

機関番号：32612  
 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

## 研究成果の概要（和文）：

数理最適化、金融工学、機械学習の3つの異なる分野の知見を活かして、機械学習の判別モデル（主に Ev-SVM）に対する理論的な妥当性評価を行った。“判別能力の良さは経験的に知られていたが、理論的に示せたのは初めてである”と評価された研究成果である。また、Ev-SVMの解法を提案した。さらに、機械学習で知られる“正則化”と呼ばれるアイデアを金融工学へ取り入れて、将来へのあてはまりのいいポートフォリオ（資産配分）決定方法の提案を行った。

## 研究成果の概要（英文）：

Using the existing studies on mathematical optimization, financial engineering and machine learning, I theoretically evaluated the prediction performance of a classification method known as Ev-SVM. The SVM has been quite successful in practice. However, no satisfactory theoretical background existed so far. We provided such background and also explain how this nonconvex optimization problem can actually be solved. Moreover, we adopted the concept of “regularization term” that is often used in machine learning for portfolio optimization problems in financial engineering and succeeded in enhancing the prediction performance of portfolio optimization models.

## 交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	800,000	0	800,000
2008年度	800,000	240,000	1,040,000
2009年度	900,000	270,000	1,170,000
2010年度	700,000	210,000	910,000
年度			
総計	3,200,000	720,000	3,920,000

## 研究分野：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. 研究開始当初の背景

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、v-SVM [Scholkopf 等,2000]は他のSVMに比べてパラメータ選択が容易なため、注目を集めている。

SVM の学習モデルでは、訓練データへのオーバーフィットを避けてテストデータに対する分類性能を高めるため、“正則化”と呼ばれるテクニックを取り入れている。これは、識別誤りに対して罰金を与える損失関数  $f(x)$  に正則化項 (例えば、二次項  $|x|^2$ ) を加えることを意味する。パラメータ  $C$  を用いて “ $f(x)+C \times$  正則化項” の目的関数を最小化することで、テストデータをうまく識別するような識別関数が得られることを期待している。それらのモデルは凸二次計画問題として定式化され、問題の特徴を利用した高速解法の提案、そしてその解法を実装したソフトウェアの提供もなされている。

一方で、正則化項導入の理論的な根拠はあまり示されていない。研究開始当初、申請者らによって、 $v$ -SVM モデルを内包する広いクラスの Extended  $v$ -SVM (E $v$ -SVM) モデル [Perez-Cruz 等,2003] が、金融業界でよく知られたリスク尺度  $\beta$ -CVaR (Conditional Value-at-Risk) の最小化モデルと等価であることが示されていた。その研究成果を踏まえて、“CVaR リスク最小化” という新しい解釈を用いて、SVM モデルの正則化項の妥当性を理論的に示せるのではないかとという予想が得られていた。そこで、本研究課題として、CVaR リスク最小化という新しい解釈を用いて、E $v$ -SVM モデルの妥当性を理論的に示すことに取り組むことにした。

## 2. 研究の目的

本研究課題では、申請当初から (1)、(2) の目標を掲げていたが、4 年の研究期間の間に、さらに (3)、(4) の研究目的を掲げることができた。

(1)  $v$ -SVM の拡張モデルである E $v$ -SVM [Perez-Cruz 等,2003] に対し、そのモデルの妥当性を示す。具体的には、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 E $v$ -SVM から得られることを理論的に示す。

(2) E $v$ -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。E $v$ -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、E $v$ -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  $v$ -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。ポートフォリオ (資産配分) 最適化問題に対して汎化誤差を考慮したモデルは初めてのものであり、汎化誤差の評価可能なモデルとして注目を集めることが期待される。

(4) 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。よって、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、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。

## 3. 研究の方法

「2. 研究の目的」で挙げた項目について、研究の方法を挙げる。

(1) 既存の汎化誤差 (新たなデータに対する予測誤差) の評価式を、金融分野のリスク尺度  $\beta$ -CVaR を用いて再評価することにより、CVaR 最小化モデルと等価である E $v$ -SVM モデルに対しての妥当性を示す。

(2) 非凸計画問題として定式化される E $v$ -SVM モデルに対して、Perez-Cruz 等 [2003] は局所最適解を求めるためのアルゴリズムを提案した。しかし、有限回の反復で終わる保証はなく、求解に時間がかかっていた。より規模の大きな問題が解けるようにするため、反復解法を有限回の反復で終了するように改良する。

また、E $v$ -SVM モデルに対して提案されていない、大域的最適解を求めるための解法を提案する。具体的には、切除平面法を取り入れた解法を考案する。数値実験により、大域最適化による解と局所最適解を比較し、予測精度に有意な差があるかを確認する。

(3) (1) の研究において、E $v$ -SVM モデルと金融分野のリスク指標 CVaR との関連を示し、CVaR に関する理論研究成果を用いて E $v$ -SVM の妥当性を示した。ここでは、CVaR 最小化を用いた、既存のポートフォリオ最適化モデルに対して、SVM で使われる正則化項を加えることを提案する。統計的学習分野では、正則化項は予測精度を高める効果が知

られている。正則化項をポートフォリオ最適化モデルに取り入れることにより、将来の予測精度が高まるかどうかを調べる。

(4) 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する。

#### 4. 研究成果

「2. 研究の目的」で挙げた項目について、研究成果を報告する。

(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。

得られた研究成果は、2 本の査読付き論文誌と 1 本の査読付き国際会議論文誌で報告済みである。

(2) Ev-SVM は、v-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。また、大域的最適化アルゴリズムを考案し、時間はかかるものの厳密解を求めることのできるソフトウェアを開発した。

図 1 はデータ集合から肝臓疾患を判別するため、提案アルゴリズムを適用した結果である。(a) の図は、非凸最適化問題を解くことにより、予測精度が高まることを示している。また、(b) は大域最適解を得るために、大域最適化アルゴリズムは何回の切除平面が構築したかを示している。パラメータ  $n$  を小さく設定することにより、Ev-SVM の定式化の非凸性が増し、計算の手間がかかることが分かった。また、大域最適解でなくても、最初に得られた局所最適化で十分精度よい判別がなされることが分かった。

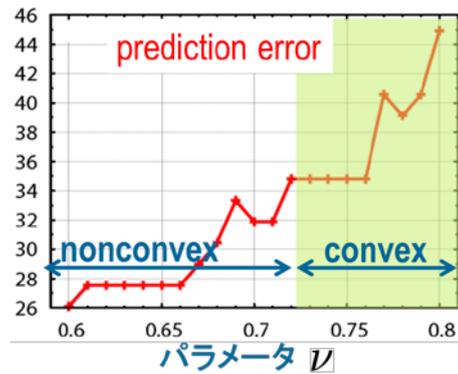


図 1 (a)

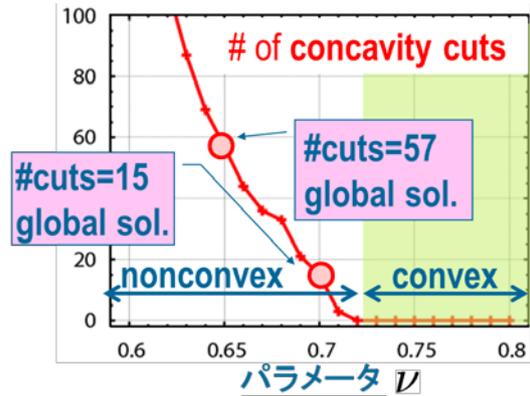


図 1 (b)

得られた研究成果は、1 本の査読付き論文誌と 1 本の国際会議論文誌で報告済みである。

(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。

図 2 は提案モデル (NCCVaR) と他の標準モデル (CVaR, Abs, Sqr) との比較を、MSE (mean squared error) の 95%tile と平均値 (mean) を用いて示している。95%tile, mean のどちらの指標を用いても、NCCVaR は他モデルに比べて予測精度が高いことがわかる。

得られた研究成果は、2 本の査読付き論文誌に投稿中、1 本の国際会議論文誌で報告済みである。

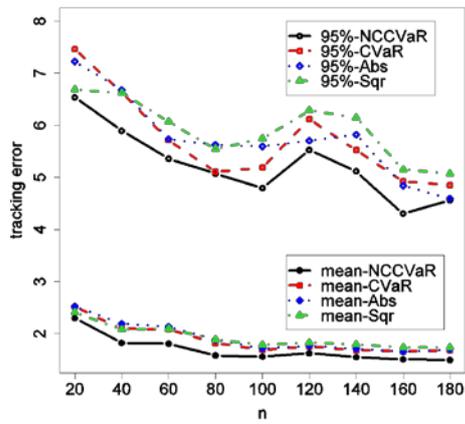


図 2

(4) 2009～2010 年度は VaR 最小化モデルの構築、アルゴリズムの考案を中心に研究を行った。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、それなりによい近似解が早く得られるようなアルゴリズムを考案した。分類・回帰・外れ値検出問題に対して、VaR 最小化モデルと CVaR 最小化モデルの予測精度比較を行った。

現在は論文としてまとめつつ、近似解の精度がよりよくなるよう、アルゴリズムを再検討している段階である。

### 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 11 件)

(1) A. Takeda, S. Taguchi and T. Tanaka, "A Relaxation Algorithm with a Probabilistic Guarantee for Robust Deviation Optimization", *Computational Optimization and Applications*, 47(1), pp. 1-31 (2010), 査読有.

(2) A. Takeda and M. Sugiyama, "On generalization performance and non-convex optimization of extended  $\nu$ -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 21 件)

(1) A. Takeda, J. Gotoh and M. Sugiyama, "Support Vector Regression as Conditional Value-at-Risk Minimization with Application to Financial Time-series Analysis", 2010 IEEE International Workshop on Machine Learning for Signal Processing, (Kittila, Finland) 2010年 8月 31日.

(2) A. Takeda and M. Sugiyama, "Non-convex Optimization of Extended  $\nu$ -Support Vector Machine", 20th International Symposium of Mathematical Programming, (Chicago, USA) 2009年 8月 24日.

[その他]

ホームページ等

<http://www.ae.keio.ac.jp/lab/soc/takeda/takeda/research-j.html>

### 6. 研究組織

#### (1) 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

#### (2) 研究分担者

該当なし

#### (3) 連携研究者

該当なし