

交通事故分析へのサポートベクターマシンの適用に関する基礎的検討*

A Basic Study on Traffic Accident Analysis by Support Vector Machine *

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

我が国の交通事故は昭和 45 年に交通事故死者数が最悪を記録したが、その後、道路構造面、車両面、法制度面などでの交通安全対策が逐次推進されたことから死者数は年々減少し、平成 17 年には昭和 45 年の約 4 割迄減少した。一方、昭和 44 年以降減少傾向を示していた事故件数は、昭和 52 年から再び増加に転じ、平成 17 年には昭和 44 年の 1.3 倍にまで増加している。このように、死者数は減少傾向にあるものの、死亡事故の引き金となる事故件数は年々増加傾向を示し、交通事故は依然として大きな社会問題である。交通事故の要因としては、運転者を取り巻く環境や車両の安全性、走行環境に起因するものなど多種多様な要因が挙げられる。このうち、運転者面や車両面については法制度の充実、取り締まりの強化や車両の安全性の向上により一律に事故軽減が期待できる。しかし、気象環境、地形条件、交通需要、トリップ特性など、地域特性が大きく関わる走行環境面については、その都度、効果的・効率的な対応策が求められているのが現状である。

交通安全対策を検討する上で、死傷事故率の高い事故危険箇所を抽出し、その要因を詳細に分析することは効果的ではあるが、要因が複雑多岐に絡み合っている大量な交通事故データを効率的・効果的に扱うのは難しい問題であった。そこで、本研究では事故分析に関する既往の研究をレビューし、データマイニングの一手法として近年注目されているサポートベクターマシン (Support Vector Machine:以下 SVM を記す) を取り上げ、交通事故データ分析への適用を試み、その有効性と課題について検討を行った。

*キーワード: 交通事故分析,SVM

**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、TEL0143-46-5289、FAX0143-46-5289)

***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

****正員,工博,(株)ドーコン 交通部(札幌市厚別区厚別中央1-5-4-1 TEL 011-801-1520,FAX 011-801-1521)

*****正員,工博,室蘭工業大学工学部建設システム工学科

2. 既往研究のレビューと本研究の位置づけ

交通事故の分析手法としては、衝突実験によるものと実際の交通事故データに基づく統計解析によるものがあるが個別具体的な事故の要因と事故発生箇所における対策について検討するうえでは前者は比較優位である。しかし、多大な労力を要すること、さらには大量のデータの中に存在する、特有な傾向や因果関係などの情報を見つけ出すことは難しく、類似した事故の発見なども含め交通事故原因の包括的理解と対策の立案にはそぐわない。このため大量の事故データの包括的な判別分析といった観点からは統計的解析による方法がとられている。

交通事故総合データなどのマクロデータを用い、交通事故分析を行っている既存研究事例としては、森らによる研究¹⁾²⁾、後藤らによる研究³⁾、平澤らによる研究⁴⁾が挙げられる。このうち森らは、交通事故対策と事故削減との関連性分析¹⁾や、交通事故の発生状況(事故密度)を歩行者、自転車、自動車交通を説明変数とした予測を試みている²⁾ものの、いずれも道路構造や沿道用途毎の各カテゴリー別での分析であり包括的に判別する分析とはなっていない。また、後藤らによるリスクマップを使用した研究³⁾では、地域ごとの特性を反映した交通事故発生リスクは明示されているものの、事故対策に結びつく事故発生要因とリスクの関連性は定量化されていない。平澤らは事故対策を立案するためのツールとしてGISを活用した事故分析システムを開発している。事故発生箇所に着目し、個別・具体的な分析を行う上では有効ではあるが、類似した事故への対応も含め地域の実情を勘案した事故対策方針立案等の分析ツールとしては課題が多い⁴⁾。

一方、交差点面積、交通安全対策の実施内容など交通事故の多様な要因を整理した交通事故要因データベースを用いて、事故発生メカニズムを明示的に定式化した既存研究事例としては、岐阜社会基盤研究所によりファジーニューロモデルを活用した交通事故類型区分別推計モデルが提案されている⁵⁾。このモデルは交通事故発生に関する知識の蓄積を言語変数により多様に表現される推論ルール群として表現するファジー推論により交通事故件数と交通事故要因との因果関係系をある程度明確にす

ることは期待できるが、交通事故類型毎の発生要因のメカニズムを知識表現面からルール化するため、ルールが恣意的になりやすいことが課題として挙げられている。

日々蓄積される事故データは、殆どの場合、例えば事故類型データのシェアから重点施策が検討される等、集計的に用いられているものと推測される。しかし、マクロ的な交通事故減少の動向を伺うことができても、ミクロな事故対策の立案には現在の事故データは質的に不十分であることはいうまでもない。

データマイニング手法を交通事故分析に用いることを研究する直接の動機は、交通事故データとして道路構造や運用方法、事故当時の車両挙動記録等、ミクロな事故データが収集される場合の効率的な分析方法を用意しておくことにある。本研究において適用を検討する SVM は、Vapnik らが、1960 年代に提案した Optimal Separating Hyperplane を基礎とする二値判別手法であり 1990 年代になって Vapnik 自身により、カーネル関数を組み込むことで非線形判別にも対応できるモデルとして拡張された。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る⁹⁾。未知データに対する識別性能の高さにより、SVM は事故対策診断モデルの基本アルゴリズムとなる可能性を持つ。本研究の目的は、SVM による交通事故分析への適用方法の基礎的な検討、及び課題整理にある。

3 . SVMの概要

SVM は、線形しきい素子を用いて 2 クラスのパターン識別器を構成する手法である。SVM の概念図を図 - 1 に示す。データ群 と を分離する識別関数 $f(x)$ を求める問題になるが、これはデータの存在する領域の限界面 H_1 、 H_2 とデータを分離する超平面間の距離 $1/\|w\|$ を最大化させるような $f(x)$ を求める問題を意味する。このとき、データを 2 クラスに完全分離できる場合をハードマージン、一部分離できない場合をソフトマージンという。後者について以下に簡単に説明する。

まず、 $x_i (i=1-\ell)$ で表される N 個の成分 (入力) と、クラスラベル $y_i \{-1$ あるいは $1\}$ から成るトレーニングデータが与えられているとする。

このとき、ソフトマージン最適化の問題は (1) 式で定義される。

$$\begin{aligned} \min_{w, \xi} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\ell} \xi_i, w \in R^N, b \in R, \xi \in R^{\ell} \\ \text{subject to} \quad & y_i(w \cdot x_i + b) \geq 1 - \xi_i, i=1, \dots, \ell \\ & \xi_i \geq 0, i=1, \dots, \ell \end{aligned} \quad (1)$$

ここで、 w は入力に対する重みベクトルである。

この最適化問題 (主問題) に Lagrange 乗数 α_i および μ_i を導入することで (2) 式で表される双対問題が得

られる。

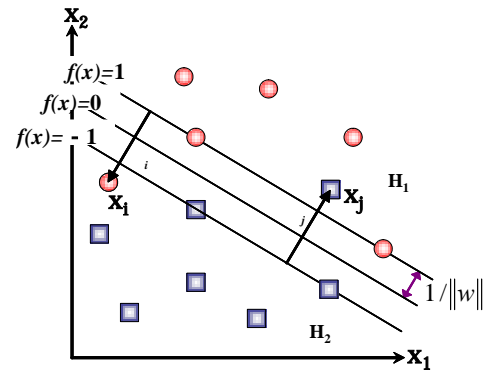


図 - 1 SVM の概念図

$$\begin{aligned} \text{Maximize}_{\alpha} \quad & L_D(\alpha) = \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{subject to} \quad & \sum_{i=1}^{\ell} \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, i=1, \dots, \ell \end{aligned} \quad (2)$$

より複雑な識別を可能とするために、曲面による分離を考える。

まず、入力データ x_i を高次元の特徴空間に写像する。

$$x \leftrightarrow \phi(x) = (\phi_1(x), \phi_2(x), \dots) \quad (3)$$

(2) 式内の特徴空間に対応する量はベクトルの内積で表される値であり、カーネル関数で置換することができる。この置換をカーネルトリックという。これを (2) 式に適用し、以下の最適化問題が得られる。

$$\begin{aligned} \text{maximize}_{\alpha} \quad & W(\alpha) = \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j K(x_i, x_j), \alpha \in R^{\ell} \\ \text{subject to} \quad & \sum_{i=1}^{\ell} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i=1, \dots, \ell \end{aligned} \quad (4)$$

このとき、バイアス b は (5) 式で得ることができる。

$$b^* = y_i - \sum_{j \in S_V} \alpha_j^* y_j K(x_j, x_i) \quad (5)$$

ここで、 S_V はサポートベクターの集合、 j は任意のサポートベクターを表す。結局、識別関数 $f(x)$ は (6) 式となる。

$$f(x) = w^{*T} x_i + b^* = \sum_{j \in S_V} \alpha_j^* y_j K(x_j, x) + b^* \quad (6)$$

さて、バイアス b を固定値とすると (4) 式の等号条件は消え、問題は 2 次関数の最大化問題となる。最急降下法を用いて、次式で を更新すれば最適解を求めることが出来る。

$$\begin{aligned} \alpha_i & \leftarrow \min \left(C, \max \left(0, \alpha_i + \eta \frac{\partial W(\alpha)}{\partial \alpha_i} \right) \right) \\ \alpha_i & = \frac{1}{K(x_i, x_i)} \quad (0 \leq \alpha_i \leq 2), i=1, \dots, \ell \end{aligned} \quad (7)$$

ここで、 α は収束比を表す。この最適化問題の収束判定は、主問題と双対問題の目的関数の値の比較により行う。

$$Proportion = \frac{\text{主目的関数値} - \text{双対目的関数値}}{\text{主目的関数値} + 1}$$

$$= \frac{\sum_{i=1}^{\ell} i - 2W(\alpha) + C \sum_{i=1}^{\ell} i}{\sum_{i=1}^{\ell} i - W(\alpha) + C \sum_{i=1}^{\ell} i + 1} \leq \quad (8)$$

$$i = \max \left\{ 0, 1 - y_i \left(\sum_{j=1}^{\ell} y_j K(x_i, x_j) + b \right) \right\}, (i=1, \dots, \ell)$$

4. SVMを用いた事故分析

(1) 分析方法の概要

本研究で提案する事故分析フローの概要を図-2に示す。まずターゲットとなる事故を抽出し、その上で主成分分析により次元を集約した上でSVMにより重大事故を判別する。

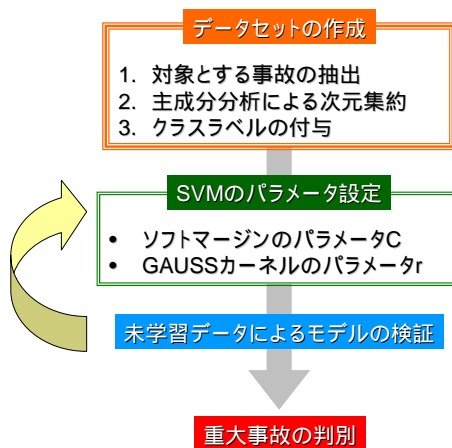


図-2 事故分析フロー

(2) データの前処理

a) 交通事故データ

交通事故を集計したデータとしては、(財)交通事故総合分析センターが提供する交通事故統計データ(ITARDAデータ)が一般に利用可能である。ITARDAデータでは、事故・行動類型、事故による被害の程度、当事者に関する情報、事故発生地点の道路形状や路面状態、沿道環境、地域特性、などが集計されている。

b) データセットの作成

本研究では、平成11年度から16年度に北海道渡島・松山支庁で発生した交通事故約7500件の中から、1) DID地区の交差点部で発生、2) 加害者年齢65歳以上、を条件として、エラーデータを除いた149件を抽出し、SVMの判別問題の例題として設定した。各事故に関連付けられた項目全てが事故発生要因となる訳ではない。そのた

め、要因となりうる10項目を選択した。さらに、道路の問題点を直感的に把握するために主成分分析を用いて要因を縮約した(表-1)。分析の結果、第3主成分までで累積寄与率60%を超えたため、要因数3とした。

表-1 主成分分析の結果

要因および主成分の解釈	主成分 1	主成分 2	主成分 3
主成分 1: 走行環境			
事故発生月	0.003	0.067	-0.587
路面状態	-0.065	-0.319	0.429
平日昼夜率	0.046	0.222	-0.325
主成分 2: 道路構造			
車線数	-0.414	-0.398	-0.124
車道幅員	-0.459	-0.262	-0.113
歩道代表幅員	-0.470	-0.130	-0.078
中央帯設置延長	-0.397	0.317	0.191
平日昼間 12 時間交通量	-0.420	0.218	-0.01
主成分 3: 利用特性			
指定最高速度	-0.188	0.568	0.407
平日昼間 12 時間	-0.148	0.361	-0.359
大型車混入率			
寄与率 (%)	35.177	13.804	12.170

本研究での識別クラスは重大事故(死者、重傷者発生)とそれ以外の軽傷事故の2クラスとした。

主成分分析によって得られた主成分得点を入力 x_i とし、各データにクラス分けのための指標 $y_i \{-1, 1\}$ を付与した(図-3)。

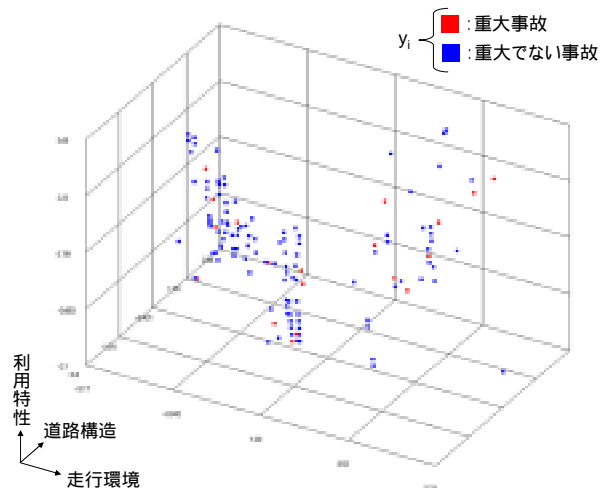


図-3 データの分布

(3) モデルパラメーターの設定

前節で作成したデータセットの中から、 $y=1$ のデータ10個、 $y=-1$ のデータ10個、合計20個を抽出しトレーニングデータとした。収束比 $=1.0$ 、収束条件 $=10^{-4}$ 、最大繰り返し数は 8×10^6 とした。

なお、カーネル関数 $K(x_i, x_j)$ は(9)式で表されるGaussカーネルを用いた。

$$K(x_i, x_j) = \exp\left[-\frac{\|x_i - x_j\|^2}{2r^2}\right], \quad i, j = 1, \dots, \ell \quad (9)$$

SVMの識別能力にはソフトマージンのパラメータ C とGaussカーネルのパラメータ r が影響するが、これは繰り返し計算によって適当な値を求める必要がある。本研究では、試行した中で最も良い結果が得られた $C=50$ 、 $r=0.1$ を用いることとした。

(4) 未学習データによるモデルの検証

データセットからトレーニングデータを除いた129データを検証用データとし、作成したSVMを用いて識別した。結果を図-4に示す。図中の緑色から赤色のデータが死亡・重傷事故、緑色から青色のデータが軽傷事故を示す。

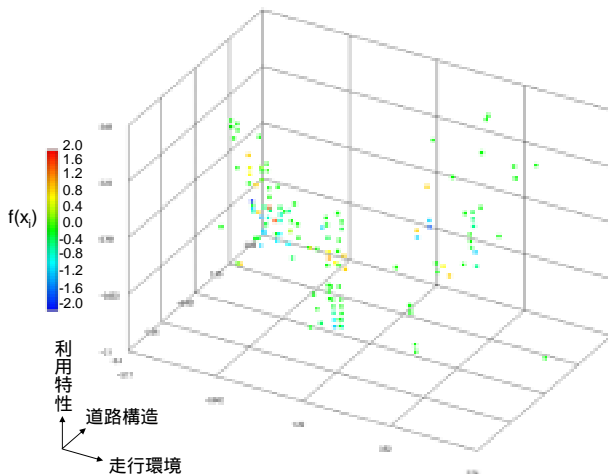


図-4 SVMによるデータの分離結果

未学習データに対するSVMの判別率は例題サンプル全129データ中73%の94データに関して正しく判別できた。死亡・重傷事故については全10サンプル中7サンプルについて判別できた。

5. おわりに

本研究では例題としてITARDAデータを用いて死亡事故・重大事故とそれ以外の軽微な事故についてSVMにより判別を試みた。本研究で明らかとなった点としては、まず、事故データは離散量と連続量が混在するデータであり、単純にSVMを適用することは望ましくないことが挙げられる。本研究においては、まず分析対象を高年齢者DID交差点事故データとして抽出し、その上で事故データの分布に対して主成分分析を実施し、走行環境・道路構造・利用特性の三軸の連続量に変換した。また死

亡・重大事故の割合はサンプルのデータセットの約1割程度であり、トレーニングデータセットを作成するには少ないため、死亡・重大事故以外のデータを分割し、死亡事故・重大事故のデータ数と同数の軽微な事故データセットを作成した。その上で、複数のトレーニングデータを作成し、SVMにより判別関数を得た。本研究においては、未学習データに対するSVMの判別率は約7割程度であり、少ない死亡・重大事故データに関しても同様の判別率を示す判別関数を得ることができた。この識別関数に対して主成分量による感度分析を行うことで、死亡・重大事故の発生可能性のある領域を推測することが可能であり、これは今後の課題である。

本研究においては既存の事故統計データに対してSVMをデータマイニングの一手法として適用しているが、マクロデータに対して分析を試みても、マクロな施策の立案にしか使えないジレンマを筆者らは感じている。交差点“形状”の特徴量や、車両の交差点進入速度等、連続量として収集可能なデータはSVMによる分析によく合うものと考えられ、ITS社会におけるデータ分析手法としてのSVMの可能性は大きいものと考えられる。

参考文献

- 1) 森,池田: 道路事業による交通事故削減効果推計手法の検討, 第4回 道路新技術会議資料3-13, pp.17~20,2003.
- 2) 森,池田,宮下: 交通事故データ等による事故要因の分析,国土交通省国土技術政策総合研究所: 「平成17年度国総研研究マップ」, (<http://www.nilim.go.jp/lab/bcg/siryou/tnn/tnn0253pdf/ks025319.pdf>)
- 3) 後藤,斎藤,平澤: 交通事故リスクの分析・評価方法に関する研究,平成17年度土木学会北海道支部論文報告集,第62号, 2006.
- 4) 平澤,高田,浅野: 交通事故分析システムの開発について, 第47回北海道開発局技術研究発表会論文集,2003年度
- 5) 岐阜社会基盤研究所: 平成17年度交通安全に着目した道路構造および動連施設整備に関する評価報告,岐阜大学工学部社会基盤工学科 (<http://www.crcr.pref.gifu.jp/kibanken/kenkyu/H17seika/h17koutuu.pdf>)
- 6) N. Cristianini & J. Shawe-Taylor: An Introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, 2000. (邦訳:大北剛:サポートベクターマシン入門,共立出版, 2005.)
- 7) 国土交通省自動車交通局 自動車運送事業に係わる交通事故要因分析検討会: 自動車運送事業に係わる交通事故要因分析,平成17年度, (<http://www.mlito.go.jp/kisha/kisha06/09/090501/02.pdf>)