

# ドライブレコーダデータを利用した安全運転支援システム構築に向けて

横山 大作<sup>†</sup> 豊田 正史<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: †{yokoyama,toyoda}@tkl.iis.u-tokyo.ac.jp

あらまし ドライブレコーダの普及により、ドライバーの運転挙動に関するセンサデータを長期的にわたって収集し、解析することが可能となりつつある。我々は、この運転挙動とドライバーの過去の事故歴との間の関係性について解析を行い、両者の間に関係があることを明らかにした。運転管理者が安全運転支援を行う際、この知見を利用するために必要となる事柄について整理し、展望について述べる。

キーワード ドライブレコーダ, 安全運転支援, 運転操作特徴

## 1. はじめに

様々な技術的・制度的改良などにより、日本における交通事故件数は年々減少してきてはいるが、依然として多くの被害が発生している。警察庁交通局の統計<sup>(注1)</sup>によると、2016年において499,201件の交通事故が発生し、死傷者数は622,757人の上っている。このうち、事業用の自動車による事故は16.8%を占めており、事業者にとって重要な問題となっている。これに対し、運輸業などを中心として、事故が発生した際に前後数秒間の画像を自動的に保存するドライブレコーダが普及を始めており、ドライブレコーダの導入によって事故率が低下したという事例も報告されている[1]。近年では、GPSによる走行経路記録、ブレーキなどの運転操作時の加速度などを運転中全てにわたって記録する、高機能なドライブレコーダが開発されており、運転後に急ブレーキなどの好ましくない運転がなかったか検証するなどの取り組みを通して、安全運転教育に活かせず試みも始まっている。しかし、急ブレーキは良くない、など経験的に理解しやすい事柄に関する教育が行われているのみであり、運転の安全性を客観的に判断できるほどの理解には至っていない。

我々は、大規模なドライブレコーダデータを利用し、ドライバーの現在の運転操作からそのドライバーの安全運転度の指標を得て、安全運転支援に役立てるシステムの構築を目指している。安全運転度は定義が難しいため、我々は、ドライバーの過去の事故歴の有無を指標とし、機械学習によってこれを運転操作の特徴から判別することを試みる。本論文では、ドライバーごとの運転操作分布と、道路幅や天候などの運転状況情報を組み合わせて得られる特徴量を用いることで、過去に事故を起こしているドライバー群を分類できる判別器を生成できることを示す。また、不均衡な問題に対応するため、オーバーサンプリングを利用することで有用な判別器が得られることを示す。

## 2. 関連研究

運転時の速度や位置情報などを記録し、様々な用途に役立

てようという研究が提案され始めている[2][3]。The 100-Car Naturalistic Driving Study [4]は大規模に収集されたドライブレコーダデータを利用した研究の代表例である。このアーカイブは専用の機材を用いた詳細な運動記録と、性別などのドライバー属性からなり、詳細な統計的解析を加えている。Guoら[5]はこのアーカイブを用い、運転の安全性に関するリスクが高いドライバーを判定するモデルを提案した。彼らのモデルはドライバー属性とまれな危険運転イベントの出現に依存しており、通常運転時の運転操作に着目したものではない。Zhengら[6]は中国において自然な運転時の運動記録を収集し、危険度との関係を解析した。彼らの解析もまた、高い加速度を持つ運転操作のみに限定されている。

運転挙動に着目してドライバーを分類する研究もいくつか存在する。Higgsら[7]は、3人のドライバーを対象に、車の後ろに付いている時の挙動を分類し、ドライバーごとに違いがあることを示した。Dangら[8]は、レーン変更時の挙動に着目し、12人のドライバーの間に変更頻度などの違いが存在することを示した。しかし、これらの研究では、違いがはっきりしていると想定される少数のドライバーの分類にとどまっている。また、運動に関する詳細な情報が利用できることを想定しているが、これはコストの面から一般的に利用できるとは限らない。

## 3. ドライブレコーダデータの解析実験

### 3.1 ドライブレコーダデータ

株式会社データ・テックの協力の下、運送会社で東京都内の営業所に所属するドライバー2200人強、2014年7月から2016年11月まで2年強の運転動作記録を用いて解析を行った。株式会社データ・テックが開発した高性能ドライブレコーダである、セーフティレコーダ“SRVideo”を搭載した営業車の運転操作記録を収集した。対象車両は一般家庭への配達・集荷を行っており、区外にある営業所と担当地区間の移動、ならびに担当地区内の集配における運転が主に記録されている。

運転操作の特徴を分析する際には、ある程度長い期間にわたって運転を行っているドライバーを対象にする必要がある。今回は、記録された総運転時間が10時間以上のドライバーを対象とすることにした。また、10時間以上運転時間があつて

(注1): [http://www.npa.go.jp/publications/statistics/koutsuu/H28\\_zennjiki](http://www.npa.go.jp/publications/statistics/koutsuu/H28_zennjiki)

表 1 解析対象ドライブレコーダデータ  
全収集データ

ドライバー数	2,236
合計運転時間	209,343 時間
運転時間 10 時間以上、基本運転操作を 300 回以上記録 運転経験 3 年以上	
ドライバー数	1,336
合計運転時間	198,915 時間

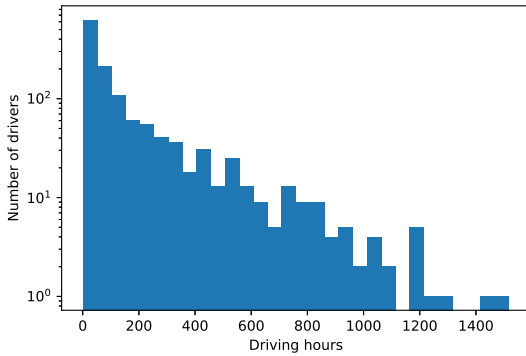


図 1 解析対象ドライバーの運転時間分布

も運転操作の回数が極めて少ないドライバーが存在したため、ブレーキとハンドル操作の双方とも 300 回以上記録されたドライバーのみを解析対象とした。これは、拠点間の長距離運送など、業務形態が異なるドライバーが混在しているために起きたと考えられる。さらに、3 年以上の運転経験を持つドライバーのみに解析対象を限定した (詳細は 3.2 章に記す)。

全データ、並びに解析対象としたデータの統計量を表 1 に示す。また、解析対象ドライバーの運転時間の頻度分布を図 1 に示す。

### 3.1.1 ドライブレコーダの記録内容

ドライブレコーダは前後加速度計、左右加速度計、方位ジャイロ、GPS を搭載している。このセンサを利用して、ブレーキなどの基本的な運転動作を検出するとともに、ブレーキの急激さなどの情報を記録している。今回の論文では、以下の 4 動作を解析対象とする。

- ブレーキ動作

速度を調節するためにブレーキを踏んだ時の動作。ブレーキ時の前後方向加速度とその時間微分などを記録している。

- 停止動作

停止するときの動作。前述のブレーキ動作とは必ずしも重複しない。前後方向加速度、ブレーキ開始から停止までの所要時間などを記録している。

- ハンドル動作

カーブ、車線変更などにおけるハンドル操作。方位角速度とその時間微分などを記録している。

- 右左折動作

交差点で右左折を行った時の動作。前述のハンドル動作とは必ずしも重複しない。交差点への進入速度、曲がった後の速度、

表 2 運転操作レコード数統計

操作種別	ドライバーごとの操作レコード数 (min, max)		全レコード数
	min	max	
ブレーキ	302	92371	8458181
ハンドル	301	105747	12132843
右左折	69	53278	5393134
停止	223	84630	8832775

表 3 分類実験に用いる正例負例数

3 年間事故あり	248
事故なし	1088

方位角速度のゆらぎなどを記録している。

ドライブレコーダは、これらの動作が 1 回行われるごとに 1 エントリを記録する。運転 10 時間以上、ブレーキ・ハンドル操作記録 300 回以上のドライバーに限定した解析対象データの統計量を表 2 に示す。

### 3.2 運転経歴データ

運送会社の協力により、ドライバーの過去 10 年にわたる事故履歴と運転経歴を利用することができた。

事故履歴のみの情報では、あるドライバーが事故を起こしていなかったとしても、必ずしも無事故期間が長いことを意味しない。例えば、そのドライバーが運転を始めたばかりであれば、事故記録がない場合であっても安全なドライバーである保証はない。そこで、ドライバーの運転経歴を利用し、最も古い運転経歴記録、及びドライバーの免許取得日のいずれか古い方をドライバーが運転を開始した時点とみなして、ドライバーの運転経歴年数を推定した。今回の実験では、推定した運転経歴年数が 3 年以上のドライバーについて、直近の 3 年以内に事故を起こしているかいないかの情報を利用した。なお、事故の記録には、ドライバーに過失のない軽微な事故も含まれている。

### 3.3 機械学習を用いた分類

ドライバーの運転の安全度を直接定義することは難しい。ここでは、ドライバーが過去に事故を起こしているか否かを、そのドライバーの運転の安全度の指標として用いることとした。運転経歴が 3 年以上あるドライバーに関して、過去 3 年以内に 1 回も事故を起こしていない場合、その人は安全運転をしているのではないかとある程度考えられる。そこで、この情報を利用し、運転操作記録の特徴からドライバー事故歴の有無を判別する問題であると定式化した。表 3 に、利用したドライバー群における事故歴の有無の割合を示す。事故ありのドライバーは 18.6% の割合であり、正例の割合が偏った判別問題である。

### 3.4 特徴量設計

ドライブレコーダに記録された運転操作記録の分布には、ドライバーの特徴が現れていることが予想される。そこで、それぞれの要素について平均値と標準偏差を求め、これを特徴量として機械学習による判別器の作成を試みた。

また、運転操作は状況によって大きく影響を受ける。例えば、ハンドル操作時の角加速度は運転速度によって影響されることが考えられる。これらの運転状況を考慮するため、以下の項目について情報を付加し、状況が異なる運転操作を別々に集計し、

それぞれの平均と標準偏差を求めて特徴量とした。

#### a) 動力学制約の考慮

高速でハンドル操作を行うときは、小さい角速度でなければ曲がれない、というように運転操作は動力学制約を持つ。これを考慮するため、制約を最も強く規定すると考えられる速度に着目し、運転操作が行われた時の速度を考慮した特徴量を作成した。いくつかの速度帯で操作ログを分け、別々に分布の統計量を導出こととする。

#### b) 道路特徴の考慮

運転操作が行われた場所の GPS データを数値地図とマッチングすることで、操作が行われた道路に関する情報を得ることができる。数値地図データとしては、東京大学空間情報科学研究センター研究用空間データ基盤より、住友電設による Advanced Digital Road Map Database の提供を受けて利用している。マッチング方法は、単純に最近傍の道路セグメントを探索するものとした。ただし、最近傍道路からの距離が 30m 以上の場合は、道路情報なしとして扱った。道路が対応付けられない状況は主に、大規模な施設内の構内道路など、数値地図に掲載されていない道路を走行している場合に発生する。

ここでは、特に操作に大きな影響を与えとされる特徴として、道路の幅について考慮することとした。数値地図の道路には以下の情報が付加されている。

- 道路幅: 5.5m 未満、5.5~13m、13m 以上、不明
- 一方通行か否か
- 中央分離帯があるか否か

ここで、一方通行でない、または中央分離帯が存在しない場合には道路幅を専有できないことを考慮して、そのような場合に道路幅を 1 区分小さいものとして扱うこととした。

#### c) 気象の影響

路面の状況は運転に大きな影響を与えとされる。ここでは、国土交通省が運用する X-band radar による降雨情報 (XRAIN)<sup>(注2)</sup> を利用し、降雨の有無に着目して特徴量を作成した。XRAIN は 250m メッシュ、1 分単位という細粒度の降雨情報が得られ、近年都市に多く発生しているゲリラ豪雨のような局所的な降雨を検出することが可能である。

運転操作が行われた場所の GPS データと時刻情報から、該当する時点の降雨量を抽出し、5mm/h 以上、0mm/h から 5mm/h、降雨なし、の 3 通りで分割して特徴量化を行った。

#### d) 交通量の影響

朝夕のラッシュ時など、交通量は運転に大きな影響を与えとされる。ここでは、運転操作が行われた時刻に着目し、1 日を 5 つの時間帯 (6~9 時、9~12 時、12~18 時、18~21 時、21~翌 6 時) に分割して特徴量化を行った。

### 3.5 特徴量選択

3.4 章に記述した運転操作に関する特徴に加え、以下の項目を特徴量とした。

- 年齢
- 運転経験年数

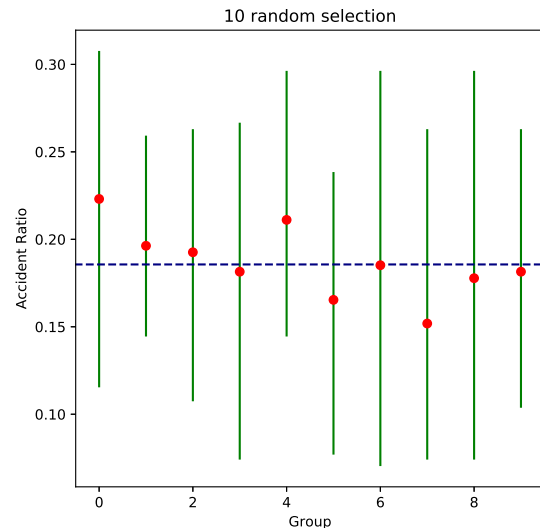


図 2 不均衡なデータで学習した判別結果に基づいたドライバー群の事故率分布

- ゴールド免許の有無
- 運転時間

生成した特徴量には、交差点の進入速度と最高速度など、相関の高い項目が存在すると考えられる。そこで、相関係数の絶対値が 0.7 を超える特徴量については片方を削除するという方針で特徴量の選択を行った。結果、最終的に利用する特徴量は 196 項目となった。

### 3.6 事故履歴判別実験

得られた特徴量を用い、ドライバーの事故歴の有無を判別する判別器を学習してその判別性能を評価した。

判別には Gradient Tree Boosting アルゴリズムの XGBoost [9] を利用した。正例割合を保存したランダムサンプリングを行い、80%のドライバーのデータを学習とパラメータチューニングに使い、20%のドライバーをテスト用として判別スコアを導出した。このスコア順にテスト用ドライバーを並べ、上位から 10%ごとに分割した 10 個のグループを作成し、それぞれのグループに含まれたドライバーが実際に事故を起こしていたかどうか、から得られる事故率を求めた。この実験を 10 回繰り返した結果を図 2 に示す。各グループの事故率の平均値を赤い点で、実験結果の 80 パーセンタイル値を緑の棒で示している。グループは 0 が上位 10%(すなわち、判定器が最も事故ありらしいと判断したグループ)、1 が上位 10~20%、という順で並び、右に行くほど判定器が事故無しらしいと判断したグループとなる。青い点線がドライバー全体の事故率 18.6%を示している。上位グループの方が若干実際の事故率が高い傾向にはあるが、上位と下位のグループ間の差異は誤差内にとどまっていることが見て取れる。

このデータセットは、正例の割合が低い不均衡な集団であるため、これを考慮することが必要であると考えた。そこで、80%の学習用ドライバーを抽出した後、オーバーサンプリングを

(注2): XRAIN: <http://www.river.go.jp/kawabou/ipXAreaMap.do>

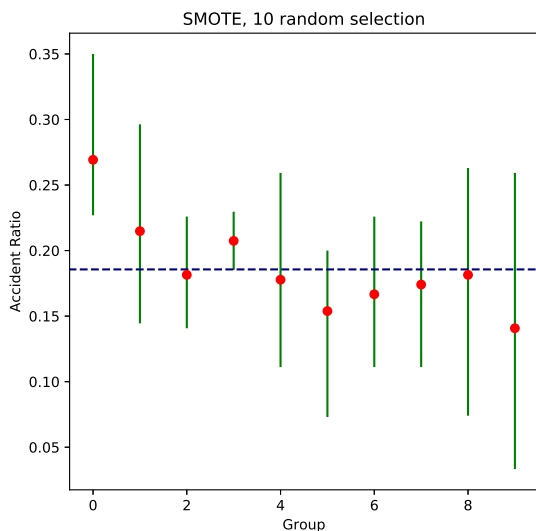


図3 リサンプリングを行った場合の判別結果に基づいたドライバー群の事故率分布

行ってから学習することを試みた。オーバーサンプリングには Synthetic Minority Over-sampling Technique(SMOTE) [10] を利用した。実験結果を図3に示す。トップ10%に相当するグループ0は、27%程度の事故ありドライバーを含んでおり、これはドライバー全体の事故率より9ポイント程度高く、80パーセントの誤差を越えた差異を示している。すなわち、学習によって得られた判定器を利用すると、事故ありと考えられる上位10%に含まれるドライバーは実際に事故を起こしている割合が高い。また、事故率の平均値は右へ行くほど減少傾向にあり、最下位グループと最上位グループとの差異は12ポイント程度あることが見て取れる。オーバーサンプリングを行うことで、学習した判定器の判別性能が向上し、事故を起こしたドライバーの割合が高いグループを分類できるようになったと言える。

得られた判定器は、ドライバーの教育に利用することが可能である。判定器によって事故ありの可能性が高いグループに分類されたドライバーは、その運転操作に、事故ありドライバーに共通する何らかの特徴があると考えられる。そのドライバーの運転操作分布を平均的な分布と比較することで、特徴的な運転操作を発見し、ドライバーへの助言に役立てるシステムの構築を検討している。

#### 4. 終わりに

我々は、大規模なドライブレコーダデータを利用し、ドライバーの現在の運転操作からそのドライバーの安全運転度の指標を得て、安全運転支援に役立てるシステムの構築を目指している。本論文では、ドライバーごとの運転操作の分布に着目し、道路幅や天候などの運転状況情報を組み合わせて特徴量化することで、過去に事故を起こしているドライバー群を分類できる判別器を構築できることを示した。事故の発生率は低く、事故

の有無を判別するデータセットは不均衡になるが、オーバーサンプリングを用いることで有用な判別結果が得られることも確認された。

今後、事業者に対してドライバーに関する情報のフィードバックを行い、安全運転支援システム構築の可能性を評価していくことを計画している。また、事故の危険性が高いと判定されたドライバーに対し、どのような運転操作の差異が存在しているのかをわかりやすく示すのか、どのように助言に役立てるのか、を検討していく予定である。

#### 文 献

- [1] 石川博敏. ドライブレコーダの効果と課題. 予防時報, Vol. 250, pp. 30–35, July 2012.
- [2] D.A. Johnson and M.M. Trivedi. Driving style recognition using a smartphone as a sensor platform. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, pp. 1609–1615, Oct 2011.
- [3] Wei Wu, Wee Siong Ng, S. Krishnaswamy, and A. Sinha. To taxi or not to taxi? - enabling personalised and real-time transportation decisions for mobile users. In *Mobile Data Management (MDM), 2012 IEEE 13th International Conference on*, pp. 320–323, July 2012.
- [4] Sheila G Klauer, Thomas A Dingus, Vicki L Neale, Jeremy D Sudweeks, and David J Ramsey. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. Technical Report DOT HS 810 594, National Highway Traffic Safety Administration, 2006.
- [5] Feng Guo and Youjia Fang. Individual driver risk assessment using naturalistic driving data. *Accident Analysis & Prevention*, Vol. 61, pp. 3–9, 2013.
- [6] Yang Zheng, Jianqiang Wang, Xiaofei Li, Chenfei Yu, Kenji Kodaka, and Keqiang Li. Driving risk assessment using cluster analysis based on naturalistic driving data. In *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, pp. 2584–2589. IEEE, 2014.
- [7] B. Higgs and M. Abbas. A two-step segmentation algorithm for behavioral clustering of naturalistic driving styles. In *Intelligent Transportation Systems - (ITSC), 2013 16th International IEEE Conference on*, pp. 857–862, Oct 2013.
- [8] Ruina Dang, Fang Zhang, Jianqiang Wang, Shichun Yi, and Keqiang Li. Analysis of chinese driver's lane change characteristic based on real vehicle tests in highway. In *Intelligent Transportation Systems - (ITSC), 2013 16th International IEEE Conference on*, pp. 1917–1922, Oct 2013.
- [9] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, pp. 785–794, New York, NY, USA, 2016. ACM.
- [10] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Int. Res.*, Vol. 16, No. 1, pp. 321–357, June 2002.