On Migratable Traffic Risk Estimation in Urban Sensing: A Social Sensing based Deep Transfer Network Approach

Yang Zhang, Daniel Zhang, Dong Wang

Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN 46556

Abstract

This paper focuses on the *migratable traffic risk estimation* problem in intelligent transportation systems using the social sensing. The goal is to accurately estimate the traffic risk of a *target area* where the ground truth traffic accident reports are not available by leveraging an estimation model from a *source area* where such data is available. Two important challenges exist. The first challenge lies in the discrepancy between source and target areas and such discrepancy would prevent a direct application of a model from the source area to the target area. The second challenge lies in the difficulty of identifying all potential features in the migratable traffic risk estimation problem and decide the importance of identified features due to the lack of ground truth labels in the target area. To address these challenges, we develop DeepRisk, a social sensing based migratable traffic risk estimation scheme using deep transfer learning techniques. The evaluation results on a real world dataset in New York City show the DeepRisk significantly outperforms the state-of-the-art baselines in accurately estimating the traffic risk of locations in a city.

Keywords: Social Sensing, Urban Sensing, Intelligent Transportation, Traffic Risk Estimation, Deep Learning

1. Introduction

Social sensing has emerged as a new networked sensing paradigm that uses humans as sensors to report the states of the physical world [1, 2, 3]. Examples of social sensing applications include tracking the real-time traffic condition using mobile crowdsensing [4], obtaining the real-time disaster and emergency awareness using online social media [5], and monitoring the urban air quality using reports from citizens [6]. Compared to traditional sensing paradigms that use physical sensors, social sensing has been shown to be more pervasive, scalable and economic in its applications [7]. In this paper, we focus on a *migratable traffic risk estimation* problem in intelligent transportation systems using social sensing. Our goal is to accurately estimate the traffic accident rate of locations in a *target area* where the ground truth traffic accident reports are *not available* by leveraging an estimation model from a *source area* where such data is available.

Previous efforts have been made towards addressing the traffic risk estimation problem in data mining, networked sensing, and intelligent transportation systems [8, 9, 10, 11, 12]. Those solutions primarily rely on historic ground truth labels of traffic accidents in the studied area to build reliable estimation models [13]. However, such ground truth data is not always available in many places due to various resource constraints and privacy/legal concerns [14, 15]. For example, less than 1% of cities in United States have open web portals to access their traffic accident data [1]. The traffic mon-

[1] https://www.forbes.com/sites/metabrown/2017/06/30/quick-links-to-municipal-open-data-portals-for-85-us-cities/#274a68962290

itoring devices are prohibited by law in 10 states in United States [2]. On the other hand, the rich traffic information collected by some widely deployed mobile crowdsensing applications (e.g., Waze) is privately owned by the companies and not available to the public access [16]. The lack of publicly available ground truth data on traffic accidents presents a fundamental challenge to the traffic risk estimation problem.

To address the above challenge, this paper develops a social sensing based deep transfer learning solution to estimate the traffic risk of locations in an area where the ground truth traffic accident data is not available. For example, consider the traffic risk estimation problem in two cities near our campus: South Bend and Mishawaka. The two cities are both located in the northern Indiana region with similar population densities (i.e., 2,457/sq mi vs. 2,765/sq mi) and weather conditions (i.e., long snow season). However, the city of South Bend has an open data portal for accessing traffic accident data of the city [3] motivated by its smart city initiative but Mishawaka does not. In this example, our goal is to estimate the traffic risk of locations in Mishawaka (target area) by "migrating" the traffic risk estimation model learned in South Bend (source area). Such a migratable traffic estimation problem is not trivial to solve due to several challenges elaborated below.

Discrepancy Between Source and Target Areas. A simple solution to address the migratable traffic risk estimation problem is to directly apply the estimation model learned from the source area to estimate the traffic risks in the target area. However, a major issue of this solution is that the

[2] https://www.iihs.org/iihs/topics/laws/automated_enforcement/enforcementtable?topicName=speed

[3] https://data-southbend.opendata.arcgis.com/

target and source areas may be different in many aspects (e.g., layouts, road conditions, traffic volumes, and local regulations) that would prevent a direct application of a model learned from the source area to the target area [17]. Such discrepancy between the source and target areas can potentially lead to the undesirable overfitting problems in the estimations (i.e., the model learned from the source area might be an overfitted model to estimate traffic risk in the target area) [18]. Therefore, the migrated estimation model needs to explicitly accommodate the discrepancy between different areas.

Complex and Latent Risk Features. The second challenge refers to the fact that it is extremely difficult (if possible) to identify all relevant features in the migratable traffic risk estimation problem due to the arbitrarily large feature space and the latent nature of certain features (e.g., cognitive conditions of drivers, the driving habits of an area) [9]. Furthermore, it is also challenging to decide the importance (e.g., weights) of the identified features in an estimation model when the training data set is insufficient or not available at all (e.g., target areas in our problem) [19]. Therefore, it is a not trivial task to automatically identify a critical set of complex and latent traffic risk features and effectively encode them into the migratable estimation model.

To address the above challenges, we develop DeepRisk, a social sensing based migratable traffic risk estimation scheme using deep transfer learning techniques. To address the challenge of discrepancy between source and target areas, we develop a principled deep transfer learning framework to effectively migrate the risk estimation model from the source area to the target area through a novel transformation neural network design. To address the complex and latent risk feature challenge, DeepRisk judiciously learns the traffic risk related features through a novel adversarial learning algorithm

and explicitly encodes learned features into deep transfer learning network for the model migration. To the best of our knowledge, the DeepRisk is the first deep transfer learning based approach to address the migratable traffic risk estimation problem in intelligent transportation systems. We evaluate DeepRisk on a real-world dataset from New York City. The results show that our scheme significantly outperforms the state-of-the-art baselines in terms of accurately estimating the traffic risk of locations in a city.

We choose the deep transfer learning framework to address the migratable traffic risk estimation problem in intelligent transportation systems for two main reasons. First, the source and target areas often have very different area-specific traffic risks (e.g., road conditions, traffic volumes, and speed limits), which could lead to a significant accuracy drop when a traffic risk estimation model is directly applied across areas [20]. The transfer learning solution is a natural solution for such application scenarios because it is capable of transferring the traffic risk estimation model learned from the source area to the target area by accommodating the discrepancy between the two areas using the cycle-consistent network design. Second, our deep neural network solution provides a capable data-driven solution that effectively identifies a critical set of complex and latent traffic risk features for accurate traffic estimation and effectively encodes the identified risk features using multi-layer neural representation. Our deep learning solution is in sharp contrast to the conventional estimation techniques (e.g, regression and supervised learning) which often assume well-defined traffic risk features in their models. However, such a well-defined risk feature set is not available in our problem setting due to the arbitrary large latent feature space of traffic risks.

A preliminary version of this work was published in the [21]. This pa-

per is a significant extension of the previous work in the following aspects. First, we compare our DeepRisk scheme with a new set of recent traffic risk estimation baselines and demonstrate the performance gains achieved by the DeepRisk scheme compared to the state-of-the-art solutions (Section 5). Second, we extend the evaluation in the conference paper by explicitly studying the performance of all compared schemes on traffic risk estimation across a different number of sensing cycles. The new results demonstrate the robustness of our scheme with various time windows (Section 5). Third, we extend the related work by adding a new discussion on the recent progress and challenges in urban sensing, which is closely related to the topic of this paper (Section 2). Fourth, we add a detailed justification on why deep and transfer learning techniques are adopted to address the migratable traffic risk estimation problem in the solution section (Section 4). Finally, we add a new discussion section to discuss the limitations and future work of our DeepRisk framework (Section 7).

2. Related Work

2.1. Urban Sensing

In urban sensing applications, both infrastructure-based and human sensors report the measurements about the urban environment for sustainable city monitoring and management [22, 23]. Examples of urban sensing application include real-time air quality monitoring using smart sensors [24], urban transportation demand prediction using crowdsourcing [25], and malfunctioning urban infrastructure report using geotagging [26]. Several important challenges exist in urban sensing applications. Examples include data quality and reliability [27], privacy preservation [28], energy

savings [29], and incentives design [30]. However, the migratable traffic risk estimation using social sensing data remains to be an open and challenging problem in urban sensing that has yet to be fully addressed. To goal of this paper is to accurately estimate the traffic risks at a fine-grained spatial scale in a target area using social media data. In particular, we develop a novel deep transfer learning framework to address this problem.

2.2. Social Sensing

Social sensing has emerged as a new sensing paradigm in networked sensing that uses humans as sensors to report the states of the physical world [31, 32, 7, 33]. This new sensing paradigm is a motivated by the proliferation of potable devices for individuals (e.g., smartphone), the ubiquitous wireless communication technology (e.g, 4/5G), and the mass information dissemination media (e.g., Twitter, Facebook) [1]. Examples of social sensing applications include smart urban environment and facility monitoring [6, 34], truth discovery on social media [35, 36], disaster and emergency response systems [37, 38], and intelligent transportation systems [39, 40]. This paper focuses on the migratable traffic risk estimation problem by leveraging the massive publicly available social sensing data on traffic. Compare to traditional sensing paradigms that collect traffic data from infrastructure-based sensors, social sensing provides a more pervasive, scalable, and economic approach to the problem studied in this paper.

2.3. Traffic Risk Estimation

A significant amount of efforts have been made towards addressing the traffic risk estimation problem in data mining, networked sensing, and intelligent transportation systems [41, 9, 10]. For example, Sun developed

a dynamic Bayesian network based model to predict the traffic accident risk using the data collected from speed sensors deployed on urban expressways [41]. Yuan *et al.* proposed a deep learning based model to predict the probability of traffic accidents using long-term motor vehicle crash data and traffic camera data [9]. Qin *et al.* developed a large-scale data-driven framework for urban traffic sensing using vehicle GPS and cellular signaling data [10]. However, those approaches cannot be applied to our migratable traffic risk estimation problem because they primarily rely on a rich set of historic ground truth data on traffic accidents or accurate traffic information collected from infrastructure-based physical sensors in the studied area to build reliable estimation/prediction models [14]. In contrast, we develop a novel deep transfer learning approach to estimate the traffic risk of locations in the areas where the ground truth traffic accident data is not available.

3. Problem Definition

In this section, we formulate the migratable traffic risk estimation problem using social sensing data. We first define the terms that will be used in the problem statement and then formally present our problem.

Definition 1. ***Source Area*** *(S): We define a source area to be an area (e.g., borough, district, city) where the ground truth traffic accident reports are available for traffic risk estimation.*

Definition 2. ***Target Area*** *(T): We define a target area to be the studied area where the ground truth traffic accident records are not available.*

Definition 3. ***Sensing Cell*** *(C): We divide both the source and target areas into disjoint sensing cells (e.g.,* $60\text{m} \times 60\text{m}$ *squares as shown in Figure 1) where each cell represents a subarea of interest. In particular, we*

define A and B to be the number of cells in the source and target area, respectively. In particular, we denote c_a^S as the a^{th} sensing cell in the source area ($a = 1, 2, \cdots, A$), and c_b^T as the b^{th} sensing cell in the target area ($b = 1, 2, \cdots, B$).

(a) A Traffic Circle (b) A Highway Exit (c) An Intersection

Figure 1: Examples of Sensing Cells

Definition 4. ***Sensing Cycle****: A sensing cycle is a period of time (e.g., a week) where the estimation model identifies the traffic risk in the target area. In particular, we define w to be the w^{th} sensing cycle and W to be the total number of sensing cycles in the studied duration of the application.*

Definition 5. ***Ground Truth Traffic Accident Records*** *(GT): We define GT to be the ground truth traffic accident record data, which is often published by the government authorities (e.g., police department). In our problem, the ground truth traffic accident records are only available in the source area, which we refer to as GT^S.*

Definition 6. ***Social Sensing Data*** *(SD): We define SD to be the social sensing data (e.g., social media posts) on traffic accidents. In particular, we denote social sensing data collected from the source and target areas as SD^S and SD^T, respectively.*

Definition 7. ***Traffic Risk Index*** *(Y): We define the traffic risk index of a location as the traffic accident rate (i.e., number of accidents per sensing cycle) in that location. In particular, we define* $Y_w^S = \{y_{1,w}^S, y_{2,w}^S, ..., y_{A,w}^S\}$ *and* $Y_w^T = \{y_{1,w}^T, y_{2,w}^T, ..., y_{B,w}^T\}$ *as the traffic risk indexes for source and target areas at sensing cycle* w*, respectively.* $y_{a,w}^S$ *and* $y_{b,w}^T$ *represent the traffic risk indexes of cell* c_a^S *and* c_b^T*, respectively. Finally, we define the* $\widehat{y_{b,w}^T}$ *to be the* estimated *traffic risk index for* c_b^T *in the target area.*

The goal of the migratable traffic risk estimation problem is to correctly estimate the real traffic risk index of the sensing cells in the target area by "migrating" the traffic risk estimation model learned from the source area. Using the definitions above, our problem is formally defined as:

$$\arg\min_{\widehat{y_{b,w}^T}} \left(\frac{1}{W} \cdot \sum_{w=1}^{W} \frac{1}{B} \cdot \sum_{b=1}^{B} abs(\widehat{y_{b,w}^T} - y_{b,w}^T) \mid SD^T, SD^S, Y^S\right) \tag{1}$$

where $abs()$ is function to generate the absolute value of a given number. Y^S is the set of traffic risk indexes for all sensing cells in the source area.

4. Solution

In this section, we present the DeepRisk scheme to address the migratable traffic risk estimation problem formulated in the previous section. We first present an overview of the DeepRisk scheme and then discuss its components in detail. Finally, we summarize DeepRisk using a pseudocode.

4.1. Overview of the DeepRisk scheme

The overview of DeepRisk is shown in Figure 2. It consists of three major components: i) Traffic Risk Feature Extraction (TRFE); ii) Deep Transfer Network Construction (DTNC); iii) Adversarial Deep Network Learning

(ADNL). First, the TRFE component extracts the traffic risk features from the unstructured social sensing data in both source and target areas. Second, the DTNC component constructs a principled deep transfer learning network to project the extracted traffic risk features from the target area to the source area. Finally, the ADNL component learns the optimal instance of the deep transfer learning network to make accurate traffic risk estimation in the target area using a novel adversarial learning algorithm.

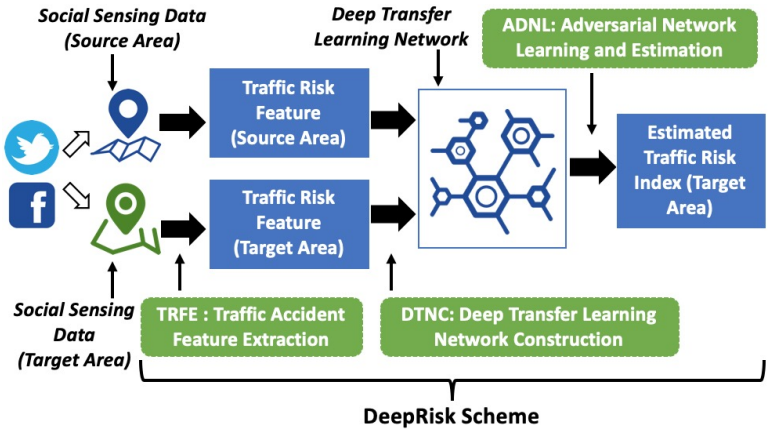


Figure 2: Overview of DeepRisk Scheme

We design the deep transfer learning framework to address the migratable traffic risk estimation problem in intelligent transportation systems for two main reasons. First, the source and target areas often have different traffic characteristics in many aspects that prevent a direct application of a model learned from the source to the target area. Our transfer learning network addresses this problem by explicitly transferring the extracted traffic risk features from the target to source area and avoid the undesirable overfitting problem in the estimation. Second, it is challenging to automatically identify all relevant traffic risk features and decide the importance of

identified features given the arbitrarily large feature space. However, such a problem can be effectively addressed by leveraging the multi-layer neural representation design in the DeepRisk framework.

4.2. Traffic Risk Feature Extraction (TRFE)

In this subsection, we describe the TRFE component that extracts the traffic accident related features (i.e., locations and time of accidents) from the unstructured social sensing data collected in both source and target areas for the migratable traffic risk estimation.

The raw social sensing data (e.g., tweets) are usually unstructured (e.g., text) and can not be directly used for traffic risk estimation [42]. For instance, human sensors often provide a high-level description of the accident location instead of the accurate GPS location (e.g., a tweet saying "Incident on #I278 EB at Hunts Point Avenue"). Therefore, the objective of TRFE is to extract the traffic risk features (i.e., locations and time of accidents) and represent the extracted features as feature vectors. In particular, we characterize the traffic risk feature extraction process as follows:

$$X_c^{S/T} = \{(\alpha_p, \beta_p) | \alpha_p \in c, \forall p \in SD^{S/T}\}, \text{ where } (\alpha_p, \beta_p) = \mathcal{F}(p), \forall c \in C^{S/T} \tag{2}$$

where $X_c^{S/T}$ is the traffic risk feature vector to represent the extracted traffic risk features in sensing cell c from either source area or target area $C^{S/T}$. In addition, we define X^S and X^T to be the sets of traffic risk feature vectors in source area and target area, respectively. p is a piece of social sensing data (e.g., a tweet) collected from source area SD^S or target area SD^T. α_p and β_p are location and time of the accident reported in p. $\mathcal{F}$ is a feature extraction function that extracts the traffic risk features from the raw social sensing

data. In particular, traffic accident locations α_p can be extracted from the social sensing data by examining the associated geo-tags (e.g., "coordinates" field of a tweet [4]) or utilizing the traffic accident location extraction tools that analyze the content of social sensing data as described in [43]. The accident time β_p can be extracted by checking the timestamp of the data sample (e.g., "created_at" field of a tweet).

4.3. Deep Transfer Network Construction (DTNC)

Section 4.3: In this subsection, we present the DTNC component that migrates the risk estimation model from the source area to the target area through a deep transfer learning network inspired by the cycle-consistent network design (e.g., Cycle-GAN, CyCADA) from the computer vision community [17, 44]. The DTNC component takes the traffic risk feature vectors generated by the TRFE component as inputs and constructs a set of neural networks for the traffic risk estimation in the target area. Different from the CycleGan/CyCADA that focuses on unpaired image-to-image translation tasks using convolutional operations, our deep transfer learning network utilizes the fully-connected layers to judiciously project the extracted traffic risk features from the target area to the source area for effective model migration and incorporates an estimation network to simultaneously obtain an effective traffic risk estimation model from the source area. An overview architecture of the deep transfer network is shown in Figure 3. In general, the deep transfer network design consists of three types of neural network architectures using fully-connected layers: estimation neural network (E_0), transformation neural network (F_1 and F_2), and examination neural network

[4] https://developer.twitter.com/en/docs/tutorials/filtering-tweets-by-location.html

(G_T and G_S). In general, E_0, F_1, F_2, G_T, and G_S work collaboratively to learn a migratable traffic risk estimation model for the target area. In particular, the E_0 is first used to learn an effective traffic risk estimation model in the source area. Then F_1 and F_2 work collaboratively to transfer the estimation model learned by E_0 from source to target area. In particular, F_1 and F_2 effectively transform the traffic risk features from the target area to fit in with the traffic risk feature distribution of the source area to avoid the undesirable overfitting problem given the insufficient ground-truth traffic accident records in the target area. Finally, the G_T and G_S are used to regulate the F_1 and F_2 to generated the transformed feature vectors that fit in well with the traffic risk feature distributions of the source area to ensure the desirable estimation accuracy of the migratable traffic risk estimation model.

We first define three types of neural networks that will be used in our solution.

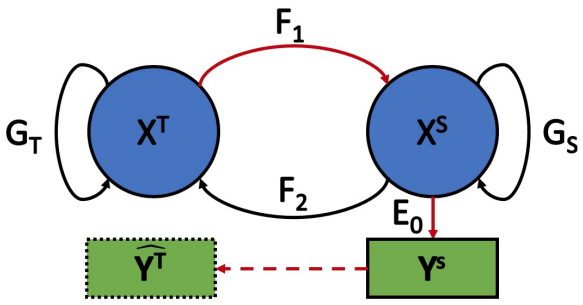


Figure 3: Deep Transfer Learning Network

Definition 8. ***Estimation Neural Network*** *E: We define E as the estimation neural network:*

$$E : \mathcal{M} \rightarrow \mathcal{N} \tag{3}$$

where $\mathcal{M}$ is the set of input features (e.g., traffic risk features) and $\mathcal{N}$ is the estimation results (e.g., traffic risk index).

Definition 9. ***Transformation Neural Network*** *F: We define F as the transformation neural network:*

$$F : m \in \mathcal{P}(\mathcal{M}) \rightarrow n \in \mathcal{P}(\mathcal{N}) \tag{4}$$

where $\mathcal{P}(\mathcal{M})$ and $\mathcal{P}(\mathcal{N})$ are the distributions of $\mathcal{M}$ and $\mathcal{N}$ defined above, respectively. In particular, the transformation neural network F is able to map a data point m from the distribution $\mathcal{P}(\mathcal{M})$ to a unique data point n from the distribution $\mathcal{P}(\mathcal{N})$.

Definition 10. ***Examination Neural Network*** *G: We define G as the examination neural network:*

$$G : \mathbf{1} : \{m \in \mathcal{P}(\mathcal{M})\}, \mathbf{0} : \{m \notin \mathcal{P}(\mathcal{M})\} \tag{5}$$

The network returns "1" (i.e., true) if m belongs to the distribution $\mathcal{P}(\mathcal{M})$ and "0" (i.e., false) otherwise.

The first part of our deep transfer learning network is to instantiate an estimation neural network for the traffic risk estimation in the source area (i.e., E_0) as follows:

$$E_0 : X^S \rightarrow Y^S \tag{6}$$

where X^S is the set of traffic risk feature vectors generated by the TRFE component in the source area. Y^S is the set of traffic risk indexes (i.e., traffic accident rate) in the source area, which can be obtained from the ground

truth traffic accident records GT^S. The objective of E_0 is to minimize the difference between the estimated and real traffic risk indexes in the source area as follows:

$$\mathcal{L}_E : \arg\min_{E_0} ||E_0(X^S) - Y^S||_2 \tag{7}$$

where $\mathcal{L}_E$ is the objective for estimation network E_0. $E_0(X^S)$ and Y^S are the estimated and real traffic risk indexes of the source area, respectively. $||\cdot||_2$ donates the L2-norm of a given matrix [45]. We develop an efficient adversarial deep network learning scheme to learn the optimal instance (i.e., optimal weight of each node in the network) of E_0, which will be discussed in the next subsection.

The input data to the learned estimation network E_0 has to share the same underlying distribution as the training data (i.e., X^S) to avoid the undesirable overfitting problem [46]. However, the discrepancy between the source and target area naturally lead to different traffic risk distributions between the two areas [17]. Please note that it is not feasible to directly modify E_0 (e.g., re-tuning the weights of the network, etc.) to fit the traffic risk distributions of the target area due to the complex and latent nature of the risk features as well as the insufficient ground truth traffic accident records in the target area [19]. Therefore, we transform the traffic risk features from the target area (i.e., (X^T)) to fit in with the traffic risk feature distribution of the source area (i.e., $\mathcal{P}(X^S)$) using the transformation networks:

$$F_1 : x^T \in \mathcal{P}(X^T) \rightarrow x^S \in \mathcal{P}(X^S), F_2 : x^S \in \mathcal{P}(X^S) \rightarrow x^T \in \mathcal{P}(X^T) \tag{8}$$

where $\mathcal{P}(X^T)$ and $\mathcal{P}(X^S)$ are the traffic risk feature distributions in target area and source area, respectively. In particular, we define two transfor-

mation networks F_1 and F_2 to ensure the consistency of the transformation process (e.g., $x^T = F_2(F_1(x^T))$, $x^S = F_1(F_2(x^S))$) as follows:

$$\mathcal{L}_F : \arg\min_{F1,F2} \sum_{x^T \in \mathcal{P}(X^T)} (||F_2(F_1(x^T)) - x^T||_1) + \sum_{x^S \in \mathcal{P}(X^S)} (||F_1(F_2(x^S)) - x^S||_1) \tag{9}$$

where $\mathcal{L}_F$ is the objective for transformation networks F_1 and F_2. $||\cdot||_1$ donates the L1-norm of a given matrix [45].

The above objective function ensures the one-to-one transformation between the original and transformed feature vectors (e.g., x^T and $F_1(x^T)$). However, the transformed feature vector (e.g., $F_1(x^T)$) has not been regularized to the desired distribution (e.g., $\mathcal{P}(X^S)$). To regularize the transformed feature vectors, we introduce two examination neural networks G_S and G_T for the traffic risk feature distributions in source and target areas (i.e., $\mathcal{P}(X^S)$ and $\mathcal{P}(X^T)$) as:

$$\begin{aligned} G_S &: \mathbf{1} : \{x \in \mathcal{P}(X^S)\}, \mathbf{0} : \{x \notin \mathcal{P}(X^S)\} \\ G_T &: \mathbf{1} : \{x \in \mathcal{P}(X^T)\}, \mathbf{0} : \{x \notin \mathcal{P}(X^T)\} \end{aligned} \tag{10}$$

Such a loss function design for the symmetric examination neural networks is to ensure G_S and G_T can accurately identify if the traffic risk feature vector fits in with the underlying traffic risk feature distribution of source and target area, respectively. Our goal is to ensure the G_S and G_T can be explicitly used to regulate the $F1$ and $F2$ in establishing effective model migration between the source and target area.

Consider a traffic risk feature vector x^T and its transformed feature vector $F_1(x^T)$ generated by the transformation network F_1. The transformation network F_1 is effective if $F_1(x^T)$ belongs to $\mathcal{P}(X^S)$. In particular, the transformed feature vector $F_1(x^T)$ should be verified by the examination network

G_S (i.e., returning 1). We update the objective for F_1 and F_2 by leveraging the examination networks as follows:

$$\mathcal{L}_F : \arg\min_{F1,F2} \sum_{x^T \in \mathcal{P}(X^T)} (||F_2(F_1(x^T)) - x^T||_1) + \sum_{x^S \in \mathcal{P}(X^S)} (||F_1(F_2(x^S)) - x^S||_1) + \sum_{x^T \in \mathcal{P}(X^T)} ||\mathbf{1} - G_S(F_1(x^T))||_E + \sum_{x^S \in \mathcal{P}(X^S)} ||\mathbf{1} - G_T(F_2(x^S))||_E \tag{11}$$

where $||\cdot||_E$ donates the cross-entropy loss of a given matrix [47]. $\mathcal{L}_F$ is the final objective for transformation network F_1 and F_2. The above process requires examination networks to clearly i) identify imperfect transformed feature vectors (e.g., $F_1(x^T)$) that do not fit in with the desired distributions (e.g., $\mathcal{P}(X^S)$); and ii) differentiate identified feature vectors from the original ones (e.g., x^S). Hence, we define the objective for examination networks G_S and G_T as follows:

$$\mathcal{L}_G : \arg\min_{G_S,G_T} \sum_{x^T \in \mathcal{P}(X^T)} ||\mathbf{0} - G_S(F_1(x^T))||_E + \sum_{x^S \in \mathcal{P}(X^S)} ||\mathbf{1} - G_S(x^S)||_E + \sum_{x^S \in \mathcal{P}(X^S)} ||\mathbf{0} - G_T(F_2(x^S))||_E + \sum_{x^T \in \mathcal{P}(X^T)} ||\mathbf{1} - G_T(x^T)||_E \tag{12}$$

where $||\cdot||_E$ donates the cross-entropy loss of a given matrix.

4.4. Adversarial Deep Network Learning (ADNL))

In this subsection, we describe the Adversarial Deep Network Learning (ADNL) component. The goal of ADNL is to obtain the optimal instances (e.g., optimal weights of nodes in the network) of the neural networks (i.e., E_0, F_1, F_2, G_S, G_T) in DTNC component to make accurate traffic risk estimation for the target area.

We formulate the problem of learning the optimal deep transfer learning network as an adversarial learning problem. In particular, we observe that the transformation neural networks (i.e, F_1, F_2) and examination neural networks (i.e, G_S, G_T) formulated in the DTNC component have opposite objectives (i.e., $\mathcal{L}_F$ in Equation 11 and $\mathcal{L}_G$ in Equation 12). On one hand, the transformation networks target at effective transformations so the examination networks believe all transformed feature vectors belong to the desired distributions. On the other hand, the examination networks target at identification of imperfect feature vectors generated by transformation networks. To capture such conflicting objectives, we divide the neural networks into two adversarial groups as follows:

$$\begin{aligned} \mathcal{L}_A &: \arg\min_{E_0,F_1,F_2}(\gamma_E \cdot \mathcal{L}_E + \gamma_F \cdot \mathcal{L}_F) \\ \mathcal{L}_B &: \arg\min_{G_S,G_T}(\gamma_G \cdot \mathcal{L}_G) \end{aligned} \tag{13}$$

where $\mathcal{L}_A$ and $\mathcal{L}_B$ represent the objectives for the two adversarial groups, respectively. In particular, $\mathcal{L}_A$ includes the networks (E_0, F_1, F_2) that aim to make effective feature transformations. $\mathcal{L}_B$ includes the networks (G_S, G_T) that aim to identify any imperfect transformed feature vectors (e.g., $F_1(x^T)$). $\gamma_E, \gamma_F, \gamma_G$ are the weights for different neural networks, which are usually set to be a small positive number with the order of $\gamma_F > \gamma_G > \gamma_E$ to ensure the consistency and effectiveness of the feature transformation process [17]. Given the above adversarial objective functions, we can alternatively minimize $\mathcal{L}_A$ and $\mathcal{L}_B$ using gradient descent techniques (e.g., Adam optimization [48]) to obtain the optimal instances of all neural networks, which ensure the effective feature transformation through F_1 and accurate traffic risk estimation through E_0. Finally, we apply the optimal neural

networks to estimate the traffic risk indexes in the target area as follows:

$$x^T \rightarrow F_1^*(x^T) \rightarrow E_0^*(F_1^*(x^T)) \rightarrow \widehat{y^T}, \forall x^T \in X^T \tag{14}$$

where E_0^* and F_1^* are the optimal instances of E_0 and F_1, respectively. In the above estimation, we first transform the feature vector x^T from the target area to fit in with the distribution of the source area as the transformed feature vector $F_1^*(x^T)$. We then apply the estimation neural network E_0^* to obtain the traffic risk estimation result $E_0^*(F_1^*(x^T))$ as the estimated traffic risk index $\widehat{y^T}$ for the feature vector x^T.

4.5. Summary of the DeepRisk Scheme

We summarize the DeepRisk scheme in Algorithm 1. The inputs to the algorithm are the social sensing data collected from both source and target areas (SD^S, SD^T) and traffic risk indexes in source area Y^S. The output is the estimated traffic accident rate for each sensing cell in target area $\widehat{y^T}$.

5. Evaluation

In this section, we evaluate the performance of the DeepRisk scheme using the real world traffic datasets collected from New York City. We compare the performance of DeepRisk with state-of-the-art traffic risk estimation baselines in the literature. The evaluation results show that DeepRisk significantly outperforms the baselines in terms of the estimation accuracy.

[5] threshold is usually set to be a small value (i.e., less than 0.1) to ensure the accuracy of the learned neural networks

Algorithm 1 Summary of the DeepRisk Scheme

initialize F_1,F_2,G_S,G_T,E_0
set γ_E, γ_F, γ_G
extract X^S and X^T from SD^S and SD^T using TRFE component and obtain Y^S from GT^S
calculate initial $\mathcal{L}_E$, $\mathcal{L}_F$, $\mathcal{L}_G$ using Equation 7, Equation 11, Equation 12, respectively
calculate initial $\mathcal{L}_A$ and $\mathcal{L}_B$ using Equation 13
set $\Delta_{\mathcal{L}_A} \leftarrow +\infty$ and $\Delta_{\mathcal{L}_B} \leftarrow +\infty$
while $\Delta_{\mathcal{L}_A}$ or $\Delta_{\mathcal{L}_B} \geq$ *threshold*[5] **do**
    optimize $\mathcal{L}_A$ using Adam optimizer
    calculate updated $\mathcal{L}_E$, $\mathcal{L}_F$ using Equation 7, Equation 11 as $\mathcal{L}'_E$, $\mathcal{L}'_F$, respectively
    calculate updated $\mathcal{L}_A$ using Equation 13 as $\mathcal{L}'_A$
    $\Delta_{\mathcal{L}_A} \leftarrow |\mathcal{L}'_A - \mathcal{L}_A|$
    $\mathcal{L}_E \leftarrow \mathcal{L}'_E$, $\mathcal{L}_F \leftarrow \mathcal{L}'_F$,$\mathcal{L}_A \leftarrow \mathcal{L}'_A$
    optimize $\mathcal{L}_B$ using Adam optimizer
    calculate updated $\mathcal{L}_G$ using Equation 12 as $\mathcal{L}'_G$
    calculate updated $\mathcal{L}_B$ using Equation 13 as $\mathcal{L}'_B$
    $\Delta_{\mathcal{L}_B} \leftarrow |\mathcal{L}'_B - \mathcal{L}_B|$
    $\mathcal{L}_G \leftarrow \mathcal{L}'_G$,$\mathcal{L}_B \leftarrow \mathcal{L}'_B$
end while
for each sensing cell in target area **do**
    output $\widehat{y^T}$ with optimized neural network F_1^*, E_0^* using Equation 14
end for

5.1. Dataset

We study the migratable traffic risk estimation problem using the real world traffic datasets collected from four different boroughs in New York City (i.e., Manhattan, Bronx, Queens, and Brooklyn). These boroughs are observed to have different layouts, road conditions, traffic volumes, and population density, which create a challenge scenario for the DeepRisk scheme. We choose the four boroughs in our evaluation because we can obtain the detailed ground truth labels on traffic accidents in these boroughs as discussed below. We use the ground truth labels in the target area for the evaluation purpose only.

Twitter Traffic Report Dataset: We collected a dataset by using the Twitter API on traffic accidents as our *social sensing data*. This dataset consists of 239,734 traffic-related tweets from the four different boroughs (i.e., Manhattan, Bronx, Queens, and Brooklyn) in New York City over the time period from Jan. 1st, 2016 to Jun. 30th, 2018.

Motor Vehicle Accident Report Dataset: We use a dataset published by New York City Police Department (NYPD) [6] to obtain the ground truth labels of traffic risk index in the studied areas. This dataset consists of 568,051 reports of traffic accidents that happened in the four major boroughs in New York City between Jan. 1st, 2016 and June 30th, 2018. Each report contains the accurate time and location of an accident.

[6] https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95

5.2. Baselines

We choose several representative traffic risk estimation schemes as the baselines in our evaluations. To ensure fairness, the inputs to the baselines and DeepRisk are the same (i.e., social sensing data from both source and target areas and the accident reports from the source area).

- **Gaussian Process (Gaussian)**: it is a generic supervised learning method that utilizes *Gaussian Processes* to minimize the difference between the *true* and *estimated* traffic risk indexes [49].
- **Ridge Regression (Ridge)**: it adds a ridge regularizer on top of the linear regression model to avoid the underlying multicollinearity problem in the traffic risk estimation [50].
- **Stochastic Gradient Descent (SGD)**: it leverages an optimization process driven by the stochastic gradient descent to obtain an optimized estimation model to infer accurate traffic accident risk [51].
- **Estimation Neural Network (ENN)**: it trains a multi-layer neural network model with the same network architecture (i.e., layers and nodes) and activation function (i.e., ReLU) as the estimation network E_0 (defined in Definition 6). The network learns a non-linear function approximator to minimize the difference between the *true* and *estimated* traffic risk indexes [52].

5.3. Evaluation Metrics

In our evaluation, we define the following metrics to evaluate the performance of all compared schemes.

- **Mean Absolute Error (MAE)**: We define the MAE for all sensing cells across all sensing cycles as follows:

$$MAE = \frac{1}{W} \cdot \sum_{w=1}^{W} \frac{1}{B} \cdot \sum_{b=1}^{B} abs(\widehat{y_{b,w}^{T}} - y_{b,w}^{T}) \tag{15}$$

where B is the number of the sensing cells in the target area and W is number of the sensing cycles.

- **Root Mean Square Error (RMSE)**: We define the RSME for all sensing cells across all sensing cycles as follows:

$$RMSE = \sqrt{\frac{1}{W} \cdot \sum_{w=1}^{W} \frac{1}{B} \cdot \sum_{b=1}^{B} (\widehat{y_{b,w}^{T}} - y_{b,w}^{T})^2} \tag{16}$$

5.4. Evaluation Results

5.4.1. Estimation Accuracy

In this subsection, we present the results of our DeepRisk scheme and all compared baselines on the real world datasets discussed above. We study the performance of all schemes by varying the combinations of *source* and *target* areas. Specifically, in each set of the experiment, we set the source area to be one of the four boroughs in NYC and set each of the remaining three boroughs as the target area. In our experiment, we set the length of the sensing cycle to be one week by considering the frequency of the accidents in the studied area. In our experiments, we vary the number of sensing cycles from 20 to 40 in each source and target area combination to establish a comprehensive evaluation of the performance for all compared schemes. In addition, we set the structures of the neural networks in our evaluation for

network F_1, F_2, E_0 to be 10 layers with 64 nodes in each layer for complex multi-variable generation tasks. We set G_S, G_T to be 2 layers with 3 nodes in each layer for the simple binary classification tasks by following the training procedures of adversarial networks proposed in [53]. The reasons for such hyperparameter settings are multi-fold. First, for F_1, F_2, E_0, we choose a relatively larger number of nodes and layers in F_1, F_2, E_0 to ensure the desirable accuracy of the learned migratable traffic risk estimation model while preventing the undesirable overfitting issue of the learned models. Second, we choose a relatively smaller number of layers and nodes for G_S, G_T to achieve an effective adversarial training of the transformation networks F_1, F_2 and the examination networks G_S, G_T in our adversarial learning framework. In particular, compared to F_1, F_2 that focus on the complex generative task of traffic risk feature transformation, G_S, G_T are designed to handle the less complex binary classification tasks to identify imperfect transformed traffic risk feature vectors. Hence, we reduce the number of nodes and layers in G_S, G_T to keep the examination networks from becoming too dominant to prevent us from learning an optimized F_1, F_2 during the adversarial training process. We also set the activation function in each network to be ReLU function for effective transformation between layers [54].

The results are shown in Table 1 to Table 8. In Table 1, we observe that the performance gains achieved by DeepRisk scheme compared to the best performed baseline are 25.8%, 18.1%, 28.2% respectively in MAE on three different target areas with 40 sensing cycles. Such performance gains of DeepRisk are achieved by judiciously constructing a principled deep transfer network to effectively migrate the risk estimation model from the source area to the target area. In particular, the DeepRisk scheme explicitly transforms the traffic risk features from the target area to fit in with the feature

distribution of source area and obtains the optimal risk estimation function through an adversarial deep network learning process. In addition, we continue to observe that our DeepRisk scheme to consistently outperform the compared baselines in Table 3 to Table 8 when we shuffle the source and target area combinations. We also observe that our DeepRisk scheme outperforms all compared baselines across different number of sensing cycles. The results demonstrate the robustness of DeepRisk with various time windows in real-world traffic risk estimation applications.

Table 1: Performance Comparisons (MAE) on Different Source and Target Area Settings (Source = Bronx)

		Bronx → Queens			Bronx → Manhattan			Bronx → Brooklyn		
Category	Algorithm	W=20	W=30	W=40	W=20	W=30	W=40	W=20	W=30	W=40
Est-Alg	Gaussian	1.553	1.546	1.538	1.478	1.430	1.428	1.569	1.513	1.528
	Ridge	1.590	1.520	1.531	1.338	1.289	1.292	1.392	1.399	1.405
	SGD	1.435	1.410	1.403	1.495	1.439	1.468	1.272	1.340	1.359
	ENN	1.398	1.293	1.336	1.253	1.216	1.224	1.195	1.255	1.272
Ours	**DeepRisk**	**1.034**	**1.054**	**1.062**	**1.022**	**1.032**	**1.036**	**1.008**	**0.975**	**0.992**

Table 2: Performance Comparisons (RMSE) on Different Source and Target Area Settings (Source = Bronx)

		Bronx → Queens			Bronx → Manhattan			Bronx → Brooklyn		
Category	Algorithm	W=20	W=30	W=40	W=20	W=30	W=40	W=20	W=30	W=40
Est-Alg	Gaussian	1.931	1.962	1.944	1.844	1.792	1.810	1.987	1.917	1.921
	Ridge	2.212	2.214	2.189	1.843	1.773	1.804	1.936	1.951	1.919
	SGD	1.921	1.880	1.880	2.008	1.958	1.993	1.597	1.713	1.736
	ENN	1.916	1.781	1.834	1.739	1.753	1.738	1.489	1.526	1.562
Ours	**DeepRisk**	**1.274**	**1.303**	**1.315**	**1.255**	**1.257**	**1.276**	**1.191**	**1.209**	**1.223**

Table 3: Performance Comparisons (MAE) on Different Source and Target Area Settings (Source = Manhattan)

		Manhattan → Bronx			Manhattan → Queens			Manhattan → Brooklyn		
Category	Algorithm	W=20	W=30	W=40	W=20	W=30	W=40	W=20	W=30	W=40
EST-Alg	Gaussian	2.123	2.084	2.081	1.464	1.466	1.446	1.408	1.442	1.448
	Ridge	2.659	2.781	2.729	1.493	1.534	1.562	1.408	1.502	1.479
	SGD	2.234	2.593	2.714	1.386	1.421	1.431	1.501	1.355	1.386
	ENN	2.225	2.238	2.266	1.341	1.346	1.369	1.263	1.206	1.237
Ours	**DeepRisk**	**1.090**	**1.096**	**1.118**	**0.969**	**1.006**	**1.019**	**1.007**	**0.938**	**0.966**

Table 4: Performance Comparisons (RMSE) on Different Source and Target Area Settings (Source = Manhattan)

		Manhattan → Bronx			Manhattan → Queens			Manhattan → Brooklyn		
Category	Algorithm	W=20	W=30	W=40	W=20	W=30	W=40	W=20	W=30	W=40
Est-Alg	Gaussian	2.679	2.631	2.613	1.897	1.896	1.861	1.812	1.871	1.861
	Ridge	4.227	4.480	4.386	2.192	2.237	2.267	1.971	2.334	2.238
	SGD	3.832	4.140	4.320	1.914	1.939	1.946	2.006	1.758	1.826
	ENN	3.440	3.589	3.669	1.796	1.838	1.838	1.704	1.698	1.685
Ours	**DeepRisk**	**1.492**	**1.487**	**1.484**	**1.248**	**1.284**	**1.295**	**1.262**	**1.218**	**1.229**

Table 5: Performance Comparisons (MAE) on Different Source and Target Area Settings (Source = Queens)

		Queens → Bronx			Queens → Manhattan			Queens → Brooklyn		
Category	Algorithm	W=20	W=30	W=40	W=20	W=30	W=40	W=20	W=30	W=40
Est-Alg	Gaussian	2.123	2.109	2.109	1.343	1.361	1.370	1.568	1.487	1.514
	Ridge	3.272	3.101	2.960	1.247	1.263	1.268	1.510	1.433	1.459
	SGD	2.197	2.283	2.252	1.322	1.269	1.309	1.305	1.326	1.319
	ENN	2.141	2.091	2.053	1.200	1.185	1.207	1.244	1.262	1.171
Ours	**DeepRisk**	**1.091**	**1.092**	**1.110**	**1.023**	**1.024**	**1.029**	**1.032**	**1.021**	**0.949**

5.4.2. Deviation and Significance Studies of DeepRisk

Finally, we conduct the deviation and significance study of our DeepRisk scheme by performing the widely adopted ANOVA test on the estimation errors generated by all compared schemes. The results of the ANOVA test

Table 6: Performance Comparisons (RMSE) on Different Source and Target Area Settings (Source = Queens)

		Queens → Bronx			Queens → Manhattan			Queens → Brooklyn		
Category	Algorithm	W=20	W=30	W=40	W=20	W=30	W=40	W=20	W=30	W=40
Est-Alg	Gaussian	2.647	2.630	2.624	1.731	1.736	1.755	2.037	1.912	1.945
	Ridge	5.677	5.323	5.012	1.779	1.827	1.819	2.281	2.086	2.164
	SGD	3.205	3.329	3.311	1.999	1.914	2.009	1.810	1.739	1.920
	ENN	3.756	3.542	3.467	1.644	1.609	1.647	1.585	1.723	1.732
Ours	**DeepRisk**	**1.482**	**1.474**	**1.473**	**1.274**	**1.265**	**1.283**	**1.353**	**1.317**	**1.217**

Table 7: Performance Comparisons (MAE) on Different Source and Target Area Settings (Source = Brooklyn)

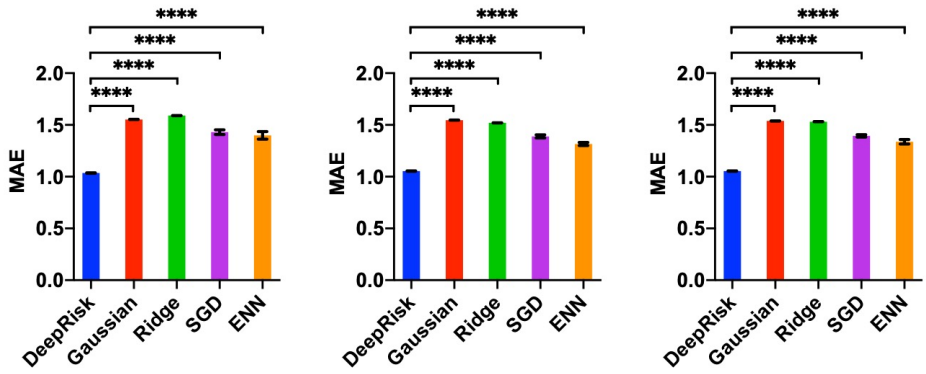
		Brooklyn → Queens			Brooklyn → Manhattan			Brooklyn → Bronx		
Category	Algorithm	W=20	W=30	W=40	W=20	W=30	W=40	W=20	W=30	W=40
Est-Alg	Gaussian	1.378	1.330	1.346	1.388	1.363	1.331	2.242	2.127	2.165
	Ridge	1.325	1.228	1.260	1.321	1.264	1.251	2.571	2.421	2.471
	SGD	1.412	1.207	1.275	1.289	1.273	1.247	2.741	2.602	2.648
	ENN	1.224	1.133	1.163	1.223	1.206	1.191	2.467	2.430	2.300
Ours	**DeepRisk**	**1.042**	**1.045**	**1.016**	**1.003**	**0.999**	**0.991**	**1.131**	**1.146**	**1.124**

Table 8: Performance Comparisons (RMSE) on Different Source and Target Area Settings (Source = Brooklyn)

		Brooklyn → Queens			Brooklyn → Manhattan			Brooklyn → Bronx		
Category	Algorithm	W=20	W=30	W=40	W=20	W=30	W=40	W=20	W=30	W=40
Est-Alg	Gaussian	1.884	1.867	1.916	1.771	1.739	1.705	2.669	2.656	2.632
	Ridge	2.216	2.212	2.179	2.236	2.030	1.991	4.052	3.854	4.032
	SGD	2.039	2.050	1.969	1.945	1.921	1.867	4.025	3.844	3.937
	ENN	2.236	2.191	2.036	1.884	1.855	1.836	3.969	4.086	3.891
Ours	**DeepRisk**	**1.245**	**1.285**	**1.302**	**1.230**	**1.227**	**1.244**	**1.478**	**1.465**	**1.467**

are shown in Figure 4. Given the space limit, we only present the ANOVA figures on Bronx→Queens. The results of other scenarios are similar. We observe that the performance gains achieved by DeepRisk are statistically significant in comparison with all baselines. In addition, we also observe

the standard deviation of estimation errors generated by our scheme is very small, which indicate its stable performance. The results further validate that our DeepRisk scheme clearly outperforms all compared baselines in accurately and reliably estimating the traffic risk.



(a) W=20 (b) W=30 (c) W=40

"****" in above figures indicates extremely statistical significance (i.e,. P-Value < 0.0001) between two compared errors generated by the ANOVA Test

Figure 4: Deviation and Significance Studies of DeepRisk

6. Conclusion

In this paper, we develop a DeepRisk scheme to solve the migratable traffic risk estimation problem in intelligent transportation system using social sensing. The DeepRisk addresses two critical challenges, namely discrepancy between source and target areas and complex and latent risk features. In particular, we develop a principled deep transfer learning network to effectively migrate the risk estimation model from the source area to the target area. We also design a novel adversarial learning algorithm to learn the traf-

fic risk related features for effective risk estimation model migration. The evaluation results on the real world case study demonstrate that the DeepRisk achieves significant performance gains compared to the state-of-the-art baselines in accurately estimating the traffic risk of locations in a city.

7. Discussion and Future Work

First, while our DeepRisk scheme can effectively identify the reported locations of traffic accidents from unstructured and noisy online social media data, it can also make some mistakes. For example, a tweet reports an accident in NYC as "minor accident and stopped traffic near 30th St". In this example, we cannot accurately geo-locate the accident because multiple boroughs (e.g., Manhattan, Brooklyn) in New York city have streets with the same name as 30th St. To address this challenge, we plan to explore the context information of the social media users [55]. For instance, the knowledge of a user's profile information can help us better identify the correct traffic accident locations reported by that user (e.g., the traffic accident is more likely to happen in 30th St Brooklyn if the user lives in Brooklyn area in the above example).

Second, our DeepRisk scheme provides accurate traffic risk estimation results at a fine-grained spatial granularity. Such results can be used to alert people (e.g., drivers, pedestrians) to avoid the high traffic risk locations and take the alternative travel routes with low risks. However, we also observe that the decisions people make (e.g., changing their route) will in turn change the traffic conditions in the transportation systems, and consequently affect the estimation results of DeepRisk. We plan to leverage control theory and feedback systems [56] to model the dynamic interaction between the

DeepRisk and the relevant changes in the transportation systems. One important challenge here is how to model people's response to the results from DeepRisk and evaluate the closed-loop system in a realistic scenario.

Third, we evaluate our DeepRisk model on a real world dataset from New York City. It will also be interesting to further investigate the robustness of our model across cities with different scales, population densities, and road conditions. Moreover, our current solution only utilizes the online social media data to accomplish the traffic risk estimation task. However, it is possible to further improve its performance by explicitly considering external or background knowledge that is publicly available (e.g., social events, road construction alerts, and weather reports).

Acknowledgement

This research is supported in part by the National Science Foundation under Grant No. CNS-1845639, CNS-1831669, CBET-1637251, Army Research Office under Grant W911NF-17-1-0409. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] D. Wang, T. Abdelzaher, L. Kaplan, Social sensing: building reliable systems on unreliable data, Morgan Kaufmann, 2015 (2015).

[2] D. Wang, D. Zhang, Y. Zhang, M. T. Rashid, L. Shang, N. Wei, Social edge intelligence: Integrating human and artificial intelligence at the edge, in: 2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI), IEEE, 2019, pp. 194–201 (2019).

[3] D. Zhang, N. Vance, D. Wang, When social sensing meets edge computing: Vision and challenges, in: 2019 28th International Conference on Computer Communication and Networks (ICCCN), IEEE, 2019, pp. 1–9 (2019).

[4] J. Wan, J. Liu, Z. Shao, A. V. Vasilakos, M. Imran, K. Zhou, Mobile crowd sensing for traffic prediction in internet of vehicles, Sensors 16 (1) (2016) 88 (2016).

[5] M. T. Rashid, D. Zhang, Z. Liu, H. Lin, D. Wang, Collabdrone: A collaborative spatiotemporal-aware drone sensing system driven by social sensing signals, in: 2019 28th International Conference on Computer Communication and Networks (ICCCN), IEEE, 2019, pp. 1–9 (2019).

[6] L. Wang, D. Zhang, A. Pathak, C. Chen, H. Xiong, D. Yang, Y. Wang, Ccs-ta: Quality-guaranteed online task allocation in compressive crowdsensing, in: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2015, pp. 683–694 (2015).

[7] D. Wang, B. K. Szymanski, T. Abdelzaher, H. Ji, L. Kaplan, The age of social sensing, Computer 52 (1) (2019) 36–45 (2019).

[8] R. Yu, M. Abdel-Aty, Analyzing crash injury severity for a mountainous

freeway incorporating real-time traffic and weather data, Safety science 63 (2014) 50–56 (2014).

[9] Z. Yuan, X. Zhou, T. Yang, Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, 2018, pp. 984–992 (2018).

[10] Z. Qin, Z. Fang, Y. Liu, C. Tan, W. Chang, D. Zhang, Eximius: A measurement framework for explicit and implicit urban traffic sensing, in: Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems, ACM, 2018, pp. 1–14 (2018).

[11] L. Lin, Q. Wang, A. W. Sadek, A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction, Transportation Research Part C: Emerging Technologies 55 (2015) 444–459 (2015).

[12] Y. Zhang, X. Dong, L. Shang, D. Zhang, D. Wang, A multi-modal graph neural network approach to traffic risk forecasting in smart urban sensing, in: international conference on sensing, communication, and networking (SECON)(IEEE, 2020), 2020 (2020).

[13] A. Najjar, S. Kaneko, Y. Miyanaga, Combining satellite imagery and open data to map road safety., in: AAAI, 2017, pp. 4524–4530 (2017).

[14] Q. Chen, X. Song, H. Yamada, R. Shibasaki, Learning deep representation from big and heterogeneous data for traffic accident inference., in: AAAI, 2016 (2016).

[15] J. Schiff, M. Meingast, D. K. Mulligan, S. Sastry, K. Goldberg, Respectful cameras: Detecting visual markers in real-time to address privacy concerns, in: Protecting Privacy in Video Surveillance, Springer, 2009, pp. 65–89 (2009).

[16] Y. Gu, Z. S. Qian, F. Chen, From twitter to detector: Real-time traffic incident detection using social media data, Transportation research part C: emerging technologies 67 (2016) 321–342 (2016).

[17] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks.

[18] C. Schaffer, Overfitting avoidance as bias, Machine learning 10 (2) (1993) 153–178 (1993).

[19] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, Journal of machine learning research 3 (Mar) (2003) 1157–1182 (2003).

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

[21] Y. Zhang, H. Wang, D. Zhang, D. Wang, Deeprisk: A deep transfer learning approach to migratable traffic risk estimation in intelligent transportation using social sensing, in: 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), IEEE, 2019, pp. 123–130 (2019).

[22] N. D. Lane, S. B. Eisenman, M. Musolesi, E. Miluzzo, A. T. Campbell, Urban sensing systems: opportunistic or participatory?, in: Proceedings of the 9th workshop on Mobile computing systems and applications, ACM, 2008, pp. 11–16 (2008).

[23] D. Wang, T. Abdelzaher, L. Kaplan, Surrogate mobile sensing, IEEE Communications Magazine 52 (8) (2014) 36–41 (2014).

[24] H.-P. Hsieh, S.-D. Lin, Y. Zheng, Inferring air quality for station location recommendation based on urban big data, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2015, pp. 437–446 (2015).

[25] P. Wang, Y. Fu, G. Liu, W. Hu, C. Aggarwal, Human mobility synchronization and trip purpose detection with mixture of hawkes processes, in: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, ACM, 2017, pp. 495–503 (2017).

[26] Y.-C. Yu, A mobile social networking service for urban community disaster response, in: Semantic Computing (ICSC), 2015 IEEE International Conference on, IEEE, 2015, pp. 503–508 (2015).

[27] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira Jr, C. Ratti, Understanding individual mobility patterns from urban sensing data: A mobile phone trace example, Transportation research part C: emerging technologies 26 (2013) 301–313 (2013).

[28] I. Krontiris, F. C. Freiling, T. Dimitriou, Location privacy in urban sensing networks: research challenges and directions [security and privacy in emerging wireless networks], IEEE Wireless Communications 17 (5) (2010) 30–35 (2010).

[29] X. Sheng, J. Tang, W. Zhang, Energy-efficient collaborative sensing with mobile phones, in: 2012 Proceedings IEEE INFOCOM, IEEE, 2012, pp. 1916–1924 (2012).

[30] L. G. Jaimes, I. J. Vergara-Laurens, A. Raij, A survey of incentive techniques for mobile crowd sensing, IEEE Internet of Things Journal 2 (5) (2015) 370–380 (2015).

[31] D. Wang, L. Kaplan, H. Le, T. Abdelzaher, On truth discovery in social sensing: A maximum likelihood estimation approach, in: Information Processing in Sensor Networks (IPSN), 2012 ACM/IEEE 11th International Conference on, IEEE, 2012, pp. 233–244 (2012).

[32] D. Wang, T. Abdelzaher, L. Kaplan, C. C. Aggarwal, Recursive fact-finding: A streaming approach to truth estimation in crowdsourcing applications, in: 2013 IEEE 33rd International Conference on Distributed Computing Systems, IEEE, 2013, pp. 530–539 (2013).

[33] D. Y. Zhang, J. Badilla, Y. Zhang, D. Wang, Towards reliable missing truth discovery in online social media sensing applications, in: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2018, pp. 143–150 (2018).

[34] Y. Zhang, D. Y. Zhang, N. Vance, D. Wang, An online reinforcement learning approach to quality-cost-aware task allocation for multi-attribute social sensing, Pervasive and Mobile Computing 60 (2019) 101086 (2019).

[35] L. Shang, D. Y. Zhang, M. Wang, D. Wang, Vulnercheck: a content-agnostic detector for online hatred-vulnerable videos, in: 2019 IEEE International Conference on Big Data (Big Data), IEEE, 2019, pp. 573–582 (2019).

[36] L. Shang, D. Y. Zhang, M. Wang, S. Lai, D. Wang, Towards reliable on-

line clickbait video detection: A content-agnostic approach, Knowledge-Based Systems 182 (2019) 104851 (2019).

[37] D. Zhang, Y. Zhang, Q. Li, T. Plummer, D. Wang, Crowdlearn: A crowd-ai hybrid system for deep learning-based damage assessment applications, in: 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), IEEE, 2019, pp. 1221–1232 (2019).

[38] D. Y. Zhang, Y. Huang, Y. Zhang, D. Wang, Crowd-assisted disaster scene assessment with human-ai interactive attention., in: AAAI, 2020, pp. 2717–2724 (2020).

[39] Y. Zhang, H. Wang, D. Zhang, Y. Lu, D. Wang, Riskcast: Social sensing based traffic risk forecasting via inductive multi-view learning, in: 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2019, pp. 154–157 (2019).

[40] Y. Zhang, X. Dong, D. Zhang, D. Wang, A syntax-based learning approach to geo-locating abnormal traffic events using social sensing, in: 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2019, pp. 663–670 (2019).

[41] J. Sun, J. Sun, A dynamic bayesian network model for real-time crash prediction using traffic speed conditions data, Transportation Research Part C: Emerging Technologies 54 (2015) 176–186 (2015).

[42] C. C. Aggarwal, T. Abdelzaher, Social sensing, in: Managing and mining sensor data, Springer, 2013, pp. 237–297 (2013).

[43] Y. Zhang, Y. Lu, D. Zhang, L. Shang, D. Wang, Risksens: A multi-view learning approach to identifying risky traffic locations in intelligent transportation systems using social and remote sensing, in: 2018 IEEE International Conference on Big Data (Big Data), IEEE, 2018, pp. 1544–1553 (2018).

[44] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros, T. Darrell, Cycada: Cycle-consistent adversarial domain adaptation, in: International conference on machine learning, 2018, pp. 1989–1998 (2018).

[45] A. Y. Ng, Feature selection, l 1 vs. l 2 regularization, and rotational invariance, in: Proceedings of the twenty-first international conference on Machine learning, ACM, 2004, p. 78 (2004).

[46] Z. Yi, H. Zhang, P. Tan, M. Gong, Dualgan: Unsupervised dual learning for image-to-image translation, in: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, 2017, pp. 2868–2876 (2017).

[47] G. E. Nasr, E. Badr, C. Joun, Cross entropy error function in neural networks: Forecasting gasoline demand., in: FLAIRS Conference, 2002, pp. 381–384 (2002).

[48] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[49] C. K. Williams, C. E. Rasmussen, Gaussian processes for machine learning, Vol. 2, MIT Press Cambridge, MA, 2006 (2006).

[50] A. Alaoui, M. W. Mahoney, Fast randomized kernel ridge regression

with statistical guarantees, in: Advances in Neural Information Processing Systems, 2015, pp. 775–783 (2015).

[51] S. Mandt, M. D. Hoffman, D. M. Blei, Stochastic gradient descent as approximate bayesian inference, The Journal of Machine Learning Research 18 (1) (2017) 4873–4907 (2017).

[52] J. Tang, C. Deng, G.-B. Huang, Extreme learning machine for multilayer perceptron, IEEE transactions on neural networks and learning systems 27 (4) (2016) 809–821 (2016).

[53] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, pp. 2672–2680 (2014).

[54] P. Ramachandran, B. Zoph, Q. V. Le, Searching for activation functions (2018).

[55] C. Huang, D. Wang, S. Zhu, Where are you from: Home location profiling of crowd sensors from noisy and sparse crowdsourcing data, in: INFOCOM 2017-IEEE Conference on Computer Communications, IEEE, IEEE, 2017, pp. 1–9 (2017).

[56] R. C. Carlson, I. Papamichail, M. Papageorgiou, Local feedback-based mainstream traffic flow control on motorways using variable speed limits, IEEE Transactions on Intelligent Transportation Systems 12 (4) (2011) 1261–1276 (2011).