

## サポートベクターマシンを用いた意識調査データの 分析に関する基礎的研究

その他（別言語等） のタイトル	A Basic Study on Attitude Survey Analysis using Support Vector Machines
著者	長谷川 裕修, 藤井 勝, 有村 幹治, 田村 亨
雑誌名	土木計画学研究・講演集
巻	36
号	332
ページ	1-2
発行年	2007-11
URL	<a href="http://hdl.handle.net/10258/1769">http://hdl.handle.net/10258/1769</a>

## サポートベクターマシンを用いた意識調査データの分析に関する基礎的研究

その他（別言語等）のタイトル	A Basic Study on Attitude Survey Analysis using Support Vector Machines
著者	長谷川 裕修, 藤井 勝, 有村 幹治, 田村 亨
雑誌名	土木計画学研究・講演集
巻	36
号	332
ページ	1-2
発行年	2007-11
URL	<a href="http://hdl.handle.net/10258/1769">http://hdl.handle.net/10258/1769</a>

# サポートベクターマシンを用いた意識調査データの分析に関する基礎的研究\*

## A Basic Study on Attitude Survey Analysis using Support Vector Machines\*

長谷川裕修\*\*・藤井勝\*\*\*・有村幹治\*\*\*\*・田村亨\*\*\*\*\*

By Hironobu HASEGAWA\*\*・Masaru FUJII\*\*\*・Mikiharu ARIMURA\*\*\*\*・Tohru TAMURA\*\*\*\*\*

### 1. はじめに

国民のニーズを把握し、行政運営に活用するためにアンケート調査等の意識調査が行われている。

本研究は、非線形判別が可能なサポートベクターマシン (Support Vector Machines : 以下 SVM と記す) の意識調査データ分析への適用を試みるものである。2章で SVM の概略を述べ、3章で意識調査データ分析に SVM を用いる意義と課題を整理する。その後4章でケーススタディを行い、5章で本研究のまとめを述べる。

### 2. サポートベクターマシンについて<sup>1),2)</sup>

SVM は、1960 年代に Vapnik らが提案した最適超平面識別器を基礎とする二値判別手法である。最適超平面識別器は線形分離可能な問題に対しては高い性能を示したが、非線形な問題に対応できないため実問題への応用は進まなかった。しかし、最適超平面識別器は 1990 年代に Vapnik 自身によってカーネル法と組み合わせられることで非線形判別が可能な SVM として拡張された。この拡張により SVM は最も認識性能の優れた手法の一つと言われ、画像認識、手書き文字認識、テキストマイニング、バイオインフォマティクス等多様な分野で利用が広がっている。

図-1に SVM の概念図を示す。SVM を用いたパターン識別は、2つのクラス  $y_1$ 、 $y_2$  に属するデータ (図中の赤丸と青四角) を分離する識別関数  $f(x)$  を求める問題となる。分離可能な識別関数は無数にあるが、データの存在する領域の限界面  $H1$ 、 $H2$  とデータを分離する超

\*キーワード: サポートベクターマシン, 意識調査分析, データマイニング

\*\*学生員, 工修, 室蘭工業大学大学院 工学研究科 博士後期課程 建設工学専攻 (北海道室蘭市水元町27番1号、TEL&FAX0143-46-5289)

\*\*\*正員, 工修, 室蘭工業大学大学院 工学研究科 博士後期課程 建設工学専攻

\*\*\*\*正員, 工博, (株) ドーコン 交通部

\*\*\*\*\*フェロー, 工博, 室蘭工業大学工学部 建設システム工学科

平面間の距離  $1/\|w\|$  を最大化させるような識別関数  $f(x)$  を求める。

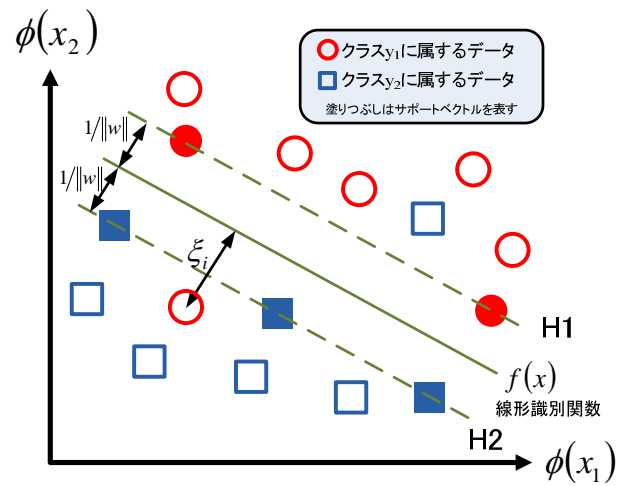


図-1 SVM 概念図

### 3. 意識調査データ分析にSVMを用いる意義<sup>1),3)</sup>

SVM を意識調査データ分析に用いる理由と問題点を以下に述べる。

1つめの理由は未知データに対する判別能力 (汎化能力) の高さである。一般に、変数を増やすと既知データに対する判別能力は向上するが、汎化能力は低下する (次元の呪いと呼ばれる)。SVM はマージン最大化基準に基づいて判別関数を求めることによりこの問題を克服している。汎化能力の高さは識別結果の一般性を向上させるものと思われる。

2つめの理由はエラーデータへの頑健性である。SVM は入出力がそろったデータ (教師データと呼ばれる) から識別関数を学習するが、教師データの中でも境界面に近い小数のデータ (サポートベクター) のみから学習する。サポートベクター以外のデータを取り除いて学習しても同じ識別関数が得られる (この性質を疎性: スパースネスという)。回答者の勘違い、記入ミス、集計ミス等による異常値が解析結果に影響を与えにくいといえる。

3つめの理由は多重共線性への対応である。意識調

査データで得られる個人属性には収入と年齢等、互いに高い相関（多重共線性）を持つものが含まれることが多い。多変量解析手法を用いて分析する場合、いずれかの説明変数を取り除かなければ解析不能となる。しかし、SVMをはじめとしたカーネル法を用いれば多重共線性があってもこれらのデータを解析可能である。すなわち、これまで変数選択によって捨てていた情報を使用することが可能となる。

4つめの理由は計算量的な効率の良さである。SVMは入力データを非線形の多次元特徴空間に写像し、特徴空間内で線形判別を試みる手法である。このため、データ数が増えると計算量が爆発的に増大し学習が困難になる。これに対して SVM はカーネルトリックを用いることで計算量の増大を抑えている。また、SVM の学習は凸二次計画問題に帰着するため、局所的最適解が大域的最適解に一致する点も優れている。

上に述べたとおり、SVM は優れた性質を持つ。しかし、SVM を意識調査データ分析に適用するに当たり解決すべき問題がある。すなわち、SVM によって得られる判別関数が判別の結果のみを与えるブラックボックスであるため、各変数の判別への寄与を解釈することが困難であるという問題である。土木計画分野における代表的な判別手法である判別分析、数量化Ⅱ類の適用目的の多くは判別に対する変数の寄与を把握することであり、この問題の解決は重要である。

筆者ら<sup>4)</sup>は SVM を用いた交通事故分類において、主成分分析によって変数を3変数まで減らし、3次元の主成分得点空間上で判別関数を可視化することで変数の影響を解釈した。可視化によって問題の直感的な理解が可能となったが、1) 判別に使用した第3主成分までの累積寄与率は60.4%に過ぎず情報損失が多い、2) 課題の把握は可能だが具体的な対策の立案には向かない等、問題が多い。一方、矢島ら<sup>5)</sup>は SVM による判別関数が特徴空間内では線形であることに着目し、特徴空間において判別への寄与度を求める方法を提案している。

#### 4. ケーススタディ

人々の行動や考え方に影響を与えるイベント（行事や施策）の前後に意識調査を実施し、属性を説明変数、行動結果や意思を判別基準として SVM を用いれば、そのイベントがどのような人々に影響を与え易いか、という傾向が把握できる。具体的にはモビリティ・マネジメントにおける施策の影響範囲の予測などのマーケティング的な利用が可能だと考えられる。

ケーススタディとして苫小牧市で実施された防災講演参加者に対して行ったアンケート調査結果の分析を行う。表-1に調査の概要を示す。講演の前後で参加者の平均知識量（13点満点）は5.95点から8.90点まで上昇した。回答者属性を説明変数、防災知識量を分類基準として SVM による判別分析を行った。なお、分析結果は発表時に示す。

表-1 アンケート調査概要

調査実施日	2005年11月29日
調査対象	苫小牧市平成17年度 特別研修 樽前山のことを知ろう 参加者 (苫小牧市職員)
アンケート形式	会場アンケート調査（講演の 前後2回実施）
調査項目	防災知識、防災意識等
配布数	42
有効回答	35
有効回答率	84%

#### 5. おわりに

本研究は意識調査データ分析への SVM 適用を検討するものである。本研究の成果は以下の2点である。

- 1) 意識調査データ分析へのSVM適用時の利点と問題点を整理した
- 2) パネルデータを用いた施策評価を提案した

#### 参考文献

- 1) N. Cristianini & J. Shawe-Taylor : An Introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, 2000.
- 2) 栗田多喜夫：サポートベクターマシン入門、産業技術総合研究所 脳神経情報研究部門 Web サイト、<http://staff.aist.go.jp/takio-kurita/index-j.html>
- 3) 石井健一郎、上田修功、前田英作、村瀬洋：わかりやすいパターン認識、オーム社、1998.
- 4) 長谷川裕修、藤井勝、有村幹治、田村亨：サポートベクターマシンの土木計画学への適用に関する研究、第35回土木計画学研究発表会・講演集、vol.35、CD-ROM、2007.
- 5) 矢島安敏、阿部哲郎：非線形SVMによる属性の抽出、年日本オペレーションズ・リサーチ学会秋季研究発表会アブストラクト集、vol.2003、pp.290-291、2003