

機関番号：17102

研究種目：若手研究 (B)

研究期間：2009～2010

課題番号：21700171

研究課題名 (和文) バイアスを用いた最大マージン疎超平面学習の高速化

研究課題名 (英文) Efficient Learning of Maximum Margin Sparse Hyperplanes with Bias

研究代表者

畑 堯平 (HATANO KOHEI)

九州大学・システム情報科学研究所・助教

研究者番号：60404026

研究成果の概要 (和文)：

1 ノルムソフトマージン最適化問題とは疎な分類器を学習する主なアプローチの1つである。この問題では、マージンを大きくするような線形分類器を重みの1ノルムを正則化することによって求める。本研究では、1 ノルムソフトマージン最適化問題における疎性に着目し、新しいブースティング手法 Sparse LPBoost を提案した。人工データおよび実データにおける準備的実験において、提案手法 Sparse LPBoost が、従来手法である線形計画ソルバや LPBoost よりも1 ノルムソフトマージン最適化問題をより高速に解くことを示した。

研究成果の概要 (英文)：

The 1-norm soft margin optimization is a popular formulation for obtaining sparse classifiers. We propose a new boosting algorithm based on linear programming. Our algorithm can take advantage of the sparsity of the solution of the underlying optimization problem. In preliminary experiments, our algorithm outperforms a state-of-the-art LP solver and LPBoost.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	2,200,000	660,000	2,860,000
2010年度	1,200,000	360,000	1,560,000
年度			
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：機械学習，計算学習理論

科研費の分科・細目：情報学・知能情報学

キーワード：機械学習，計算学習理論，オンライン予測，ブースティング，ソフトマージン最適化，サポートベクターマシン

1. 研究開始当初の背景

ソフトマージン最適化問題は、機械学習において広く用いられる定式化である。この問題は与えられた複数の正例・負例を「なるべく」分離するような超平面を求める問題の一つである。この問題の定式化は統計的学習理

論の知見に基づいており、多くの応用問題に対して予測能力の高い仮説が得られることが知られている。大きく分けてソフトマージン最適化問題は2種類ある。1つ目の2ノルムソフトマージン最適化問題は2次計画問題であり、SVMなどの効率的解法が存在す

る。また、この問題に対してはカーネル法を適用できるため、よく用いられる。

一方、1ノルムソフトマージン最適化問題は線形計画問題に属する。1ノルムソフトマージン最適化問題の解は疎になるという性質があり、特徴選択の面からも近年注目されている。この問題に対してはブースティングなどの解法が提案されているが、2ノルムソフトマージン最適化問題に比べ研究が進んでおらず、さらなる効率化の余地が残されている。また、近年この問題に対しては、半正定値カーネルを含む一般的な「類似性指標」を用いた学習の枠組にも有効であることがわかってきており、今後さらなる応用が期待される。

2. 研究の目的

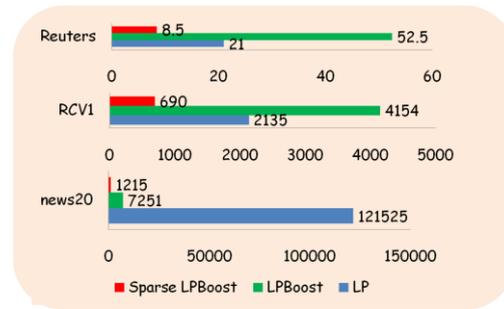
本研究の目的は、大規模データにも適用可能な効率のよい1ノルムソフトマージン最適化問題の解法の開発である。

3. 研究の方法

本研究では、1ノルムソフトマージン最適化問題の解の疎性を利用する。この問題においては、多くの場合、ごく少数のデータのみが重要で、他の大多数のデータは最適解に関与しない（つまり、大多数のデータがなくとも最適解を求めることができる）。したがって、「重要な」データのみを選択して解くことで大幅な高速化が可能になる。

4. 研究成果

(1) 1ノルムソフトマージン最適化問題の解法に関する研究を行った。本研究では、1ノルムソフトマージン最適化問題における疎性に着目し、新しいブースティング手法 Sparse LPBoost を提案した。提案手法は重要そうな事例や仮説のみを選択し、より小さい線形計画問題解くことを繰り返すことにより、高速化をはかる。人工データおよび実データにおける準備的な実験において、提案手法 Sparse LPBoost が、従来手法である線形計画ソルバや LPBoost よりも1ノルムソフトマージン最適化問題をより高速に解くことを示した。特に、事例や仮説の総数が1万から100万以上であるような大規模なデータにおいて、Sparse LPBoost は他の手法に比べて数倍から100倍の高速化を達成した（右図参照）。この結果は国際会議 Discovery Science で受理されると共に、情報論的学習理論ワークショップにおいて、プログラム委員会特別奨励賞ファイナリストに選ばれた。



(2) 楕円型のサポートベクターマシンの定式化および解法に関する研究を行った。機械学習における代表的な学習手法の一つであるサポートベクターマシン (SVM) は線形制約を満たす最大超球を求める問題の解と見なすことができる。また、この観点に沿った拡張であるベイズポイントマシン (BPM) は線形制約内の重心を求める手法であり、ベイズ的な仮定の下、最適な学習手法であることが示されている。しかし、BPM の計算量は大きいため大規模データには適さない。そこで、本研究では、線形制約内の最大超楕円を求める問題を定式化した (楕円型サポートベクターマシン)。この定式化により、SVM よりも頑健な BPM の近似を得ることを目指している。本定式化では、カーネルを用いる事も可能である。さらに、本研究では、SMO アルゴリズムに基づくカーネルを用いた解法および、オンライン凸最適化に基づくカーネルを用いない解法を提案した。実データ上において、提案手法は SVM に比べてより頑健な予測精度を示し、また BPM よりも高速であった。本研究結果は、国際会議 Asian Conference on Machine Learning に受理された。

(3) 部分文字列パターン上の1ノルムソフトマージン最適化に関する研究を行った。テキストの分類において、テキストに出現する部分文字列は大きな役割を果たす。実際、部分文字列に対応した仮説を複数組み合わせることにより、高い分類精度を得られる。しかしながら、与えられたテキスト中の部分文字列は元のテキストのサイズの2乗オーダー存在し、ナイーブな方法では分類に有効な部分文字列の探索に計算時間を要する。そこで、本研究では文字列の索引構造の一つである接尾辞配列を用いることにより効率的な探索手法を開発した。本研究結果は、国際会議 Discovery Science に受理された。

また、その他の研究として、(4) 非類似度関数を用いた機械学習アルゴリズムの研究 (Wang らとの共同研究)、(5) AIBO ロボットのスキル発見における試行の削減手法の開発 (小林らとの共同研究)、(6) ランキングのオンライン予測手法の研究、(7) 非定常データストリームのためのオンライン

予測手法の開発, を行った. (4), (5) についてはそれぞれ Neural Computation および人工知能学会論文誌に掲載された.

さらに, (8) 機械学習を用いた将棋プログラム, (9) モンテカルロ探索手法の研究を行った. さらに人工知能学会研究会優秀賞を受賞した.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 5 件)

1. Michinari Momma, Kohei Hatano, and Hiroki Nakayama,
“**Ellipsoidal Support Vector Machines**”,
Proc. of the 2nd Asian Conference on Machine Learning (ACML 2010),31-46, 2010.

1. Kazuaki Kashihara, Kohei Hatano, Hideo Bannai, and Masayuki Takeda,
“**Sparse Substring Pattern Set Discovery using Linear Programming Boosting**”,
Proc. of the 13th International Conference on Discovery Science, 132-144, 2010.

2. Kohei Hatano and Eiji Takimoto,
“**Linear Programming Boosting by Column and Row Generation**”,
Proc. of 12th International Conference on Discovery Science, 401-408, 2009.

3. Liwei Wang, Masashi Sugiyama, Cheng Yang, Kohei Hatano, and Jefu Feng,
“**Theory and Algorithm for Learning with Dissimilarity Functions**”,
Neural Computation, vol. 21, No. 5, 1459-1484, 2009.

4. 小林隼人, 畑埜晃平, 石野明, 篠原歩,
“**間引き: ロボットのスキル発見における評価の削減手法**”,
人工知能学会論文誌, vol. 24, No. 1, 191-202, 2009.

[学会発表] (計 11 件)

1. 安武翔太, 畑埜晃平, 来嶋秀治, 瀧本英二, 竹田正幸
“**Online Prediction over Permutahedron**”,
情報処理学会 第 134 回アルゴリズム研究会 (SIGAL), 2011.

2. 奥山洋平, 畑埜晃平, 瀧本英二, 竹田正幸
“**確率的評価値をもつゲーム木における最善手探索**”,
冬の LA シンポジウム, 2011.

3. 末廣大貴, 畑埜晃平, 坂内英夫, 瀧本英

二, 竹田正幸

“**カーネル法を用いたコンピュータ将棋の評価関数の学習**”,
第 15 回ゲームプログラミングワークショップ 2010 (GPW2010), 2010.

4. 安武翔太, 畑埜晃平, 瀧本英二, 竹田正幸

“**オンラインランク統合問題**”,
第 13 回情報論的学習理論ワークショップ (IBIS2010), 2010.

5. 末廣大貴, 畑埜晃平, 坂内英夫, 瀧本英二, 竹田正幸

“**SVM によるバイパータイトランキング学習を用いたコンピュータ将棋における評価関数の学習**”,
第 13 回情報論的学習理論ワークショップ (IBIS2010), 2010.

6. 樫原和昭・畑埜晃平・坂内英夫・竹田正幸

“**Sparse Substring Pattern Set Discovery using Linear Programming Boosting**”,
第 9 回情報科学技術フォーラム (FIT 2010), 2010.

7. 安武翔太, 畑埜晃平, 瀧本英二, 竹田正幸

“**Online Rank Aggregation**”,
第 9 回情報科学技術フォーラム (FIT 2010), 2010.

8. 阿南陽子, 畑埜晃平, 坂内英夫, 竹田正幸

“**楽曲の分類と文字列間の類似性指標について**”,
人工知能学会 人工知能基本問題研究会 (SIG-FPAI), 第 78 回, 2010.

9. 吉田真一, 畑埜晃平, 瀧本英二, 竹田正幸

“**重みつき窓を用いた適応型オンライン予測**”,
人工知能学会 データマイニングと統計数理研究会 (SIG-DMSM) 第 12 回, 55-64, 2010.

10. 安武翔太, 畑埜晃平, 瀧本英二, 竹田正幸

“**オンラインランク統合問題**”,
冬の LA シンポジウム, 2010.

11. 畑埜晃平, 瀧本英二,
“**行と列の生成による線形計画ブースティング**”,
第 12 回情報論的学習理論ワークショップ (IBIS2009), 2009.

〔図書〕（計0件）

〔産業財産権〕

○出願状況（計0件）

○取得状況（計0件）

〔その他〕

2010 2009年度人工知能学会研究会優秀論文賞

2009 情報論的学習理論ワークショップ IBIS2009 プログラム委員会特別奨励賞ファイナリスト

ホームページ等

<http://www.i.kyushu-u.ac.jp/~hatano/>

6. 研究組織

(1) 研究代表者

畑埜 晃平 (HATANO KOHEI)

九州大学大学院・システム情報科学研究所・助教

研究者番号：60404026

(2) 研究分担者

(3) 連携研究者

該当なし。