Yuya, Iwata Tomoharu	4. 巻 81
2. 論文標題 Neural generators of sparse local linear models for achieving both accuracy and interpretability	5. 発行年 2022年
3. 雑誌名 Information Fusion	6. 最初と最後の頁 116 ~ 128
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.inffus.2021.11.009	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Yoshikawa Yuya, Iwata Tomoharu	4. 巻 -
2. 論文標題 Gaussian Process Regression With Interpretable Sample-Wise Feature Weights	5. 発行年 2021年
3. 雑誌名 IEEE Transactions on Neural Networks and Learning Systems	6. 最初と最後の頁 1 ~ 15
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TNNLS.2021.3131234	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 吉川 友也, 岩田 具治
2. 発表標題 スパース局所線形モデルのニューラル生成器
3. 学会等名 第34回人工知能学会全国大会
4. 発表年 2020年

〔図書〕 計0件

〔出願〕 計1件

産業財産権の名称 学習装置, 方法およびプログラム	発明者 吉川友也, 岩田具治	権利者 同左
産業財産権の種類、番号 特許、特願2020017869	出願年 2020年	国内・外国の別 国内

〔取得〕 計0件

〔その他〕

本研究の成果のソフトウェアを <https://github.com/yuyay/gpx> で公開している。

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------