

# 運転状況を考慮したドライブレコーダデータからの 潜在リスク交差点検知手法

豊田 正史<sup>†</sup> 横山 大作<sup>†</sup> 伊藤 正彦<sup>†</sup>

<sup>†</sup> 東京大学生産技術研究所 〒 153-8505 東京都目黒区駒場 4-6-1

E-mail: †{toyoda,yokoyama,imash}@tkl.iis.u-tokyo.ac.jp

あらまし 交通事故の件数は近年減少傾向にあるものの依然として深刻な問題であり、運送事業者等においては交通安全への取り組みが活発に行われている。本論文では、運転者や通行者が注意を払うべき潜在リスクを持つ交差点を大規模なドライブレコーダデータから検知する手法を提案する。本手法では、大規模な運転操作データに加え、運転が行われた道路の形状や天候等の状況を考慮して実際に事故が発生した交差点と特徴が類似する交差点を潜在リスクとして検知する。数千人規模のドライブレコーダデータを用いた実験を行いその有効性を示す。

キーワード 交通, ドライブレコーダ, 潜在リスク

## 1. はじめに

内閣府の平成 28 年版交通安全白書 [1] によると、交通事故発生件数は近年減少を続け平成 27 年には年間約 53 万 7 千件となったものの、自動車保有台数 1 万台当たりの死傷者数は、平成 3 年から横ばいになり近年では漸減傾向が続いており、依然として深刻な社会的課題である。今後さらに事故や死傷者数を減少させるためには、運転者および車以外の通行者に対し、運転データ等に基づくより精密な交通安全支援が必要となることが見込まれる。

自動車に関する交通事故の証拠映像の記録などを目的としたドライブレコーダの市場は現在も拡大を続けており、交通安全に資するデータが大規模に獲得できる状況が実現している。GfK ジャパン社の調査<sup>(注1)</sup>によると、ドライブレコーダの国内販売台数は 2015 年には前年比 42% 増の 61 万台となり、自動車保有者における製品認知率も 97% へ上昇している。ドライブレコーダはもともと事故の瞬間を記録するものであったが、近年では高機能化が進んでおり、記録できる情報の量も種類も増えてきている。事故の記録だけでなく、ブレーキやハンドル等の運転操作を常時記録できるようになってきており、大規模な運転操作データが蓄積されている。

本論文では、大規模な業務用ドライブレコーダデータを用いて、道路上での運転操作の特徴に関して詳細に分析し、さらに運転が行われた道路や天候の状況などの異種データを統合して潜在的リスクのある道路を検知する手法を提案する。ここでは交差点の事故に着目し、実際の交通事故発生交差点における運転操作の特徴を学習し、特徴の類似した交差点を潜在リスクとして検知する。1 つの事故の背後には、多数のヒヤリハット事例が隠れており、運転操作の分析から交通事故に関するヒヤリハットを抽出し、潜在リスクとして示すことができれば事故防

止への活用が期待できる。例えば、運送・運輸事業者におけるドライバーの教育や、自治体におけるヒヤリハット地図の提供、道路修繕の優先順位付けなどへの利用が考えられる。現在、多くの自治体では、住民にアンケートを取るなどして道路の危険箇所を示した「ヒヤリハット地図」を作成しているが、主に人手で作成されているためカバーできる範囲や正確性に問題がある。また警視庁では実際に事故が発生した箇所を「交通事故発生マップ」として公開している<sup>(注2)</sup>が、実際に事故が発生した場所のみを注意するだけでは不十分である。

我々は、ドライブレコーダのメーカーと協力し、バス事業者の大規模なドライブレコーダデータ、道路の数値地図、雨量データを用いた実験を行った。これらの異種データを統合した特徴量を設計するためのデータ統合化方法を示し、操作数やその属性分布、異常操作数などの特徴量を用いた潜在リスク交差点の検知手法を設計した。実験の結果、交通事故リスクに寄与する特徴量に関するいくつかの知見が得られたためこれを報告する。

本論文の構成は以下の通りである。2. 節では関連研究を述べる。3. 節では研究に用いたデータセットの詳細を説明する。4. 節ではドライブレコーダデータと道路や天候状況を考慮した、データ統合化手法ならびに潜在リスク交差点検知手法を提案する。大規模な業務用ドライブレコーダデータを用いた実験結果を 5. 節で述べる。6. 節でまとめと今後の課題について述べる。

## 2. 関連研究

カーナビなどから得られた危険運転が行われた場所を提供するサービスとして本田技研工業株式会社のインターナビ<sup>(注3)</sup>や東京都トラック協会の G Map<sup>(注4)</sup>がある。実装の詳細は不明であるが、これらのサービスでは急ブレーキや急ハンドルの発生

(注1): 「ドライブレコーダの 2015 年販売動向と消費者動向」(2016 年 3 月)  
<http://www.gfk.com/jp/insights/press-release/2015-17/>

(注2): <http://www3.wagamachi-guide.com/jikomap/>

(注3): <http://www.honda.co.jp/internavi/>

(注4): <http://gmap.dgis.jp/dc/ngt.html>

等，単純な指標しか用いておらず，本研究で用いているような詳細な運転操作の属性については考慮されていないようである．

ドライブレコーダデータから得られる詳細な運転操作データを分析した研究としては，[2-4] などがあるが，いずれもドライバーの分類を目的としており，道路のリスクに関する分析は行われていない．[2,3] では少数のドライバーを詳細に分析することを試みており実験から得られるデータを道路リスクの分析に用いることは難しい．我々の研究グループは [4] において本研究と同種のドライブレコーダから得られた大規模なデータを用いてドライバーのリスクを推定する手法を研究している．このような 1,000 人規模のデータになると道路の網羅率も高くなり，道路リスクを分析することが可能となる．

道路の事故リスクに関する研究では，これまで主に道路の形状や環境および交通量や事故データを用いて分析が行われてきた [5-7]．これらの研究では，交差点等における事故の発生率を様々な道路特徴から推定する関数を Safety Performance Function (SPF) と称し，道路の幅やレーン数，曲率などの特徴や交通量，事故発生データなどから事故率を回帰等によって求め，事故の発生原因などを分析している．しかし，交通データに関しては，交通量や平均速度など基本的な統計量のみを用いており，詳細な運転操作データを用いたものは見当たらない．

交通事故発生データの視覚的な分析を行った研究としては，[8-10] などがあるが，主に事故の統計的な分析や閲覧を目的としたもので，道路の潜在的なリスクを把握しようとする試みはなされていない．我々は，ブレーキ操作やハンドル操作における車両の加速度や躍度といった詳細な情報を可視化し，要注意道路を視覚的に探索するシステムを開発してきたが [11]，手動による探索には限界があり，潜在リスク箇所を自動的に検知する手法による支援が必要不可欠である．

### 3. データセット

本節では，本研究で用いたデータセットの詳細について述べる．本研究では業務用ドライブレコーダを設置しているバス事業者から提供を受けたドライブレコーダデータおよび交通事故データを用いている．さらに道路状況，運転状況を考慮するため道路数値地図データ，および雨量データを用いている．

#### 3.1 ドライブレコーダデータ

ドライブレコーダのメーカーである株式会社データテックの協力の元，顧客であるバス事業者から実際に運行されているバスのドライブレコーダの提供を受けた．本データは，2700 人強のドライバーの約 1 か月分の走行データを含む大規模なものである．本データは，GPS から取得した 0.5 秒毎の車両の軌跡データと，ブレーキやハンドル操作といった操作の際に記録される運転操作データを用いている．

本実験で用いたドライブレコーダは，危険運転時に限らず全ての運転操作を記録している．各種運転操作を走行中に検知し，操作毎に操作が行われた時の時間，位置，速度や加速度等の代表値を記録する．例えば，ブレーキ操作が検知されると，そのブレーキ操作が行われていた時間帯中の最大速度，最大前後加速度，最大前後加速度変化（ジャーク）の値がブレーキ操作の

代表値として記録される．本実験では，ブレーキ操作とハンドル操作の 2 種類の操作記録を用いている．記録されているデータの属性の詳細を表 1 に示す．

#### 3.2 数値地図データ

運転操作の行われた道路状況を取得するために，道路の数値地図データを用いた．本実験では，住友電工の拡張版全国デジタル道路地図 DB を用いてバス事業者の営業地域における交差点（ノード）の緯度経度情報，および交差点に接続している道路（リンク）の属性情報を用いている．リンク属性としては，道幅，交通制限などの情報が利用可能であるが，今回は主に道幅の属性を用いている．拡張版全国デジタル道路地図 DB においては，各リンクに 4 段階（0~3m，3m~5.5m，5.5m~13m，13m~）の道幅情報が付されており，道幅と運転挙動の関係を考慮したリスク交差点検知のための特徴量を設計する．

#### 3.3 事故データ

ドライブレコーダデータの提供を受けたバス事業者において実際に起こった事故の記録，及び交通事故総合分析センター（ITARDA）が提供しているイタルダ区間別データ（平成 24~26 年事故版）を用いて，正例となる交差点データを抽出した．バス事業者の事故記録については，交差点で起こった事故の記録から，緯度経度を特定し上記数値地図の交差点にマップした．そのうち運転操作が十分に行われている交差点を使用した．事故の記録には，ドライバーに責任がないものも含まれているが，今回はドライバー自身のリスクではなく，場所のリスクを検知することが目的であるため全ての事故を正例として用いた．ITARDA 区間別データでは，警察庁からの交通事故データが道路地図にマップされており，平成 24 年から 26 年の間に 1 回以上交通人身事故が発生した交差点を正例として用いる．

#### 3.4 雨量データ

運転操作の行われた時の状況として天候も考慮する．国土交通省 X バンド MP レーダーネットワーク（XRAIN）から得られた 250m メッシュ，1 分単位の雨量情報を用いて運転操作の行われた場所および時間における雨量を天候状況情報として用いる．

## 4. 潜在リスク交差点検知手法

本節ではまず，ドライブレコーダデータ，数値地図データ，雨量データの 3 つの異種データを統合するための方法について述べ，その後，統合したデータを用いて潜在リスク交差点を検知するための手法について述べる．

#### 4.1 ドライブレコーダデータ，数値地図データ，および雨量データの統合化

提案手法では，交差点のリスクに関する特徴を，その交差点において行われた運転操作の集合，および各運転操作が行われた時の道路および天候状況を用いて表現する．このため本手法では，運転操作を基本データとし，各運転操作に対して，その運転が行われた時の道路および雨量を紐づける統合化処理を行う．その詳細を以下に述べる．

（1）各運転操作の前後の時点での道路を特定するため，操作時の座標，および前後 3 秒の時点における座標を GPS の軌

表 1 運転操作データの詳細

運転操作	共通属性	操作特有の属性
ブレーキ操作	時刻, 緯度, 経度	最大速度, 最大前後加速度, 最大前後加速度変化
ハンドル操作	時刻, 緯度, 経度	最大速度, 最大角加速度, 最大角加速度変化, 最大左右加速度

跡データから取得する。

(2) 数値地図データについて、座標から近傍の道路リンクを取得するインデックスを作成し、これを用いて上記3地点を最近傍の道路リンクへ射影する。この3地点に対応する緯度経度、リンクID、道幅、交通規制などの必要情報を当該運転操作に付加する。

(3) 雨量データについて、座標および時刻から雨量を取得するインデックスを作成し、運転操作が行われた位置、時刻における雨量を雨量データから取得し、当該運転操作に付加する。

この処理により、各運転操作から道路、雨量の情報を参照することが可能となる。データの冗長性は増すことになるが、実験の過程において運転操作のスキャンおよび運転状況データの参照を繰り返し行うことになるため、統合化を行う方が効率的である。運転状況が付加された本運転操作データは、道路リスクの分析のみならずドライバーの分析にも利用可能であり汎用性も高い。

#### 4.2 潜在リスク交差点検知のための特徴量

上記の統合化データを用いて、交差点のリスク検出のための特徴量を設計した。まず、数値地図を用いてバス事業者の営業エリア内の交差点をすべて抽出し、各交差点に関して30m以内で行われた全ての運転操作を抽出する。さらに、操作数が一定以上ある交差点のみを実験対象として抽出する。この運転操作の集合を用いて交差点の特徴を表現することになる。本手法で用いる特徴量の詳細を以下に述べる。

**操作数** 交通量の多い交差点は事故の発生する可能性が高いことを反映するため、交差点で行われた操作回数を疑似的な交通量として用いる。ブレーキ操作およびハンドル操作の2種類の回数を使用する。

**操作属性値の分布** 交差点で行われた操作の属性値の分布を特徴量として用いる。ここでは、各属性値（最大速度、最大加速度等）の平均値および標準偏差を使用する。

**異常操作数** 急ブレーキ、急ハンドル等の異常操作が行われた回数が多いと事故が発生する可能性が高くなるという仮説を反映するため、異常操作を検出しその数を特徴量として用いる。最大速度と最大加速度、最大加速度と最大加速度変化など属性値の間にはある程度の相関関係が認められるため、マハラノビス距離を用いた異常値検出手法を使用する。特徴量としては、マハラノビス距離が閾値以上となる操作の数をを用いる。今回の実験では複数の閾値(2,4,6,8)を用い、各閾値を超える操作の数を特徴量として用いた。

**道幅毎の操作数** 交差点では異なる道幅の道路が交わることもあり、太い道から細い道へ右折で進入したり、交通量の多い太い道へ細い道から合流したりする場合などに事故リスクが増加する。こうした運転状況を考慮するため、操作3秒前の道幅と3秒後の道幅のペアをパターンとして用い、パターンごとに

操作集合を分割し、各パターンの操作数を特徴量として用いる。今回はバスのデータを用いているため3m以下の道路での操作はなく、それ以外の3段階(3m~5.5m, 5.5m~13m, 13m~)の組み合わせ(計9通パターン)を使用する。

**道幅毎の操作属性値の分布** 道幅のパターンで分割した操作集合それぞれに対して、属性値の平均および標準偏差を特徴量として用いる。

**道幅毎の異常操作数** 道幅のパターンで分割した操作集合それぞれに対して、上記の方法で異常操作を検出し、同じ閾値のセットを用いた異常操作数を特徴量として用いる。

**雨天時の操作数** 雨量による運転挙動の変化を考慮するため、雨天時(雨量1mm以上)の運転操作を抽出し、その操作数を特徴量として用いる。

**雨天時の操作属性値の分布** 雨天時の運転操作に関する属性値の平均および標準偏差を特徴量として用いる。

**雨天時の異常操作数** 雨天時の運転操作に対して、上記の方法で異常操作を検出し、同じ閾値のセットを用いた異常操作数を特徴量として用いる。

#### 4.3 潜在リスク交差点の検知

上記で設計した特徴量を用い、交差点における事故発生データを教師データとした教師あり機械学習により、潜在リスク交差点を検知する。交差点を事故発生のありなしで分類する分類器を用いる場合は、事故未発生であるにも関わらず事故ありと判定された交差点を潜在リスク交差点と考えることができる。また、交差点の事故回数に対する回帰を用いれば、事故の確率の高い順に交差点を順序付けすることで、潜在リスク交差点のランキングとして用いることができる。

### 5. 実験結果

今回の実験では、事故が発生した交差点のラベルを事故回数に関わらず1とし、事故の発生していない交差点を0として、ロジスティック回帰を行うことで潜在リスク交差点の検知を試みた。このような設定とした理由には、事故データ中の交差点のうち、十分な運転操作数が得られたものが非常に少ないことがある。ドライブレコーダデータの期間が1月程度であることから、1日平均5回程度の運転操作がある交差点を対象としたところ(150以上のブレーキ操作を閾値として使用)、事故発生交差点数は2,061箇所であった。それに対し、対象となる交差点数は23,164箇所あり、正例が8.9%と少ない。このため事故回数での回帰は難しく、また負例中にバス以外の事故が実際に起こっている交差点が多数あることが想定されることから、負例に関する精度にはほぼ意味がない状況である。

こうした理由から、潜在リスク交差点検知の評価方法としては、ロジスティック回帰による交差点のランキングにおいて、事故発生交差点が上位にランキングされることを目標として、

表 2 潜在リスク交差点検知実験の精度

特徴量	平均適合率 (Ave.P.)
ランダム	0.089
操作数	0.156
操作数+属性分布	0.214
操作数+属性分布+異常操作数	0.224
操作数+属性分布+異常操作数+道幅毎操作数	0.232
操作数+属性分布+異常操作数+道幅毎操作数+道幅毎属性分布	0.236
操作数+属性分布+異常操作数+道幅毎操作数+道幅毎属性分布+道幅毎異常操作数	<b>0.242</b>
上記全特徴量+雨天時操作数・属性分布・異常操作数	<b>0.244</b>
上記全特徴量+雨天時操作数・属性分布・異常操作数+雨天時道幅毎操作数・属性分布・異常操作数	0.242

平均適合率 (Average Precision) を用いる。上記特徴量の組み合わせに対して、L2 正則化付ロジスティック回帰を用い、平均適合率が最適となるように C パラメタのチューニングを行い、10 フォールドのクロスバリデーション時の平均適合率を評価に用いる。また、サンプルの順番による性能のばらつきが見られたため、サンプルのシャッフルとクロスバリデーションを 20 回繰り返しその平均を評価した。

実験結果を表 2 に示す。平均適合率はランダムにランキングした場合、正例の比率と同じになるため 0.089 が最低の基準値となる。操作数のみを用いたものが実質的なベースラインで 0.156 となっており、交通量が事故発生に寄与するという仮定がある程度裏付けられた形である。

それ以降は、設計した特徴量を追加していった場合の平均適合率を示している。運転操作の属性分布、及び異常操作数を追加することにより、0.224 まで平均適合率が向上しており、交差点での運転操作における速度、加速度等の平均やその標準偏差、および異常操作数が、精度向上に寄与していることが確認できる。さらに、道幅毎の特徴量についても操作数、属性分布、道幅毎異常操作数の追加に関して、さらなる精度向上が見られ、最大で 0.242 の平均適合率となっていることから、道幅の通過パターンを考慮することが重要であることがわかる。

雨量に関する特徴量に関しては、ここまでの全特徴量に加えて雨天時の操作数・属性分布・異常操作数を加えたものと、さらに雨天時の道幅を考慮した特徴量を加えたもので精度を測定している。表 2 の下段にその結果を示す。道路幅を考慮しない、雨天時のすべての運転操作に関する特徴量の追加に関してはわずかながら性能向上が見られた。しかし雨天時の道幅毎の特徴量に関しては精度向上は見られなかった。雨天時の運転操作数は全体の 1 割程度に留まっており、さらに道幅毎にデータを分割したことにより操作数がスパースになることが原因として考えられる。より長期のデータを用いることによりスパースネスの問題はある程度緩和されると考えられ、今後実験を継続して行う予定である。

また、設計した特徴量が精度にどの程度寄与しているか調べるため、雨量関連以外のすべての特徴量を用いた分類器から、1 種類ずつ特徴量を除いて精度の変化を見たものが表 3 である。すべての特徴量を用いた分類器の平均適合率 0.242 と比較すると、各特徴量を除いた場合の精度の下落はほとんどの場合非常に小さい。これは特徴量同士にある程度重複した情報が含まれ

ており互いを補完しているからと考えられる。その中でも最も下落の度合いが大きかったものが道幅毎の属性分布であることから、異常操作の影響よりも通常の運転時の各種属性の分布の方がより重要であることがわかる。これは、急ブレーキや急ハンドルのみを指標として用いていた従来の方法に限界があることも示唆している。

表 3 各特徴量を除いた時の精度 (雨量関連特徴量以外)

除いた特徴量	平均適合率 (Ave.P.)
なにも除かない場合	0.242
-操作数	0.239
-属性分布	0.239
-異常操作数	0.239
-道幅毎操作数	0.240
-道幅毎属性分布	<b>0.236</b>
-道幅毎異常操作数	0.239

## 6. まとめと今後の課題

本論文では、運転者や通行者が注意を払うべき潜在リスクを持つ交差点を大規模なドライブレコーダデータに加え、運転が行われた道路の形状や天候等の状況を考慮して検知する手法を提案した。バス事業者の大規模なドライブレコーダデータ、道路の数値地図、雨量データという異種データを統合した特徴量を設計するためのデータ統合化方法を示し、操作数やその属性分布、異常操作数などを用いた特徴量を用いた、潜在リスク交差点の検知手法を設計し、実験を行った。実験結果から、交差点での運転操作数および異常操作数が、事故リスクに寄与することが確認でき、さらに道幅や雨量を考慮することにより検知性能が向上することが示された。今後は潜在リスク交差点の検知結果を可視化し、バス事業者や自治体へ提供するため、リスクの原因を詳細に分析し、わかりやすく提示する方法についても検討を進めている。

## 文 献

- [1] 内閣府. 平成 28 年版交通安全白書. 2016.
- [2] Y. L. Murphey, R. Milton, and L. Kiliaris. Driver's style classification using jerk analysis. In *2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems*, pp. 23–28, March 2009.
- [3] D. A. Johnson and M. M. Trivedi. Driving style recognition

- using a smartphone as a sensor platform. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1609–1615, Oct 2011.
- [4] D. Yokoyama and M. Toyoda. Do drivers' behaviors reflect their past driving histories? - large scale examination of vehicle recorder data. In *2016 IEEE International Congress on Big Data (BigData Congress)*, pp. 361–368, June 2016.
- [5] Ketong Wang, Jenna K. Simandl, Michael D. Porter, Andrew J. Graettinger, and Randy K. Smith. How the choice of safety performance function affects the identification of important crash prediction variables. *Accident Analysis and Prevention*, Vol. 88, pp. 1 – 8, 2016.
- [6] Francesca Russo, Mariarosaria Busiello, and Gianluca DellaAcqua. Safety performance functions for crash severity on undivided rural roads. *Accident Analysis and Prevention*, Vol. 93, pp. 75 – 91, 2016.
- [7] Alfonso Montella and Lella Liana Imbriani. Safety performance functions incorporating design consistency variables. *Accident Analysis and Prevention*, Vol. 74, pp. 133 – 144, 2015.
- [8] Anna Fredrikson, Chris North, Catherine Plaisant, and Ben Shneiderman. Temporal, geographical and categorical aggregations viewed through coordinated displays: A case study with highway incident data. In *Proceedings of the 1999 Workshop on New Paradigms in Information Visualization and Manipulation in Conjunction with the Eighth ACM International Conference on Information and Knowledge Management, NPIVM '99*, pp. 26–34, New York, NY, USA, 1999. ACM.
- [9] M. L. Pack, K. Wongsuphasawat, M. VanDaniker, and D. Filippova. Ice-visual analytics for transportation incident datasets. In *2009 IEEE International Conference on Information Reuse Integration*, pp. 200–205, Aug 2009.
- [10] A. Anwar, T. Nagel, and C. Ratti. Traffic origins: A simple visualization technique to support traffic incident analysis. In *2014 IEEE Pacific Visualization Symposium*, pp. 316–319, March 2014.
- [11] M. Itoh, D. Yokoyama, M. Toyoda, and M. Kitsuregawa. Visual interface for exploring caution spots from vehicle recorder big data. In *2015 IEEE International Conference on Big Data (Big Data)*, pp. 776–784, Oct 2015.