

ドライブレコーダデータに対するヒヤリハット危険度推定

小平美沙季[†] 山本 修平^{††} 倉島 健^{††} 戸田 浩之^{††} 加藤ジェーン^{†††}

[†] 名古屋大学情報学研究科 〒464-8601 愛知県名古屋市千種区不老町

^{††} 日本電信電話株式会社 NTT サービスエボリューション研究所 〒239-0847 神奈川県横須賀市光の丘 1-1

^{†††} 立命館大学情報理工学部 〒525-8577 滋賀県草津市野路東 1 丁目 1-1

E-mail: [†]kodaira.misaki@a.mbox.nagoya-u.ac.jp,

^{††}{yamamoto.shuhe, kurashima.takeshi, toda.hiroyuki}@lab.ntt.co.jp, ^{†††}ijen@fc.ritsume.ac.jp

あらまし ドライブレコーダで記録されるデータのうち、交通違反や交通事故などの危険な状況が記録された「ヒヤリハット」に関するデータは、ドライバーの安全運転教育や運行管理などに活用されている。ヒヤリハットデータの実用性を高める上で、各データに危険度を表すアノテーションを付与し、危険度に応じたデータの検索ができるようになることが望まれる。本論文では、ドライブレコーダデータに対して、非ヒヤリハットを含めた危険度を推定する手法を提案する。提案手法では、映像とセンサ、及び映像からの物体検出結果を組み合わせ変換を行い、危険度の推定に有用な特徴表現を獲得する。また、クラス分類と回帰を組み合わせたモデルを用いて算出されるスコアに基づいて危険度を推定することで、推定精度の向上を実現する。

キーワード 深層学習, 危険度推定, ヒヤリハット, ドライブレコーダ, マルチモーダルデータ

1. はじめに

ドライブレコーダとは、車内に設置され、前方映像、加速度、ウinker操作、ブレーキ操作などの車両運行状況を記録する装置である [1]。全時刻の情報を記録することは、記憶容量などの問題もあるため、実際には加速度に一定の閾値を設け、車両に何らかの衝撃が見られたと考えられるタイミングを検知し、その前後十数秒を記録することが多い。このように記録されたデータを、本論文ではイベントデータと呼ぶ。イベントデータの内、交通違反や交通事故に繋がる危険な状況を記録した「ヒヤリハット」に関するデータは、実際にドライバーが視聴して、自身の運転行動を省察したり [2]、危険予知訓練を行うなど [3]、安全運転教育に活用されている。また、法人営業車両などを対象に、走行中の運行車両からヒヤリハットシーンをリアルタイムに検出し、運行管理者に通知することによって、頻繁に危険運転をするドライバーの事故を未然に防ぐことにも利用されている [4]。

一方、ドライブレコーダで記録される全てのイベントデータがヒヤリハットに関するものではない。東京農工大学スマートモビリティ研究拠点の調査によると、例えば、道路の激しい起伏が原因で車両に大きな衝撃が加えられ記録されたものや、狭い道路での方向転換のために行う小刻みな加速が原因で記録されたもの等、非ヒヤリハットに該当するイベントデータも全体の 70% 近く含まれている [5]。

また、ヒヤリハットに該当するイベントデータにおいても、その危険度は様々である。例えば、図 1 のように、ヒヤリハットの発生対象（歩行者や自転車など）に対して、一定の距離を置いて余裕ある状況で衝突を回避した場合には相対的に危険度が低く、あと僅かで衝突しそうだった場合は相対的に危険度が高いといえる。このため、イベントデータに危険度のラベルを



図 1 実際のヒヤリハットシーン及び危険度ラベルの例。左図では、飛び出してきた歩行者に対して急ブレーキを踏んで停車している。右図では、運転者が早い段階で歩行者を認識し、速度を緩めている。右例に比べて左例の方が交通事故に近い危険なヒヤリハットシーンであり、左例に危険度「高」、右例に危険度「低」などのラベルを付与したい。

付加することで、安全運転教育に資するデータを効率良く抽出したり、ドライバーが危険運転をしているか否かをより正確に判定したりすることが可能となると考えられる。しかしながら、収集された膨大なイベントデータについて、人手で危険度を判定することは、多大な労力や注意力を要することに加え、判定者によって危険度の判定に個人差が生じることが問題である。したがって、危険度を自動的に判定し、ラベリングをするシステムを導入することにより、ラベリングコストを削減し、ドライブレコーダデータの実用性を高めることが期待される。そこで本論文では、イベントデータの集合に対して、非ヒヤリハットを含む危険度のラベルを自動的に付与するタスクに取り組む (図 2 参照)。

本論文の提案手法により、実際の活用現場において、ヒヤリハットを自動的に大量に判別できることによるメリットが見込まれる。ヒヤリハットデータの具体的な活用事例として、多くの地方自治体で「ヒヤリハットマップ」作成の取り組みが行われている [6-8]。ヒヤリハットマップは、地域住民が個々人の経

験や知識を基にして、交通の危険箇所を地図に表示することによって作成される [7]。これらは公共機関や各家庭で活用され、危険箇所を人々に周知させることで、交通事故に対する意識を向上させ、事故の未然防止、再発防止に貢献している [6]。本研究で取り組む、ヒヤリハットの有無及び危険度の自動判別を利用し、大量のドライブレコーダデータからヒヤリハットの発生地点を集計することによって、ヒヤリハットマップを自動的に作成することが可能となる。これにより、人手による作業コストを削減できるほか、大量のデータから、より網羅性の高い情報を得ることができる。更に、危険度を付加することで、特に危険な箇所や、未だ事故として現れていない潜在的に危険な箇所を把握することが可能となるため、事故の未然防止、再発防止への有効性が高まることが期待される。

著者らは、これまでにドライブレコーダに含まれるセンサと映像からなるマルチモーダルデータを用いた、教師あり深層学習に基づくヒヤリハット検出手法 [9] 及びヒヤリハット発生対象分類手法 [10] を提案している。これらの手法では、センサと映像をそれぞれ Fully-Connected Neural Network (FC) と Convolutional Neural Network (CNN) [11] で特徴変換し、Recurrent Neural Network (RNN) [12] で時間的な遷移をモデル化している。更に [10] では、前方映像からの物体検出結果をグリッド空間へ埋め込んだ入力特徴量の追加手法と、クラス間の階層関係を考慮したマルチタスク学習の枠組みを提案している。実際のイベントデータを用いた評価実験の結果、提案手法が高い精度でヒヤリハット検出（ヒヤリハットの有無の判定）並びにヒヤリハット発生対象の分類ができることを明らかにしている。従来手法 [9] はヒヤリハット危険度を推定するための多クラス分類に容易に拡張可能であるが、素朴な拡張では精度が低下することが予想される。その理由として、以下の 2 項目が挙げられる。

課題 1. ヒヤリハットと非ヒヤリハットの間には、発生対象が自車の進行方向に存在するか否か、といった見かけ上の明確な差異があるため、単純な CNN によって推定に必要な特徴量を獲得することが可能である。一方で、ヒヤリハットの危険度を推定する場合、対象物体—自車間の相対的な位置、距離、接近速度及び自車の加速度をそれぞれ関連づけて判断することが求められる。従来手法ではこのような要件を満たす特徴量は得られないため、推定精度が低下すると考えられる。

課題 2. ヒヤリハットの危険度には連続的な関係性がある。一方で、データに対してアノテーションを行う際に、危険度に相当する連続値のスコアを人手で判定することは困難であるため、{ 高, 中, 低 } のような離散的なラベルを付与することが想定される。このようなラベルを教師データとして、従来手法 [9] のように危険度推定を行うための多クラス分類器を学習した場合、各危険度がそれぞれ独立の関係とみなされることにより、推定精度が低下すると考えられる。

そこで本研究では、課題 1 に対するアプローチとして、従来手法 [10] で提案されているグリッド特徴量の利用方法を検討し、各時刻毎にセンサデータと組み合わせて新たな特徴量に変換することで、異なる特徴量同士の関係性を表現するエンコーディ

ング手法を提案する。また、課題 2 に対するアプローチとして、危険度を連続値のスコアとして扱い、クラス分類と回帰の手法を組み合わせた手法を提案する。また、チューニングデータを用いて、各危険度ラベルが取り得るスコアの範囲を分布として考慮することで、出力されるスコアを適切な危険度ラベルに変換する。

本論文の貢献は以下の 3 項目である。

(1) ドライブレコーダによって記録される映像データ、センサデータ及びフレーム画像に対する物体検出結果を効果的に組み合わせて時系列モデリングを行う危険度推定用フレームワークを提案した。

(2) 多クラス分類において一部のクラス間に連続性がある場合に、分類と回帰を組み合わせたフレームワークによって連続値のスコアを決定し、クラス毎のスコア分布からクラスラベルを決定することで、クラスの持つ特性に沿った分類を実現する手法を提案した。

(3) 上記の 2 つの提案要素に関して、それぞれの有効性と、それらを組み合わせたときの有効性を、ヒヤリハット危険度推定タスクを通じて明らかにした。

以下、本論文の構成を示す。2 章では、ヒヤリハットの危険度推定の関連研究について説明する。3 章では、データとその前処理についての説明、及び本論文で取り組むタスクの定式化を行う。4 章では、提案手法について説明する。5 章では、提案手法の評価実験について説明する。最後に、6 章でまとめる。

2. 関連研究

本章では、ドライブレコーダからのヒヤリハット検出及びヒヤリハット危険度推定を扱う関連研究と、ドライブレコーダを用いたその他の関連研究について述べ、本研究の位置付けを整理する。

ヒヤリハット検出を対象とした研究。 ドライブレコーダからのヒヤリハット検出を対象とする研究として、速度と加速度に経験的にルールを設定し、閾値調整によってヒヤリハットを検出する研究 [13] やセンサデータを適切な特徴量に変換して右折時に限定したヒヤリハット発生形態を分析し、右折時のヒヤリハット検出に取り組む研究 [14] がある。また、前方映像データを扱ったヒヤリハット検出としては、前方映像から歩行者検出を行い、自車位置と映像中の検出領域との距離を計算して、歩行者とのヒヤリハットを検出する研究 [15] がある。これらの研究では、センサデータや映像データを用いてヒヤリハットを検出しているが、本研究が目指すヒヤリハットの危険度推定は扱っていないことから、本研究とは異なる。

ヒヤリハット危険度推定を対象とした研究。 ヒヤリハット危険度推定を対象とする研究としては、歩行者検出と危険度推定を同時に学習するモデルを提案している研究 [16] がある。この研究では、ヒヤリハットの発生対象を歩行者に限定し、ResNet [17] ベースのネットワークを利用して、歩行者のバウンディングボックスの回帰と危険度に関するクラス分類を同時に最適化する。入力にはフレーム画像を利用し、各フレーム画

像に対してバウンディングボックス及び危険度の推定を行う。しかしヒヤリハットを検出するだけではなく、その危険度を推定する場合、対象との位置・距離関係と自車の速度・加速度の値、及びそれらの時間的な遷移が重要な意味を持つため、それらの情報を表現可能な特徴を利用することが望ましい。本研究では歩行者以外の対象にも対応し、映像データに加えて、センサデータ及びフレーム画像に対する物体検出結果を落とし込んだグリッド特徴量を効果的に組み合わせた特徴量を利用する。また、フレーム画像毎ではなく、一つのイベントデータに対して一つの危険度ラベルを付与する。

ドライブレコーダを用いたその他の研究。ヒヤリハット検出や危険度推定以外にも、ドライブレコーダは様々なタスクで用いられている。センサデータに含まれるドライバの運転操作を用いたものでは、ヒヤリハットの起こりやすい潜在リスクを持つ道路の推定をする研究 [18, 19] や、ヒヤリハットを起こしやすい危険運転ドライバーの分類に関する研究がある [20, 21]。車両走行中の前方映像データは、深層学習技術の発展により、自動車の自動運転に関する研究で幅広く利用されている。自動運転タスクでは、前方映像に基づいてハンドルの操舵角を予測する研究 [22, 23]、車の直進、左折、右折、車線変更、停止などの運転操作を予測する研究 [24]、同様のタスクに車両映像も用いた研究 [25, 26]、交通事故を回避するため、前方で発生する交通事故を予測する研究 [27] などがある。これらはいずれも深層学習に基づく手法を提案しており、画像解析は CNN、時系列モデリングは RNN において有効性の知られている、Long Short-Term Memory(LSTM) [28] を用いている。本論文でも CNN や LSTM を用いたモデル化に取り組むものの、上記タスクでは各時刻での運転操作やハンドルの操舵角、交通事故の起こりやすさを予測することに対して、本研究では入力の特徴時刻でヒヤリハットの発生対象を分類する点で、これらの研究とタスクや提案するモデルが異なる。

3. 問題設定

本章では、本論文で扱うデータとその前処理について説明する。また、本論文で扱うタスクの定式化を行う。本論文で用いる主な記号を表 1 に示す。

3.1 データ

本論文で想定するドライブレコーダで記録されるイベントデータは、衝撃が加わった（加速度に大きな変化が生じた）時点を中心として前 10 秒、後 5 秒の計 15 秒の前方映像及びスピード・加速度等のセンサデータから構成される。1 つのイベントデータにつき、1 つの危険度ラベルが付与されており、危険度ラベルは C 段階で表現される。

3.2 準備

物体検出の適用。ヒヤリハットの発生対象の自車に対する位置を考慮するため、本論文では、前方映像のフレーム画像に対して物体検出を行い、その結果を利用する。物体検出には、事前学習された YOLOv3 [29] を利用する^(注1)。画像 I^t から得られる

表 1 本論文で用いる主な記号

記号	説明
C	危険度のクラス数
I^t	t フレーム目の画像
s^t	t フレーム目のセンサデータ
G^t	t フレーム目のグリッド特徴
T	イベントデータのフレーム数
N^t	t フレーム目の画像から検出された総物体数
$\mathbf{o}_n^t, \mathbf{b}_n^t, p_n^t$	t フレーム目の画像から n 番目に検出された物体の 1-of-K 表現ベクトル, 領域ベクトル, 検出確率
y	危険度の推定結果を表すラベル $y \in \{0, 1, \dots, C-1\}$
H, W	オリジナル画像の縦幅, 横幅
G_h, G_w	物体検出結果を埋め込むグリッド空間の縦, 横のグリッド数
t_a, t_b	2 クラス分類, 回帰スコアに対する教師信号
y_a, y_b	2 クラス分類, 回帰の出力スコア
L_a, L_b	t_a, t_b に対する出力スコアの誤差
L	L_a, L_b の結果を組み合わせた目的関数

物体検出結果は N^t 個の検出されたオブジェクトを持ち、各オブジェクトの ID を総物体種類数 V 次元の 1-of-K 表現で持つベクトル $\mathbf{o}_n^t$ 、各オブジェクトの領域情報を持つベクトル $\mathbf{b}_n^t$ 、及び各オブジェクトの検出確率 $p_n^t \in (0, 1]$ からなる ($t = 1, 2, \dots, T$)。ここで $\mathbf{b}_n^t$ は、その左端, 上端, 右端, 下端の 4 つの画像中の座標を持つ ($\mathbf{b}_n^t = \{b_n^t.left, b_n^t.top, b_n^t.right, b_n^t.bottom\}$)。

危険度の数値表現。ヒヤリハットの危険度は、それぞれ整数値に置き換えて表現する。表 2 に、ヒヤリハット/非ヒヤリハットの 2 クラス分類を行う場合、及び高/中/低/非ヒヤリハットの 4 クラス分類を行う場合のそれぞれについて、危険度ラベルと数値の対応付けを示す。

表 2 危険度ラベルの数値表現

	非ヒヤリハット	ヒヤリハット		
		低	中	高
2 クラス分類	0	1		
4 クラス分類	0	1	2	3

3.3 問題

前方映像のフレーム画像 $I^t (t = 1, 2, \dots, T)$ と対応する時刻のスピード・縦加速度・横加速度に関するセンサデータ $s^t (t = 1, 2, \dots, T)$ 及び次章で説明するグリッド特徴 $G^t (t = 1, 2, \dots, T)$ を入力とし、危険度の推定結果 $y \in \{0, 1, 2, 3\}$ を出力する。

4. 提案手法

本章では、本論文で提案する深層学習を活用したヒヤリハット危険度の推定手法について述べる。第 4.1 節では、提案手法における入力から出力までのデータの処理過程を説明する。第 4.2 節ではモデルに入力するグリッド特徴量について、第 4.3 節では回帰を利用した危険度の分類手法について説明する。

4.1 全体像

図 2 に提案手法の全体像を示す。提案ネットワークは、画像

(注1) : 次の URL の実装を用いた。 <https://pjreddie.com/darknet/yolo/>

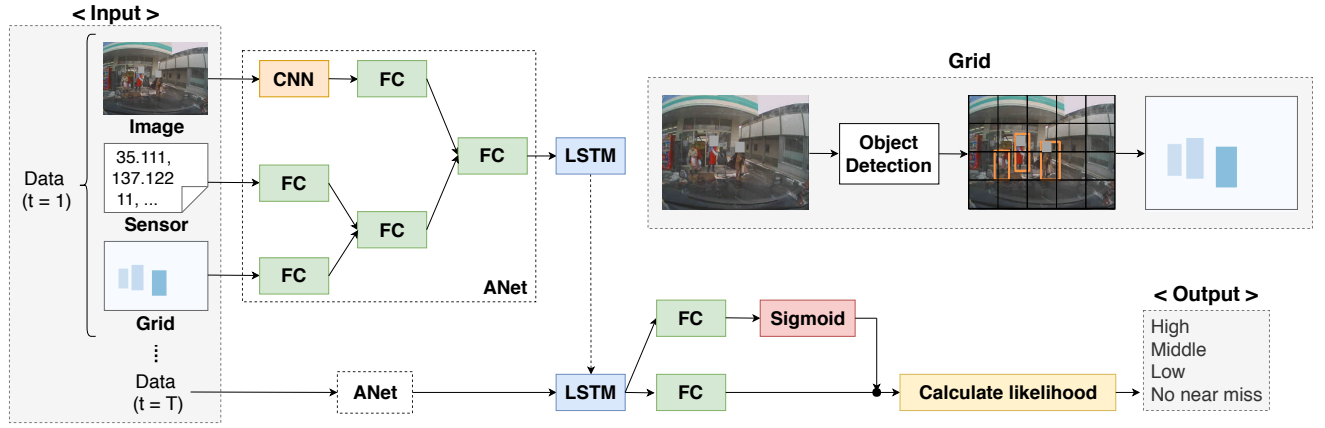


図 2 提案手法の概要図

データ処理ユニット、センサデータ・グリッド特徴処理ユニット、時系列モデリングユニット、危険度分類ユニットの4つのユニットから構成される。以下、各ユニットについて説明する。

画像データ処理ユニット。画像データはCNNによってエンコードする。提案手法のCNNは3層の畳み込み層から構成し、各畳み込み層の後には、学習の高速化や過学習の抑制に効果のあるBatch Normalization [30]、情報の圧縮による計算コストの低減や位置変化に対する頑健性の向上に効果のあるMax Pooling [31]を行う。画像はRGBチャネルを持ち、 64×64 の解像度にリサイズしたデータを入力することを想定し、畳み込みのStrideはいずれの層でも1とし、フィルタサイズは1層目で 7×7 、2層目で 5×5 、3層目で 3×3 とする。Max Poolingは、いずれの層でもStrideを2、フィルタサイズを 2×2 とする。また、全てのBatch NormalizationとMax Poolingの間で用いる活性化関数にはReLUを用いる。CNNによって得られる特徴マップは全結合層によってエンコードし、特徴ベクトル h_i^t を得る。

センサデータ・グリッド特徴処理ユニット。センサデータ及びグリッド特徴は、全結合層によってエンコードし、特徴ベクトル h_s^t, h_g^t を得る。更に、得られた特徴ベクトルを連結した後、再度全結合層によってエンコードすることで、新たな特徴ベクトル h_{sg}^t を得る。本ユニットの構成においては、ヒヤリハットの危険度が「自転車速度」、「自転車-対象物体間の距離」という2要素の関係に依存して決定されることに着目する。例えば、自転車速度が大きく、且つ対象物体との距離が近いときには危険度が高いといえるが、自転車速度が小さいまたは停車している状態で、且つ対象物体との距離が近いときには危険度が低くなる。提案手法では、画像データの特徴ベクトルより前の段階でセンサデータ・グリッド特徴の特徴ベクトルを連結し、エンコードすることによって、上述のような2要素間の関係性を表現する特徴量の獲得が期待できる。

時系列データモデリングユニット。本ユニットでは、画像データの特徴ベクトル h_i^t とセンサデータ・グリッド特徴の特徴ベクトル h_{sg}^t を連結し、再度全結合層によってエンコードすることで h_{isg}^t を得る。このようにして得られる T フレーム分の時系列データをLSTMを用いてエンコードする(h_l)。LSTM

では系列データの長期的な依存関係を学習することができるため、ヒヤリハット危険度推定に必要な特徴量の時間的な変化のモデル化を期待できる。

危険度分類ユニット。LSTMによって獲得した特徴ベクトル h_l を全結合層によって1次元にエンコードし、危険度スコアを算出する。このとき、クラス分類と回帰を組み合わせたフレームワークを利用する。得られた危険度スコアは、正規分布の対数尤度に基づいて $\{0, 1, 2, 3\}$ に離散化し、推定結果として出力する(4.3節参照)。

4.2 グリッド特徴

本節では、提案ネットワークに入力するグリッド特徴の詳細について説明する。なお、この手法は先行研究[10]の手法を拡張したものである。本論文で取り組むヒヤリハットの危険度推定においては、前方映像中に存在する物体の、自転車に対する位置・距離を考慮することが重要であると考えられる。そこで本論文では、適当な縦と横のグリッド数 G_h, G_w からなる行列 $G^t \in \mathbb{R}^{G_h \times G_w}$ を用意し、物体検出結果をグリッド空間に埋め込むことによって、物体の領域情報を考慮した特徴行列を得る。 t フレーム目のグリッド特徴行列 G^t の生成手順をアルゴリズム1に示す。

G^t の生成では、入力として t フレーム目の画像から検出された特定の物体の検出数 N^t と、物体の領域情報の系列 $B^t = \{b_n^t\}_{n=1}^{N^t}$ 、画像の縦幅 H と横幅 W 、縦と横のグリッド数 G_h, G_w を用いる。出力する行列 G^t を0で初期化した後、ステップ幅 S_h, S_w を計算する。時刻 t 、検出物体 n に関して、その領域が画像中のどのグリッドに含まれているかを得るため、左端、上端、右端、下端インデックス $left, top, right, bottom$ を計算する。また、当該グリッドに埋め込むスコアとして、物体領域の画像サイズに対する面積比 r を計算する。これは、物体領域の面積が大きいほど自転車に対して物体は近い距離にあり、また近い距離にあるほど危険度が高くなる要因となっていると考え、その特性を考慮したスコアをグリッド空間に埋め込むためである。そして、対象グリッドについて、縦方向を top から $bottom$ 、横方向を $left$ から $right$ の範囲で走査し、該当するグリッドに対してスコア r を $g_{i,j}$ に加算する。

先行研究[10]では全時刻の情報を足し合わせて1つのグリッ

Algorithm 1 Calculate grid**Input:** N^t, B^t, H, W, G_h, G_w **Output:** G^t

```

1:  $G^t \in \mathbb{R}^{G_h \times G_w} \leftarrow 0$ 
2:  $S_h \leftarrow \frac{H}{G_h}$ 
3:  $S_w \leftarrow \frac{W}{G_w}$ 
4: for each detected object  $n$  in 1 to  $N_t$  do
5:    $left \leftarrow \lceil \frac{b_n^t.left}{S_w} \rceil$ 
6:    $top \leftarrow \lceil \frac{b_n^t.top}{S_h} \rceil$ 
7:    $right \leftarrow \lceil \frac{b_n^t.right}{S_w} \rceil$ 
8:    $bottom \leftarrow \lceil \frac{b_n^t.bottom}{S_h} \rceil$ 
9:    $height \leftarrow b_n^t.top - b_n^t.bottom$ 
10:   $width \leftarrow b_n^t.right - b_n^t.left$ 
11:   $r \leftarrow \frac{height \times width}{H \times W}$ 
12:  for each vertical grid index  $i$  in  $top$  to  $bottom$  do
13:    for each horizontal grid index  $j$  in  $left$  to  $right$  do
14:       $g_{i,j} \leftarrow g_{i,j} + r$ 
15:    end for
16:  end for
17: end for

```

ド特徴を生成している。しかし、危険度推定においては自転車一対象物体間の距離の時間的遷移や、自転車速度との関係が重要であることから、本論文の提案手法では、各時刻毎にグリッド特徴を生成している。

4.3 回帰を利用した危険度分類

本節では、提案ネットワークにおける回帰を利用した危険度分類の手法について説明する。本論文では、ヒヤリハットの危険度を C クラスに分類することを想定しているが、本来、ヒヤリハットの危険度には自転車の速度やヒヤリハットの発生対象との近さなどによって決定される連続的な関係性がある。一方で、非ヒヤリハットシーンには危険な状況が含まれていないことから、ヒヤリハットと非ヒヤリハットの間には、明確な差異が存在することが考えられる。このとき、ヒヤリハット（例：高/中/低）と非ヒヤリハットを連続的な関係性として扱うことは望ましくない。以上の点を考慮して、本論文ではヒヤリハット/非ヒヤリハットの2クラス分類と、ヒヤリハットのスコアの回帰を組み合わせた危険度分類手法を提案する。

このとき、提案手法のモデルはマルチタスク学習の枠組みで最適化される。深層学習におけるマルチタスク学習は、メインタスクに関連するサブタスクを教師信号として与え、同一のニューラルネットワークで最適化することにより、汎化性能の高いモデルパラメータを学習できることが知られている [32, 33]。本論文では、メインタスクであるヒヤリハットの危険度推定を、「ヒヤリハット/非ヒヤリハットの2クラス分類」「ヒヤリハットのスコアの回帰」という2つの簡単なタスクに切り分け、サブタスクとして設定する。これらのタスクからなる目的関数を最小化することで、汎化性能の高いモデルパラメータを学習することが期待できる。

提案手法では図2に示すように、LSTMの出力を2層の並列な全結合層によってエンコードし、それぞれ1次元のスコア

に変換する。ここで、一方ではヒヤリハット/非ヒヤリハットの2クラス分類スコア、もう一方ではヒヤリハット（高/中/低）の回帰スコアを計算する。2クラス分類スコアの算出においては、シグモイド関数を適用することによって $[0, 1]$ の範囲の値 y_a に変換する。 y_a に対する教師信号は $t_a \in \{0, 1\}$ とし、ヒヤリハット（高/中/低）の場合を $t_a = 1$ 、非ヒヤリハットの場合を $t_a = 0$ と定める。誤差 L_a の計算は式(1)に示す交差エントロピーを用いて行う。

$$L_a = -\frac{1}{D} \sum_d \{t_a \log y_a + (1 - t_a) \log(1 - y_a)\} \quad (1)$$

ここで、 D はデータ数、 d はデータを走査するためのインデックスであり、各々の教師信号と推定スコアに付随するが、本論文では省略して記述する。

2クラス分類のスコア y_a は、式(2)のように回帰スコア y_c に乗算して利用することにより、ヒヤリハットである場合にのみ値を保持するゲートのような役割を果たす。

$$y_b = y_a y_c \quad (2)$$

最終的に得られる回帰スコア y_b に対する教師信号は $t_b \in \{0, 1, \dots, C-1\}$ とし、危険度が大きいほど、教師信号の値が大きくなるように対応付ける。誤差 L_b の計算は式(3)に示す二乗和誤差を用いて行う。

$$L_b = \frac{1}{D} \sum_d (t_b - y_b)^2 \quad (3)$$

なお、提案ネットワークを学習する際の目的関数 L は、式(4)のように L_a と L_b の和をとることによって計算する。

$$L = L_a + L_b \quad (4)$$

得られた連続値の回帰スコア y_b は、離散的な推定ラベル {非ヒヤリハット, 低, 中, 高} に変換して出力する。変換は、次の手順にしたがって行う。まずはじめに、学習データの2割をチューニングデータとし、チューニングデータの回帰スコアに対する平均、分散を各ラベル毎に計算することで、 μ_{label} , σ_{label}^2 を得る。但し、 $label \in \{0, 1, 2, 3\}$ である。次に、得られた μ_{label} と σ_{label} をパラメータとする正規分布の所属確率が最大となる $label$ の値をモデルの出力 y とする（式(5)）。

$$y = \arg \max_{label \in \{0, 1, 2, 3\}} \frac{1}{\sqrt{2\pi\sigma_{label}^2}} \exp\left\{-\frac{(y_b - \mu_{label})^2}{2\sigma_{label}^2}\right\} \quad (5)$$

5. 評価実験

本章では、提案手法の有効性を検証するために、実際のドライブレコーダデータを用いて行った実験について述べる。

5.1 データセットと設定

実験に用いるドライブレコーダデータには、東京農工大学スマートモビリティ研究拠点の提供する「ヒヤリハットデータベース」を利用した。このデータセットでは、1つのイベントデータにつき1つの危険度ラベルが付与されており、{事故, 高, 中, 低, 非ヒヤリハット}の5段階で表現される。本論文

では、画像データとしては、450 フレームから成る 1 つのイベントデータのうち、ヒヤリハットシーンが含まれているフレームを中心に分析するため、210 フレーム目から 450 フレーム目までの中から等間隔に 15 フレームサンプリングして使用した。画像の解像度は 640×400 で提供されているが、CNN で分析するにあたって、各フレーム画像は RGB 形式で 64×64 の解像度に線形変換した。物体検出はオリジナル画像に対して直接実行した。また、センサデータとしては、前後加速度、横加速度、スピードの 3 種類を使用した。各ラベル毎のデータ件数を表 3 に示す。

表 3 各ラベル毎のデータ件数

	訓練	テスト	合計
高	208	92	300
中	412	188	600
低	627	273	900
非ヒヤリハット	1273	527	1800

提案手法の有効性を確認するため、本研究では、次の 2 つの比較手法を用意した。

Baseline. Baseline の手法は、ヒヤリハット/非ヒヤリハットの 2 クラス分類を行う先行研究 [9] のモデルを、危険度推定を行うため、4 クラス分類を行うモデルに拡張したものである。

Improved Dense Trajectories. Improved Dense Trajectories(IDT) は映像中の人物の動作認識タスクに利用される特徴表現で、密な点群を追跡し、その軌跡上の局所特徴量に基づいて高次特徴量を記述する手法である [34]。本論文で扱う危険度推定タスクでは、映像データ中の物体の動きや自転車から見た周囲環境の動きを用いるため、IDT の有効性が期待される。IDT は合計 426 次元の特徴量を映像中の局所特徴毎に抽出する。本論文では IDT で得られた特徴量を K-means でクラスタリングした後、各局所特徴に割り当てられたクラスタ ID の頻度を映像ごとに求める。その頻度分布を映像の特徴ベクトルとして、SVM でモデルを学習した後、分類結果を得た。

5.2 評価尺度

ヒヤリハットの危険度推定タスクにおいて、提案手法の有効性を議論するため、推定結果がどれだけ正解しているかという正確性と、推定結果が全ての正解のうちどれだけ網羅しているかという網羅性の、2 つの観点から評価する。本論文では、正確性を適合率 (Precision)、網羅性を再現率 (Recall)、またこれらの評価指標の調和平均である F 値 (F1-score) によって表し、提案手法の性能を評価する。これらの指標を計算するため、評価データの集合と推定結果の集合に関して、次のような 3 種類のパラメータを定義する。

TP: 正例に対して推定結果が正であったデータの件数

FP: 負例に対して推定結果が正であったデータの件数

FN: 正例に対して推定結果が負であったデータの件数

このとき、Precision, Recall, F1-score は次のように計算される。

$$Precision = \frac{TP}{TP + FP}, \quad (6)$$

$$Recall = \frac{TP}{TP + FN}, \quad (7)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}. \quad (8)$$

5.3 実験結果

本節では、実験結果の分析を通じて、提案手法の各要素、及びそれらを組み合わせた場合の有効性を検証する。

検証 1. グリッド特徴を入力として追加することにより、危険度推定の性能が向上するか？

ここでは、前方映像に対する物体検出の結果を表現する「グリッド特徴」の使用により、危険度推定の性能が向上するか検証する。なお、グリッド特徴を生成する際には、グリッドの分割 $G_h \times G_w$ が細かい設定 (64×40)、粗い設定 (6×4) の 2 通りについて実験を行い、それぞれの精度を比較した。また、グリッドに埋め込むスコアとして、4.2 節に示したように面積比を利用する場合に加え、面積比にガウスカネルを乗算した場合の影響を検証した。ガウスカネルを適用することにより、物体領域の中心に近いほど、スコアが大きくなるように重み付けされる。式 (9) に、ガウスカネルを乗算する場合のスコアの計算式を示す。但し、 l は物体領域の中心からの距離を表し、 σ は (検出領域の縦幅)/2 とする。

$$r_{gauss} = r \times \frac{1}{\sqrt{2\pi}\sigma^2} \exp \frac{l^2}{2\sigma^2} \quad (9)$$

図 3 に、面積比のスコアを利用した場合と、面積比 $\times$ ガウスカネルのスコアを利用した場合のそれぞれについて、グリッド特徴のイメージを示す。

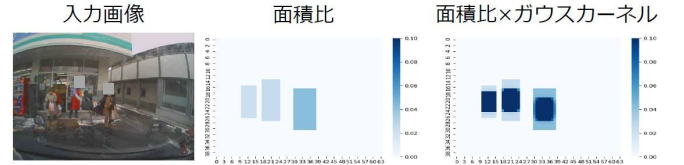


図 3 同一のフレーム画像に対して計算された 2 種類のグリッド特徴のイメージ。スコアが大きい部分ほど、濃い色で表現されている。

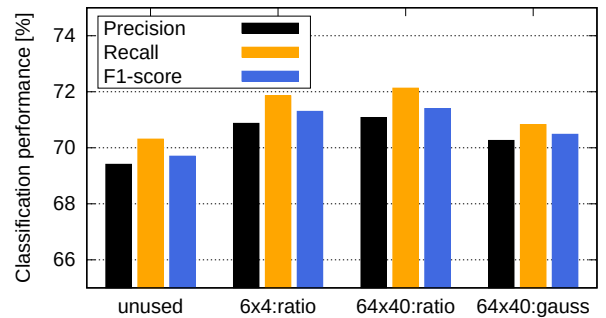


図 4 グリッド特徴の有無、及びグリッド特徴の計算方法に関する異なる設定における精度比較

図 4 に検証結果を示す。結果から、グリッド特徴を使用しない場合 (unused) と比較して、グリッド特徴を使用した場合に

精度が向上していることが確認できる。また、グリッドの分割数については、粗い場合 (6×4 : ratio) と比較して、細かい場合 (64×40 : ratio) の方が僅かに精度が高いことが確認できる。これは、分割が細かい方が物体の検出位置をより正確に表現できるためであると考えられる。グリッドに埋め込むスコアについては、ガウスクERNELを適用する場合 (64×40 : gauss) と比較して、適用しない場合 (64×40 : ratio) の精度が高い。この結果から、同じ物体の領域は一樣に考慮すべきであると結論づけられる。

検証2. 回帰手法を利用することにより、危険度推定の性能が向上するか？

ここでは、モデルの出力部において回帰手法を利用した場合に、危険度推定の性能が向上するか検証する。なお、検証にあたって、予め次の2つの仮説を立てる。

仮説1. ヒヤリハット {高・中・低} と非ヒヤリハットの間には離散的な関係がある。

仮説2. ヒヤリハット {高・中・低} と非ヒヤリハットの間には連続的な関係がある。

仮説3. ヒヤリハットと非ヒヤリハットには明確な差異 (離散的な関係) があるが、ヒヤリハット {高・中・低} の間には連続的な関係がある。

仮説1, 2 はそれぞれ、交差エントロピーと二乗和誤差によって実現される。仮説3は式 (4) によって実現される。実験では、仮説1, 2, 3のそれぞれに基づくモデルについて比較を行った。このとき、仮説1のモデルでは分類のみ、仮説2のモデルでは回帰のみを利用し、仮説3のモデルでは分類と回帰を利用する。

図5に検証結果を示す。結果から、分類のみを利用した場合 (Cla) と比較して、回帰を利用した場合に精度が向上していることが確認できる。また、回帰のみを利用した場合 (Reg) と比較して、分類と回帰を利用した場合 (Cla+Reg) に精度が向上していることから、仮説3が妥当であると結論づけられる。

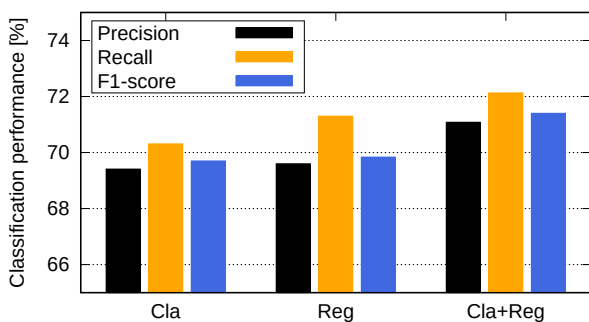


図5 クラス分類, 回帰, クラス分類+回帰の精度比較

検証3. 各要素の組み合わせにより、危険度推定の性能が向上するか？

提案手法では、(1) グリッド特徴, (2) 分類+回帰という2つの工夫を施している。ここでは、(1)(2)を組み合わせた場合の効果を検証する。また、提案手法と他手法を比較することで、提案手法の有効性を検証する。

図6に検証結果を示す。結果から、工夫点 (1)(2) のいずれかのみを利用した場合 (Grid, Cla+Reg) と比較して、工夫点 (1)(2) を組み合わせた場合 (Proposed) に精度が向上していることがわかる。このことから、各要素を組み合わせることの有効性が確認できる。また、比較手法 (Baseline, IDT) に対して提案手法 (Proposed) の精度が高いことから、提案手法の有効性が確認できる。

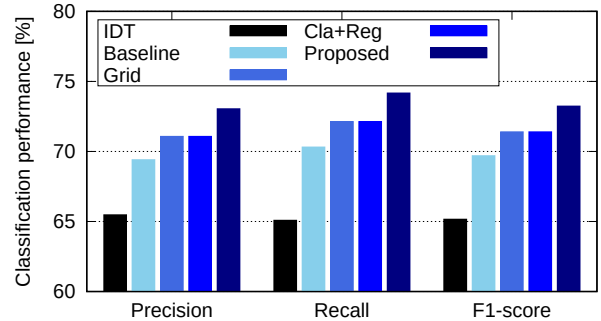


図6 比較手法と提案手法の各要素, 及びそれらの組み合わせの精度比較

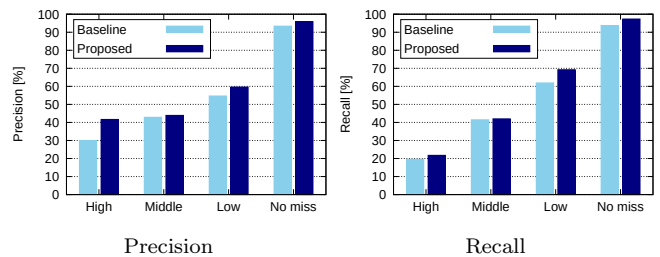


図7 ベースラインと提案手法の各クラス別の Precision と Recall

次に、ベースラインの手法と提案手法の性能をクラス別に比較した結果を図7に示す。結果から、提案手法ではベースラインの手法と比較して、特にクラス「高」の適合率、及びクラス「低」の再現率が改善されていることがわかる。

6. まとめ

本論文では、ドライブレコーダーデータで記録されたイベントデータに対してヒヤリハットの危険度を推定することを目的に、映像とセンサ、物体検出結果を複合的に活用した推定手法を提案した。提案手法では、ヒヤリハット発生対象の自転車に対する位置関係を表現するグリッド特徴、及び危険度の連続性を考慮する回帰手法を取り入れたモデルを提案し、評価実験においてそれぞれの有効性を確認した。また、双方を組み合わせるときに、より高精度でヒヤリハットの危険度を推定できることを明らかにした。

今後は、更に推定精度を向上させるため、ヒヤリハット発生対象の属性に関する情報を活用することを検討している。

謝 辞

本研究を遂行するにあたり、国立大学法人東京農工大学スマートモビリティ研究拠点の提供する「ヒヤリハットデータベース」を利用した。ここに記して謝意を示す。

文 献

- [1] M. Maeda, T. Uetani, and M. Takagi. Development of drive recorder (obvious recorder). *Fujitsu Ten Technical Journal*, Vol. 27, , 2006.
- [2] ドライブレコーダーの活用について | 警察庁 web サイト. <https://www.npa.go.jp/bureau/traffic/anzen/drive_recorder.html>, (参照 2018-11-23).
- [3] Driving safety promotion center honda motor co. ltd.: Honda driving safety promotion activities 2016. <http://www.honda.co.jp/safetyinfo/global/safetyinfo_2016_E.pdf>, (参照 2018-11-23).
- [4] 日本カーソリューションズ株式会社: 全国の 38 拠点から安全でエコな自動車リースを展開, ntt 技術ジャーナル. <<http://www.ntt.co.jp/journal/1110/files/jn201110040.pdf>>, (参照 2018-11-23).
- [5] Smart mobility research center - research. <<http://web.tuat.ac.jp/~smrc/drcenter.html>>, (参照 2018-11-23).
- [6] さいたま市/ヒヤリハットマップを作製しました. <<http://www.city.saitama.jp/chuo/001/001/003/p036723.html>>, (参照 2018-11-23).
- [7] ヒヤリハットマップをつくりませんか — 和歌山県. <<https://www.pref.wakayama.lg.jp/prefg/031300/documents/hiyarihatto.html>>, (参照 2018-11-23).
- [8] 子ども目線の交通安全マップを作製しました 松山市ホームページ. <<https://www.city.matsuyama.ehime.jp/kurashi/kurashi/seibi/keikaku/childmap/kodomomosen.html>>, (参照 2018-11-23).
- [9] 山本修平, 遠藤結城, 戸田浩之. 映像とセンサ信号を用いたドライブレコーダーデータからのヒヤリハット検出手法. 情報処理学会論文誌データベース (TOD), Vol. 10, No. 4, pp. 26–30, dec 2017.
- [10] 山本修平, 倉島健, 戸田浩之. ドライブレコーダーデータに対するヒヤリハット発生対象分類. In *DICOMO*, 2018.
- [11] K. Fukushima and S. Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets*, pp. 267–285. Springer, 1982.
- [12] T. Mikolov, M. Karafiát, L. Burget, J. Černocký, and S. Khudanpur. Recurrent neural network based language model. In *Eleventh Annual Conference of the International Speech Communication Association*, 2010.
- [13] 菊池理人, 日景由華, 御室哲志. ドライブレコーダーデータからの自動分別の試み. 計測自動制御学会東北支部第 290 回研究集会, 2014.
- [14] 森村哲郎, 悠輔谷澤, 山崎慎也, 井手剛. 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. 自動車技術会論文集, Vol. 43, No. 2, pp. 573–578, 2012.
- [15] R. Ke, J. Lutin, J. Spears, and Y. Wang. A cost-effective framework for automated vehicle-pedestrian near-miss detection through onboard monocular vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 25–32, 2017.
- [16] T. Suzuki, Y. Aoki, and H. Kataoka. Pedestrian near-miss analysis on vehicle-mounted driving recorders. In *Machine Vision Applications (MVA), 2017 Fifteenth IAPR International Conference on*, pp. 416–419. IEEE, 2017.
- [17] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [18] S. Ravada, M. Ali, S. Newsam, M. Renz, and G. Trajcevski, editors. *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS 2016, Burlingame, California, USA, October 31 - November 3, 2016*. ACM, 2016.
- [19] 豊田正史, 横山大作, 伊藤正彦. 運転状況を考慮したドライブレコーダーデータからの潜在リスク交差点検知手法. *DEIM*, 2017.
- [20] 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.
- [21] D. Yokoyama, M. Toyoda, and M. Kitsuregawa. Understanding drivers' safety by fusing large scale vehicle recorder dataset and heterogeneous circumstantial data. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 734–746. Springer, 2017.
- [22] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. Jackel, M. Monfort, U. Muller, J. Zhang, et al. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016.
- [23] J. Kim and J. Canny. Interpretable learning for self-driving cars by visualizing causal attention. In *ICCV*, pp. 2961–2969, 2017.
- [24] H. Xu, Y. Gao, F. Yu, and T. Darrell. End-to-end learning of driving models from large-scale video datasets. *arXiv preprint*, 2017.
- [25] A. Jain, H. Koppula, B. Raghavan, S. Soh, and A. Saxena. Car that knows before you do: Anticipating maneuvers via learning temporal driving models. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3182–3190, 2015.
- [26] A. Jain, A. Singh, H. Koppula, S. Soh, and A. Saxena. Recurrent neural networks for driver activity anticipation via sensory-fusion architecture. In *Robotics and Automation (ICRA), 2016 IEEE International Conference on*, pp. 3118–3125. IEEE, 2016.
- [27] F. Chan, Y. Chen, Y. Xiang, and M. Sun. Anticipating accidents in dashcam videos. In *Asian Conference on Computer Vision*, pp. 136–153. Springer, 2016.
- [28] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [29] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018.
- [30] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [31] D. Scherer, A. Müller, and S. Behnke. Evaluation of pooling operations in convolutional architectures for object recognition. In *Artificial Neural Networks-ICANN 2010*, pp. 92–101. Springer, 2010.
- [32] A. Bordes, X. Glorot, J. Weston, and Y. Bengio. Joint learning of words and meaning representations for open-text semantic parsing. In Neil D. Lawrence and Mark Girolami, editors, *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, Vol. 22 of *Proceedings of Machine Learning Research*, pp. 127–135, La Palma, Canary Islands, 21–23 Apr 2012. PMLR.
- [33] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, Vol. 12, No. Aug, pp. 2493–2537, 2011.
- [34] H. Wang and C. Schmid. Action recognition with improved trajectories. In *IEEE International Conference on Computer Vision*, pp. 3551–3558, 2013.