

氏 名 篠原 靖志

学位（専攻分野） 博士（情報学）

学位記番号 総研大甲第 1199 号

学位授与の日付 平成 20 年 9 月 30 日

学位授与の要件 複合科学研究科 情報学専攻
学位規則第 6 条第 1 項該当

学位論文題目 サポートカーネルマシン近似解に基づく高速・高精度な
能動学習

論文審査委員	主 査 教授	高須 淳宏
	教授	佐藤 健
	教授	佐藤 真一
	教授	山田 誠二
	准教授	北本 朝展

論文内容の要旨

本論文は、事例データの学習に基づいて自動判定システムを効率的に構築するために、専門家とシステムのインタラクションによって、学習用事例データベースの効率的構築、判別に有効な属性(カーネル)の組み合わせの選択、高精度の判別関数の学習を平行して行う能動学習手法を提案するものである。

近年の機械学習技術の進歩により、機器状態とその診断結果など「入力」と「ラベル」とが対となったラベル付き学習用事例データベースが存在すれば、「サポートベクトルマシン」等により高精度の判別関数を学習することは比較的容易になっている。しかし、機器状態などの「入力」のデータベース化は容易に行なえても、ラベル付き学習用データベースは用意することが困難な場合が少なくない。各機器状態に対する診断結果など各「入力」の「ラベル」は、専門家等に「ラベル付け」をしてもらう必要があるためである。そこで、極力少数の入力に対するラベル付けで高精度の判別関数を学習できる手法として「能動サポートベクトルマシン」などの能動学習が提案されている。これは、現在までに得られたラベルによる学習結果(判別関数)に基づいて、コンピュータがデータベース中からラベル付けが有用な入力を能動的に選び出し、専門家にラベル付けをしてもらうことを繰り返し、ラベル付き学習用事例データベースを効率的に構築しながら高精度の判別関数の学習を行なう手法である。

ただし、従来の能動学習では、判別に有効な属性(例えば、色か形か)は事前にわかっていると前提であった。しかし、十分な事前知識がなければ、能動学習を始める前に高精度な判別に有効な属性を選定することは本来難しく、能動サポートベクトルマシンでは、適切な属性(厳密には、カーネル)が選ばれていれば非常に効率的に能動学習が進むのに、不適切な属性(カーネル)が選ばれた場合には学習が途中で止まり、効率的な能動学習ができなくなるという課題があった。

そこで本研究では、各入力のラベルだけでなく、判別に有効な属性(カーネル)も不明な状況から出発し、単に高精度の判別結果を得るためだけでなく、判別に有効な属性(カーネル)を見極めるために専門家によるラベル付けを用いる高速で精度の高い能動学習手法を提案する。

本研究での基本的なアプローチは、属性選択を一般化し高精度の判別に有効なカーネルの最適な組み合わせを見つけ出す多カーネル学習の問題と捉え、多カーネル学習アルゴリズムを能動学習化する。これにより、属性選択(正確には、カーネルの組み合わせの選択)と高精度な診断基準の構築の両者に有効な能動学習を実現する。

1章では本研究の背景と目的を述べる。2章では、本研究の基礎となるサポートベクトルマシン、サポートカーネルマシン、能動学習とそのためのサンプリング戦略についての既存研究、サポートベクトルマシンによる能動学習の課題である「カーネル事前選択問題」について述べる。

3章では、「カーネル事前選択問題」解決の基本手法として代表的多カーネル学習手法である「サポートカーネルマシン」を能動学習化した「能動サポートカーネルマシン学習」を提案し、さらに、能動サポートカーネルマシン学習の能動学習時の専門家とのインタラ

クシヨンの高い応答性と最終的に得られる判別関数の判別精度の高さの両立方法について検討する。

サポートカーネルマシンではカーネルの組み合わせの最適化も同時に行なうため判別関数の学習に計算時間を要する。この結果、能動サポートカーネルマシン学習で、能動学習時に毎回、最適な判別関数を再学習してラベル付け対象を選択すると、専門家に次のラベル付け対象を提示するまでの時間が掛かり過ぎる場合が生じる。そこで、能動学習段階ではサポートカーネルマシンの近似解による判別関数を用いてラベル付け対象を選択することで能動学習時の応答性を確保し、能動学習が終了した段階でサポートカーネルマシンを完全に最適化して判別関数の判別精度の高さを確保する2段階能動サポートカーネル学習手法を提案する。さらに、このような近似を行なっても、最適な単一のカーネルを用いたサポートベクトルマシンとほぼ同等の判別精度を持つ判別関数が得られることを実験的に示す。

4章では、能動サポートカーネルマシン学習におけるサンプリング戦略について、サポートベクトルマシンで有効性が知られているRBFカーネルを用いた場合についての検討を行なう。

半径の異なるRBFカーネルを用いる能動サポートカーネルマシン学習において、ラベル付け対象の選択戦略として能動学習で通常用いられるマージン戦略よりも、適切なカーネルの組み合わせにより少ないラベル付けで収束し易い戦略として、SKM-SHIFT戦略を提案する。その基本的考え方は、最適なカーネルの組み合わせが絞り込めていない初期段階では極力多様な入力に対するラベルを得てカーネルの組み合わせの絞込みを重視するのに対して、カーネルの組み合わせが最適値に近づいた後半では判定精度向上に重点を置くというものである。そして、2段階サポートカーネルマシン学習とSKM-SHIFTを組み合わせた能動学習手法の有効性を標準的ベンチマーク問題により示す。

5章では、実問題への適用として、配電柱につけられた腕金と呼ばれる金属部材の表面写真に基づく再利用判定への適用結果の分析により、SKM-SHIFTを用いることで選択されたカーネルの特性から入力データの偏り等を早期に発見できるなど、カーネル選択をも含めて能動学習を行なう利点を示す。

最後に第6章で本論文の成果をまとめる。

本研究の最大の成果は、従来は既知としていた判別に有効な属性（カーネル）の選択を含めて能動学習を行ない、かつ、専門家との円滑なインタラクションを妨げない能動学習時の高い応答性と最終的な判別関数の判別精度の高さとを両立させた実用性の高い能動学習手法を開発した点にある。

論文の審査結果の要旨

篠原靖志君は、機械学習の技術を用いた分類器の一つであるサポートカーネルマシンに対する能動学習の研究を行い、効率の良い学習法を提案するとともに、考案した学習法を配電柱の腕金が再利用可能であるかどうかを判定するという実問題に適用しその有効性を示した。

サポートベクターマシンは、さまざまな問題に対して精度の高い分類を行うことができる分類器として利用されているが、高い精度を実現するためには、適切なカーネルを選択しなければならない。また、サポートベクターマシンを構築するためには、訓練データを事前に用意する必要があり、このコストを低減することが望まれる。前者の問題に対しては、複数のサポートベクターマシンを組み合わせたサポートカーネルマシンと呼ばれる分類器が提案されている。本論文は、サポートカーネルマシンを効率よく学習するための能動学習法を提案している。

能動学習においては、学習の過程でシステムがサンプルを選択し、そのラベルを専門家に問い合わせることが繰り返し行われるが、専門家に問い合わせるべきサンプルを短時間で選択するとともに、少ないサンプルに対する問い合わせで分類性能の高い分類器を得ることが重要になる。サンプルの選択には、それまでに得られた訓練データに基づいて作られた分類器を用いることが多いが、サポートカーネルマシンは、構築のための計算コストが大きく、結果としてサンプル選択に時間がかかり応答性能が劣化するという問題が生じる。そこで、本論文では、サンプル選択のための暫定的な分類器の近似解を高速に求めるとともに、学習の最終段階では得られる分類器の性能に重点をおいて最適解を求める2段階学習アルゴリズムを提案することによって能動学習の応答性と分類性能を確保した。また、サポートベクターマシンの評価のためによく用いられる13種類のデータベースより構成されるベンチマークデータを用いて、提案アルゴリズムが応答性の向上に有効であることを実験的に示した。

能動学習では、分類性能の向上に寄与するサンプルを見つけるサンプリング戦略も重要になる。サポートベクターマシンでは、いくつかのサンプリング戦略が提案されているが、本論文では、これらの戦略のサポートカーネルマシンでの効果について考察し、サポートカーネルマシンの能動学習に適したサンプリング戦略を提案した。この戦略の基本的な考え方は、学習の初期段階では幅広い領域からサンプルを選択し、ある程度学習が進んだ段階で分類器の判別面に近いサンプルを選択し分類器の精緻化を行うものである。この考え方はサポートベクターマシンでも用いられているが、サポートカーネルマシンに適した戦略の切り替え方法を提案しているところに新規性がある。提案したサンプリング戦略を人工データおよびこの分野で評価によく用いられる手書き数字認識のベンチマークデータを用いて評価し、その有効性を示した。

電力会社では、配電網のメンテナンスにおいて電柱の腕金を定期的にチェックして、劣化した腕金を交換している。劣化した腕金は、さらに再利用可能なものと廃棄されるものに分類される。現在は、目視による判定が行われているが、その効率化が求められてい

る。本論文では、提案したサポートカーネルマシンの学習法を腕金の画像データを用いた再利用判定問題に適用し、高い精度で判定できることを示した。特に、サポートカーネルマシンを能動的に学習することによって、学習の早い段階で判定に有効な特徴を見つけることができることを示した。

以上、篠原靖志君の研究は、機械学習の主要な問題のひとつである分類問題において、精度の高い分類器を効率良く獲得することを可能にし、その有効性も標準的なベンチマークデータを用いて示されており、新規性、有効性の高い研究である。本研究の成果の一部は、学術雑誌論文1篇および国際会議論文2篇にまとめられており研究業績も十分である。よって、篠原靖志君の論文は学位（情報学）を授与するに値する、と審査委員会一致で判断した。