

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

IBM東京基礎研究所

谷澤悠輔

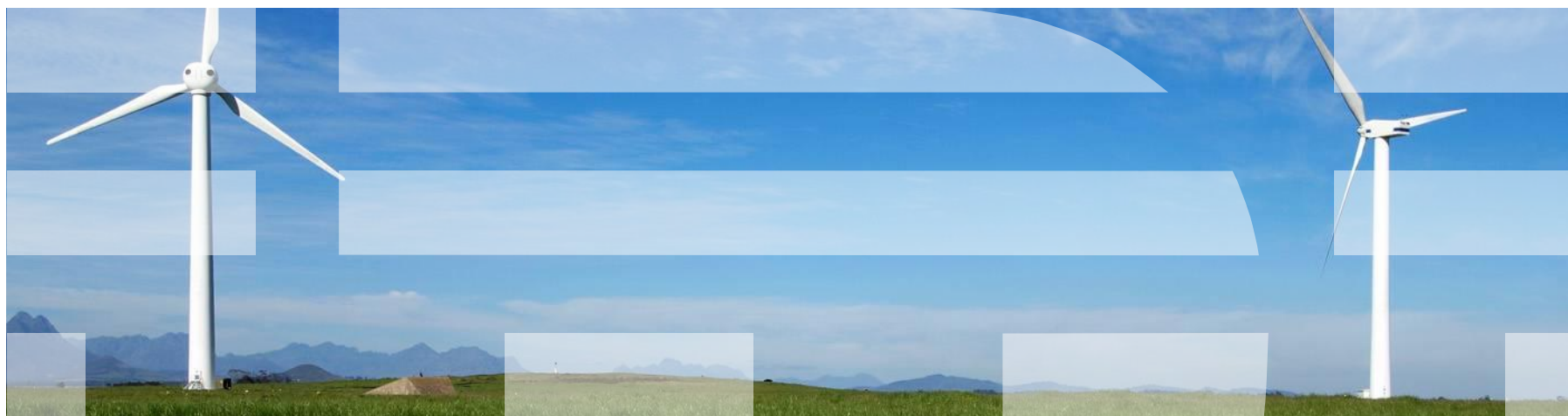
マツダ株式会社

山崎慎也

マツダ株式会社

井手剛

IBM東京基礎研究所

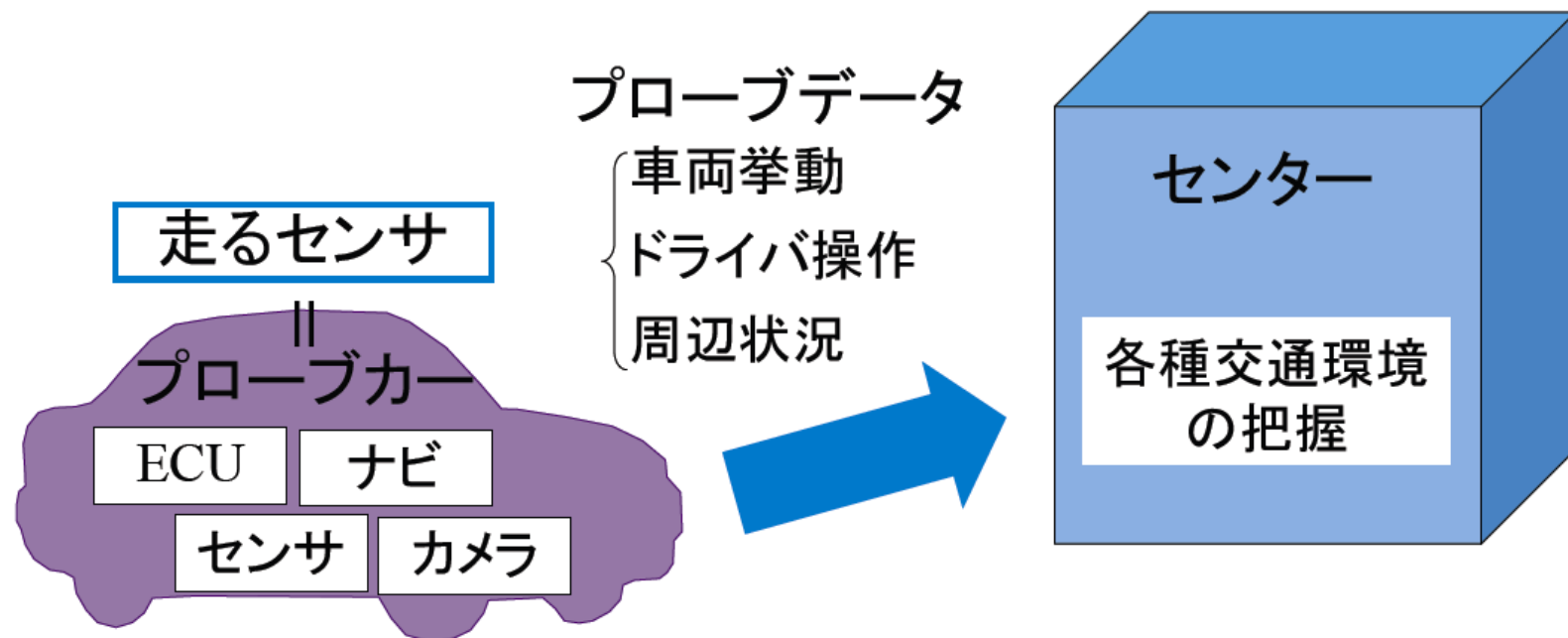


目次

- **背景・目的**
- **プローブデータの概要**
- **ヒヤリハット発生形態分類手法の提案**
 - 統計的機械学習手法
 - 領域知識に基づく方法
- **実験結果**
- **まとめ**

背景

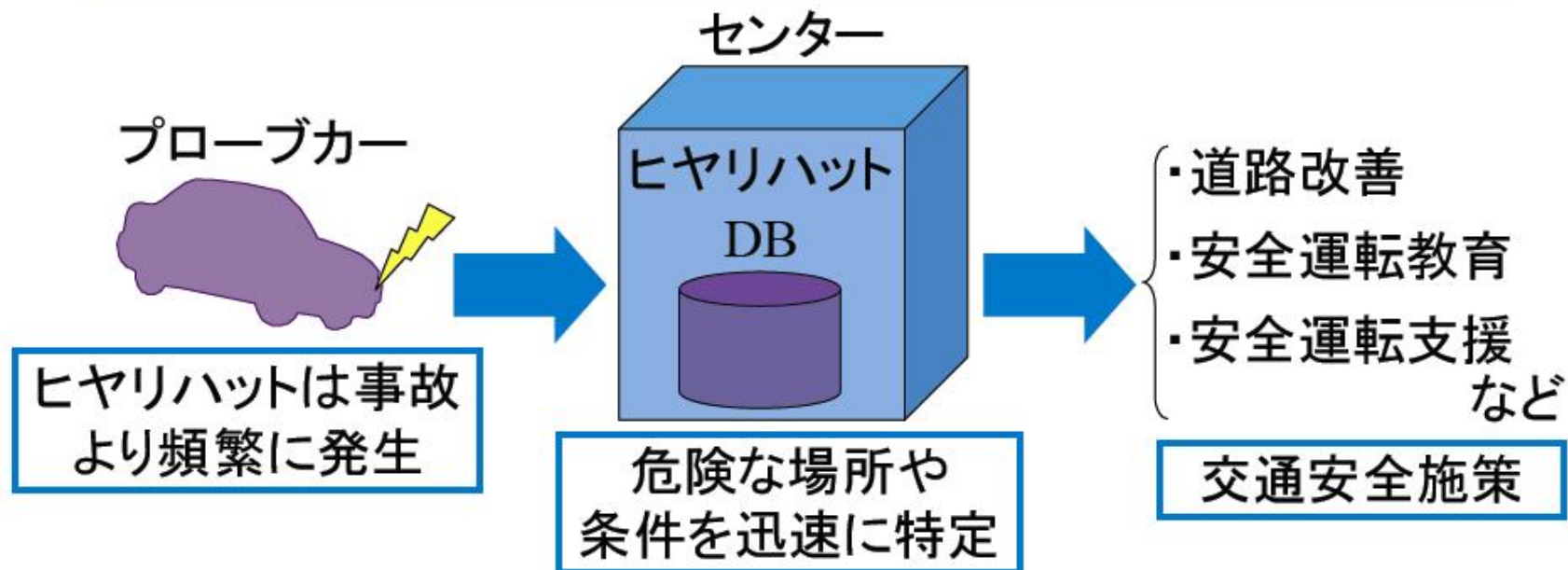
- 車を走るセンサ(プローブカー)として活用して、様々な交通環境を把握することが可能になった



目的

- プローブカーを活用して危険な場所や条件を特定して、それを適切な形でドライバなどに提供することで交通安全に役立てる

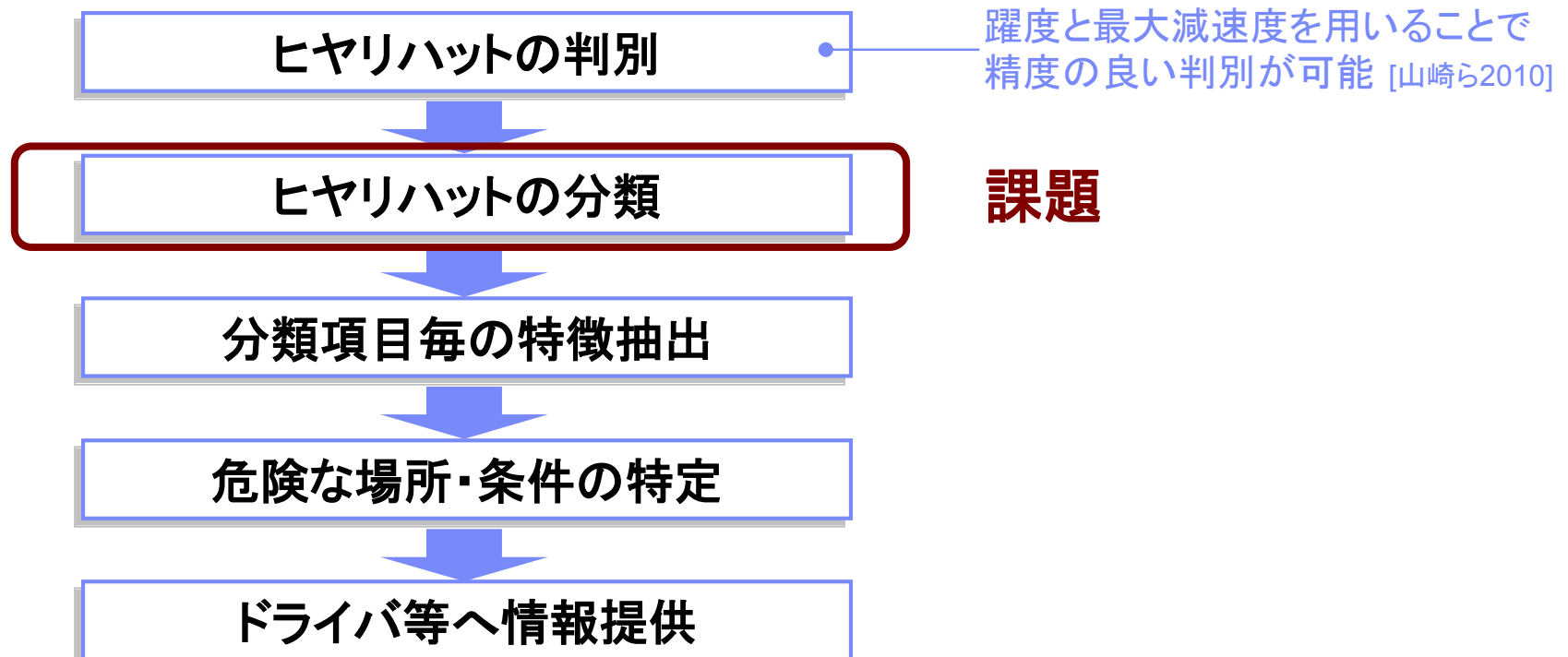
ヒヤリハットとは、危うく事故になりかけたが回避できた事象



課題

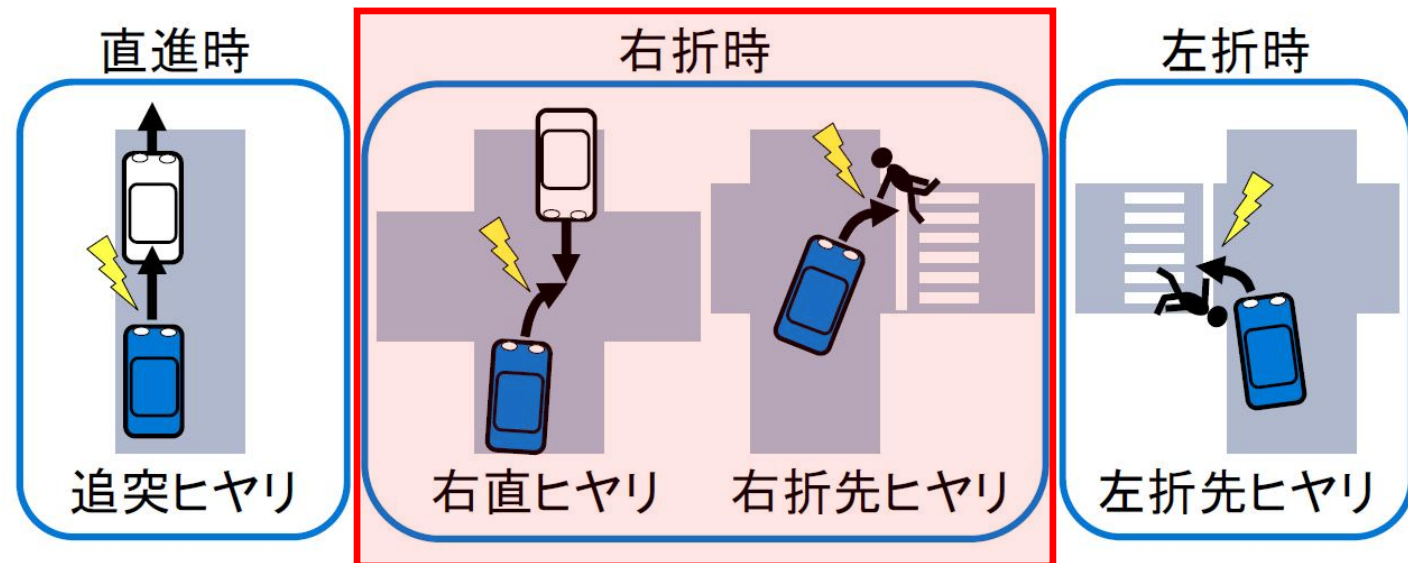
- ヒヤリハットの発生を精度良く判別できるようになってきたが、その発生形態の分類が課題となる

＜システム処理フロー＞



ヒヤリハット発生形態

- GPSデータ等を用いて、大まかなヒヤリハット発生形態分類はできるが、詳細な分類は難しい
 - ドライバに価値のある情報を伝えるためには、詳細な分類が必須
 - 現在は手作業で分類するしかないので、自動化が望まれる
- 本報告では、右折時のヒヤリハット発生形態の分類問題を解く
 - 右直ヒヤリ vs. 右折先ヒヤリ



目次

- 背景・目的
- プローブデータの概要
- ヒヤリハット発生形態分類手法の提案
 - 統計的機械学習手法
 - 領域知識に基づく方法
- 実験結果
- まとめ

プローブデータの概要

■ 標本数全160件

- 右直ヒヤリ: 80件
- 右折先ヒヤリ: 80件

※自動車技術会のヒヤリハットデータベースを利用

■ 15秒間のドライブレコードによる時系列データ

- ヒヤリハット発生タイミングの前10秒～後5秒
- 1/30秒刻み ⇒ 450の観測点からなる

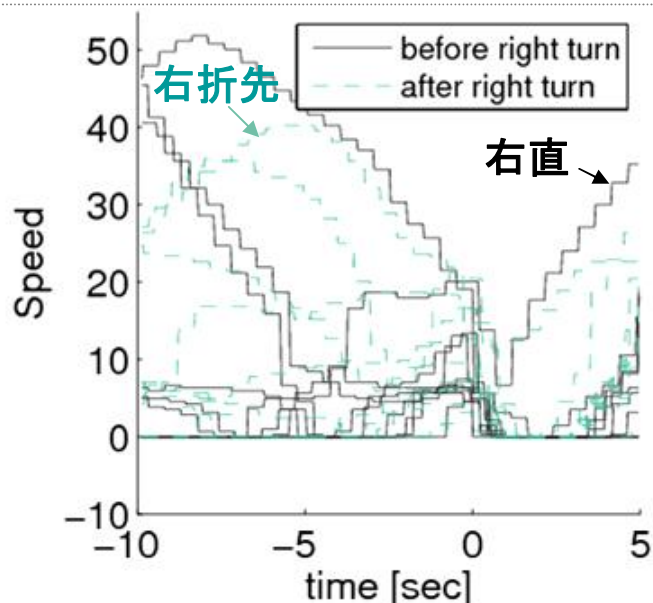
■ 用いたセンサーデータ: 計5種 ⇒ 次ページ以降で可視化

- 速度 ✓
- 縦加速度
- 横加速度 ✓
- ブレーキ (On / Off 情報)
- 右側ウィンカー (On / Off 情報)

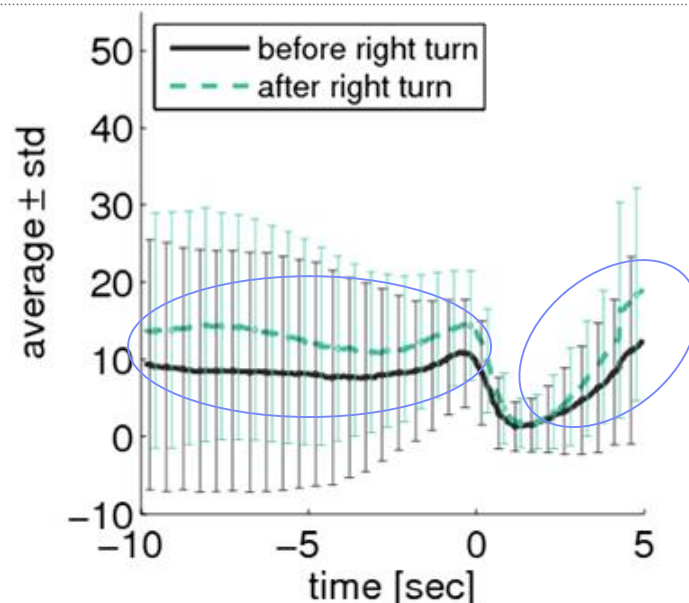
速度

• 時系列プロット

右直、右折先ヒヤリの各15標本の速度を図示

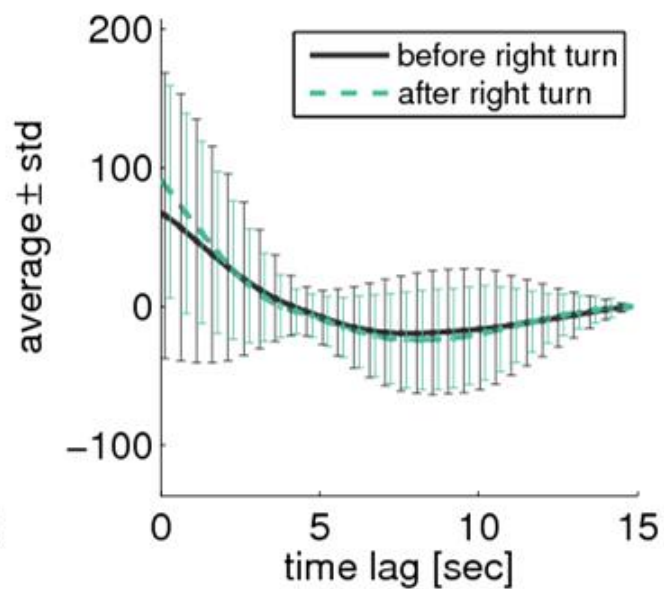
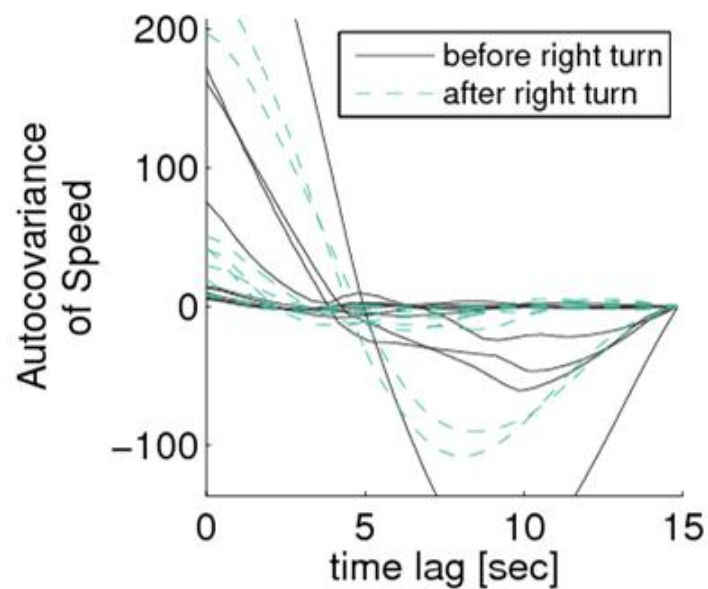


右直80個、右折先80個をそれぞれまとめたもの



• 自己共分散係数列プロット

✓時系列解析でよく用いられる標準的な統計量

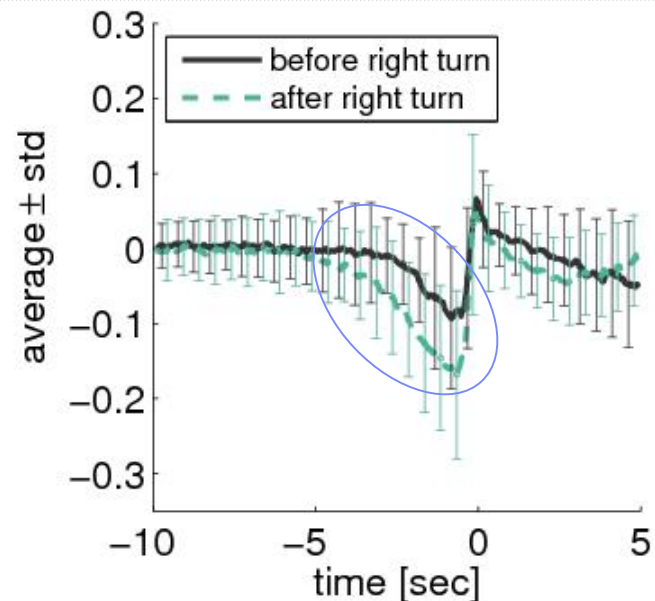
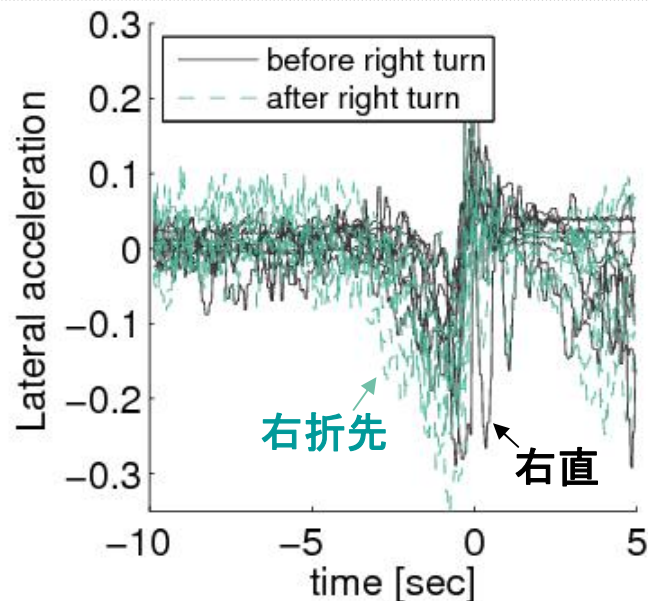


横加速度

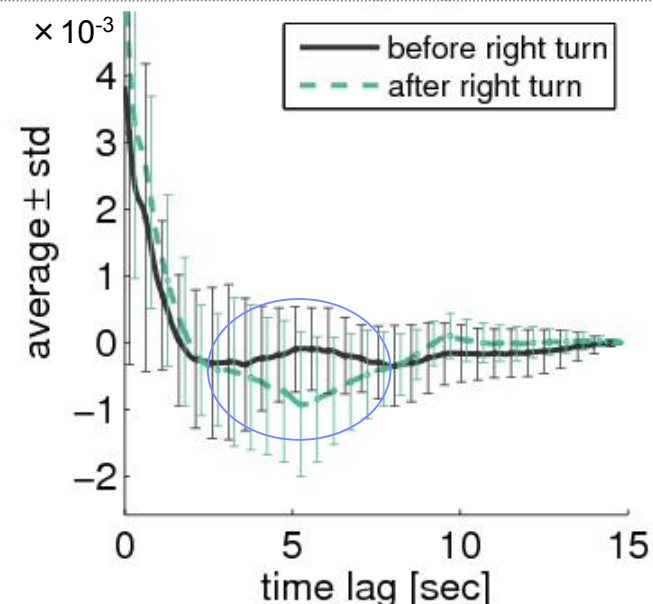
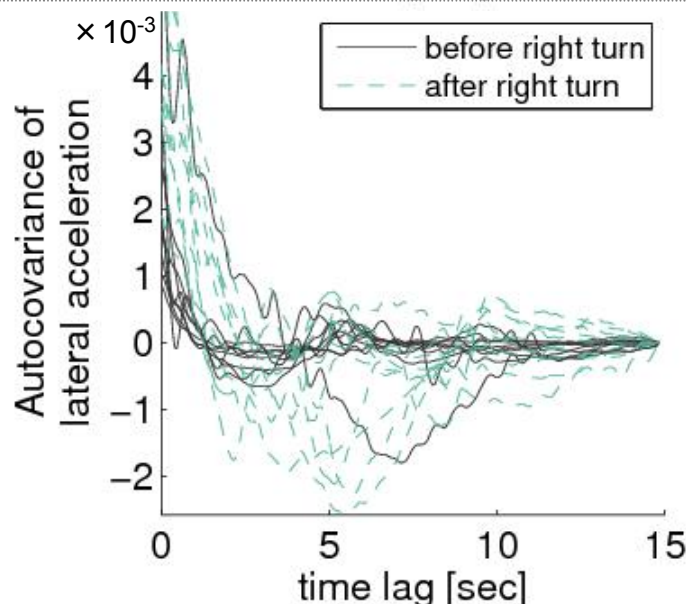
右直、右折先ヒヤリの各15標本の横加速度を図示

右直80個、右折先80個をそれぞれまとめたもの

・時系列プロット



・自己共分散係数列プロット



データの観察:まとめ

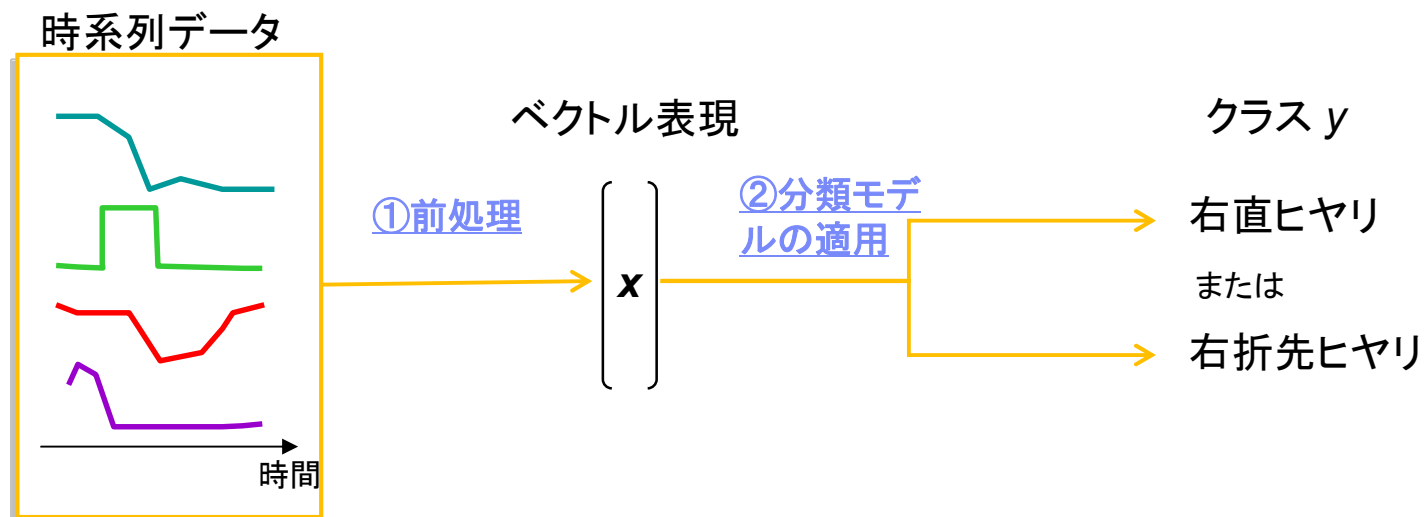
- 時系列データの可視化から、
 - クラス間の若干の傾向の違いは確認できる
 - 特に、速度と横加速度の時系列プロット
横加速度の自己共分散係数列プロット
 - しかし、その違いは顕著でなく、目視や単純な閾値での分類は困難
- 統計的機械学習等に基づいた系統的な解析が必要

目次

- 背景・目的
- プローブデータの概要
- **ヒヤリハット発生形態分類手法の提案**
 - 統計的機械学習手法
 - 領域知識に基づく方法 (比較手法)
- 実験結果
- まとめ

機械学習による時系列クラス分類

- 任意の時系列データが与えられた時に、そのクラス(右直か右折先か)を判別する関数を求めることが目標
 - = 時系列データ x が与えられた時に、クラス y を返す関数を知りたい
- 入力が多変数の時系列データであるため、通常のカテゴリ分類問題とは異なるやや高度な問題
 - 通常は入力 x が普通のベクトルだが、今は多変量時系列
- 時系列データから「特徴ベクトル」を求めることで、通常のカテゴリ分類問題に帰着させる



機械学習による時系列クラス分類

ステップ1: 前処理

- 時系列データから特徴ベクトルの変換には、データの性質に応じた処理が必要

– 2つのアプローチを試行

- 時系列をそのまま特徴ベクトルとする
- 滑走窓で部分時系列を作り、それらの平均・標準偏差を特徴ベクトルの要素とする

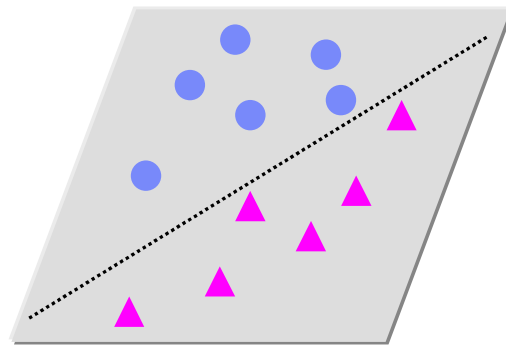
– スケーリング方法は3種類を試行

- スケーリング無し
- 単純スケーリング: 最小値を0、最大値を1
- 分位点スケーリング: 外れ値に頑健
 - ✓ 0.02-分位点を0、0.98-分位点を1

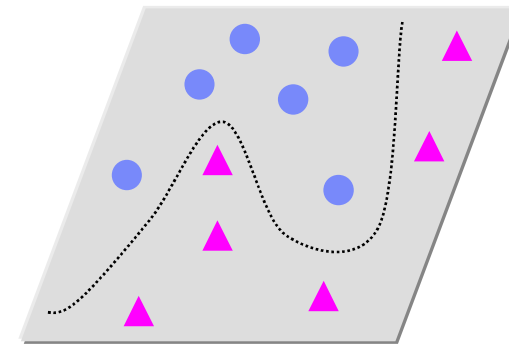
機械学習による時系列クラス分類 ステップ2: クラス分類手法の適用

■ 線形/非線形の代表的な機械学習分類法を適用

- 正則化付きロジスティック回帰(LR)
 - 線形クラス分類手法
- サポートベクトルマシン(SVM)
 - 非線形クラス分類手法



- 汎化能力が高い
- 学習結果(モデル)の解釈が容易
- 非線形な分離平面は学習できない



- 過学習を起こし易い
- 学習結果の解釈が困難
- 複雑な分離平面の学習が可能

[比較手法] 領域知識に基づく分類方法

■ヒヤリハット発生時の交差点内位置から、右直/右折先ヒヤリを分類

– 右直/右折先ヒヤリとでは交差点内でのヒヤリハット発生位置が異なる

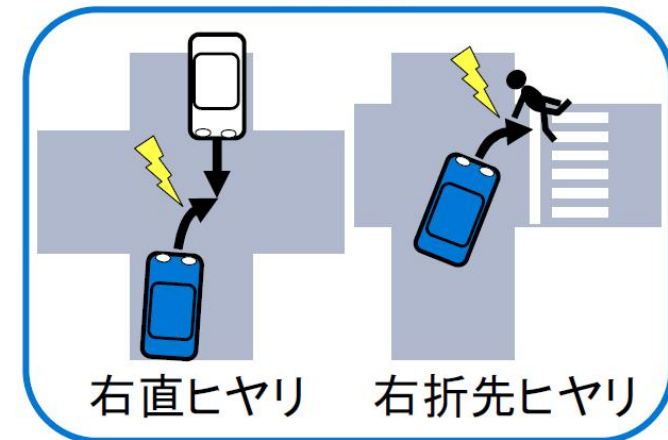
– 「**右折旋回開始からヒヤリハット発生までの時間**」で交差点内の位置を推定して、その大小で分類

- 回転半径 r を指標として、右折開始時を同定:

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

v: 速度
a: 横加速度

- 継続的に値が負になり始めたタイミングを右折開始時とした



- 右直ヒヤリ:
交差点**入り口**付近で発生しやすい
- 右折先ヒヤリ:
交差点**出口**付近で発生しやすい

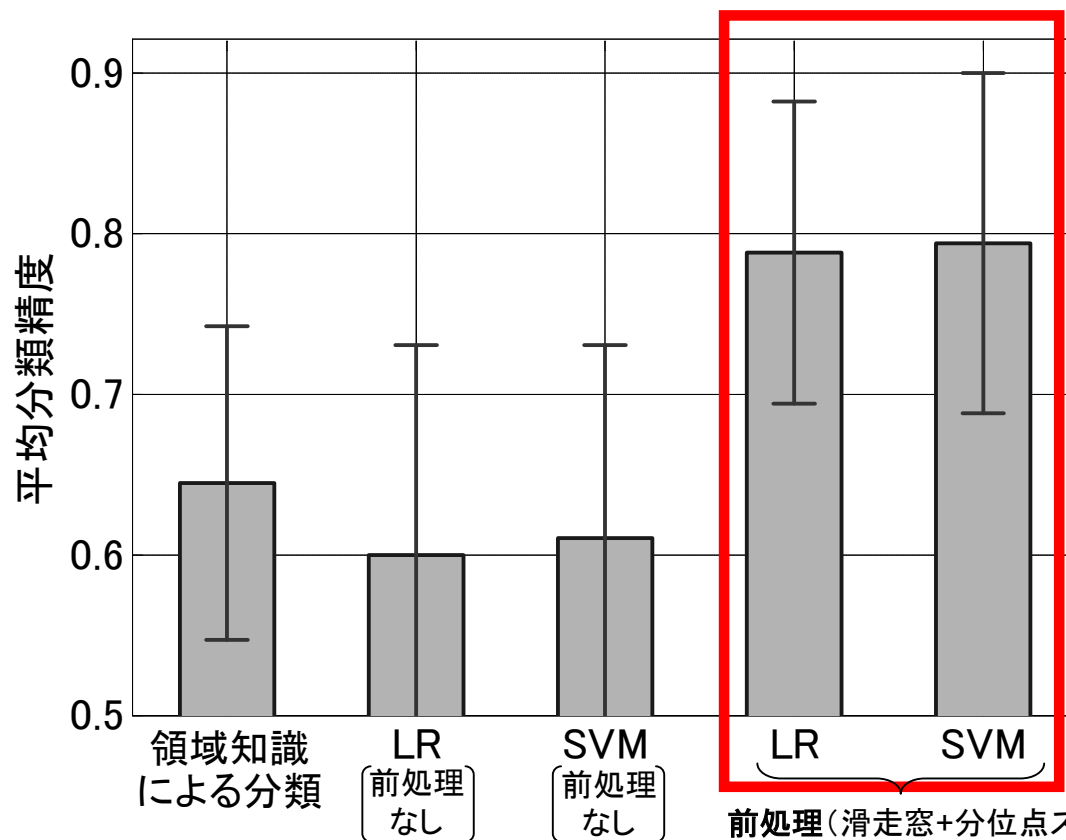
目次

- 背景・目的
- プローブデータの概要
- ヒヤリハット発生形態分類手法の提案
 - 統計的機械学習手法
 - 領域知識に基づく方法
- **実験結果**
- まとめ

分類精度の比較結果

■ 交差検定法により評価

– 標本を訓練用と評価用に分けて、モデルの学習と評価に同じ標本は使わない



- 適切な前処理と統計的クラス分類手法であるLRやSVMを適用することで、ヒヤリハット分類において高分類精度を達成できた
- 領域知識に基づく方法では考慮できていない有用な特徴を、統計的機械学習法により発見できる

前処理(滑走窓+分位点スケールング)
+ 統計的機械学習法

※ 単純スケールングの
精度は 0.7 ± 0.1 程度

考察： ロジスティック回帰による分類モデルからの知見

■ 分類モデルの解析により判明した右直ヒヤリの特徴

–ヒヤリハット発生**2秒前付近で横加速度が低い**

- [解釈] 右直ヒヤリの場合、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

–ヒヤリハット発生**2秒後付近でブレーキが踏まれている**

- [解釈] 右直ヒヤリの場合、発生後もそのまま2秒以上停止していることが多い

–ヒヤリハット発生**5秒後付近で右ウィンカーが点灯している**

- [解釈] 右直ヒヤリの場合、発生から右折を完了までの時間が比較的長いいため、ウィンカーが消えるまでの時間も長い

■ 統計的機械学習法はこれらの特徴を総合的に用いて分類判定を行うため、分類性能が高くなるものと考えられる

目次

- 背景・目的
- プローブデータの概要
- ヒヤリハット発生形態分類手法の提案
 - 統計的機械学習手法
 - 領域知識に基づく方法
- 実験結果
- まとめ

まとめ

- ヒヤリハット発生形態の分類に統計的機械学習の適用を検討した
- 右直/右折先ヒヤリハットの時系列分類問題において、機械学習法は領域知識による方法より有意に有効であることがわかった
 - 領域知識に基づく解析では見落としてしまうような指標であっても、機械学習では抽出可能であり、それらの指標から総合的に判断して分類するため、分類性能が高くなることが示唆された

今後の課題

- 他の発生形態のヒヤリハット分類など、適用領域の拡大
- ドライバへの効果的なヒヤリ情報の提供方法など、プローブデータ解析結果の活用方法の検討

ご静聴ありがとうございました。

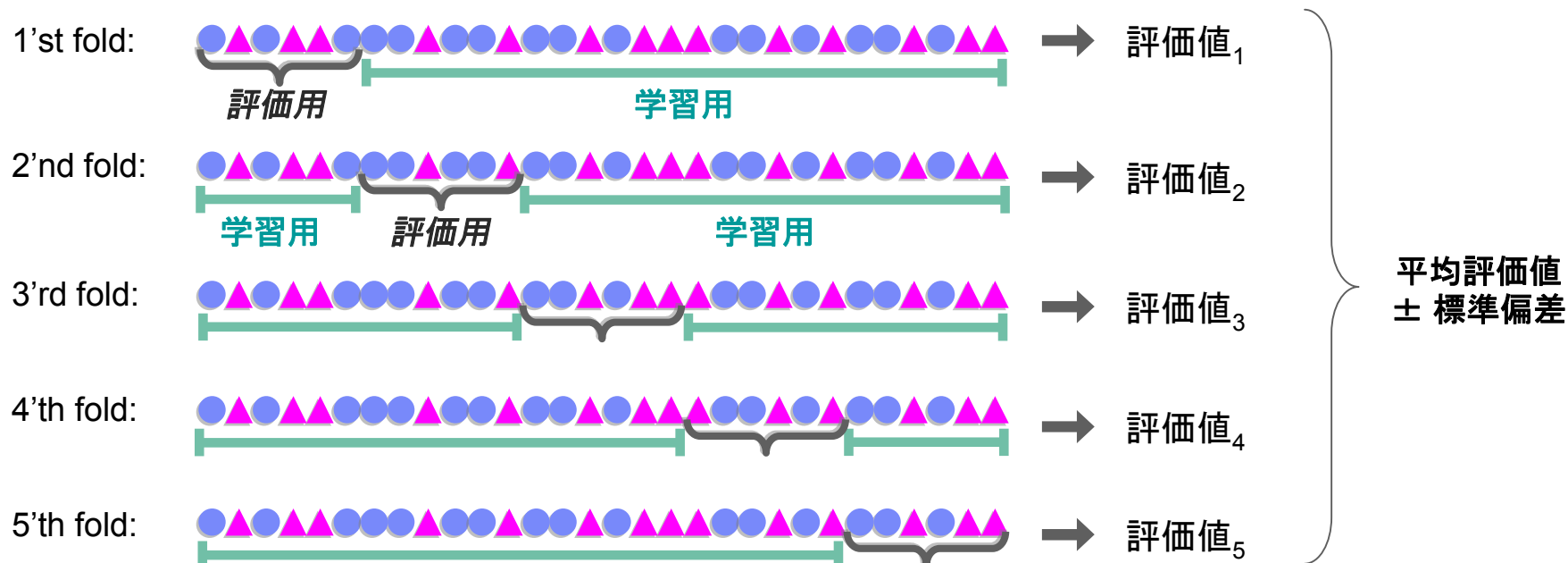
補足スライド

交差検定法(クロスバリデーション)について

- 平等な評価とするために、標本を学習用と評価用にわける検定方法
- 5フォールドの交差検定例(30標本の場合)

– 標本を5個のグループに分ける ⇒ 

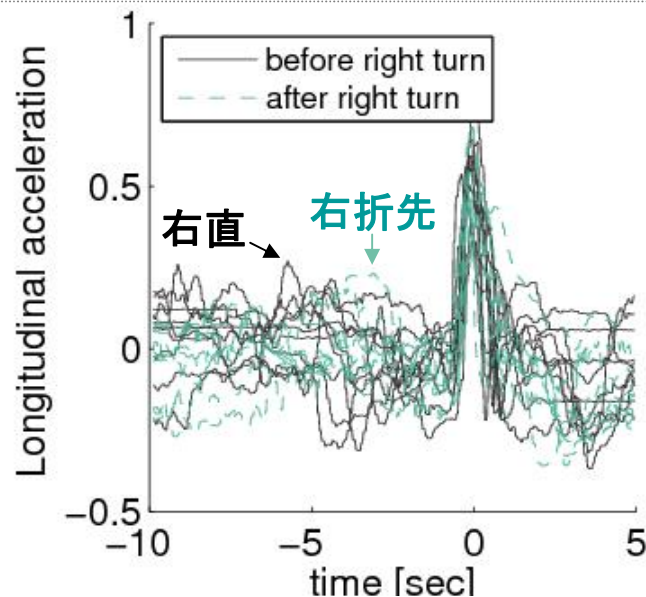
– 4グループの標本からモデルを学習し、残りの1グループの標本でモデルを評価



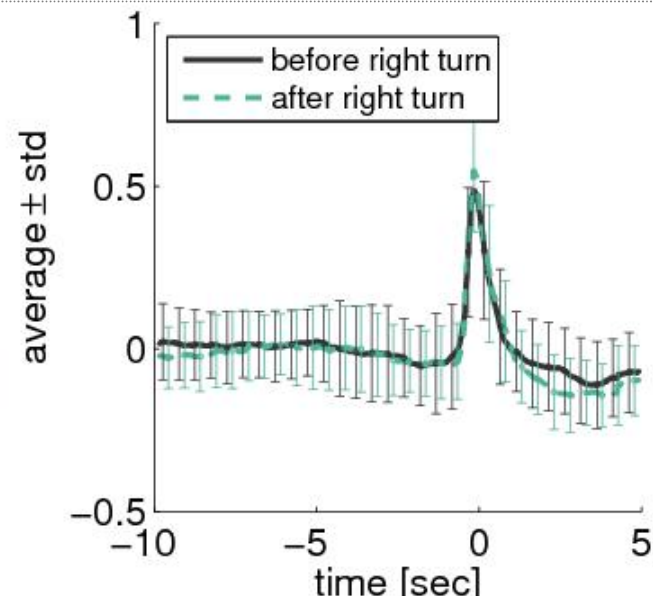
縦加速度

• 時系列プロット

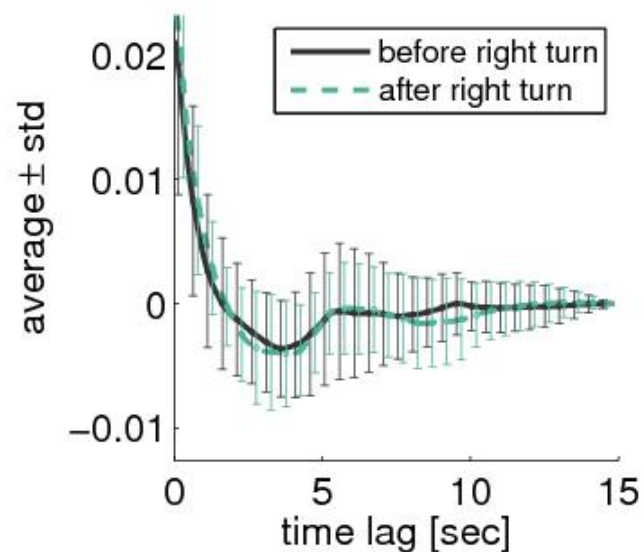
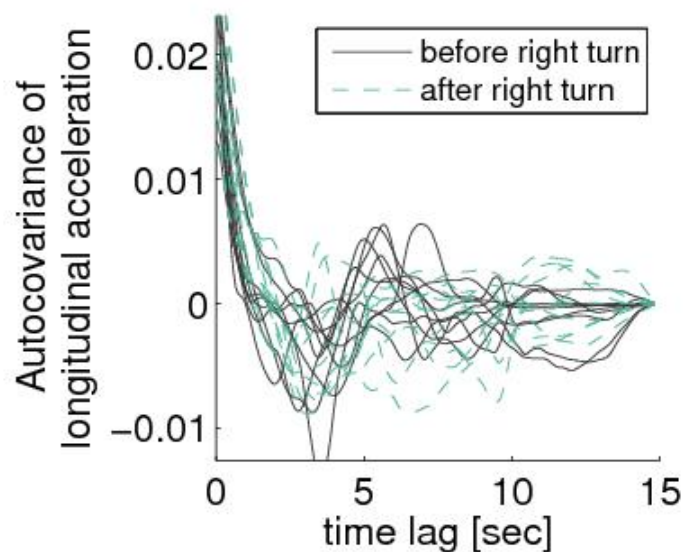
右直、右折先ヒヤリの各15標本の縦加速度を図示



右直80個、右折先80個をそれぞれまとめたもの



• 自己共分散係数列プロット

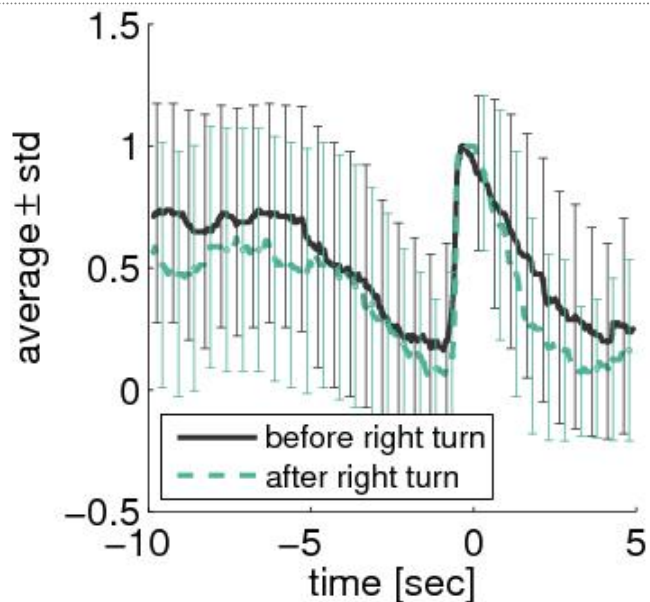
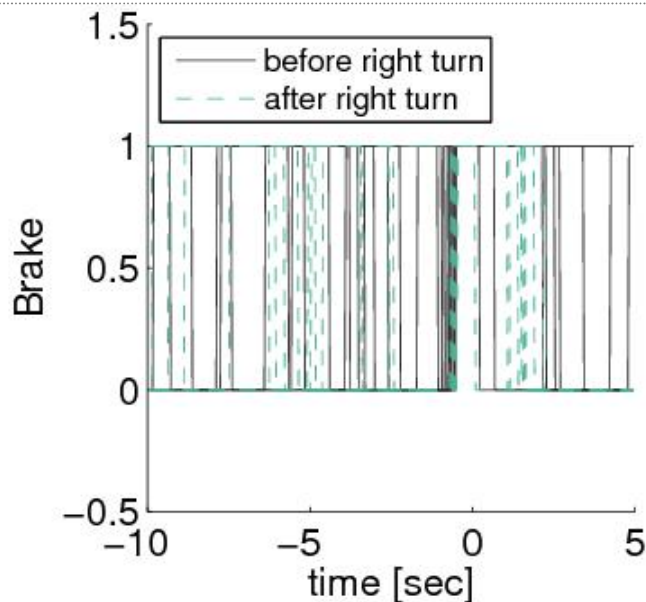


ブレーキ

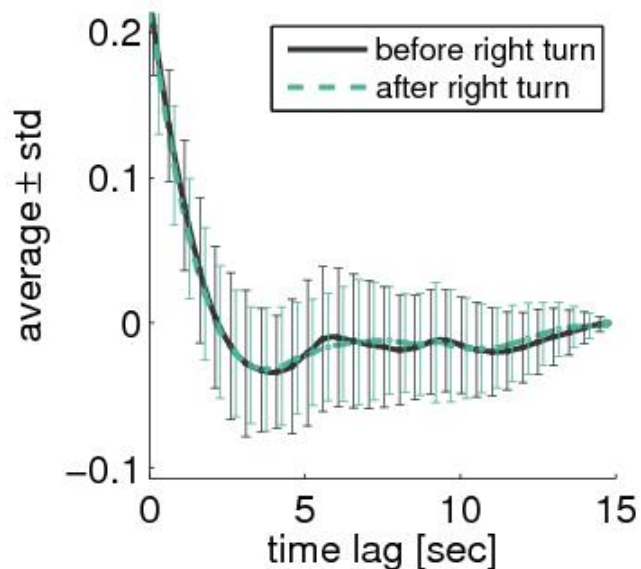
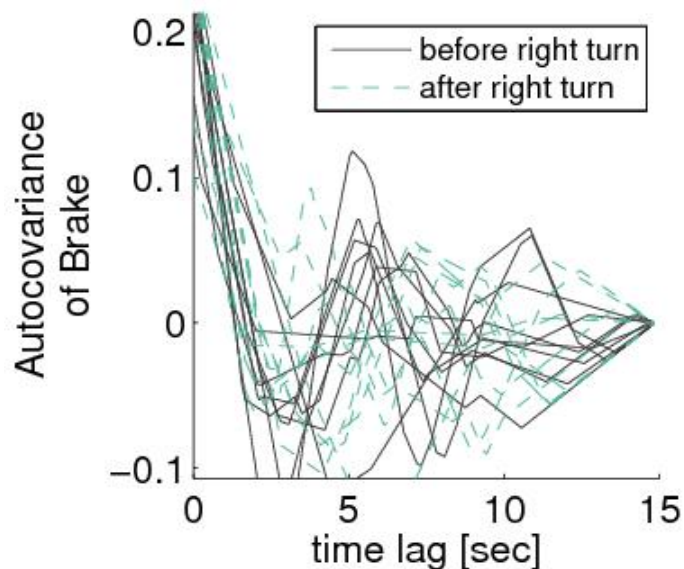
右直、右折先ヒヤリの各15標本の
ブレーキon/offを図示

右直80個、右折先80個を
それぞれまとめたもの

• 時系列プロット



• 自己共分散係 数列プロット

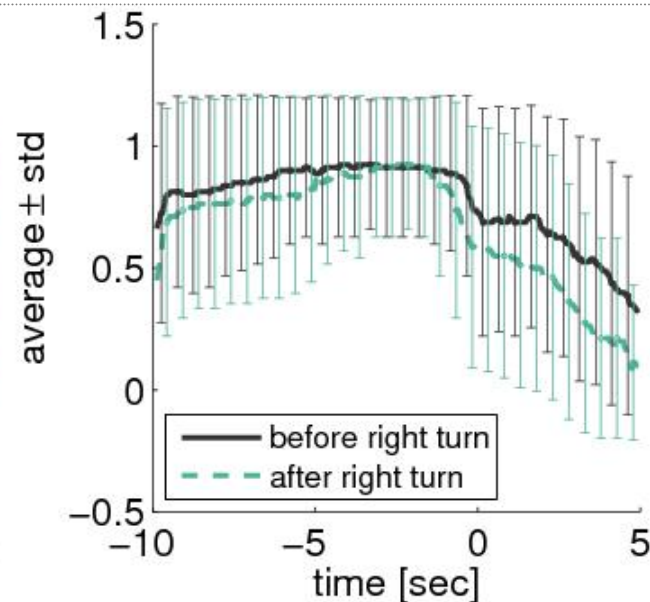
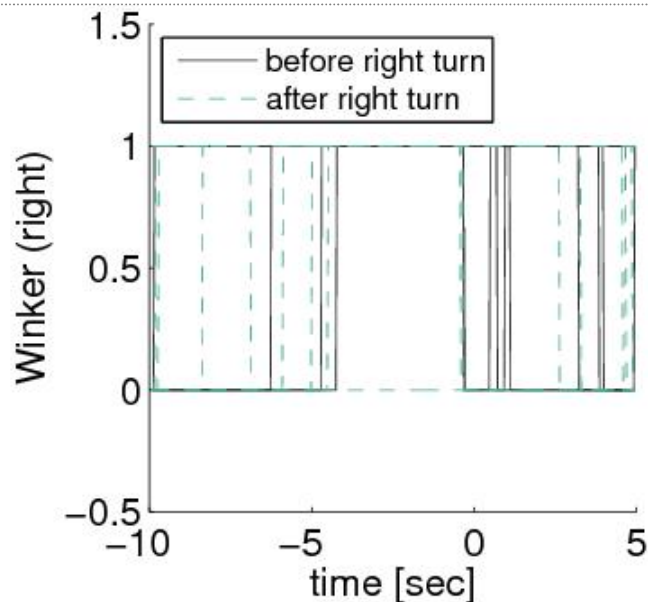


右ウィンカー

右直、右折先ヒヤリの各15標本の右ウィンカーon/offを図示

右直80個、右折先80個をそれぞれまとめたもの

• 時系列プロット



• 自己共分散係数列プロット

