

事故防止に向けたドライブレコーダデータからの運転挙動分析

横山 大作[†] 豊田 正史[†]

[†] 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

あらかし 事故や危険運転時のみならず、通常の運転状況を全て自動的に記録する、高機能なドライブレコーダが普及しつつある。我々は、業務用の車に取り付けられた高機能ドライブレコーダデータを長期間にわたってアーカイブし、ドライバーの事故経歴などを利用して運転操作の傾向とドライバー特性との関連を理解することを試みている。ブレーキやハンドルを操作している時点での車の速度、加速度等の運転操作記録に加え、その運転操作が行われた道路の幅など、運転環境の違いを考慮することで、スキルの違いなどのドライバー特性がより明確に表現できることが期待される。本論文では、業務用車両のドライバーを対象に、記録された運転操作と事故防止に関係するドライバー特性との関係性を分析する。

キーワード ドライブレコーダ, 運転動作特徴分類, ドライバー分類

1. はじめに

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、タクシー、運送業などの業種を中心に普及を始めており、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す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、ドライバーの運転の安全性と運転操作との関連を理解することを目指すため、運送会社の営業車両が記録したドライブレコーダデータを大規模・長期間にわたって収集し、そこに記録された運転挙動からドライバーの事故履歴の有無を判別することを試みた。多くの企業や保険会社では、年齢や性別などといったドライバーの属性を用いてドライバーの安全性に関する指標を作成し、運行管理などに役立てている。しかし、ブレーキ、ハンドルなどの運転操作にはドライバーごとに特徴が現れると考えられ、これを考慮すればドライバーの運転の実情に合わせた、より精度の良い安全性の指標が構築できる可能性がある。

ドライバーの運転時の挙動に着目した研究はいくつか存在する [2] [3] [4] が、それらは運転中にまれに現れる危険な挙動のみを解析対象にしたり、少数のドライバーの判別に関する研究であった。我々は、運転中の全ての運転操作における GPS 軌跡、速度、加速度を利用することで、通常の運転時の運動特性も考慮したドライバー特徴の表現方法の構築を試みる。ドライバーごとの、速度、加速度などの運転操作分布に着目し、これに操作時の速度や道路幅などの運転環境に関する考慮を加えた特徴

量作成手法を設計した。この特徴量を利用して、SVM を用いた判別実験を行った結果、最高で適合率 0.83、再現率 0.81 という高い性能で過去の事故歴の有無を判別できることが示された。また、年齢、性別など、従来安全に関係することが知られているドライバー属性を用いるのみでは高い判別性能が得られないことも示された。

本論文の構成は以下の通りである。2. 章では関連研究を述べる。3. 章ではドライブレコーダデータの解析実験について記述する。このうち、3.1 章では利用したドライブレコーダデータの詳細を、3.2 章ではドライバーの事故履歴情報の詳細をそれぞれ示し、3.3 章以降では操作記録分布に着目した特徴量を設計し、SVM を用いてドライバーの事故履歴に関する判別実験を行う。4. 章ではまとめを行い、今後の課題について述べる。

2. 関連研究

運転時の速度や位置情報などを記録し、様々な用途に役立つという研究が提案され始めている [5] [6]。The 100-Car Naturalistic Driving Study [7] 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 [8] 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 [9] 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 [2] 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 [3] 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

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	1469
合計運転時間	77450 時間
運転日数 20 日以上、運転時間 20 時間以上	
ドライバー数	320
合計運転時間	60190 時間

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
	コード数 (min, max)	全レコード数	
ブレーキ	(114, 45861)	1993341	
ハンドル	(239, 46452)	2783723	
右左折	(121, 21027)	1218957	
停止	(418, 40625)	2221166	

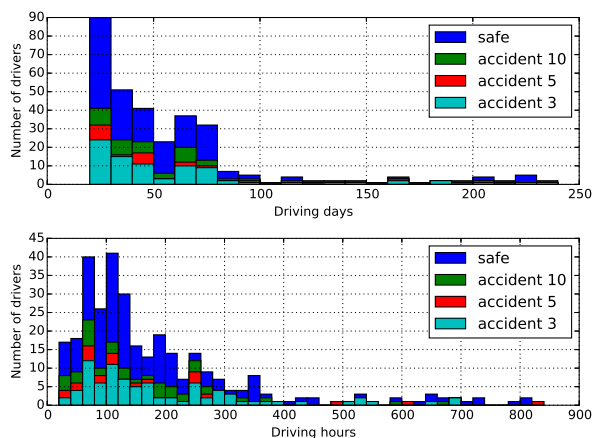


図 1 解析対象ドライバーの運転日数、運転時間に関する頻度分布

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 1400 人強、2014 年 7 月から約 1 年分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、セイフティレコーダ “SRVvideo” 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、運転を行った日数が 20 日以上、総運転時間が 20 時間以上のドライバーを解析対象とした。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日数、運転時間の頻度分布を 1 に示す。多くのドライバーについて、100 時間程度以上の運転操作が記録されている。なお、過去 3 年以内、1 回でも事故を起こしたことがあるドライバーの分布を accident 3 の系列に、5 年以内、10 年以内の事故ありドライバーを accident 5, 10 の系列にそれぞれ示している。これは、3.2 節で後述する事故履歴記録により分類されている。

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、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 20 日以上、20 時間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験年数を推定した。図 2 は、ある年数以上の運転経験を有しているドライバー数を示したものである。また、その年数以内に 1 回以上の事故を起こしているドライバー数を accident という系列で示す。

3.3 機械学習を用いた分類

ドライブレコーダに記録された運転操作には、ドライバーの安全運転度合いに関する何らかの特徴が現れていると考えられる。そこで、ドライバーの過去の事故履歴と現在の運転操作の特徴との関係性を調査し、運転の安全度合いに関する知見を得ることを試みた。

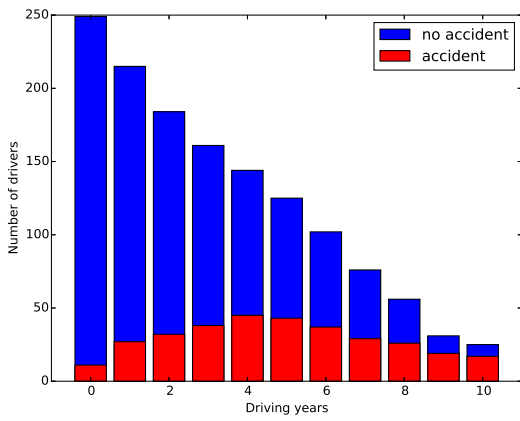


図 2 ある年数以上の運転経験を有するドライバー数

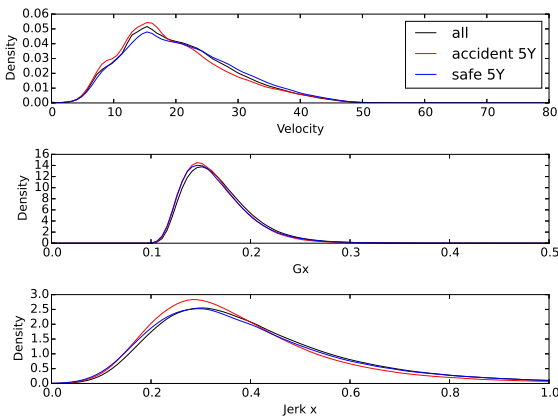


図 3 ブレーキ操作に関する操作分布

図 2 に示したように、事故の発生率は低く、あまり短い期間では安全かどうかを判断する指標にはならないと考えられるため、本論文では 5 年以上の運転経歴を持つドライバーを対象に、5 年間の事故の有無に関する解析を行う。機械学習の手法として Support Vector Machine (SVM) [10] を利用し、様々なカーネルを用いて分類実験を行った。

図 3 はブレーキ操作時の速度、前後加速度、前後躍度の頻度分布を示している。ドライバー全体の頻度分布 (系列 all) に対し、5 年以内に 1 回以上事故を起こしているドライバー (系列 accident 5Y) と、その間無事故だったドライバー (系列 safe 5Y) とは、おおむね一致するものの差異が現れる箇所も見られる。

このように、運転操作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、頻度分布を利用した特徴量を設計し、機械学習手法を適用することで過去の事故履歴の有無を判別することを試みる。

3.4 特徴量設計

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。これを考慮するため、運転操作が行われた速度帯、及びその操作が行われた道路の特徴別に集計して、分布を特徴量化する。

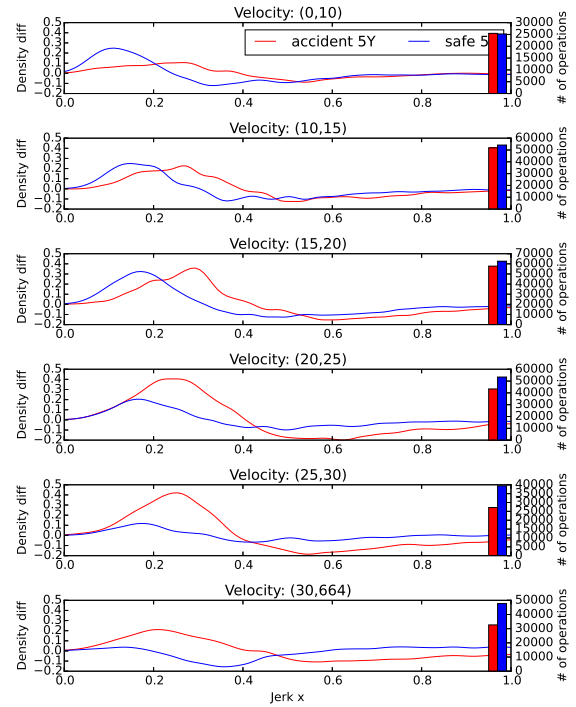


図 4 速度帯を考慮したブレーキ操作時前後方向躍度の確率分布差

a) 操作時の速度帯の考慮

ブレーキ操作時前後方向躍度について、運転操作が行われた時の速度を、10km/h 未満、10~15km/h、...、30km/h 以上、の 6 つの速度帯に分けて集計した時の分布の違いを図 4 に示す。Y 軸は、事故の有無で分割したドライバーグループごとの、全ドライバーの分布からのずれを表している。速度帯ごとにずれの特徴が異なり、特に低速から中速域において事故ありのドライバーが高い躍度の操作を多く行っていることがわかる。また、右側の軸近くの棒グラフは、集計対象操作のエントリ数を示している。数値目盛りは右側の Y 軸となる。操作数の割合も速度帯ごとに異なっていることがわかる。

このような違いを考慮するために、運転操作を速度帯ごとに分割して集計し、分布を求めることとする。

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株式会社ゼンリンによる Zmap TOWN II 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られられる特徴として、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

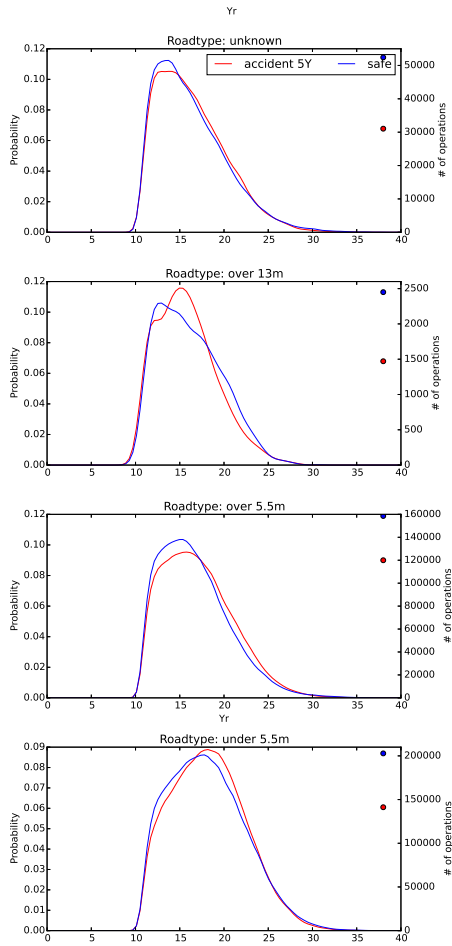


図5 道路幅を考慮したハンドル操作時の角速度分布

- 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。道路幅ごとに運転操作を分割集計した時の分布の違いを図5に示す。

図5は、ハンドル操作時の角速度分布を道路幅ごとに集計したものを示している。上から順に、道路幅不明、13m以上、5.5~13m、5.5m未満の分布である。道路幅ごとに分布の形が異なっていることがわかる。また、5.5m未満の細い道では、事故ありのドライバーの方が角速度が高い操作を行っている頻度が多いなど、事故の有無の違いによる分布の違いも見取れる。

右側の軸近くにプロットしてある点は、集計対象となった操作数を示している。13m以上の太い道路の運転操作数は他の道路と比較して少なく、分布を考慮する際に誤差が大きいが予想される。

c) 出現頻度を考慮した特徴量

運転操作は、速度帯と道路幅の組み合わせ別に分割し、10個のビンに分けて出現頻度を求めた。最大と最小のビンの分割点は分布の形をもとに決定し、その間は均一幅で分割した。

このビンごとの頻度をSVMの特徴量とする際に、2つの方法を用いて比較した。

- probability method

表3 分類実験に用いる正例負例数

5年間事故あり	43
事故なし	88

ビンに入った頻度をドライバーごとに正規化して出現確率を求め、その値を用いる。

- entropy-like method

通常とは異なる、出現頻度が低い運転操作を行った場合には、ドライバーの特徴がより強く現れたと考えることができる。これを考慮する手法としてentropy-like methodを提案する。

全ドライバーの運転操作を集計し、ビンごとに出現確率 Q を求める。これは平均的なドライバーの運転操作分布であると考えられる。各ビンに入るような運転操作の出現は、 $-\log(Q)$ の情報量を持っていると考えられるため、あるドライバーの運転操作出現確率を P とすると、そのピンは $-P\log(Q)$ と重み付きの確率で表現する。このような技法は、出現頻度が低い事象を強調することが可能であり、外れ値検出などにおいてしばしば利用されている。

また、平均的なドライバーと比較した時に頻度が多いか少ないかに関しても重要な特徴となることが考えられるため、 $P > Q$ の時に1、それ以外に-1の値を取る $sign$ を用いて、 $-P\log(Q) \times sign$ を特徴量とする。

d) その他のドライバー属性

運転の安全さは、ドライバーの性別や年齢によって影響を受けることが知られている。そこで、ドライバーの性別、年齢、運転免許を取得してからの期間、ゴールド免許の有無、の4つの属性を特徴量として利用した。性別とゴールド免許に関しては0,1の2値、年齢と運転期間は長さをそのまま特徴量とした。

3.5 分類実験

5年間以内の事故の有無を正解データとし、SVMを用いて分類する実験を行い、特徴量ごとに分類性能を比較した。正解データの数を表3に示す。

SVMは様々なカーネルが選択でき、それぞれにハイパーパラメータが存在する。本実験では表4に示した組み合わせについて網羅的に探索し、F値を基準に最良の設定を求めた。ここで、 $w_{事故あり}$ は事故ありの正例ラベルに対する相対的な重みである。

また、速度帯別に分類した場合の特徴量の素性数は548、速度帯と道路幅属性の組み合わせで分類した場合の素性数は2168となる(双方ともドライバー属性に関する素性4つを含む)。この素性を選択して、素性数を減らした場合の性能も網羅的に測定した。素性選択は、素性ごとにラベルとのカイ2乗統計量を求め、カイ2乗値が大きい素性から順に選択するという方法を用いた。素性数は2から400までの範囲で設定した。

各特徴量は素性ごとに正規化して利用した。分類性能は10-foldの交差検定で求めた。

3.6 分類結果

利用する特徴量の素性を以下のように変化させ、判定性能を比較した。

- 年齢・運転歴・性別のみを利用したもの(a)

表 4 SVM の設定

カーネル	ハイパーパラメータ
線形	$C: 2^{-5} \sim 2^{10}, w_{\text{事故あり}}: \{1, 2, 5, 10\}$
多項式	$C: 2^{-5} \sim 2^{10}, \gamma: 2^{-10} \sim 2^3, \text{次数}: \{2, 3\}, w_{\text{事故あり}}: \{1, 2, 5, 10\}$
ガウシアン	$C: 2^{-5} \sim 2^{10}, \gamma: 2^{-10} \sim 2^3, w_{\text{事故あり}}: \{1, 2, 5, 10\}$

表 5 判定器の性能

手法	feature 数	適合率	再現率	F 値
年齢・運転歴・性別 (a)	3	0.41	0.88	0.56
(a) + ゴールド免許有無 (b)	4	0.45	0.88	0.60
(b) + 運転挙動 (probability)	40	0.57	0.91	0.70
(b) + 運転挙動 (entropy)	100	0.70	0.77	0.73
(b) + 運転挙動 + 道路情報 (probability)	75	0.91	0.72	0.81
(b) + 運転挙動 + 道路情報 (entropy)	100	0.83	0.81	0.82
ベースライン (全て正例に分類)				0.51

- (a) に加え、ゴールド免許の有無を利用したもの (b)
- (b) に加え、速度帯別に分類した運転操作分布を利用したもの。特徴量の表現方法には probability と entropy-like の 2 つの手法を用いる。
- (b) に加え、速度帯と道路幅の組み合わせで分類した運転操作分布を利用したもの。2 通りの特徴量の表現形式を用いる。

それぞれの場合での判別性能を表 5 に示す。SVM の設定を網羅的に変え、最高の F 値が得られる設定での性能を採用している。また、それぞれの場合で得られた F 値を図 6 に示す。

従来知られている年齢・運転歴・性別のドライバー属性のみを用いた場合でも、全て正例 (事故あり) に判定するような判別器よりは良い F 値が得られているが、それほど高い性能であるとは言えない。これは、対象としているドライバーが全て運転を業務としており、専門的な教育を受けているためであると考えられる。ゴールド免許の有無に関する情報を追加したとき、F 値は若干改善するが、やはり同様の理由でそれほど高い性能ではない。

運転操作の分布に関する特徴量を追加することで、F 値は 0.7 強まで改善する。さらに道路特徴も考慮することで、最高 0.82 の F 値が得られ、運転操作分布に着目することで過去の事故履歴を高い精度で判別できることが示された。なお、最良の F 値が得られている (道路幅考慮、entropy-like) の設定では、年齢・運転歴・性別・ゴールド免許有無に関する素性は特徴量選択によって選ばれていなかった。

それぞれの手法における判別結果の ROC 曲線を図 7 に示す。年齢等のドライバー属性、並びにゴールド免許の有無に関する情報のみでは、ROC 曲線はランダムな判定結果 (点線に対応) と同等の判別性能しか得られていない。AUC (ROC 曲線の下側の面積) も 0.5 程度である。運転操作の分布に関する特徴量を追加することで、AUC は 0.8 強まで改善する。さらに、道路特徴を追加した entropy-like method 適用時には、AUC は 0.91 と高い判別性能が得られた。ROC 曲線も急激に立ち上がる曲線を描いており、事故ありのドライバーの大部分を高い分類スコアで判別できている、すなわち、得られた分類器が事故

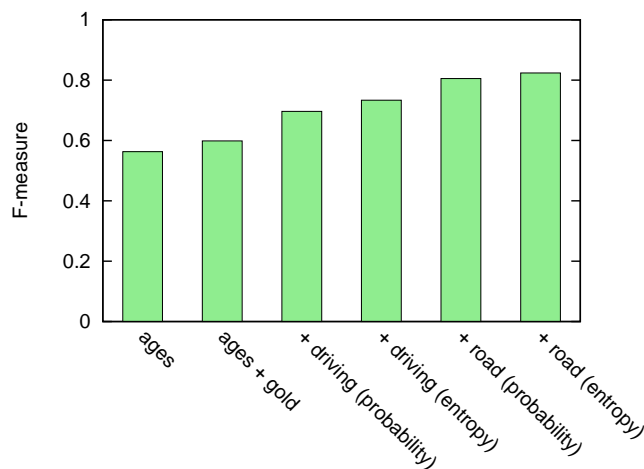


図 6 特徴量設計ごとの事故ありドライバー判定の F 値

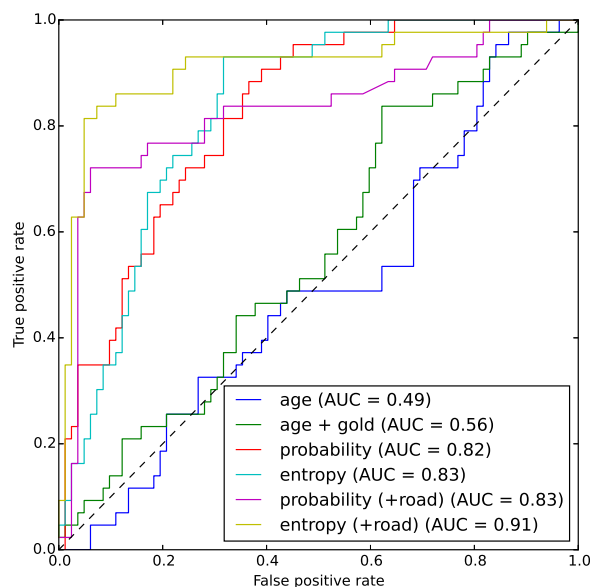


図 7 特徴量設計ごとの ROC 曲線

ありの可能性が高いと判断したドライバーは高い確率で事故ありドライバーであったことがわかる。

なお、道路特徴を追加した時の probability と entropy-like

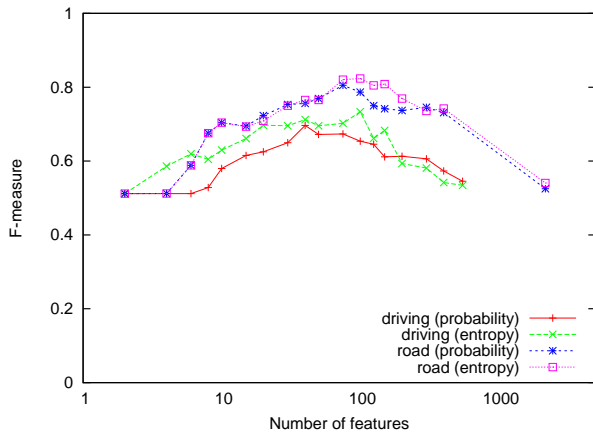


図 8 特徴量選択による性能変化

の 2 手法について、得られる F 値は同程度であったが、ROC 曲線を比較すると entropy-like method の方が良好な判別性能を示していると言える。

速度帯別に分類した運転操作特徴を用いた場合、及び道路幅を考慮した場合において、特徴量選択を行った時の F 値の変化を図 8 に示す。driving の系列が速度帯のみを考慮した特徴量、road の系列が速度帯と道路幅の組み合わせを考慮した特徴量である。いずれの場合も、おおむね 100 個程度の特徴量を選択した場合に良好な性能を示している。また、全ての特徴量を利用した場合にはそれほど高い F 値は得られていない。

probability method と entropy-like method を比較した場合、entropy-like method の方が高い性能を示している。操作の出現頻度を考慮することが判別に有効であると考えられる。

4. 終わりに

本論文では、大規模なドライブレコーダデータを利用してドライバーの運転操作と運転の安全さとの関係を理解することを試みた。ドライバーごとの運転操作の分布に着目し、特徴量として利用することで SVM を用いて過去の事故歴の有無を判別できることが示された。運転操作が行われた時の速度帯、ならびに道路幅を考慮することで判別性能が向上し、最高で適合率 0.83、再現率 0.81 の高い性能が得られた。年齢、性別など、従来安全に関係することが知られているドライバー属性を用いるのみでは高い判別性能が得られないことも示された。

今後、継続的にデータを取得し、運転操作の長期間・大規模な観察につなげていくことを計画している。また、道路幅以外にも運転操作が行われた状況を組み合わせることで、運転動作の違いをより精度よく表現できる手法を探っていく。例えば、天候データと組み合わせることで、雨の日と晴れた日の運転動作の違いが考慮できるようになることが期待される。

謝 辞

本研究の一部は東京大学空間情報科学研究センターとの共同研究に基づき研究用空間データ基盤よりデータ提供を受けて実施している。

- [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] 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.
- [3] 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.
- [4] G. Castignani, R. Frank, and T. Engel. Driver behavior profiling using smartphones. In *Intelligent Transportation Systems - (ITSC), 2013 16th International IEEE Conference on*, pp. 552–557, Oct 2013.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] Feng Guo and Youjia Fang. Individual driver risk assessment using naturalistic driving data. *Accident Analysis & Prevention*, Vol. 61, pp. 3 – 9, 2013.
- [9] 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.
- [10] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92*, pp. 144–152, New York, NY, USA, 1992. ACM.