

敵対的学習に基づくドメイン適応による ドライブレコーダを用いたヒヤリハット検出及び分類

瀧本 祥章^{1,a)} 山本 修平^{1,b)} 松林 達史^{1,c)} 倉島 健^{1,d)} 戸田 浩之^{1,e)}

概要: ドライブレコーダデータから交通事故やヒヤリハットなどの危険な状態を有用なデータとして抽出、分類し、その利活用が行われている。しかしながら、それらのデータを自動的に抽出、分類するには多量の教師データを人手で用意する必要があり、高コストである。また、既存の教師ありデータを用いて学習したモデルを適用することも考えられるが、車種やカメラの種類などの収集時の条件の違いによって、データ集合全体の傾向が大きく変化し、その精度は制限される。そこで本稿では、分類対象のデータと異なる環境で収集されたデータセットとその教師データを元に、高精度な分類モデルの構築を目指す。具体的には、既存の Convolutional Recurrent Neural Networks ベースのヒヤリハット分類手法をベースに、収集した環境を推定する層を追加することで、敵対的学習によるドメイン適応を行う。これによって、従来は収集した環境に大きく依存した特徴量を抽出する CNN 部分について、環境に依存しない特徴量の抽出が期待できる。実験では、実際のドライブレコーダデータを用いて検証を行い、ドメイン適応によって教師データがない環境下のデータ集合に対しても、高い分類精度を持つモデルを構築できることを示した。

1. はじめに

ドライブレコーダの普及に伴い、前方映像や加減速度、ブレーキ操作などの車両運行状況が膨大に収集されるようになった。例えば [1] では、運転中の急加減速をトリガーとして、その前後数秒間を記録することにより、交通事故やヒヤリハット（事故には至らなかったものの、事故に繋がる危険な状態）を膨大に収集している。

このように収集されたデータを安全運転教育^{*1}などに活用する機運が高まっている。しかし、ドライブレコーダによって記録されたデータには、交通事故やヒヤリハットでない、即ち、安全な状態も多く記録されることがわかっており [1]、効率的に有用なデータを抽出することが求められている。そのため、膨大に収集されたドライブレコーダデータから、危険な状態のみを抽出し、衝突しそうになった相手（歩行者や自転車）や危険度に応じて分類する手法 [2,3] が提案されている。

しかしながら、これらの手法の適用には予め多くの同一ドメインの教師データを用意する必要がある。ここでドメインとは、特定の条件下で収集されたデータの集まりを表す。例えば、[2] や [3] では 2000 件以上の十数秒の映像データを閲覧して、ヒヤリハットの有無や、その要因のラベル付けを人手で行うことを想定している。この作業は注意力や多くの時間を要するため高コストである。また、既存の教師ありデータを用いて学習したモデルによってヒヤリハットデータを抽出することも考えられるが、ドメインが異なる、即ち、車種や収集地域、カメラの種類や設置箇所が異なるなどの様々な要因によって、収集されるデータ全体の性質が異なるためその精度は制限される。

そこで本稿ではあるソース、ソースドメイン ($\mathcal{D}_S$) で収集された既存の教師ありデータセットと、それとは異なるターゲットドメイン ($\mathcal{D}_T$) で収集された教師なしデータセットの双方を用いて学習を行い、ドメイン $\mathcal{D}_T$ で収集されたデータからヒヤリハットの抽出及び分類を行う手法を提案する。具体的には、既存の Convolutional Recurrent Neural Networks (CRNN) ベースのヒヤリハットクラス分類モデル [2] に敵対的学習によるドメイン適応手法である Domain-Adversarial Neural Network (DANN) [4] を組み込む。この際、CNN 部分で得られる特徴量においては、前述の要因によって、複数ドメイン間の差異が大きく表れる一方で、RNN 部分では、ヒヤリハット発生の過程など

¹ 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
NTT Service Evolution Laboratories, NTT Corporation,
Yokosuka, Kanagawa 239-0847, Japan

a) yoshiaki.takimoto.ar@hco.ntt.co.jp

b) shuhei.yamamoto.ea@hco.ntt.co.jp

c) tatsushi.matsubayashi.tb@hco.ntt.co.jp

d) takeshi.kurashima.uf@hco.ntt.co.jp

e) hiroyuki.toda.xb@hco.ntt.co.jp

*1 https://www.npa.go.jp/bureau/traffic/anzen/drive_recorder.html

を学習しており、ドメイン間で共通の特徴量を持つと想定できる。そこで、CNN 部分についてのみ敵対的学習を行うことによって、ドメイン間で共通の特徴を効率的に抽出することが期待できる。最後に、実際のドライブレコーダデータを用いた実験は、ヒヤリハットの抽出及び分類タスクにおける DANN に基づくドメイン適応の有効性を示す。

2. 関連研究

2.1 ヒヤリハットクラス分類

ドライブレコーダを用いてヒヤリハットの検出やクラス分類を行う手法として、様々な手法が提案されている [2,3,5]. [5] では、前方映像から自転車と歩行者との距離を計算して、その距離に基づいてヒヤリハットを検出する。[3] では、two-stream CNN [6] をベースとする手法によって、前方映像から周囲の物体の動きなどを捉え、ヒヤリハットの検出と対象物（歩行者、自転車、車）と危険度（高、低）に基づく分類を行う。[2] では、CRNN ベースの手法によって前方映像及びセンサ（速度や加速度）を複合的に考慮し、ヒヤリハットを検出し、対象物（歩行者、自転車、車、バイク、単独）の分類を行う。これらの手法はあるドメイン $\mathcal{D}$ で教師あり学習を行い、同一のドメイン $\mathcal{D}$ に対し手法を適用する。これに対し、本稿では教師データがあるドメイン $\mathcal{D}_S$ と教師データがないドメイン $\mathcal{D}_T$ で学習を行い、ドメイン $\mathcal{D}_T$ に対し手法を適用するという点で異なる。また、本手法は [2] をベースとしているが、CRNN ベースであれば任意のニューラルネットの構造に対し、本手法を適用可能であり、ヒヤリハット発生予測 [7] など別のタスクにも用いることが可能である。

2.2 ドメイン適応

ドメイン適応は2つ以上のドメイン（ソースとターゲット）が共有する特徴空間を保有するが、異なる分布を持つことを想定する転移学習の一種であり、従来から多くの研究が行われてきた [8,9]。その手法は様々であるが、深層学習ベースのシンプルかつ有効な手法に DANN [4] がある。DANN では、ソースドメインとターゲットドメインのデータ双方を同時に学習を行い、2つの損失関数に基づく同時最適化を行う。具体的には、ソースドメインのデータに対し通常のクラス分類に基づく損失関数と、双方のドメインのデータに対してドメインをクラスとみなしたクラス分類に基づく損失関数を用意し、前者は通常通り、後者は敵対的に学習を行う。これにより、クラス分類には有効で、かつ、ドメインに依存しない特徴を抽出することが期待できる。DANN をドライブレコーダデータに適用した手法として [10] が存在する。しかし、[10] は運転操作の予測タスクを解いており、本稿が取り組むタスクとは異なる。そのため、想定するドメイン間の差分が異なる。例えば、特定の操作前の挙動やその過程は運転手ごとによってある程

度異なることが考えられることから、運転手をドメインとして捉え、運転手間の違いをドメインの差分としてそれに依存しない特徴を抽出する。これらに対し本稿では、ヒヤリハットの検出及び分類タスクに取り組み、ヒヤリハット発生過程はドメイン間で共通であると考えられ、[10] とは適応したいドメイン間の差分の性質が異なる。

3. 問題設定

本稿では、入力空間 V とクラス集合 $Y = \{1, \dots, C\}$ についての C クラス分類におけるソースドメイン $\mathcal{D}_S$ からターゲットドメイン $\mathcal{D}_T$ への教師なしドメイン適応学習を考える。ここで、ドメイン $\mathcal{D}$ は、 $V \times Y$ 空間上の分布を表す。また、入力空間上のドライブレコーダデータ $\mathbf{v}_i = \{(\mathbf{x}_{il}^i, \mathbf{x}_{il}^s, \mathbf{x}_{il}^o)\}_{l=1}^L \in V$ は前方画像 $\mathbf{x}_{il}^i$ 、センサ $\mathbf{x}_{il}^s$ 、前方画像に対する物体検出結果 $\mathbf{x}_{il}^o$ の3つ組（以下、フレーム）を L 個持つ集合である。なお、 i はドライブレコーダデータ固有の識別子であり、 l はドライブレコーダ内において記録された順序を表すフレームの識別子である。

本稿における具体的な問題設定は次のようになる。独立同分布によって収集されたクラスラベル付きの n 個のソースドメインのデータ集合 $S = \{(\mathbf{v}_i, y_i)\}_{i=1}^n \sim (\mathcal{D}_S)^n$ と、クラスラベルがない n' 個のターゲットドメインデータ集合 $T = \{\mathbf{v}_i\}_{i=n+1}^{n+n'} \sim (\mathcal{D}_T^V)^{n'}$ の合計 N 個のデータが与えられた際に、 $\Pr_{(\mathbf{v}, y) \sim \mathcal{D}_T}(f(\mathbf{v}) \neq y)$ を最小化、即ち、ターゲットドメインのデータを正しいクラスに分類する分類器を求める。ここで、 $\mathcal{D}_T^V$ は V の周辺分布を表す。

4. 提案手法

提案手法は CRNN ベースのヒヤリハット分類モデルである [2] に DANN を取り入れる。[2] は図 1 の Feature Extraction Layer (FEL), Temporal Layer (TL), Classifier Layer (CL) の3つの部分要素から構成される。各部分要素の役割は以下のように捉えられる。FEL は画像やセンサなどから得られた生の情報からそれぞれの特徴量を抽出しそれらの特徴量を組み合わせる ANet を用いて、その瞬間の状態を捉える役割を持つ。TL は FEL で捉えた瞬間の状態の変化、即ち、ヒヤリハット発生時の過程を捉える役割を持つ。CL は TL で捉えた過程に基づいてクラスを分類する役割を持つ。これらの役割の違いに着目すると生の情報を取り扱う FEL はドメイン間で大きな差異があり、ヒヤリハット発生時の過程やクラス分類を行う TL, CL ではドメイン間の差が軽微であることが想定される。

これを踏まえ本稿では図 1 のように、FEL の出力を入力とする Domain Classifier Layer (DCL) を追加して、FEL に関してドメイン適応を行う。DCL は Gradient Reversal Layer (GRL) と Full Connect (FC) 層で構成され、ドメインをクラスとみなしたクラス分類を行う。ここで、GRL とは逆伝播時の勾配に -1 をかける層であり、敵対的な学

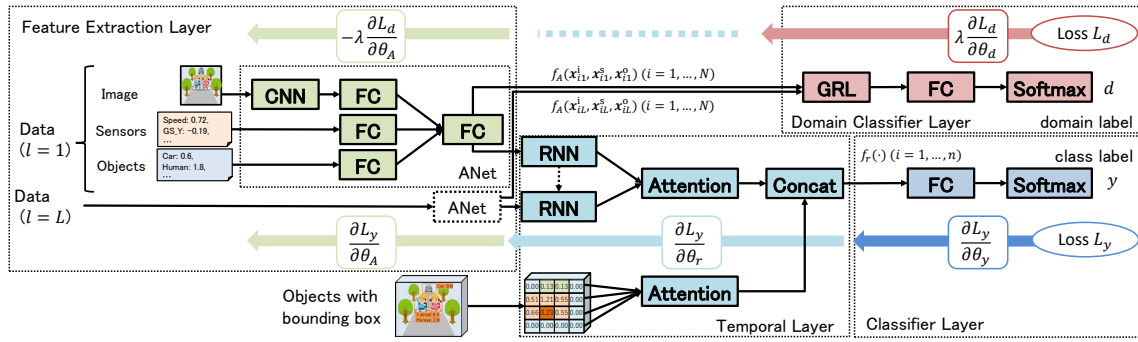


図 1: ネットワーク構成図 (黒矢印は順伝播時のデータの流れを, 色付きの矢印は逆伝播時の流れを表す)

習を行うための層である。即ち, GRL 以前の層ではドメインの分類が行えないように特徴を抽出し, GRL 以降の層ではドメインの分類が行えるように学習を行う。

本手法における損失関数は, DANN と同様に従来の CL での損失関数を L_y , DCL による損失を L_d とすると $L_y + \lambda L_d$ と書ける。ただし, λ は 2 つの損失関数間のスケールを調整するためのハイパーパラメータである。なお, 各部分要素によって学習する関数 f と, そのパラメータ θ を FEL, TL, CL, DCL の順に $(f_A, \theta_A), (f_r, \theta_r), (f_y, \theta_y), (f_d, \theta_d)$ とすると, 学習時の順伝播, 及び, 逆伝播の流れは図 1 に示すとおりである。

5. 評価実験

5.1 実験条件

実験で用いるデータセットには, 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」(以下, 農工大) [1] と, 日本カーソリューションズ株式会社^{*2}から提供して頂いたイベントデータ (以下, NCS) を異なるドメインとして用いた。農工大データは日本国内のタクシーに設置されたドライブレコーダで収集され, 設置位置や車種が統制されているのに対し, NCS データは日本国内の法人車両に設置されたドライブレコーダで収集され, 設置位置や車種が多様である。各イベントデータは加速度トリガが反応した時刻を中心に, 前後十数秒の前方映像とセンサ系列からなり, 共に 30 fps で記録される。なお, センサには前後加速度, 横加速度, 速度の 3 種類のデータが含まれ, 物体検出には YOLOv2 [11] を用いた。また, 各イベントデータは熟覧した人間によって, 事故やヒヤリハットの有無, ヒヤリハットの発生対象 (車, 自転車など) のラベルが付与されている。本稿では, { 安全, ヒヤリハット, 事故 } の 3 クラス分類タスクと, { 安全, 車, 自転車, 歩行者 } の 4 クラス分類タスクで評価を行った。各ラベルの訓練データと評価データの件数は表 1 のとおりである。なお, 実装には Chainer^{*3}を用いて, FC 層のユニット数は 256 とし, CNN は 3 層, RNN には LSTM [12], 最適化に

表 1: データセットの内訳

Label	農工大		NCS	
	Train	Test	Train	Test
安全	5600	350	2765	350
3 クラス ヒヤリハット	2000	125	987	125
事故	400	25	198	25
安全	3584	224	1792	224
4 クラス 車	512	32	256	32
自転車	512	32	256	32
歩行者	512	32	256	32

は Adam [13] (学習率などは Chainer のデフォルトの値) を用いた。また, バッチサイズは 50 とした。

5.2 比較手法

本稿では以下の 4 種の手法について評価を行った。

SourceModel: ソースドメインのデータのみで教師あり学習を行ったモデル。

Supervised: ターゲットドメインのデータのみで教師あり学習を行ったモデル。

DARNN: [10] で提案された DANN の適用方法を採用したモデル。図 1 の GRL の入力 ANet でなく TL の Concat 層であり, RNN 部分にもドメイン適応を行う。

Proposed: 4 章で述べた提案手法。

なお, DARNN と Proposed の学習には, ソースドメインのデータ及びラベルと, ターゲットドメインのデータのみを用いる。

5.3 実験結果

各タスクにおける各モデルの精度は表 2, 表 3 のとおりである。これらの結果から, ヒヤリハット検出及び分類タスクにおいて, 敵対的学習によるドメイン適応が有効であり, 通常の教師あり学習と同等か, それ以上の精度を実現可能であることがわかる。これは, ターゲットドメインのデータのみでは学習することが困難なデータ量のためであると考えられる。また, DARNN と Proposed を比較すると, Proposed の方が精度が良い, または同等の精度を達成しており, 2 つのドメイン間の差異は FEL 部分に存在し,

^{*2} <https://www.ncsol.co.jp>

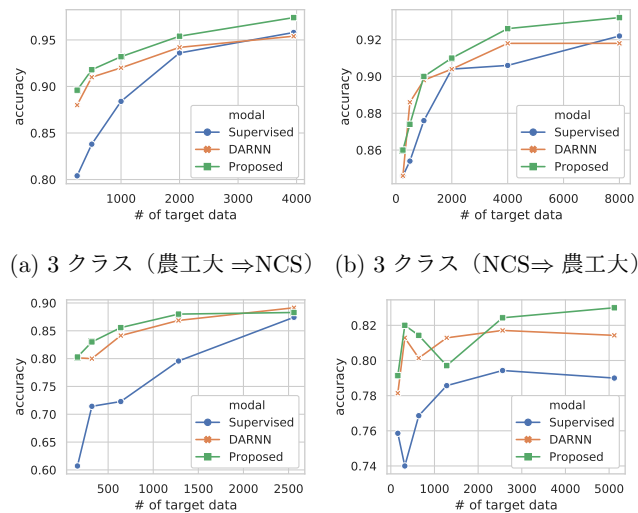
^{*3} <https://chainer.org>

表 2: 各タスク・各モデルの Accuracy の比較

	農工大 ⇒ NCS *4		NCS ⇒ 農工大	
	3 クラス	4 クラス	3 クラス	4 クラス
SourceModel	0.362	0.351	0.406	0.444
Supervised	0.958	0.874	0.922	0.790
DARNN	0.954	0.891	0.918	0.814
Proposed	0.974	0.883	0.932	0.830

表 3: 各タスク・各モデルの F 値 (重み付き平均) の比較

	農工大 ⇒ NCS		NCS ⇒ 農工大	
	3 クラス	4 クラス	3 クラス	4 クラス
SourceModel	0.311	0.702	0.411	0.511
Supervised	0.954	0.891	0.911	0.822
DARNN	0.951	0.882	0.900	0.794
Proposed	0.972	0.888	0.930	0.840



(a) 3 クラス (農工大 ⇒ NCS) (b) 3 クラス (NCS ⇒ 農工大)

(c) 4 クラス (農工大 ⇒ NCS) (d) 4 クラス (NCS ⇒ 農工大)

図 2: ターゲットデータ数による Accuracy の変化

TL で捉えるヒヤリハットの発生の過程にはドメイン間で共通であることが示唆される。

また、ターゲットドメインのデータ数をクラス間の比率を変えずに変化させて実験を行い、 x 軸に学習に用いたターゲットデータ数をとった結果は図 2 のようになった。これらの結果から、ターゲットドメインのデータ数が多いほど、精度が向上することが確認できるが、特にターゲットドメインのデータ数が少ないときに、DARNN 及び提案手法が有効であることが読み取れる。

6. まとめ

本稿においては、教師データがないデータ収集条件下でのヒヤリハット検出及び分類に取り組んだ。このタスクを解くため、敵対的学習に基づくドメイン適応に着目し、深層学習のネットワーク構造からドメイン間の差異が大きい部

分のみへの適用を行った。また、実際のドライブレコーダデータを用いた実験では、ドメイン適応によって同ドメインでの教師あり学習と同等以上の精度を実現可能であることを示した。今後は、情報源の性質を考慮した手法や [14] が主張するような、クラス集合がソースとターゲット間で異なる場合におけるドメイン適応を検討している。

謝辞 本研究は、日本カーソリューションズ株式会社から提供頂いたドライブレコーダデータと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。

参考文献

- [1] Pongsathorn Raksincharensak. Drive recorder database for accident/incident study and its potential for active safety development, 2013. <http://web.tuat.ac.jp/~smrc/pdf/Driverrecorder.pdf>.
- [2] 山本修平, 倉島健, and 戸田浩之. ドライブレコーダデータに対するヒヤリハット発生対象分類. *DICOMO*, 2018.
- [3] Hirokatsu Kataoka, Teppei Suzuki, Shoko Oikawa, Yasuhiro Matsui, and Yutaka Satoh. Drive video analysis for the detection of traffic near-miss incidents. *arXiv:1804.02555*, 2018.
- [4] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *JMLR*, 2016.
- [5] Ruimin Ke, Jerome Lutin, Jerry Spears, and Yinhai Wang. A cost-effective framework for automated vehicle-pedestrian near-miss detection through onboard monocular vision. In *CVPRW*, pages 25–32, 2017.
- [6] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *NIPS*, pages 568–576, 2014.
- [7] Tomoyuki Suzuki, Hirokatsu Kataoka, Yoshimitsu Aoki, and Yutaka Satoh. Anticipating traffic accidents with adaptive loss and large-scale incident db. *arXiv:1804.02675*, 2018.
- [8] Gabriela Csurka. Domain adaptation for visual applications: A comprehensive survey. *arXiv:1702.05374*, 2017.
- [9] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *arXiv:1802.03601*, 2018.
- [10] Michele Tonutti, Emanuele Ruffaldi, Alessandro Cattaneo, and Carlo Alberto Avizzano. Robust and subject-independent driving manoeuvre anticipation through domain-adversarial recurrent neural networks. *ROBOT AUT S*, 115:162–173, 2019.
- [11] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *CVPR*, pages 7263–7271, 2017.
- [12] Haşim Sak, Andrew Senior, and Françoise Beaufays. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In *ISCA*, 2014.
- [13] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.
- [14] Kaichao You, Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Universal domain adaptation. In *CVPR*, pages 2720–2729, 2019.

*4 ソースドメイン ⇒ ターゲットドメイン (以下同様)