

研究種目：若手研究（B）
 研究期間：2007～2010
 課題番号：19710124
 研究課題名（和文） リスク最小化に基づく非凸型識別手法の開発と与信審査・医療診断問題への実証的適用
 研究課題名（英文） Nonconvex classification method based on risk minimization and its application to credit approvals and medical diagnosis
 研究代表者
 武田 朗子（TAKEDA AKIKO）
 慶應義塾大学・理工学部・講師
 研究者番号：80361799

研究代表者の専門分野：複合新領域

科研費の分科・細目：社会・安全システム科学 社会システム工学・安全システム

キーワード：OR, 数理最適化, 金融工学

1. 研究計画の概要

サポートベクターマシーン（Support Vector Machine: SVM）は、あらかじめ分類されたデータに基づいて分類ルールを学習し、そのルールを適用して新たなデータを分類する手法であり、現実の問題への応用において優れた性能を持つことが報告されている。特に、 ν -SVM [Scholkopf 等,2000]はパラメータ選択が他の SVM に比べて容易なため、注目を集めている。

本研究では、統計的学習分野で広く研究が進められている SVM について、数理最適化の知見を生かした研究を行なっている。

(1) ν -SVM の拡張モデルである Ev-SVM [Perez-Cruz 等,2003]に対し、汎化誤差（新たなデータに対する予測誤差）を最も小さくするような判別関数が Ev-SVM から得られることを理論的に示す。

(2) Ev-SVM モデルは非凸計画問題として定式化されるため、厳密解を求めるのは難しく、単純な局所最適解法が提案されているに留まっている。Ev-SVM を近似的に解くためのソフトウェア、厳密に解くためのソフトウェア、両方の開発を行なう。分類手法のためのデータベースが整っており、小・中規模データ集合を用いて、Ev-SVM の有効性の検証とともに考案解法の性能評価を行いたい。最終的には、データベースから特に大規模な実データ集合（医療データやクレジット会社の顧客データ）を用いて、与信審査・医療診断問題への実証的適用を行なう。

(3) Extended ν -SVM を金融データに適用することを試みる。観測データと同じ分布（未知の分布）より得られた新しいデータに対して、もっとも信頼性の高い予測結果を与

えるモデルを構築することを目的にしている。

2. 研究の進捗状況

「1. 研究計画の概要」で述べた項目について、進捗状況を報告する。

(1) 2007年度は Ev-SVM の理論的性能評価を中心に研究を行った。Ev-SVM に対して、金融業界でよく知られたリスク尺度（CVaR: Conditional Value-at-Risk）に基づく新しい解釈を与えた。また、Rockafellar & Uryasev 等による CVaR の理論的研究成果を Ev-SVM に適用することにより、Ev-SVM の汎化誤差の上界式を導出し、Ev-SVM による CVaR 最小化がその上界式を最小化する解を与えることを証明した。

(2) Ev-SVM は、 ν -SVM よりも性能のよい分類器を構築することが示されているが、問題が非凸二次計画問題となるため、扱えるデータ集合の規模は、現時点では実用上十分な大きさに達していない。そこで、2007～2008年度は、既存の反復解法を有限回の反復で終了するように改良し、より規模の大きな問題が解けるようになった。また、アルゴリズムの局所最適性についても示した。さらに、大域的最適化アルゴリズムを考案し、厳密解を求めることのできるソフトウェアを開発した。

(3) 2008～2009年度は Ev-SVM の金融データへの適用を中心に研究を行なった。

統計的学習の理論に基づいて、将来、予測されるリスクを最小化するようなポートフォリオ最適化モデルを構築し、効率的な解法の提案を行なった。

多くの従来モデルは、いずれも、過去のデ

ータに対して最も良い意思決定が行われるように構築されていた。しかし、過去と同じ状況がこれから生じるわけではないため、過去のデータについて最適なモデルがこれからのデータにうまく適合するとは限らない。そこで、統計的学習の知見を取り入れて、汎化能力の高いポートフォリオ最適化モデルを構築し、日経225の過去データ10年分を用いて、他の標準モデルに比べて平均収益が高く、収益のぶれの低いポートフォリオが得られたことを確認した。

3. 現在までの達成度

当初の計画以上に進展している

(理由)

当初の研究計画に挙げていた項目にすべて取りかかり、様々な成果が得られた。当初は大域的最適化アルゴリズムの考案までは目的としていなかったが、共著者との研究が進み、アルゴリズムの考案・ソフトウェアの開発まで行なうことができた。

4. 今後の研究の推進方策

昨年度の研究から、CVaRよりもVaR (value-at-risk) リスク尺度を最小にする方が汎化誤差の上界式が小さくなることが分かっており、VaR最小化モデルの方がv-SVMのCVaR最小化モデルよりもよい予測精度を達成するのではないかと期待できる。そこで、今後は分類モデルとしてVaR最小化モデルを提案したい。VaR最小化問題はCVaR最小化問題に比べて、解くのが難しいことが知られている。そこで、今後は、モデルの提案、問題を解くためのアルゴリズムの考案、ソフトウェアの開発を行い、既存モデルと比べてどれ位予測精度が向上するか検証を行いたい。

(1) VaR最小化モデルの提案:

VaRリスク尺度を用いた分類モデルを提案する。今まで分類のみを研究対象としていたが、回帰、外れ値検出に対しても同様に、汎化誤差の上界値 (VaRを用いたもの) の導出が可能であり、VaR最小化問題を構築できると考えている。よって、研究対象を分類だけでなく回帰、外れ値検出まで広げて、VaR最小化モデルを提案したい。

(2) アルゴリズムの考案:

VaR最小化問題はCVaR問題に比べて、解くのが非常に難しい。そこで、厳密に解くのではなく、それなりにより近似解が早く得られるようなアルゴリズムを考案したい。近似誤差を理論的に評価できるようなアルゴリズム構築を目標にしている。

5. 代表的な研究成果

(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 7件)

(1) A. Takeda, S. Taguchi and T. Tanaka, "A Relaxation Algorithm with a Probabilistic Guarantee for Robust Deviation Optimization", *Computational Optimization and Applications*に掲載予定, 査読有.

(2) A. Takeda and M. Sugiyama, "On generalization performance and non-convex optimization of extended v-support vector machine", *New Generation Computing*, 27, pp. 259-279 (2009), 査読有.

(3) A. Takeda, "Generalization Performance of nu-Support Vector Classifier Based on Conditional Value-at-Risk Minimization", *Neurocomputing*, 72 (10-12), pp. 2351-2358 (2009), 査読有.

(4) A. Takeda and T. Kanamori, "A Robust Approach Based on Conditional Value-at-Risk Measure to Statistical Learning Problems", *European Journal of Operational Research*, 198 (1), pp. 287-296 (2009), 査読有.

(5) J. Gotoh and A. Takeda, "Conditional Minimum Volume Ellipsoid with Applications to Multiclass Discrimination", *Computational Optimization and Applications*, 41 (1), pp. 27-51 (2008), 査読有.

[学会発表] (計 16件)

[図書] (計 0件)

[産業財産権]

○出願状況 (計 0件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

○取得状況 (計 件)

名称:
発明者:
権利者:
種類:
番号:
取得年月日:
国内外の別:

[その他]