

237-20115548 統計的機械学習を用いたプローブカーデータからの ヒヤリハット発生形態の推定*

森村 哲郎¹⁾谷澤 悠輔²⁾山崎 慎也³⁾井手 剛⁴⁾

Vehicle near-miss situation prediction from probe-car data using statistical machine learning

Tetsuro Morimura

Yusuke Tanizawa

Shinya Yamasaki

Tsuyoshi Idé

Providing a driver with information on risky events in advance is crucial for safer and more comfortable driving. This paper presents an approach to automated classification of near-miss situations for providing more useful information to drivers. Specifically, based on probe-car data, we introduce a statistical machine learning approach to the classification task. Experimental results show that our method is capable of identifying effective features from the data and is promising in near-miss pattern classification.

KEY WORDS: safety, drive recorder, near-miss analysis, probe-car data, statistical machine learning [C1]

1. はじめに

周囲のあらゆる対象物に対して常に細心の注意を払いながら運転し続けることは非常に困難であり、ドライバは何にどの程度注意を払うべきかを過去の経験や知識、周囲の状況などから予測的に判断しつつ運転していると考えられる。しかし、見通し不良地点や複雑な交差点などの特殊な道路環境においては、注意を向けるべき場所や対象物を適切に判断することが難しいため、事故やヒヤリハット（事故になりかけたが回避できた事象）が発生しやすくなる可能性がある。従って、そのような危険性の高まる場所において、どのような事象に注意しなければならないかを予め教示してくれるサービスがあれば、ドライバはより安心、安全に運転を楽しむことができると考えられる。

近年、車の ICT が進化し、走行中の様々な情報を収集可能なクルマ（プローブカー）が実現可能となり、事故やヒヤリハットが発生した時のドライバ操作や車両挙動等の情報を収集できるようになった。特にヒヤリハットは事故よりも多く発生していると考えられることから、ヒヤリハット発生時の情報（ヒヤリハット情報）をプローブカーで収集して解析すれば、事故の危険性が高まる場所やその場所で注意すべき事象を迅速かつ網羅的に特定できる可能性があり、上記サービスを実現する上で有効と考えられる。

2. 本稿の目的

プローブカーで収集したヒヤリハット情報を安全運転へ活用するためには、まず、それらの発生状況を分類、整理する必要がある。先行研究^{(1),(2),(3),(4)}においては、専任の担当者によって一件ずつ発生状況を確認しながら分類するということが行われる。この方法は、複雑な状況に対しても柔軟に対応できる反面、客観的な判断基準の統一化が難しい点や、処理工数が大きい点が課題である。また、山崎ら⁽⁵⁾はヒヤリハットの発生状況をプローブカーの走行軌跡を用いて直進時、左折時、右折時に分類し、その発生頻度を解析したが、ドライバにとって受容性の高い支援情報を提供する上では、より詳細な発生状況の分類が必要と思われる。特に、右折時の対向直進車とのヒヤリハット（右直ヒヤリハット、Fig. 1 (A)）と、右折先横断歩道上の歩行者や自転車とのヒヤリハット（右折先ヒヤリハット、Fig. 1 (B)）については、同じ右折中のヒヤリハットであっても注意すべき対象や場所が大きく異なることから分離の必要性が高いと考えられるが、現状ではヒヤリハット発生時の映像を目視して判断する以外に有効な分類方法は無い。

しかし、右直ヒヤリハットと右折先ヒヤリハットでは、右折行動の中での発生タイミングが若干異なることから、ヒヤリハット発生前後の車両挙動やドライバ操作等の時系列データを解析すれば、双方を分類する上で有効な知見や特徴を見出し高精度な分類モデルを構築できる可能性がある。このようにデータから何らかの特徴を見つけ出し分類する手法として統計的機械学習法^{(6),(7)}がある。これは、領域知識を用いた取り組みや試行錯誤で閾値を設定し分類するといった手段と

* 2011年10月14日自動車技術会秋季学術講演会において発表。

¹⁾⁻⁴⁾ IBM 東京基礎研究所 (242-8502 神奈川県大和市下鶴間 1623-14)

²⁾⁻³⁾ マツダ (株) (221-0022 神奈川県横浜市神奈川区守屋町 2-5)

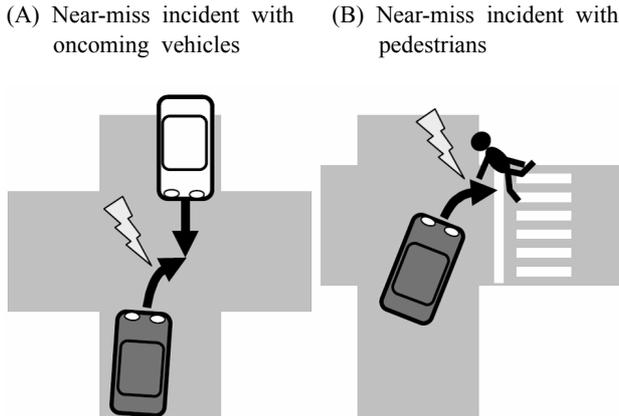


Fig. 1 Near-miss situations in right turn

は異なり，データから統計的に分類モデルを構築することを目指している．統計的機械学習による分類では，モデル構築用のデータに含まれない事例や特徴までは考慮できないが，データ数が増えるに従い，そのような特徴は少なくなり，さらに領域知識からは説明できない事象や想定外の事象等に対して頑強になることが期待される⁽⁶⁾．そのため，プローブカーデータのようにデータ数の多い，もしくは今後増加の見込まれる対象においては，統計的機械学習による取り組みは有効な手段であると考えられる．

そこで本稿では，ヒヤリハット発生前後の時系列データを用いて右直ヒヤリハットと右折先ヒヤリハットとを分類する手法として，統計的機械学習法に着目し，その適用可能性について評価する．本稿の構成は，以下の通りである．まず3節では解析に用いたプローブカーデータについて解説し，4節では右折ヒヤリハット発生形態の分類方法として領域知識に基づく手法と統計的機械学習による手法を導出する．5節では，数値実験により各手法を比較し統計的機械学習法が有効であることを示し，特に有効であったロジスティック回帰法により抽出された特徴についても議論する．

3. 解析に用いたヒヤリハット情報

3.1. 節では解析に用いるプローブカーで収集したヒヤリハット情報について説明し，3.2. 節ではそのプローブカーデータの様相を可視化することで調査する．

3.1. プローブカーデータの概要

本稿では，(社)自動車技術会が構築した「ヒヤリハットデータベース」⁽¹⁾のヒヤリハット情報を用いた．これは東京都内と静岡県内のタクシー約200台にドライブレコーダを搭載して収集されたもので，専門スタッフにより一定のルールに従って発生状況や危険性の度合いに応じて分類，整理されている⁽¹⁾．

ヒヤリハットデータベースの各標本（もしくはインシデントや事例とも呼ばれる）には下記のヒヤリハットレベルや，

ヒヤリハット発生前後の車両挙動や運転操作の時系列データが記録されている．ヒヤリハット情報の危険性の度合は，高いものから「高レベル」「中レベル」「低レベル」「反応レベル」に分類されている．今回は明確に事故の危険性が確認できる「高レベル」「中レベル」の右直ヒヤリハットと右折先ヒヤリハットをそれぞれ80件ずつ計160件を抽出し，解析および評価に用いることとした．

車両挙動や運転操作のデータとして，ヒヤリハット発生の10秒前から5秒後までを1/30秒間隔で記録された種々のセンサーデータ値が記録されている．つまり，データの記録時間は15秒であるため，各センサーデータは450(=15[sec]×30[sec])の観測点からなる時系列データである．センサーデータの種類は，GPSや，速度(speed)，縦加速度(longitudinal acceleration)，横加速度(lateral acceleration)，ブレーキ(brake)，左ウィンカー(winker-L)，右ウィンカー(winker-R)である．本解析では，右折とは関係のない左ウィンカー情報は使用していない．また，GPSデータについても，その観測精度から右折時のヒヤリハット発生形態の分類には利用できないと考えられるため用いていない．

以下では，簡単のため，変数(センサーデータ)名をラベル $s \in \{A, B, C, D, E\}$ を用いて表記する． $s = A$ は速度， $s = B$ は縦加速度， $s = C$ は横加速度， $s = D$ はブレーキ， $s = E$ は右ウィンカーとする．全160件の標本は $n \in \{1, \dots, 160\}$ で区別し，第 n 番目の標本の変数 s の時系列データを $\mathbf{x}^{(n,s)} \triangleq [x_1^{(n,s)}, \dots, x_T^{(n,s)}]^\top$ と表記する．ただし $^\top$ は転置である． T は時系列データの観測点数であるため $T = 450$ である．また，ヒヤリハット発生形態はクラスラベル $c_n \in \{1, 2\}$ で区別する． $c = 1$ は右直ヒヤリハット， $c = 2$ は右折先ヒヤリハットとする．従って，各標本 n のプローブカーデータは $\{X_n, c_n\}$ となる．ただし $X_n \triangleq [x^{(n,A)}, \dots, x^{(n,E)}] \in \mathbb{R}^{450 \times 5}$ である．

3.2. プローブカーデータの可視化

各変数の時系列データについて，標本それぞれもしくは各クラス内の平均と標準偏差を可視化し，Fig. 2 に示した．また，時系列データの時間差に関する傾向を調べるため，自己共分散係数列⁽⁸⁾についても計算し，Fig. 3 に示した．時系列データ $\mathbf{x}^{(n,s)}$ の時間差 $k/30$ 秒の自己共分散係数は

$$\sigma(k, \mathbf{x}^{(n,s)}) = \frac{1}{T} \sum_{t=k+1}^T (x_t^{(n,s)} - \bar{x}^{(n,s)})(x_{t-k}^{(n,s)} - \bar{x}^{(n,s)})$$

で与えられる．ここで， $\bar{x}^{(n,s)}$ は $x_t^{(n,s)}$ の標本平均 $\bar{x}^{(n,s)} \triangleq \frac{1}{T} \sum_{t=1}^T x_t^{(n,s)}$ である．自己共分散係数列 $[\sigma(0, \mathbf{x}^{(n,s)}), \dots, \sigma(T-1, \mathbf{x}^{(n,s)})]$ は時系列データの周波数成分と関連付けられるため，ヒヤリハット発生形態の差異による周波数領域における傾向の違いの有無を調べるのに適している^{(8),(9)}．

Fig. 2 と Fig. 3 の可視化結果から、右直ヒヤリハットと右折先ヒヤリハットの右折時ヒヤリハット発生形態の違いにより、速度や横加速度の時系列の様相、および横加速度の自己共分散係数列の様相は特に異なることがわかる。よって、これらの特徴は、右折時ヒヤリハット発生形態分類において、有効な特徴になると考えられる。一方で、速度やブレーキの自己共分散係数列については、右直ヒヤリハットと右折先ヒヤリハットによる傾向の違いはほとんど確認できない。

4. ヒヤリハット発生形態の分類方法の導出

4.1. 節で閾値設定による右折時のヒヤリハット発生形態の分類方法、4.2. 節で統計的機械学習による分類方法を与える。

4.1. 領域知識に基づく指標による分類

右直ヒヤリハットと右折先ヒヤリハットとでは交差点内でのヒヤリハット発生位置が異なることから、ヒヤリハット発生時の交差点内位置を同定できれば、ヒヤリハット発生形態の推定が可能になる。つまり、交差点内の位置の違いによって差異が表れる運転行動に着目し、そのような運転行動と相関が高い指標を求め、適切な閾値により判別することで、精度良く分類できる可能性がある。このような指標の内、プローブカーデータから導出可能なものとして「右折旋回開始からヒヤリハット発生までの時間」が考えられる。これは速度 v と横加速度 a を用いて車両の回転半径 r

$$r = \frac{v^2}{a}$$

を求め、右折旋回に伴う有効な回転半径が継続して発生し始めてからヒヤリハットが発生するまでの時間として見積もることが可能である。右折旋回をしていない場合の回転半径は、車両の振動や計測誤差等で正と負の値がランダムに発生している為、ヒヤリハット時に最も近い継続的に負の値が発生し始めたタイミングを右折旋回開始として考える。

具体的には、各標本 n のヒヤリハット発生までの各時間ステップ $t \in \{1, \dots, 299\}$ の前 1 秒間の平均回転半径

$$r_n(t) = \begin{cases} 0, & \text{if } t = 1 \\ \sum_{k=\max(1, t-30)}^{t-1} \frac{(x_k^{(n,A)})^2}{30 x_k^{(n,C)}}, & \text{otherwise} \end{cases}$$

を計算し、 $r_n(t) \leq 0$ を満たす最も大きい時間ステップ t^* からヒヤリハット発生時間ステップ $t^h = 300$ までの所要時間

$$\tau_n = \frac{t^h - t^*}{30} = 10 - \frac{t^*}{30} \text{ [sec]}$$

を「右折旋回開始からヒヤリハット発生までの時間」として計算する。そして、 τ に関する閾値として、(学習用標本を)最も精度良く分類する値を決定し、評価したい標本の τ がその閾値より小さければクラス $c = 1$ (右直ヒヤリハット)、そ

れ以外であればクラス $c = 2$ (右折先ヒヤリハット) に分類する。

4.2. 統計的機械学習による分類

代表的な統計的機械学習手法として、正則化項付きロジスティック回帰とサポートベクターマシンの二つを検討する。ロジスティック回帰は代表的な線形なクラス分類手法であり、サポートベクターマシンは非線形クラス分類手法である。

4.2.1 プロブカーデータからの特徴抽出

クラス分類の前処理として、時系列データであるプローブカーデータから、適切に特徴量を抽出することが重要である⁽¹⁰⁾。素のセンサーデータ X_n から特徴量を抽出した特徴ベクトルを

$$\phi(X_n) \triangleq \begin{bmatrix} \phi_1(X_n) \\ \vdots \\ \phi_J(X_n) \end{bmatrix} \in \mathbb{R}^J$$

と表記する。最も自明な特徴ベクトルとして、センサーデータを「素のまま」で使用 ($[x^{(n,A)\top}, \dots, x^{(n,E)\top}]^\top = \phi(X_n)$) することも考えられるが、一般に分類精度は上がらない。ここでは、標準的な方法の一つである時間窓を利用した方法を適用する。時間窓内の部分時系列を取り出し、その部分時系列における平均と標準偏差を特徴にする方法である。利用する系列は、Fig. 2 と Fig. 3 から右直ヒヤリハットと右折先ヒヤリハットとで傾向の差を確認できる、全変数の時系列、および縦加速度と横加速度、ブレーキの自己共分散係数列とした。また、時間窓の長さは、分類精度が高くなるよう試行錯誤で探索し、時系列については 0.5 秒、自己共分散係数列については 1 秒とした。また、各時間窓は重ならないようにしている。従って、特徴ベクトルのサイズ J は $J = 390$ となる。

さらに、各特徴の大きさやスケールをそろえるために、各特徴の 0.01-分位点、0.99-分位点がそれぞれ 0, 1 になるように正規化を行った。ここで分位点を用いて正規化を行ったのは、外れ値に頑強にするためである⁽¹¹⁾。

4.2.2 ロジスティック回帰による分類

ロジスティック回帰によるクラス分類は、その分類境界が(特徴ベクトル空間に関して)超平面に限られる線形なクラス分類手法である⁽⁷⁾。学習パラメータ $\theta \triangleq [\theta_0, \dots, \theta_J]$ が定めれば、第 n 番目の標本が各クラス $c_n \in \{1, 2\}$ に所属する確率は

$$\begin{cases} p(c_n = 1 | \phi(X_n), \theta) \triangleq \frac{1}{1 + \exp(\sum_{j=1}^J \theta_j \phi_j(X_n) + \theta_0)} \\ p(c_n = 2 | \phi(X_n), \theta) \triangleq 1 - p(c_n = 1 | \phi(X_n), \theta) \end{cases} \quad (1)$$

と推定される。従って、 $p(c_n = 1 | \phi(X_n), \theta) > 0.5$ であればクラス $c = 1$ 、それ以外であればクラス $c = 2$ と分類する。学習パラメータ θ は、目的関数

$$L(\theta) \triangleq - \sum_{n=1}^N \log p(c_n | \phi(\mathbf{X}_n), \theta) + \frac{1}{2} \lambda \theta^\top \theta.$$

Table. 1 Accuracy of prediction

Classification method	Average accuracy	Standard deviation
Domain-oriented method	0.644	0.098
Logistic regression	0.788	0.094
Support vector machine	0.794	0.106

を最小にするように最適化する。ここで、 $-\log p(c_n | \phi(\mathbf{X}_n), \theta)$ は対数損失と呼ばれ、誤分類に対する罰則項に対応する。 $\lambda \theta^\top \theta$ は正則化項と呼ばれ、分類モデルが学習データに過適応（過学習）することを防ぐ効果をもつ。 λ はデータから決められるべき調節パラメータであり、ここでは交差検定法⁽⁷⁾により決定する。

4.2.3 サポートベクターマシンによる分類

カーネル関数を用いたサポートベクターマシンによるクラス分類は、ロジスティック回帰とは異なり、非線形な分類手法であり、線形分離不可能な問題に対して有効である^{(6),(12)}。ここでは、カーネル関数としてガウスカーネルを用いた。

$$k(\phi(\mathbf{X}_n), \phi(\mathbf{X}_{n'})) \triangleq \exp \{ -\gamma (\phi(\mathbf{X}_n) - \phi(\mathbf{X}_{n'}))^\top (\phi(\mathbf{X}_n) - \phi(\mathbf{X}_{n'})) \}$$

γ はデータから決められるべき調節パラメータであり、交差検定法⁽⁷⁾により決定される。クラス予測は、学習に用いた N 個の標本の特徴ベクトル $\{\phi(x_1), \dots, \phi(x_N)\}$ と学習パラメータ $\alpha \triangleq [\alpha_0, \dots, \alpha_N]$ により定まるスコア関数

$$f(\phi(\mathbf{X}), \alpha) = \sum_{n=1}^N \alpha_n k(\phi(\mathbf{X}), \phi(\mathbf{X}_n)) + \alpha_0$$

に基づいて行われる。 $f(x) < 0$ であればクラス $c = 1$ 、それ以外はクラス $c = 2$ に分類される。学習パラメータ α はマージン最大化という基準で最適化される^{(6),(12)}。具体的には、次の最適化問題を解くことで α を決定する。

$$\begin{aligned} \text{minimize}_{\alpha, \xi} : \quad & L(\theta) \triangleq \frac{1}{2} \alpha^\top \alpha + C \sum_{n=1}^N \xi_n, \\ \text{subject to:} \quad & f(\phi(\mathbf{X}_n), \alpha) \geq 1 - \xi_n, \quad n = 1, \dots, N. \end{aligned}$$

ここで、 $\xi_n \geq 0$ はスラック変数である。 C は調節パラメータであり、 γ と同様に、交差検定法により決定される。

5. 実験

5.1. 節では、右折時のヒヤリハット発生形態である「右直ヒヤリハット」と「右折先ヒヤリハット」についての2クラス分類問題に対して、4節で導出した各分類方法を適用した結果を示す。5.2. 節では、得られた機械学習分類モデルに関する考察を与える。

5.1. 実験結果

交差検定法⁽¹³⁾を用いて、各分類手法の評価を行った。交差検定法とは、標本データを分類モデル学習（もしくは構築）用のデータと分類モデル評価用のデータに分割して、分類モデルの汎化性能を評価する方法である。ここでは、10分割の

交差検定法を用いた。これは、まず初めに標本データを適当に10分割する。次に、以下の手順を全10通りの組合せ全てについて繰り返し実行して、分類モデルの平均分類精度と標準偏差を評価する。この結果をTable 1に示す。

- 10組中ある9組の標本を学習用データとして、分類モデルを学習する
- 残り1組の標本で分類モデルの分類正当率を評価する

Table 1から、統計的クラス分類手法であるロジスティック回帰(Logistic regression)やサポートベクターマシン(Support vector machine)は、領域知識に基づく分類手法(Domain-oriented method)に比べ、有意に高い分類精度(正答率)を示したことが確認できる。よって、右折時のヒヤリハット発生形態の分類において統計的機械学習の方が高い推定能力をもつと考えられる。これは、今回採用した領域知識に基づく方法では考慮できていなかった分類する上で有用な特徴を、統計的機械学習法では発見し利用できていたためだと考えられる。次節5.2.で、ロジスティック回帰で得られた分類モデルを調べ、より詳細な考察を与える。

一方でTable 1から、ロジスティック回帰とサポートベクターマシンの分類精度は同程度であったことがわかる。これらの手法の違いは、判別境界が線形であるか非線形であるかであったので、今回の分類においては、判別境界は特に非線形である必要はなかったことが示唆される。つまり、時間窓を用いた特徴ベクトル空間においてほぼ線形分離可能であったと考えられる。以上より、本課題には、サポートベクターマシンより単純である、言い換えれば調節パラメータ数の少ない、ロジスティック回帰が最も適合した分類手法であると考えられる⁽¹³⁾。

最後に、特徴ベクトルに素のセンサーデータを用いた機械学習手法も適用したが、4節で予想した通り、非常に低い分類精度となった。この場合のロジスティック回帰とサポートベクターマシンそれぞれの「平均 ± 標準偏差」は 0.6 ± 0.14 、 0.61 ± 0.13 であった。従って、今回適用した時間窓による特徴抽出と分位点を用いた正規化の取り組みは単純だが、効果があったと考えられる。

5.2. ロジスティック回帰による分類モデルからの考察

ここでは、領域知識に基づく分類手法に比べ統計的機械学習法の方が高い分類精度を示した要因を探る。特に、分類モ

デルの解釈の容易性から、ロジスティック回帰によって学習した分類モデル(式(1))を分析する。式(1)より、双方のクラスに対する推定確率の対数比は

$$\log \frac{p(c=2|\phi(x), \theta)}{p(c=1|\phi(x), \theta)} = \sum_{j=1}^J \theta_j \phi_j(x) + \theta_0 \quad (2)$$

となる。つまり、 $\phi_j(x) > 0$ であれば、 θ_j が小さいほどクラス1(右直ヒヤリハット)に分類され、逆に大きければクラス2(右折先ヒヤリハット)に分類される確率が高くなる。 $\phi_j(x) < 0$ であれば、その逆の傾向となる。このことから、ロジスティック回帰による分類モデルは、右直ヒヤリハットと右折先ヒヤリハットを分類する特徴として、以下が捉えられていることが分かった。

- (a) ヒヤリハット発生の2秒前付近で横加速度が低い場合、右直ヒヤリハットの可能性が高い
- (b) ヒヤリハット発生から2秒後付近でブレーキが踏まれている場合、右直ヒヤリハットの可能性が高い
- (c) ヒヤリハット発生から5秒後付近で右ウィンカーが点灯している場合、右直ヒヤリハットの可能性が高い

これらの特徴は以下の運転行動と関連があると考えられる。

(a)は「右直ヒヤリハットの場合、発生の2秒前付近では右折待ちで速度が低いか、右折アプローチ中で操舵が発生していない」、(b)は「右直ヒヤリハットの場合、発生後もそのまま2秒以上停止していることが多い」、(c)は「右直ヒヤリハットの場合、発生から右折を完了までの時間が比較的長い、ウィンカーが消えるまでの時間も長い」といった運転行動である。このような特徴はそれだけで精度良く右直ヒヤリハットと右折先ヒヤリハットとを分類できるものではないが、運転行動面から見た傾向としては納得できるものであり、これらの指標の重み付けも学習した上で総合的に分類判定を行うことができるため、統計的機械学習法の分類性能が高くなるものと考えられる。

6. おわりに

本稿では、ヒヤリハット発生時の時系列データを用いて右直ヒヤリハットと右折先ヒヤリハットを分類する手法として、統計的機械学習法に着目し、その分類性能を評価した。その結果、領域知識を基に決定した操舵開始タイミングに着目した閾値で分類するより、統計的機械学習法を用いた分類性能の方が高いことが示された。また、領域知識を用いた解析では見落していた判別指標が、機械学習では抽出できていることがわかり、それらの指標に適切な重みをつけながら総合的に判断しているため、統計的機械学習法の分類性能が高くなっている可能性があることがわかった。以上のことから、プローブカーから収集されたヒヤリハット情報を基に、右直

ヒヤリハットと右折先ヒヤリハットを分類する上で統計的機械学習法が有効である可能性が示された。また、このことから、他の発生形態のヒヤリハットの分類においても有効である可能性があると考えられ、今後、適用領域を拡大して、安全な交通社会の実現に寄与できるよう技術開発を進めていく所存である。

謝辞

本研究では、(社)自動車技術会が収集したヒヤリハットデータベースを活用して分析を行った。自動車技術会の関係者をはじめ、関係各位の支援を受けた。森村、井手は総務省の地球温暖化対策ICTイノベーション推進事業(PREDICT)の支援を受けた。ここに謝意を表する。

参考文献

- (1) 社団法人自動車技術会：平成18年度国土交通省受託事業ヒヤリハット分析による予防安全対策の効果予測等の検討調査報告(2007)。
- (2) 藤田光伸、小竹元基、鎌田実、道辻洋平、永井正夫：ドライブレコーダーを用いたヒヤリハット分析に関する研究(第3報)-データベースを用いた追突ヒヤリハット解析-、自動車技術会論文集, Vol. 38, No. 4, p. 151-156 (2007)。
- (3) 林剣之介、青木義満：ドライブレコーダ映像を用いた交通事故の定量解析, Vol. 33(34), p. 25-26 (2009)。
- (4) 富永茂、西本哲也、中谷育夫、真下一則：映像収録型ドライブレコーダーによる事故・ニアミス現象の分析, Vol. No42-04, p. 5-10 (2004)。
- (5) 山崎慎也、船久保晃、谷澤悠輔：プローブカーのヒヤリハットデータを活用した危険経路の推定, 自動車技術会論文集, Vol. 42, No. 2, p. 639-644 (2011)。
- (6) Hastie, T., Tibshirani, R. and Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer (2001)。
- (7) Bishop, C. M.: *Pattern Recognition and Machine Learning*, Springer (2006)。
- (8) 北川源四郎：時系列解析入門, 岩波書店 (2005)。
- (9) Hamilton, J. D.: *Time Series Analysis*, Princeton University Press (1994)。
- (10) Geurts, P.: Pattern extraction for time series classification, In *Proc. Principles of Data Mining and Knowledge Discovery*, p. 115-127 (2001)。
- (11) Koenker, R.: *Quantile Regression*, Cambridge University Press (2005)。
- (12) Cortes, C. and Vapnik, V.: Support-Vector Networks, *Machine Learning*, Vol. 20, No. 3, p. 273-297 (1995)。
- (13) 小西貞則、北川源四郎：情報量規準, 朝倉書店 (2004)。

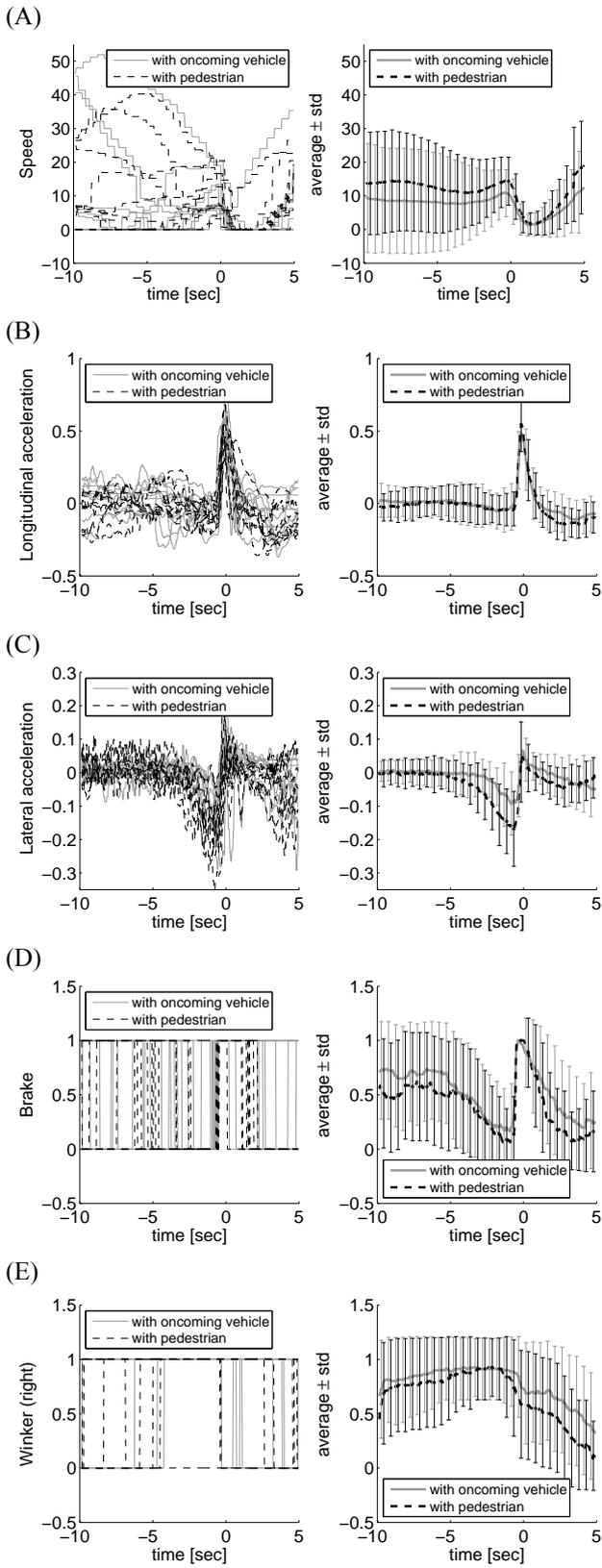


Fig. 2 Time series of the features for the near-miss incidents with oncoming vehicles and with pedestrians: On the left-hand sides, the results of each incident are shown. The mean and standard deviation for each situation are shown on the right-hand sides.

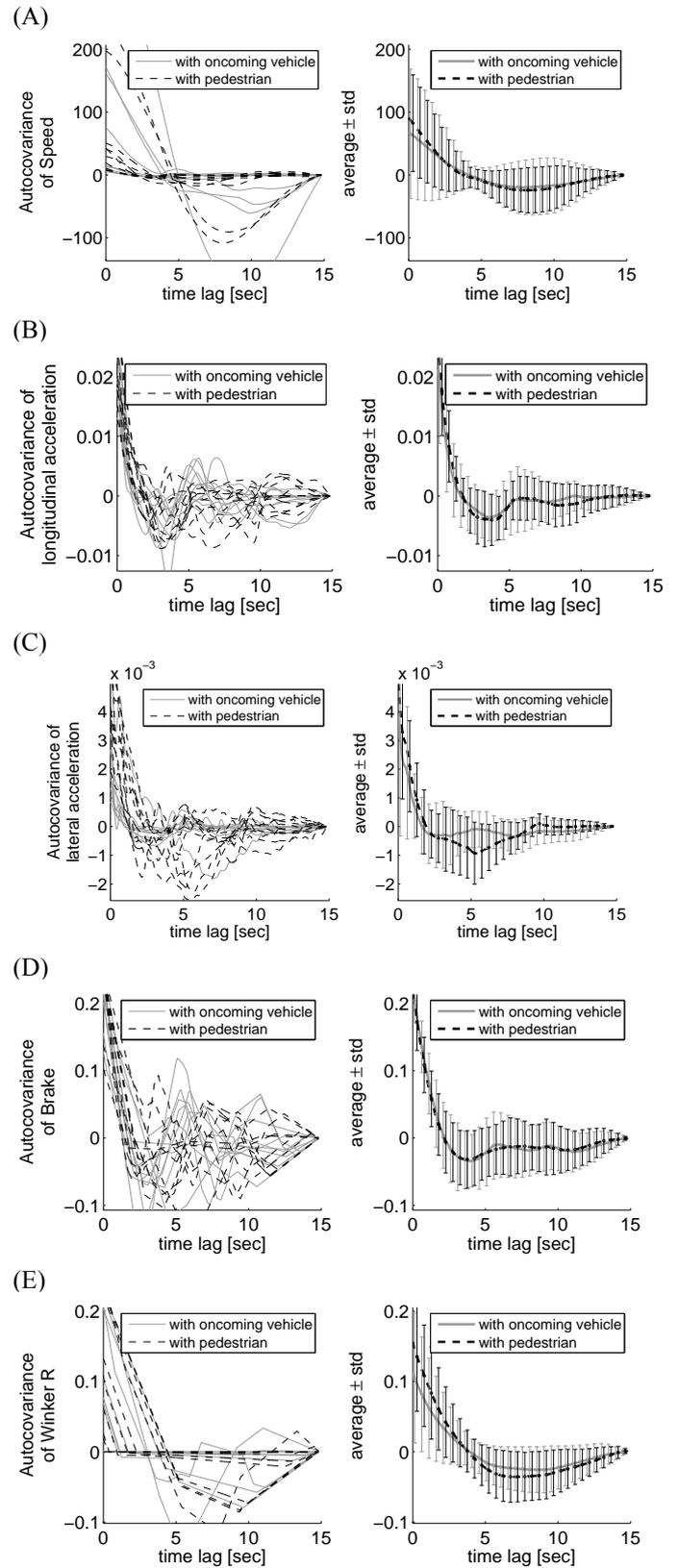


Fig. 3 Autocovariance sequence of the features for the near-miss incidents with oncoming vehicles and with pedestrians: On the left-hand sides, the results of each incident are shown. The mean and standard deviation of each situation are shown on the right-hand sides.