

令和 2 年 6 月 11 日現在

機関番号：17102

研究種目：若手研究

研究期間：2018～2019

課題番号：18K18001

研究課題名（和文）局所パターン学習とは何か：統一的定式化と理論性能保証

研究課題名（英文）Unified formulation and generalization performance of local-feature learning

研究代表者

末廣 大貴（Suehiro, Daiki）

九州大学・システム情報科学研究所・助教

研究者番号：20786967

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）：本研究の主な成果を以下に記す。

(1)局所パターン学習の枠組みを、マルチインスタンス学習と呼ばれる問題の一つとして一般定式化した。(2)局所パターンを用いた分類仮説クラスの汎化性能を理論的に示した。本理論を時系列分類問題における局所パターン学習（時系列解析分野ではShapelet学習と呼ばれる）に適用し、Shapeletと呼ばれる局所パターンを用いた仮説クラスの汎化性能を、世界で初めて示した。(3)導出した学習問題を効率的に解くアルゴリズムを提案した。(4)時系列分類を含む様々なタスクに本手法を適用し、有効性を示した。

研究成果の学術的意義や社会的意義

本研究は、時系列分類問題におけるShapelet学習のような、局所パターン学習問題と、マルチインスタンス学習問題の関連性を世界で初めて示した。従来各ドメインで独立に発展してきた様々な局所パターン学習問題に対し、統一的な定式化、汎化性能保証、解法を与えたことは、機械学習分野に大きな貢献を与えたと言える。また、本研究の骨子となった「マルチインスタンス学習の枠組みに基づく学習問題の一般化」は、局所パターン学習問題に限らない、様々な学習問題に幅広く展開可能であり、新たな研究分野の開拓に期待できる。

研究成果の概要（英文）：The main results are as follows:

(1)We proposed a general formulation of the local-feature-based learning problem by using Multiple-Instance Learning framework. (2)We showed the theoretical generalization performance of the local-feature-based hypothesis class. We applied this theory to Shapelet Learning, which is a popular task in the time-series domain, and we gave the first generalization bound of shapelet-based hypothesis class. (3)We proposed an efficient algorithm to solve the learning problem. (4)We demonstrated that our algorithm effectively works in practice.

研究分野：統計的学習理論

キーワード：局所パターン 機械学習 汎化性能保証 shapelet マルチインスタンス学習

1. 研究開始当初の背景

局所パターンは、画像認識、時系列分類といったデータの分類問題において広く用いられている非常に重要な技術である。例えば、画像内に顔があるかどうか？は、目、鼻、口などの部分的なパーツを局所パターンとすることで認識することができる。従来、局所パターンの取得は、ドメイン知識に依存した何かしらのヒューリスティクスによって学習することが多い。ところが、理論的な背景、解析が欠如しており、汎用化、実用化促進の障害となっている。

2. 研究の目的

本研究では、局所パターン学習問題に対し、次の3点に関する理論基盤 - 統一的定式化、汎化性能の理論的導出、汎化性能の限界達成を可能とする効率的アルゴリズム - を与えることで、ヒューリスティクス限界の打破を目指す。

3. 研究の方法

局所パターン学習の統一的な枠組み

局所パターン学習について、データの種類やドメインによらない統一的な枠組みを、統計的学習理論に基づき明らかにする。統計的学習理論は、多くの学習アルゴリズムの根幹となっている理論であり、具体的には、「局所パターンに基づく分類は、数式でどのように表現できるか」「局所パターンを学習する問題は、どのような最適化問題となるか」を明らかにする。

統計的学習理論に基づく汎化性能

局所パターン学習の枠組みにおいて、どのようなデータを学習すれば、未知データに対してどの程度の予測性能を発揮できるか、を統計的学習理論に基づき評価を行う。具体的には、局所パターンに基づく予測モデルの「Rademacher 複雑さ」(ノイズに対しどの程度頑健か？を測る指標)を理論的に明らかにする。統計的学習理論においては、

[未知データに対する予測誤差 学習データに対する予測誤差 + Rademacher 複雑さ]

が成り立つことが知られており、Rademacher 複雑さを示すことで予測モデルの汎化性能を理論的に保証することができる。

学習アルゴリズムの実用性(汎用性および効率性)

データから局所パターンを学習するアルゴリズムを設計する。すなわちで定式化した問題の解法を最適化理論に基づき明らかにする。また、学習アルゴリズムについて、

汎用性：どのようなデータを対象とできるか、また従来応用対象となっていたドメインをどの程度カバーできるか、

効率性：どの程度の計算量(スピードおよび計算リソース)で実行できるか、を理

論的に明らかにする．また，時系列データ，画像データ等の実データを用いた実験も行い，汎用性，理論性能を保つことができているか，評価を行う．ドメイン知識を用いている既存手法との比較も行い，実用的にも高い性能を持つかどうかを明らかにする．

4．研究成果

本研究の主な成果を以下に記す．

(1)局所パターン学習の枠組みを，マルチインスタンス学習と呼ばれる問題の一つとして一般定式化した．

(2)局所パターンを用いた分類仮説クラスの汎化性能を理論的に示した．本理論を時系列分類問題における局所パターン学習（時系列解析分野ではShapelet 学習と呼ばれる）に適用し，Shapeletと呼ばれる局所パターンを用いた仮説クラスの汎化性能を，世界で初めて示した．

(3)提案した学習問題を効率的（多項式時間）に解くアルゴリズムを提案した．

(4)時系列分類を含む様々なタスクに本手法を適用し，有効性を示した．特に時系列分類タスクにおいては，複数の最新手法と比較した結果，最も良い手法よりもはるかに効率的でありながら，匹敵する精度を残した

なお，成果の大部分は1つの論文にまとめ，Neural Computation に採録決定済みである．医療画像への適用については，国内研究会（医療画像研究会）で発表を行った．また，署名照合やデザイン解析におけるランキング学習の適用，オンライン物体追跡問題など，様々な機械学習のタスク，問題に取り組み，多くの研究成果が得られた．これらの研究を通し，本研究の骨子となった「マルチインスタンス学習の枠組みに基づく学習問題の一般化」が幅広く展開可能であることがわかった（IBIS2019で発表）．

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件／うち国際共著 0件／うちオープンアクセス 1件）

1. 著者名 Daiki Suehiro, Kohei Hatano, Eiji Takimoto, Shuji Yamamoto, Kenichi Bannai, Akiko Takeda,	4. 巻 -
2. 論文標題 Theory and Algorithms for Shapelet-based Multiple-Instance Learning	5. 発行年 2020年
3. 雑誌名 Neural Computation	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Heon Song, Daiki Suehiro, Seiichi Uchida	4. 巻 -
2. 論文標題 Adaptive Aggregation of Arbitrary Online Trackers with a Regret Bound	5. 発行年 2020年
3. 雑誌名 Proceedings of the IEEE Winter Conference on Applications of Computer Vision	6. 最初と最後の頁 681, 689
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 Xiaotong Ji, Yuchen Zheng, Daiki Suehiro, Seiichi Uchida	4. 巻 -
2. 論文標題 Optimal Rejection Function Meets Character Recognition Tasks	5. 発行年 2019年
3. 雑誌名 Proceedings of the 5th Asian Conference on Pattern Recognition	6. 最初と最後の頁 169, 183
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yan Zheng, Yuchen Zheng, Wataru Ohyama, Daiki Suehiro, Seiichi Uchida	4. 巻 -
2. 論文標題 RankSVM for Offline Signature Verification	5. 発行年 2019年
3. 雑誌名 Proceedings of the 15th International Conference on Document Analysis and Recognition	6. 最初と最後の頁 928, 933
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Takuro Karamatsu, Daiki Suehiro, Seiichi Uchida	4. 巻 -
2. 論文標題 Logo Design Analysis by Ranking	5. 発行年 2019年
3. 雑誌名 Proceedings of the 15th International Conference on Document Analysis and Recognition	6. 最初と最後の頁 1482, 1487
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計6件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 角淳之介, 末廣大貴, 加藤貴昭, 内田誠一
2. 発表標題 投手の打ちづらさとは何か ~ 機械学習に基づく投球印象解析 ~
3. 学会等名 スポーツ情報処理時限研究会
4. 発表年 2018年

1. 発表者名 八尋俊希, 末廣大貴, 本館利佳, 鈴木利治, 内田誠一
2. 発表標題 弱教師学習問題における最適局所特徴抽出および樹状突起スパイン検出への応用
3. 学会等名 医用画像研究会
4. 発表年 2019年

1. 発表者名 Daiki Suehiro, Kohei Hatano, Eiji Takimoto, Shuji Yamamoto, Kenichi Bannai, Akiko Takeda
2. 発表標題 Shapelet-based Multiple-Instance Learning
3. 学会等名 情報論的学習理論と機械学習研究会 (IBISML研究会)
4. 発表年 2019年

1. 発表者名 満尾 成亮, 末廣 大貴, 内田 誠一
2. 発表標題 オンラインエキスパート選択問題としての適応的学習率調整
3. 学会等名 画像の認識・理解シンポジウム (MIRU2019)
4. 発表年 2019年

1. 発表者名 ソン ホン, 末廣 大貴, 内田 誠一
2. 発表標題 オンライントラッカの統合について
3. 学会等名 画像の認識・理解シンポジウム (MIRU2019)
4. 発表年 2019年

1. 発表者名 末廣 大貴
2. 発表標題 マルチインスタンス学習への再定式化に基づく理論的汎化誤差導出
3. 学会等名 情報論的学習理論ワークショップ (IBIS2019)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

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
---------	---------------------------	-----------------------	----