

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の検出および分類

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)}

受付日 2019年12月9日, 採録日 2020年4月6日

概要: ドライブレコーダによって収集された膨大な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表すものを抽出, 分類し, それらのデータを利活用する取り組みが近年行われている. しかしながら, それらのデータを自動的に検出, 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セットを用いて学習したモデルの活用も考えられる. しかし, 車種やカメラの種類など, 収集時の条件が異なると, データ集合全体の傾向が大きく変化するため, そのモデルによる検出や分類精度は制限される. そこで本稿では, 分類対象のデータとそれとは異なる環境で収集された教師データがあるデータセットをもとに学習を行い, 高精度な分類を行えるモデルの構築を目指す. 具体的には, 既存の Convolutional Recurrent Neural Network 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, 敵対的学習による教師なし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 Convolutional Neural Network 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, 教師なし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.

キーワード: 教師なしドメイン適応, ヒヤリハット, CNN, RNN

Domain Adaptation based on Adversarial Learning for Detection and Classification of Near-miss Incident from Dashcam Data

YOSHIAKI TAKIMOTO^{1,a)} SHUHEI YAMAMOTO^{1,b)} TATSUSHI MATSUBAYASHI^{1,c)}
TAKESHI KURASHIMA^{1,d)} HIROYUKI TODA^{1,e)}

Received: December 9, 2019, Accepted: April 6, 2020

Abstract: Large amounts of highly useful driving data are being collected by dashboard cameras. By extracting dangerous situations from them and then classifying them into traffic accidents and near-miss incidents, many attractive applications are made possible. However, manually extraction is expensive because it requires a large amount of labeling. One method uses a model trained on an existing supervised dataset. If the collection conditions such as the model of vehicle and camera are different, the tendency of the dataset changes and the accuracy of the model is poor. In this paper, we propose a method that classifies target data with high accuracy by training based on two kinds of datasets: unlabeled target dataset, and labeled source dataset collected in an environment different from the target data. Specifically, our solution adds a layer to the existing convolutional recurrent neural network-based near-miss detection method. The added layer realizes unsupervised domain adaptation based on adversarial learning. As a result, the features extracted by the convolutional neural network are independent of the environment. An experiment on actual driving data verifies that the proposed method is effective.

Keywords: unsupervised domain adaptation, near-miss incident, CNN, RNN

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、速度や加速度などのセンサデータ、ブレーキやウィンカなどの操作データ、車両の運行状況を記録する装置である [1]。走行中のすべての情報を記録することは記憶容量などの問題から難しく、実際には一定の基準を持ってその一部のみを保存することが多い。たとえば [2] では、運転中の急加減速をトリガとして、その前後十数秒間を記録することにより、交通事故やヒヤリハットシーンの膨大な収集に成功している。ここで、ヒヤリハットとは交通事故には至らなかったものの、交通事故につながるような危険な状態を表す。このように、特定のトリガによって、その前後数十秒間を記録したデータを本稿ではイベントデータと呼ぶ。このようなイベントデータは、交通事故の証拠として利用されるほか、安全運転の促進に活用されている。たとえば、交通事故やヒヤリハットを含むイベントデータを、ドライバに視聴してもらうことにより、安全に対する意識を向上させる取り組み [3] や危険予知訓練での活用 [4] が行われている。ほかにも、法人営業車両などに対し、走行中の運行車両からリアルタイムにヒヤリハットシーンを検出、収集し、運行管理者に通知することにより、安全運転指導につながることも行われている [5]。

このように収集されたイベントデータを安全運転教育や運行管理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イベントデータには、交通事故やヒヤリハットでない、すなわち、安全な状態も多く記録されることが分かっており [6]、効率的に有用なイベントデータを抽出する手法が求められている。そのため、膨大に収集されたイベントデータから、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 [7], [8], [9] が提案されている。

しかしながら、これらの手法で得られたモデルには学習に用いたドメインのイベントデータしか高精度に分類できないという課題がある。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。ドメインが異なる場合、すなわち、車種や収集地域、カメラの種類や設置箇所の様々な収集条件が異なる場合、収集されるデータ全体の性質が異なってしまう。そのため、学習に用いたドメインに依存した特徴に基づいた分類を行うと、その分類精度は制限され、多くの分類誤りが生じる。また、同一ドメインにおける教師データを用意し、新たにモデルを用意する

ことも考えられるが高コストであり、困難である。たとえば、[7], [8], [9] では 2,000 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が、この作業は注意力や多くの時間を要する。

そこで本稿では、あるドメイン (D_S) で収集された既存の教師ありデータセットと、それとは異なるドメイン (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 D_T で収集されたイベントデータからヒヤリハットの検出および分類を行う手法を提案する。ここで、クラス分類の基準となる教師データを持つドメイン (D_S) をソースドメイン、分類を行いたい教師データがないドメイン (D_T) をターゲットドメインと呼ぶ。特に本稿では、車種、カメラの種類、設置箇所などが異なり、道路交通規則などが同一の地域から収集した 2 つのドメインについて考える。具体的には、ヒヤリハット分類に対し、有効であることが分かっている既存の Convolutional Recurrent Neural Network (CRNN) ベースのヒヤリハットクラス分類モデル [7] 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) [10] を組み込む。[7] では、イベントデータに含まれる前方映像やセンサデータから Convolutional Neural Network (CNN) などを用いて、特徴量を抽出し、Recurrent Neural Network (RNN) によって時系列変化をとらえる。そのため、入力された前方映像やセンサデータからその特徴を直接取得する CNN においては、前述の収集条件の違いによって、ドメイン間の差異が大きく表れると予想される。その一方で、RNN 以降の部分では、ヒヤリハットや安全運転などの各クラスの走行の過程やその差異を学習している。これらは、実際の走行の過程や道路交通法などのドメイン間で共通のものに依存するため、RNN 以降の部分においてはドメイン間の差異が軽微であると想定できる。そこで、CNN 部分についてのみ敵対的学習によるドメイン適応手法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。

本稿の残りの構成は以下のとおりである。まず 2 章で関連研究としてドライブレコーダデータから交通事故あるいはヒヤリハットを抽出する研究と、ドメイン適応手法について紹介する。次に、3 章で本稿で取り組むタスクを定式化し、4 章で提案手法について述べる。その後、5 章で実際のイベントデータを用いた実験について述べ、6 章にて本稿のまとめについて述べる。

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 [7], [8], [9], [11], [12], [13], [14], [15], [16], [17]。たとえ

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

ば、文献 [11] では、速度と加速度に対して、経験的に閾値を定め、条件を満たしたものをヒヤリハットとする。文献 [12] では、センサデータに適切な特徴量変換を行い、右折時のヒヤリハット検出を行っている。しかし、これらの手法は人間によって経験的に設定された特徴量に依存しており、多様な発生パターンに対応できない、ドメインが異なる場合に適切に動作することが保証されないという問題がある。文献 [13] では、前方映像から自転車と歩行者との距離を推定して、その距離に基づいてヒヤリハットを検出する。この手法は、得られた画像から実際の距離に基づいて分類を行うため、車体の大きさなどを考慮する必要はあるが、その分類精度はドメインに依存しないと考えられる。しかし、特定の発生対象（歩行者）のみを対象としており、多様な交通事故やヒヤリハットを抽出するには適していない。

深層学習を用いた手法としては、文献 [7], [8], [9], [14], [15], [16], [17] がある。文献 [14] では、歩行者検出をサブタスクで学習し、歩行者とのヒヤリハットについて危険度（高、低、安全）に基づく分類を行う。文献 [8] では、two-stream CNN [18] 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とらえ、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。文献 [15] では、CRNN ベースの手法によって前方映像およびセンサ（速度や加速度）を複合的に考慮してヒヤリハットを検出し、文献 [7] においては対象物（歩行者、自転車、車、バイク、単独）に基づく分類を行う。文献 [9] では、Dynamic Spatial Attention [16] ベースの手法によって、物体検出の結果の考慮の方法を工夫し、ヒヤリハットの検出、対象物（歩行者、自転車、車）に基づく分類を行う。文献 [16], [17] では、ヒヤリハットが発生する数秒前までの映像から数秒以内にヒヤリハットが発生するか否かを分類（予測）する。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。しかし、これらの通常の教師あり学習に基づいて学習したモデルは、異なるドメインに適用する場合には、ドメイン間の差異によってその精度が限定されることが知られている（ドメインギャップ）。これに対し、本稿では教師データがあるソースドメイン \mathcal{D}_S と教師データがないターゲットドメイン \mathcal{D}_T を用いて学習を行い、ターゲットドメイン \mathcal{D}_T に対し手法を適用するという点で異なる。

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 [19], [20]。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 [10] 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に用いて、2つの損失関数に基づく同時最適化を行う。具体的には、双方のドメインに共通する、特徴抽出用の Neural Network (NN) と、ソースドメインのデータから抽出した特徴に対する教師あり学習による損失 L_y と、双方のドメインのデータから抽出した特徴に対するドメインをクラスと見なしたクラス分類に基づく損失 L_d を用意する。この2つの損失関数に対し、前者の L_y については通常どおりの誤差逆伝播法によって学習を行う。その一方、後者の L_d については、特徴抽出用の NN の後段に Gradient Reversal Layer (GRL) を挿入することによって、敵対的に学習を行う。GRL はパラメータを持たない層であり、順伝播時には何も行わないが、逆伝播時には勾配に -1 をかける層である。これにより、GRL よりも出力側の層においては誤差逆伝播法に基づいて損失 L_d を最小化するように学習を行うが、GRL よりも入力側の層、すなわち、特徴抽出用の NN においては、勾配の符号が反転するため、損失 L_d を最大化するように学習を行う。そのため、十分に学習を行った後には、特徴抽出用の NN によって、各々のドメインに依存しない特徴を抽出することが期待でき、ターゲットドメインにおいても、ソースドメインと同様にクラス分類や目的変数値の推定が可能になる。

DANN をドライブレコーダデータに適用した手法として Domain-Adversarial Recurrent Neural Networks (DARNN) [21] が存在する。文献 [21] では運転操作の予測タスクを解くために、特徴抽出用の NN として、CRNN ベースのものを用いる。具体的には、ドライブレコーダで得られた前方映像と、車内を映したカメラから得られる運転手の動き（頭部や視線）を入力として CRNN ベースのネットワークで特徴量を抽出する。損失 L_y には、将来行う右左折や車線変更などの運転操作を分類する重み付きの交差エントロピーを用い、損失 L_d には、運転手をドメインとしてとらえ、これを分類する交差エントロピーを用いる。これは、右左折や車線変更などの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想定されるため、個々の運転手に依存しない、すべての運転手に共通する挙動を抽出し、学習データに存在しない運転手に対しても適切に運転操作を予測するためである。これに対し、本稿が取り組むタスクであるヒヤリハットの検出および分類におけるドメインギャップの主な原因は、車種、カメラの種類や設置個所によるものである。これらの条件が変化した場合、前方映像の撮像範囲などの各時刻に得られるデータはドメインによって大きく変化するが、走行の過程や道路交通法などは変化せず、ドメイン間で共通である。したがって、交通事故やヒヤリハット、安全運転の走行過程の差異に関するドメイン間の差分は軽微であると考えられる。このように、文献 [21] が適応させたいドメイン間の差分と本稿で考慮するドメイン間の差分の性質が異な

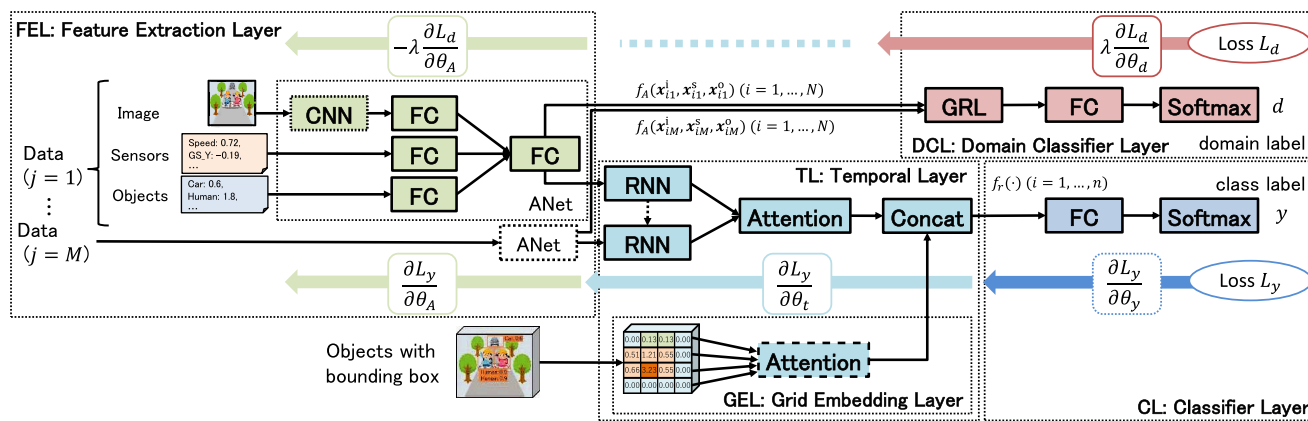


図 1 ネットワーク構成図 (黒矢印は順伝播時のデータの流れを、色付きの矢印は逆伝播時の流れを表す)

Fig. 1 Network Structure (The black arrows indicate the flow of data during the forward propagation, and the colored arrows indicate the flow during backpropagation).

るため、提案するモデルも異なる。

3. 問題設定

本稿では、入力データ空間 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 D_S からターゲットドメイン 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、入力空間上のイベントデータ $v_i = \{(x_{ij}^i, x_{ij}^s, x_{ij}^o)\}_{j=1}^M \in V$ は前方画像 x_{ij}^i 、センサ x_{ij}^s 、前方画像に対する物体検出結果 x_{ij}^o の 3 つ組 (以下、フレーム) を M 個持つ集合である。なお、 i はイベントデータ固有の識別子であり、 j は各イベントデータ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。ただし、各イベントデータはサンプリングなどにより、フレームの数 M 、および周波数、トリガのタイミングは揃えられているものとする。

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 $S = \{(v_i, y_i)\}_{i=1}^n \sim (D_S)^n$ と、クラスラベルがない n' 個のターゲットドメインのデータ集合 $T = \{v_i\}_{i=n+1}^N \sim (D_T^V)^{n'}$ の合計 N 個のデータが与えられた際に、 $\Pr_{(v,y) \sim D_T} (f(v) \neq y)$ を最小化する分類器 f 、すなわち、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、 $N = n + n'$ 、 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 [7] をベースに DANN を組み込むものである。文献 [7] 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。

個々の部分要素の機能および役割は以下のとおりである。FEL では画像やセンサなどから得られた生の情報から非線形変換によって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、各フレームに適用する。そのため、FEL は瞬間の状態をフレーム単位でとらえる役割を持つ。TL は FEL でとらえたフレーム単位の状態を入力とし、RNN と Attention を用いて時間変化をとらえる。また、TL 中の構成要素である Grid Embedding Layer (GEL) は物体検出結果 $\{x_{ij}^o\}_{j=1}^M$ に基づいて、前方画像中の注目すべき箇所をあらかじめ定めたグリッド単位で Attention によってとらえる層である。具体的には、まずあらかじめ定めた各グリッドについて、グリッドと検出領域が重なる物体検出結果を抽出する。その後、抽出した物体検出結果の前方画像中における面積比の物体の種類ごとの総和を、グリッドの特徴量とし Attention に入力する。これにより、TL はヒヤリハット発生時などの各クラスにおける過程をとらえる役割を持つ。CL は TL でとらえた過程に基づいてクラス进行分类する。なお、CL の損失 L_y は Full Connect (FC) 層の C 次元の出力に対し、ソフトマックス関数を適用し、交差エントロピーによって計算される。すなわち、

$$L_y = -\frac{1}{n} \sum_{i=1}^n \sum_{c=1}^C y_{i,c} \log \frac{\exp \hat{y}_{i,c}}{\sum_{k=1}^C \exp \hat{y}_{i,k}} \quad (1)$$

となる。ただし $y_{i,c}$ は y_i のワンホット表現中の次元 c の値を表し、 $\hat{y}_{i,c}$ はイベントデータ v_i に対する CL 中の FC の出力の次元 c の値を表す。これらの役割の違いに着目すると観測データを直接取り扱う FEL はドメイン間で大きな差異があると考えられる。その一方、ヒヤリハット発生時の過程をとらえる TL やクラス分類を行う CL においては、ドメイン間の差異が軽微であることが想定される。

これをふまえ本稿では図 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 DANN [10] に基づくドメイン適応を行う。DCL は GRL と FC 層で構成され、ドメインをクラスと見なしたクラス分類を行う。2.2 節で述べたように、GRL はその層の前後における敵対的学習を実現し、ドメインに依存しない特徴の抽出を可能にする。そのため、GRL 以前の層である FEL においては、ドメインに依存しない、かつ、クラス分類に有用な特徴量を抽出するように学習を行える。なお、DCL の損失 L_d は CL と同様に、FC 層の出力に対し、ソフトマックス関数を適用し、交差エントロピーによって計算される。すなわち、

$$L_d = -\frac{1}{N} \sum_{i=1}^N \sum_{d=1}^D y_{i,d}^{\text{DCL}} \log \frac{\exp \hat{y}_{i,d}^{\text{DCL}}}{\sum_{k=1}^D \exp \hat{y}_{i,k}^{\text{DCL}}} \quad (2)$$

となる。ただし、 $y_{i,d}^{\text{DCL}}$ はイベントデータ v_i のドメインを表すワンホット表現中の次元 d を表し、 D はドメイン数 (本稿では $D = 2$)、 $\hat{y}_{i,d}^{\text{DCL}}$ はイベントデータ v_i に対する DCL 中の FC の出力の次元 d の値を表す。これにより TL や CL にもドメイン適応を行う場合と比較すると次の 2 つの効果が期待できる。1 つ目は RNN を経由せずに、損失を FEL に伝えることができるため、勾配消失などによる学習の妨げとなる問題を避けることができ、FEL におけるドメイン適応を効率的に行える。2 つ目はドメイン間で元来共通する特徴をドメイン間の差異としてとらえてしまうことを防ぐ。これは、ヒヤリハットの発生過程の多様性に対する学習データ量の不足から生じる。

本手法における損失関数 L は、前述の CL での損失 L_y と DCL での損失 L_d の加重和であり、

$$L = L_y + \lambda L_d \quad (3)$$

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, θ_A) , (f_t, θ_t) , (f_y, θ_y) , (f_d, θ_d) とすると、学習時の順伝播、および、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) [2] と、日本カーソリューションズ株式会社*1のイベントデータ (以下、NCS) を異なるドメインとして用いた。各データセットに含まれる画像の例は図 2 のとおりである。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。実際、設置位置については図 2(a) のようにどの画像についても画像下部にボンネットがある程度

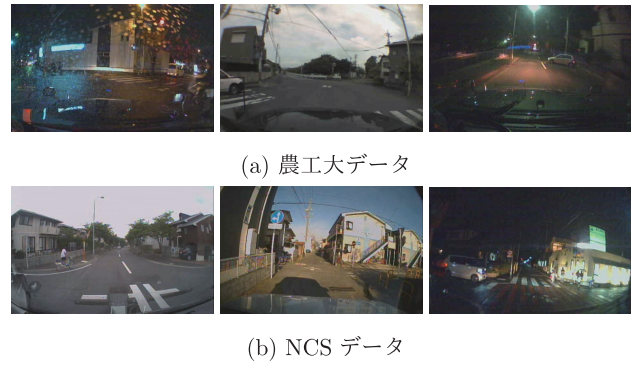


図 2 各データセットに含まれる画像の例
Fig. 2 Examples of images included in each dataset.

映っているなどの共通する特徴が読み取れる。これに対し、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。実際、図 2(b) のようにボンネットが映っている画像と映っていない画像があるなどの違いがあることが確認できる。双方の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る。本実験では [7] と同様に 2fps でフレームをサンプリングして、フレーム数 M を 30、トリガのタイミングを 20 フレーム目に統一した。なお、文献 [22], [23] によると日本国内で販売されるドライブレコーダの要件は、前方映像を 10fps 以上の周波数で記録すること、[23] において加速度トリガの前後の記録時間は前後 10 秒以上と定められていることから、任意のドライブレコーダで収集されたデータに対し、加速度トリガの前 10 秒間、後 5 秒間を 2fps の周波数でサンプリングすることが可能であり、本手法の有用性が確かめられると考える。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、ドメインごとに学習用のデータを用いて正規化し、物体検出には YOLOv2 [24] を用いた。また、各イベントデータは熟練した人間によって、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} のヒヤリハットシーンを検出する 3 クラス分類タスクと、{安全、車、自転車、歩行者} のヒヤリハットシーンの検出およびヒヤリハットの対象を分類する 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。3 クラス分類においては、農工大側の 3 つのクラスの比を NCS のデータの分布に揃え、データの分布が等しい場合を想定したデータセットになっている。これに対し、4 クラス分類においては、各々の提供されたデータの分布に基づいており、データの分布が異なる場合を想定したデータセットになっている。実装には Chainer*2を用いており、ネットワーク構造やハイパパラメータについては、先行研究 [7] に従っ

*1 <https://www.ncsol.co.jp>

*2 <https://chainer.org>

表 1 データセットの内訳

Table 1 Number of data per label in each data set.

Label	農工大		NCS	
	Train	Test	Train	Test
安全	5,600	350	2,765	350
3 クラス ヒヤリハット	2,000	125	978	125
事故	400	25	198	25
安全	6,576	411	2,799	311
4 クラス 車	928	58	927	103
自転車	192	12	396	44
歩行者	160	10	270	30

アルゴリズム 1: reverse validation

```

Input:  $S = \{(\mathbf{v}_i, y_i)\}_{i=1}^n, T = \{\mathbf{v}_i\}_{i=n+1}^N, f$ 
Output: Accuracy
1  $S^t, S^v \leftarrow S$  // 学習用と検証用に分割
2  $T^t, T^v \leftarrow T$ 
3  $f^n \leftarrow \text{train } f \text{ by } \{(\mathbf{v}_i, y_i) \mid \mathbf{v}_i \in S^t\} \cup \{\mathbf{v}_i \mid \mathbf{v}_i \in T^t\}$ 
   //  $\{(\mathbf{v}_i, y_i) \mid \mathbf{v}_i \in S^v\}$  を用いて early stopping
4  $f^r \leftarrow \text{train } f \text{ by } \{\mathbf{v}_i \mid \mathbf{v}_i \in S^t\} \cup \{(\mathbf{v}_i, f^n(\mathbf{v}_i)) \mid \mathbf{v}_i \in T^t\}$ 
   //  $\{(\mathbf{v}_i, f^n(\mathbf{v}_i)) \mid \mathbf{v}_i \in T^v\}$  を用いて early stopping
5 Accuracy  $\leftarrow \Pr_{\mathbf{v}_i \in S^v}(y_i = f^r(\mathbf{v}_i))$ 

```

た. 具体的には, FC 層のユニット数は 256 とし, RNN には Long Short-Term Memory (LSTM) [25], 最適化には Adam [26] (学習率などは Chainer のデフォルトの値) を用い, バッチサイズは 50 とした. ただし, CNN には事前学習を行っていないものを用いるため, 文献 [15] に従って 3 層とした. 本稿独自のハイパパラメータである λ はデータセットやタスク, 手法ごとに [10] と同様に, reverse validation によって決定した.

Reverse validation は教師なしドメイン適応特有の検証手法であり, ソースドメインのデータ集合 S , および, ターゲットドメインのデータ集合 T を用いて, 検証対象のモデル f のハイパパラメータの決定を行える. 具体的な手順はアルゴリズム 1 で示すとおりである. まず, ソースドメインのデータ集合 S をそれぞれ学習用と検証用のデータ集合 S^t, S^v に分割する (1 行目). 同様にターゲットドメインのデータ集合 T を T^t, T^v に分割する (2 行目). 次に, ソースドメインとターゲットドメインそれぞれの学習用データ集合 S^t, T^t を用いて, f を学習する (3 行目). この際に, ソースドメインの検証用データ S^v を用い, early stopping などを用いて f のパラメータを選択し, 得られた学習済みモデルを f^n とする. その後, 再度ソースドメインとターゲットドメインそれぞれの学習用データ集合 S^t, T^t を用いて, f を学習する (4 行目). ただし, ソースドメイン側は教師なし, ターゲットドメイン側は 3 行目で得られたモデル f^n で推定したラベルを教師として用いる. この際に, ターゲットドメインの検証用データ T^v を用い,

early stopping などを用いて f のパラメータを選択し, 得られた学習済みモデルを f^r とする. 最後に, ソースドメインの検証用データについて, 教師ラベルと 4 行目で得られた f^r で推定したラベルを比較することによって精度を算出する. なお, reverse validation で得られる各モデルの精度はハイパパラメータの決定のみに用いる.

5.2 比較手法

本稿では評価に以下の提案手法を含む 2 種の教師なしドメイン適応について, 2 種類のネットワーク構造を用いた. **DARNN w/ GEL** 文献 [21] で提案された DARNN の適用方法を採用したモデル. 2.2 節で述べたように, RNN 部分にも含めて DARNN を適用する. そのため, 図 1 の GRL の入力 ANet でなく TL の Concat 層の出力となる.

DARNN w/o GEL GEL を DARNN w/ GEL から取り除いた手法. なお, GEL は 4 章で述べたように物体検出の結果を効率良く考慮するための Attention に基づく層である.

Proposed w/ GEL 4 章で述べた提案手法.

Proposed w/o GEL DARNN w/o GEL と同様に GEL を Proposed w/ GEL から取り除いた手法.

DARNN は TL (RNN 部分) を含めてドメイン適応を行う. しかし, 4 章で述べたように, ヒヤリハットの検出および分類タスクにおいては, ドメイン間の差異は FEL でとらえる部分にそのほとんどが占められ, TL でとらえるヒヤリハットや安全運転の走行過程についてはドメイン間の差が軽微であると考えられる. そのため, FEL に対して DCL から直接勾配情報を逆伝播する Proposed の方が, TL を介して勾配情報を逆伝播する DARNN よりも効率良くドメイン適応を実現することが期待できる. なお, GEL を持つネットワークと GEL を取り除いたネットワーク双方を用いるのは, 手法の有効性が特定のネットワークに依存しないこと確認するためである.

5.3 事前分析

本節では, 3 クラス分類タスクに用いるデータセットによって, ドメイン間の差分が確かに存在することを示す.

図 3 は各々のデータセットの輝度のヒストグラムを示したものである (各データから中央のフレームの画像 $\mathbf{x}_i^i[\frac{M}{2}]$ を抽出し, 全ピクセルを集計し正規化を行った). この図から, 農工大データは比較的暗いデータが多く, 夜に走行している割合が多いことが読み取れる. これは, NCS データは営業車両などの法人車両によって収集されていることから, 勤務時間, すなわち, 日中に走行することが多いのに対し, 農工大データはタクシーによって収集されていることから, 終電後などの夜から深夜にかけて走行することが多いことによるものと考えられる. また, どちらのデー

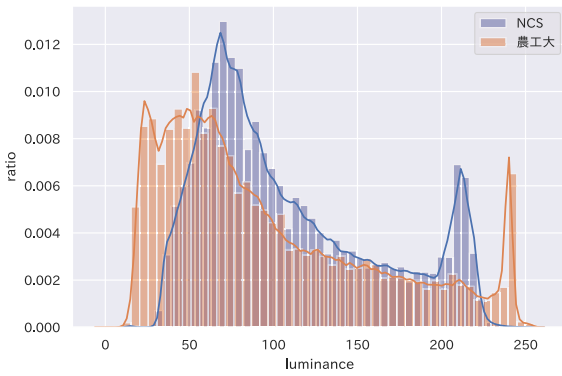


図3 各データセットの輝度のヒストグラムおよびカーネル密度推定結果*3

Fig. 3 Histogram and kernel density estimation results of luminance for each dataset.

表2 3クラス分類における Accuracy の低下

Table 2 The decline in accuracy in 3-class classifications.

	農工大 ⇒ NCS *4	NCS ⇒ 農工大
Source Model w/ GEL	0.362	0.406
Target Model w/ GEL	0.958	0.922

タセットにおいても輝度が200以上の部分の多くは空であり、太陽光の影響によって白とびが生じている場合があるなどカメラの性能や設定によって分布に違いが生じていると考えられる。

また、表2は実際にドメイン間の差異による分類精度の低下を確認するために行った3クラス分類タスクにおける予備実験の結果である。Source Modelはソースドメインで教師あり学習を行い、Target Modelはターゲットドメインで教師あり学習を行った場合のモデルを表し、表中の数値はターゲットドメインにおける分類精度 (accuracy) を表している。この結果から、ターゲットドメインで学習を行うと9割以上の精度を達成するものの、異なるドメインで学習した場合にはその精度が4割程度と大きく低下することが確認できる。

以上のことから2つのドメイン間には確かに差異があることが確認できた。

5.4 実験結果

各タスクにおける各モデルの精度は表3、表4のとおりである。なお、これらの精度は reverse validation によって得られた精度ではなく、ターゲットドメインの評価データに対する精度である。これらの結果から、ヒヤリハット検出およ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には及ばないものの9割近い精度を実現可能であることが分かる。また、

*3 ヒストグラムの bin 幅は seaborn (<http://seaborn.pydata.org/>) のデフォルトであるフリードマン=ダイアコニスの法則によって決定した。

*4 ソースドメイン ⇒ ターゲットドメイン (以下同様)

表3 各タスク・各モデルの Accuracy の比較

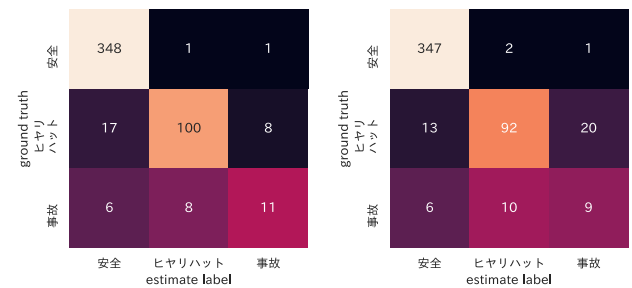
Table 3 Accuracy for each task, each model.

	農工大 ⇒ NCS		NCS ⇒ 農工大	
	3クラス	4クラス	3クラス	4クラス
DARNN w/o GEL	0.952	0.873	0.904	0.919
Proposed w/o GEL	0.950	0.906	0.916	0.933
DARNN w/ GEL	0.956	0.881	0.918	0.925
Proposed w/ GEL	0.960	0.891	0.896	0.933

表4 各タスク・各モデルの F 値 (重み付き平均) の比較

Table 4 F-measure for each task, each model.

	農工大 ⇒ NCS		NCS ⇒ 農工大	
	3クラス	4クラス	3クラス	4クラス
DARNN w/o GEL	0.946	0.867	0.900	0.915
Proposed w/o GEL	0.943	0.901	0.912	0.927
DARNN w/ GEL	0.953	0.881	0.914	0.916
Proposed w/ GEL	0.958	0.886	0.896	0.927



(a) DARNN w/ GEL

(b) Proposed w/ GEL

図4 NCS ⇒ 農工大の3クラス分類に対する混同行列*5

Fig. 4 Confusion matrix for 3-class classifications of NCS ⇒ TUAT.

DARNN と Proposed を比較すると、各モデルとも4つのタスクのうち、3つのタスクにおいて Proposed の方が精度が良くなっていることが確認できる。そのため、2つのドメイン間の差異は FEL 部分に存在し、TL でとら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、提案手法は双方のモデルで有効であることが示唆される。

次に、DARNN と Proposed 間で最も Proposed の精度が下がる NCS データをソースドメイン、農工大データをターゲットドメインとする3クラス分類タスクにおける GEL ありの場合の結果に着目する。図4は各手法による分類結果を表す混同行列である。この結果から Proposed はヒヤリハットクラスを誤って事故クラスとして推定するため、精度が低下していることが分かる。これは、ソースドメインにおける事故クラスのデータが少ないのに対し、事故の発生の仕方は多様であることから、識別境界が安定していないためと考えられる。

ここで、1章で述べた安全運転教育などにおける実用を考えた場合、危険なデータ (ヒヤリハットクラスおよび事

*5 混同行列内の数値は該当データ数

表 5 2 クラス分類タスクに基づく危険データ抽出精度
Table 5 Accuracy and F-measure based on two-class classification tasks.

	accuracy	precision	recall	F 値
農工大 ⇒ NCS				
DARNN w/ GEL	0.974	0.966	0.947	0.956
Proposed w/ GEL	0.986	0.980	0.973	0.977
NCS ⇒ 農工大				
DARNN w/ GEL	0.950	0.984	0.847	0.910
Proposed w/ GEL	0.956	0.978	0.873	0.923

故クラス)と安全なデータ(安全クラス)を識別できれば十分と考えられる。3クラス分類タスクをそのような、危険なデータを抽出する2クラス分類タスクととらえた場合の精度は表5が示すとおりである。表5から accuracy, F 値ともに Proposed の精度が良く, recall も大きいためとり逃しが少なくなっていることが分かる。

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およ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と想定できる部分に対してのみの適用を行った。また, 実際のドライブレコーダデータを用いた実験では, 複数のネットワーク構造に対し, 提案手法が適切なドメイン適応を実現し, 9割程度の精度を実現可能であることを示した。今後の課題として, 情報源の性質を考慮した手法や文献[27]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や, 交通事故などソースドメインにおいてもデータ数が少ない場合のクラス分類の精度向上があげられる。

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いただ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] Maeda, M., Uetani, T. and Takagi, M.: Development of Drive Recorder (OBVIOUS Recorder) (2006), available from <https://www.denso-ten.com/business/technicaljournal/pdf/27-1.pdf>.

[2] Raksincharoensak, P.: Drive Recorder Database for Accident/Incident Study and Its Potential for Active Safety Development (2013), available from <http://web.tuat.ac.jp/~smrc/pdf/Driverrecorder.pdf>.

[3] 警察庁交通局: ドライブレコーダーの活用について, 入手先 (https://www.npa.go.jp/bureau/traffic/anzen/drive_recorder.html).

[4] Driving Safety Promotion Center Honda Motor Co. Ltd.: Honda Driving Safety Promotion Activities 2016 (2017),

available from http://www.honda.co.jp/safetyinfo/global/safetyinfo.2016_E.pdf.

[5] 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, 入手先 (<http://www.ntt.co.jp/journal/1110/files/jn201110040.pdf>).

[6] スマートモビリティ研究拠点: ドライブレコーダデータセンター, 入手先 (<http://web.tuat.ac.jp/~smrc/drcenter.html>).

[7] 山本修平, 倉島 健, 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).

[8] Kataoka, H., Suzuki, T., Oikawa, S., Matsui, Y. and Satoh, Y.: Drive Video Analysis for the Detection of Traffic Near-Miss Incidents, arXiv:1804.02555 (2018).

[9] 山本修平, 倉島 健, 松林達史, 戸田浩之: 物体特徴量に基づく危険運転状況の推定, *DICOMO* (2019).

[10] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M. and Lempitsky, V.: Domain-adversarial training of neural networks, *JMLR* (2016).

[11] 菊池理人, 日景由華, 御室哲志: ドライブレコーダ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回研究会 (2014).

[12] 森村哲郎, 谷澤悠輔, 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, 自動車技術会秋季学術講演会 (2011).

[13] Ke, R., Lutin, J., Spears, J. and Wang, Y.: A Cost-effective Framework for Automated Vehicle-pedestrian Near-miss Detection through Onboard Monocular Vision, *CVPRW*, pp.25-32 (2017).

[14] Suzuki, T., Aoki, Y. and Kataoka, H.: Pedestrian near-miss analysis on vehicle-mounted driving recorders, *MVA*, pp.416-419, IEEE (2017).

[15] 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 (2017).

[16] Chan, F.-H., Chen, Y.-T., Xiang, Y. and Sun, M.: Anticipating accidents in dashcam videos, *ACCV*, pp.136-153, Springer (2016).

[17] Suzuki, T., Kataoka, H., Aoki, Y. and Satoh, Y.: Anticipating Traffic Accidents with Adaptive Loss and Large-scale Incident DB, arXiv:1804.02675 (2018).

[18] Simonyan, K. and Zisserman, A.: Two-stream convolutional networks for action recognition in videos, *NIPS*, pp.568-576 (2014).

[19] Csurka, G.: Domain adaptation for visual applications: A comprehensive survey, arXiv:1702.05374 (2017).

[20] Wang, M. and Deng, W.: Deep Visual Domain Adaptation: A Survey, arXiv:1802.03601 (2018).

[21] Tonutti, M., Ruffaldi, E., Cattaneo, A. and Avizzano, C.A.: Robust and subject-independent driving manoeuvre anticipation through Domain-Adversarial Recurrent Neural Networks, *ROBOT AUT S*, Vol.115, pp.162-173 (2019).

[22] 一般社団法人ドライブレコーダー協議会: ドライブレコーダー推奨ガイドライン基本要件編 (2019), 入手先 (http://www.jdrc.gr.jp/pdf/JDRC_guideline.pdf).

[23] 国土交通省: 国土交通省告示第千三百四十六号ドライブレコーダーにより記録すべき情報及びドライブレコーダーの性能要件を定める告示 (2016), 入手先 (<https://www.mlit.go.jp/common/001181710.pdf>).

[24] Redmon, J. and Farhadi, A.: YOLO9000: better, faster, stronger, *CVPR*, pp.7263-7271 (2017).

[25] Sak, H., Senior, A. and Beaufays, F.: Long short-term memory recurrent neural network architectures for large

scale acoustic modeling, *ISCA* (2014).

- [26] Kingma, D.P. and Ba, J.: Adam: A Method for Stochastic Optimization, *ICLR* (2015).
- [27] You, K., Long, M., Cao, Z., Wang, J. and Jordan, M.I.: Universal Domain Adaptation, *CVPR*, pp.2720-2729 (2019).



瀧本 祥章

2017年名古屋大学大学院情報科学研究科博士課程前期課程修了。同年日本電信電話株式会社に入社。現在、NTTサービスエボリューション研究所でデータマイニング、時空間データ分析の研究開発に従事。日本データベース

学会会員。



山本 修平

2012年筑波大学情報学群知識情報・図書館学類卒業。2014年同大学大学院図書館情報メディア研究科博士前期課程修了。2016年同大学同研究科博士後期課程修了。博士（情報学）。同年日本電信電話株式会社に入社。データ

マイニング、時空間データ分析に関する研究開発に従事。日本データベース学会会員。



松林 達史

2000年京都大学理工学部物理学科卒業。2002年10月より2年半、理化学研究所非常勤研究員。2005年東京工業大学大学院理工学研究科地球惑星科学専攻博士課程修了。同年、日本電信

電話株式会社入社。以来、NTTコミュニケーション科学基礎研究所、NTTサービスエボリューション研究所で、機械学習およびコンピュータビジョンの研究開発に従事。博士（理学）。情報処理学会会員。



倉島 健

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特別研究員。2004年同志社大学工学部知識工学科卒業。2006年京都大学大学院情報学研究科社会情報学専攻修士課程修了。同年日本電信電話株式会社入社。以

来、ユーザ行動分析、データマイニングの研究開発に従事。2014年京都大学大学院情報学研究科社会情報学専攻博士後期課程修了。博士（情報学）。2016年より1年1カ月間、米国スタンフォード大学客員研究員。ACM、情報処理学会、電子情報通信学会、日本データベース学会各会員。



戸田 浩之

1997年名古屋大学工学部材料プロセス工学科卒業。1999年同大学院工学研究科材料プロセス工学専攻博士課程前期課程修了。同年日本電信電話株式会社入社。2007年筑波大学大学院システム情報工学研究科コンピュータサ

イエンス専攻博士後期課程修了。博士（工学）。情報検索、データマイニングの研究に従事。現在、NTTサービスエボリューション研究所主幹研究員。電子情報通信学会、日本データベース学会、人工知能学会、ACM各会員。

(担当編集委員 櫻 惇志)