

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 Semi-Supervised Classification Based on
Laplacian Kernel Machines Using Quasi-Linear Kernel**

申 請 者

Yanni REN

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 2 月

パターン分類は機械学習において重要な研究課題の一つであるが、実際の分類問題では、半教師あり学習を扱わなければならないことが多い。半教師あり学習では、すべてのデータにラベルが付けられているのではなく、ラベルなしデータも多く含まれる場合を考える。どのようにラベルなしデータを活用して効果的に学習を行うかが問題である。一方、カーネルマシン(Kernel Machine)は、線形分類器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、カーネルマシンでは、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分類問題も特徴空間において線形分類問題として扱うことを可能にしている。従来のカーネルマシンでは、基本カーネルあるいは基本カーネル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、分離境界線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、分離境界線を近似する coarse-to-fine という two-step モデリング法が提案されている。そこでは、two-step モデリング法の coarse-step で、回帰ベクターの内積で定義される中間モデルを構築し、fine-step で、coarse-step で得られた中間モデルを準線形カーネル関数として利用し、カーネルマシン分類器を構築する。このように、教師あり学習または教師なし学習の場合において、体系的なカーネル関数の合成法は確立されており、その有効性も確認されているが、半教師あり学習の場合にどのようにして効果的に学習するかが問題点として残されている。

そこで本研究では、準線形カーネルという中間モデルの構築を通して、半教師あり学習に対応したラプラシアンカーネルマシン分類器の構築を行っている。まず、疑似ラベリング法の適用や局所的線形パーティションの分割を行い、ラベルなしデータを活用した潜在的分離境界線を近似するマルチローカル線形モデルの構築を行う。これによる中間モデルを準線形カーネル関数として利用したラプラシアン SVM (Support Vector Machine) カーネルマシン分類器を構築している(第2章)。また、ラベルガイド付きオートエンコーダに基づいた半教師ありゲーティングメカニズムの構築を通して、間接的に入力空間のパーティショニングを行い、潜在的分離境界線を近似する区分線形モデルの構築を行う。これによる中間モデルを準線形カーネル関数として利用したラプラシアン RLS (Regularized Least Square) カーネルマシン分類器を構築している(第3章)。そして、提案するラプラシアンカーネルマシンを実際に寄生虫画像データの分類問題に適用し、対照学習に基づく深層 CNN (Convolutional Neural Network) 特徴抽出器を含んだ高性能なラプラシアン SVM カーネルマシン分類器を構築している(第4章)。

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、半教師あり学習、カーネルマシン分類器およびカーネル法、準線形カーネルについて述べ、本研究の背景・目的を明らかにしている。

第2章では、ラベルなしデータを活用したマルチローカル線形モデリングによるラプラシアン SVM に基づく半教師あり分類器の構築について述べている。

従来のラプラシアン SVM 分類器では、カーネル関数の形が固定され、分離境界線に関する先見情報があっても利用できないという問題点がある。そこで本論文では、まず、疑似ラベリングアプローチを適用することにより、入力空間を潜在的な分離境界線に沿って複数の局所的な線形分離可能なパーティションに分割する。次に、パーティションに割り当てられた複数のローカル線形モデルを補間することにより、マルチローカル線形モデルを構築し線形回帰形式に定式化する。これによる回帰ベクターの内積として定義された準線形カーネル関数を利用したラプラシアン SVM に基づく高性能な半教師あり分類器を実現している。性能評価のための数値例では、提案された準線形カーネル関数を持つラプラシアン SVM と従来の基本カーネル関数を持つラプラシアン SVM を、ラベル付きデータが 10%しかない 8 個の UCI ベンチマークデータセットに適用して比較し、総合評価 F-score では、1 データセットにおいて両方法が同程度の結果だったが、残りの 7 データセットにおいて提案されたラプラシアン SVM が従来のラプラシアン SVM より優れていることを示している。よって、提案された準線形カーネル関数を持つラプラシアン SVM 分類器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ラプラシアン RLS に基づく半教師あり分類器の構築について述べている。まず、潜在的分離境界線を近似するための区分線形モデリングを行う。区分線形モデルのブレイクポイントを直接的に検出する代わりに、ラベルガイド付きオートエンコーダに基づく半教師ありゲーティングメカニズムを構築して、バイナリゲート制御信号を生成し、間接的にパーティショニングを実現する。区分線形モデルは線形回帰形式として定式化され、線形パラメータは、ゲート制御信号を含むカーネル関数を使用するラプラシアン RLS アルゴリズムによって最適化される。これにラベル付きとラベルなしデータの両方からデータマニホールドを推定し、準線形カーネル関数を通してラプラシアン RLS のカーネルとグラフ構築の類似性関数の両方に組み込んでいる。これにより、高性能なラプラシアン RLS カーネルマシン分類器の構築を行っている。その結果、性能評価のための数値例では、提案された準線形カーネル関数を持つラプラシアン RLS カーネルマシン分類器と従来の基本カーネル関数を持つラプラシアン RLS カーネルマシン分類器を、ラベル付きデータが 5%しかない 8 個の UCI ベンチマークデータセットに適用して比較した結果、総合評価 F-score では、すべてのデータセットにおいて提案された準線形カーネル関数を持つラプラシアン RLS カーネルマシン分類器が従来法より優れていた。特に、‘ESR’ という 5 クラスのてんかん発作認識データセットに対して、従来法のてんかん発作のクラス 1 対非てんかん発作のクラス 2、3、4、5 の F-score はそれぞれ 74.82%、74.35%、82.75%と 79.42%であるのに対して、提案法ではそれぞれ 87.30%、88.38%、92.93%と 97.62%であり、提案された準線形カーネル関数を持つラプラシアン RLS カーネルマシン分類器の有効性が明らかとなっている。

第 4 章では、実際の寄生虫画像データの分類問題に適用したラプラシアン

SVMに基づく半教師あり分類器の構築について述べている。提案された半教師あり分類器は、対照学習によってトレーニングされた半教師あり特徴抽出器と、ラプラシアン SVM によって最適化された半教師あり分類器で構成されている。まず、1) 表現レベルで構造を強化するために類似した明確な意味情報を持つ実際の画像の導入、2) 表現レベルでテクスチャを削除するためのバリエーションの外観変換、これらによって大量のペアサンプルを用意し、対照学習で高性能な深層 CNN 特徴抽出器を構築している。次に、ゲート付き線形ネットワークで分離境界線を近似する区分線形モデルを構築する。モデルの線形パラメータは、深層 CNN 特徴抽出器から生成されたゲート制御信号で構成される準線形カーネル関数を利用して、ラプラシアン SVM アルゴリズムによって最適化される。これにより、高性能な寄生虫画像データのラプラシアン SVM に基づく半教師あり分類器を実現している。性能評価のための数値例では、ラベル付きデータが 1% の場合に、提案されたラプラシアン SVM 分類器は寄生虫画像の認識精度が 95.10% であった。また、最先端の半教師あり分類器 MT(Laine et.al, 2017)、VAT (Miyato et.al, 2018) との比較では、総合評価 F-score においては、MT と VAT はそれぞれ 95.80% と 96.33% であるのに対して、提案法は 96.76% であり、提案された寄生虫画像データ分類のためのラプラシアン SVM に基づく半教師あり分類器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、準線形カーネルという中間モデルの構築を通して、ラベル付きデータとラベルなしデータの両方を利用し、効果的な学習でラプラシアン SVM とラプラシアン RLS というカーネルマシンに基づく半教師あり分類器の構築を行っている。また、提案されたカーネルマシンに基づく半教師あり分類器を複数のベンチマークデータセットや実際の寄生虫画像データの分類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 1 月 13 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