Explaining Potential Risks in Traffic Scenes by Combining Logical Inference and Physical Simulation

Ryo Takahashi, Naoya Inoue, Yasutaka Kuriya, Sosuke Kobayashi, and Kentaro Inui

***Abstract*—The automatic recognition of risks in traffic scenes is a core technology of Advanced Driver Assistance Systems (ADASs). Most of the existing work on traffic risk recognition has been conducted in the context of motion prediction of vehicles. Thus, existing systems rely on directly observed information (e.g., velocity), whereas the exploitation of implicit information inferable from observed information (e.g., the intention of pedestrians) has rarely been explored. Our previous approach used abductive reasoning to infer implicit information from observation and jointly identify the most-likely risks in traffic scenes. However, abductive frameworks do not exploit quantitative information explicitly, which leads to a lack of grounds for physical quantities. This paper proposes a novel risk recognition model combining first-order logical abduction-based symbolic reasoning with a simulation based on physical quantities. We build a novel benchmark dataset of real-life traffic scenes that are potentially risky. Our evaluation demonstrates the potential of our approach. The developed dataset has been made publicly available for research purposes.**

***Index Terms*—Advanced driver assistance system (ADAS), logical inference, physics simulation.**

I. Introduction

In the field of automotive safety, technology for advanced driver assistance systems (ADASs) and automated driving systems has received much attention [1]–[3]. One of the crucial, open problems in this field is how to enable the system to make an early prediction of potential risks from every frame of a traffic scene.

Considering the traffic scene illustrated in Fig. 1, where an individual is driving their vehicle along a street and a red taxi is driving ahead of them. In this scene, the woman in yellow may summon the taxi, and the taxi driver may suddenly stop in response. This can be a potential risk for the individual if they were to suddenly brake and the green truck was to collide with their vehicle from behind. Furthermore, this could additionally be considered a risk because the individual cannot overtake the taxi because of the purple truck approaching from the opposite direction.

The key to avoiding such risks is to predict them as early as possible. However, the early prediction of potential risks is not straightforward because it requires prediction of the behavior or intentions of pedestrians and vehicles (e.g., summoning a taxi) in a given traffic scene. Furthermore, as illustrated in Fig. 1(a), it requires reasoning about the plausible outcomes through chains of causalities. Another requirement is that it is necessary to consider physical quantities for reasoning based on such chains. In Fig. 1, the likelihood of reasoning that the green truck might collide with the individual's vehicle if the latter were to brake suddenly is dependent on the distance and the relative speed between the individual's vehicle and that of the truck. Although these requirements provide an intriguing application-oriented instance of the task of commonsense reasoning over the human-physical world, few studies regarding early prediction of potential risks exist in the field of automated driving and ADASs.

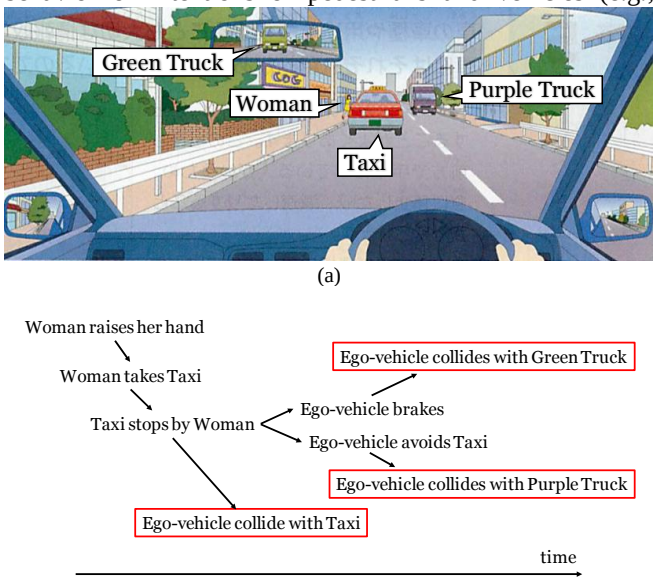


Fig. 1. What is dangerous about this traffic scene? (a) Risky traffic scene (This illustration was adapted from [4]), (b) Causality chains of above traffic scene. The red rectangles denote a potential risk.

Regarding the early prediction of potential risks, some studies on inference-based approaches have been reported [5]–[7]. However, the inference engines that were used in these studies are deductive. As described below, this approach cannot accommodate the uncertainty of observation. On the other hand, in our previous work [8], we proposed a context-aware risk prediction model that exploits first-order logic-based abductive reasoning. An abductive framework allows us to predict long-range movements of traffic objects

Manuscript received August 6, 2016. This work was supported by JSPS KAKENHI Grant Number 15H01702. The authors would like to thank Denso Corporation for funding this research.

Ryo Takahashi, Naoya Inoue, Sosuke Kobayashi, and Kentaro Inui are with the Graduate School of Information Sciences, Tohoku University (e-mail: {ryo.t, naoya-i, sosuke.k, inui}@ecei.tohoku.ac.jp).

Yasutaka Kuriya is with Denso Corporation (e-mail: YASUTAKA_KURIYA@denso.co.jp).

doi: 10.18178/ijmlc.2016.6.5.606

by using implicit contextual information and simultaneously provides deeper explanations as to why a traffic scene poses a risk. However, abductive frameworks do not exploit quantitative information explicitly and this leads to a lack of grounds for physical quantities. Our previous work also pointed out that the majority of erroneous traffic risks are derived via unreasonable inference rules, which are caused by a lack of physical information such as the precise positions and movement directions of traffic objects. For example, the system needed to understand that if a bus was currently halting at a bus stop, and a man across the street appeared to be interested in crossing the street to take the bus, then the man may suddenly cross the street to board the bus.

In this study, we integrate a symbolic inference-based approach and a simulation based on physical quantities. We rebuild our previous knowledge base in order to connect it with the physical simulation. We expect our approach to perform well for risky traffic scenes that necessitate the exploitation of quantitative information.

We evaluate our risk recognition system by building a novel benchmark dataset consisting of real-life traffic scenes that pose a potential risk as defined by the existing database of near-miss events. To the best of our knowledge, it is the largest dataset (over 3,000 scenes) of risk prediction based on real-life data. We conducted a corpus study on the dataset to select scenes that are needed to predict risks. Our preliminary evaluation results on a subset of the corpus suggest that the proposed integrated architecture provides rich information for early prediction of potential risks in real-life traffic scenes.

II. Background

A. Related Work

The majority of studies concerning inner-city risk assessment are based on detecting possible conflicts of future trajectories [3]. Broadhurst *et al.* [9] used the Monte Carlo method to generate a probability distribution for the possible future motion of every vehicle in the scene to avoid danger. Althoff *et al.* [10] proposed a stochastic approach to detect forthcoming collisions. Neither of these studies nor other related work addresses the explicit interaction among traffic participants, although its importance has been indicated in [1].

Several symbolic inference-based approaches have been proposed for understanding situations in an inner-city context. Armand *et al.* [5] formulated an ontology for inner-city traffic situation analysis and created rules that enable reasoning about the traffic participants' future behavior. Furthermore, they showed that the ontology makes it possible to identify and understand the key entities a driver should consider. Mohammad *et al.* [6] proposed an ontology-based framework for assessing the degree of risk in a road scene involving vehicles or pedestrians and indicated that the framework is capable of assessing risk with high accuracy. Zhao *et al.* [7] proposed an ontology-based knowledge base and a decision-making system capable of making safe decisions on uncontrolled intersections and narrow two-way roads.

These approaches take into account the characteristics of the environment and the interactions among them. The advantage of using a symbolic inference-based approach is the transparency of the system: the prediction result is represented by a combination of prediction rules. The rules can be used to explain the reason for the prediction, which has recently become an important research topic of ADASs. However, the inference engines employed in previous work are *deductive*, and are therefore unable to manage the uncertainty of observations. Moreover, symbolic inference-based approaches occasionally overgeneralize the physical world, as the prediction is not based on a precise physical prediction.

Grounding technologies, including image/motion recognizers and radars, are considered to have recently become advanced [11]. For object recognizers, a number of benchmark datasets are publicly available [12]–[15], and they have been extensively studied over the years. Zhang *et al.* [16] compared around 10 pedestrian detectors on the Caltech-USA pedestrian benchmark [17] and report that the best method, Checkerboards, achieves an 18.47 % miss-rate. In fact, these technologies have already been applied to traffic scene understanding [18]. Regarding other grounding technologies, such as radar and vision cameras, extensive research has also been conducted (see Bengler *et al.* [2] for a detailed overview). However, the accuracy is not always perfect: the ability to process the uncertainty of observations is important.

To the best of our knowledge, no previous work that focuses on integrating logical inference that makes maximum use of symbolic information and simulation that exploits quantitative data has been reported.

B. Abduction

Abduction is inference of the best explanation and is widely used for knowledge-based symbolic inference systems such as diagnostic systems or natural language understanding [19] in artificial intelligence research. Formally, first-order logical abduction is defined as follows:

Given: Background knowledge B, and observations O, where B is a set of first-order logical formulae, and O is a set of literals or substitutions,

Find: A hypothesis H such that, $H \cup B \models O$, $H \cup B \not\models \bot$ where H is a set of literals or substitutions.

Each hypothesis H that satisfies the condition is termed a *candidate hypothesis*, and a set of candidate hypotheses is denoted as $\mathcal{H}$. The goal of abduction is to find the best explanation [1] among candidate hypotheses by a specific evaluation measure. In this paper, we formulate abduction as the task of finding the minimum-cost explanation $\hat{H}$ among $\mathcal{H}$. Formally, we find $\hat{H} = \arg\min_{H \in \mathcal{H}} Cost(H)$, where, *Cost* is a function that maps each $H \in \mathcal{H}$ to a real number, referred to as the *abductive cost function*. We elaborate our cost function in Section IV-A2.

III. Task Definition

In this paper, we formalize the problem of traffic risk

[1] In the context of abduction, the terms *explanation* and *hypothesis* are used interchangeably.

recognition as follows:

Given: A scene description s of a traffic scene, which poses potential risks with respect to the ego-vehicle, and quantitative data E of each entity in the scene, where s is a set of literals in first-order predicate logic following the knowledge representation described in Section IV-A1, and E is a set of triples of the form $(shape, position, velocity)$,

Find: The best explanation of the risk: a set R of potential risks, where each potential risk r consists of an entity-action tuple (e, a).

This study assumes that s and E are constructed from the outputs of perception systems such as Light Detection and Rangings (LIDARs) and object recognition technologies.

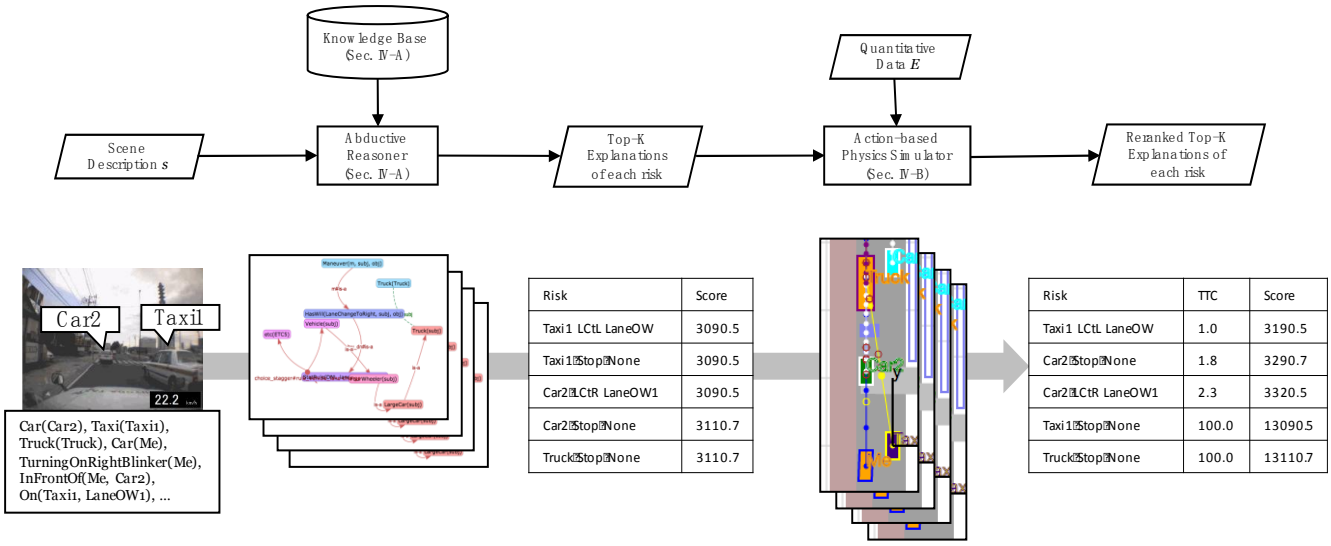


Fig. 2. Diagrammatic representation of our system and an example. LCtL and LCtR in the table denote "lane change to left" and "lane change to right," respectively. The lower the score, the more risky the action.

IV. Proposed Approach

As mentioned in Section I, the prediction of the behavior or intentions of traffic agents is crucial for the task of traffic risk prediction. Some prior studies, including our previous work [8], proposed symbolic inference-based approaches to predict such information [5]–[7]. We employ an abduction-based approach proposed in our previous work [8] as the starting point of this study because our previous methods were shown to be capable of managing the uncertainty of observation, which is considered to be more advantageous for practical situations. We then overcome a significant drawback of the previous work: this work did not exploit quantitative information such as the shape, position, or velocity of traffic agents. To solve this problem, we now propose a method whereby a simulator of physical data is plugged into a symbolic inference engine. More specifically, we rebuild the previous knowledge base such that it can predict richer information that can be utilized by the simulator, and show how to combine symbolic inference with the simulator.

We provide a brief overview of our approach. Our overall traffic risk recognition architecture is shown in Fig. 2. Firstly, an abductive reasoner predicts multiple risky entities and their actions from the point of view of the ego-vehicle with the scores, using s and a *qualitative* knowledge base as an input. Note that this module does not use precise quantitative information such as the distance between the ego-vehicle and other mobile entities or the velocity of the vehicles. Secondly, an action-based simulator of physical data simulates the former prediction by exploiting quantitative data E on a virtual space, and then outputs metrics such as the time-to-collision (TTC). We expect this module to determine whether the former prediction is indeed risky for the ego-vehicle.

For example, in the traffic scene illustrated in Fig. 2, an abductive reasoner predicts that the most dangerous risk is presented by Taxi1 in the neighboring lane, because it might stop suddenly. However, our system notices that it is not necessarily dangerous because a simulation of the situation indicates a collision of the ego-vehicle with the taxi seems unlikely; hence, it ranks lower in the scene. The remainder of this section presents a description with further detail of each component.

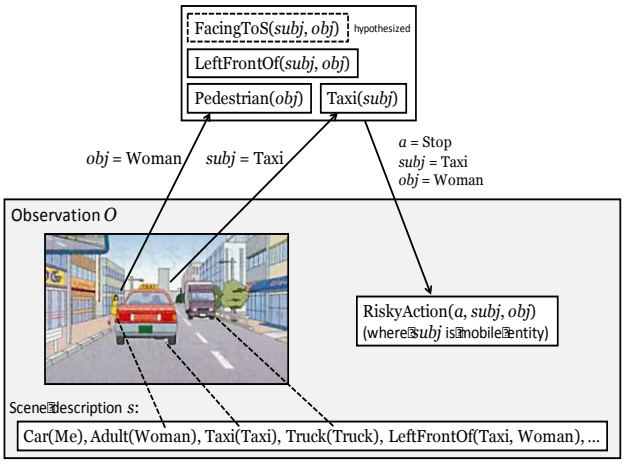


Fig. 3. Working example of action recognition as abduction.

A. Action Recognition as Abduction

Following our previous work [8], we formulate the risk prediction problem as a problem of abductive inference: the problem of explaining why an observed traffic scene and a

hypothetical observation "the observed scene has some risk" are observed, using a knowledge base. The knowledge base contains two types of axioms: (i) conceptual hierarchy and (ii) knowledge about the relation between an intention and its implied situation.

We describe the overall framework using the working example illustrated in Fig. 3. The observation O includes the logical forms of the observed scene (i.e., $Car(Me)$, $Adult(Woman)$, …) and the hypothetical observation of the existence of a risky entity (i.e., $RiskyAction(a,subj,obj)$). The literal $RiskyAction(a,subj,obj)$ indicates that a traffic agent $subj$ will take a risky action a to a target obj. We then feed this observation to the abductive inference system to obtain the best explanation, which will contain the variable binding of a, $subj$ and obj. In this example, what happens is that: (i) $RiskyAction(a,subj,obj)$ is explained by $FacingToS(subj,obj) \cup LeftFrontOf(subj,obj) \cup Pedestrian(obj) \cup Taxi(subj)$, and (ii) $Pedestrian(obj)$ and $Taxi(subj)$ are explained by the observed scene, assuming $subj = Woman$ and $obj = Taxi$. As a result, we can identify that the risky factor of this scene is the possibility of the taxi suddenly stopping beside the woman.

The main advantage of using abduction is characterized as its declarative nature. It enables us to form an abstraction independent from the inference process and to concentrate on creating a sophisticated knowledge base in the declarative fashion.

The second advantage is that by combining several types of knowledge bases (e.g., causality and ontological knowledge), our model can abductively infer implicit information from observed information and jointly identify the most-likely risks in traffic scenes. For example, in Fig. 3, $FacingToS(Taxi,Woman)$ is implicitly inferred information that is not captured by observations. Abduction-based modeling allows us to predict long-range movements of traffic objects by using implicit contextual information, and simultaneously provide deeper explanations as to why the traffic scene poses a risk.

1) Knowledge base

We now elaborate on the knowledge base used in this study. We first describe the knowledge representation of a traffic scene. In principle, our representation consists of the following concepts:

Type of object: e.g., $Vehicle(x)$, $Pedestrian(x)$;

Property of traffic objects (e.g., whether right blinker is turned on): e.g., $TurningOnLeftBlinker(x)$;

Relation between traffic objects (e.g., relative position): e.g., $InFrontOf(x,y)$, $RightOf(x,y)$;

The definition of a possible action of a pedestrian and vehicle (e.g., turning right), which are represented by constants: $Stop$, $GoLeft$.

Based on this knowledge representation, we constructed the following two types of axioms:

Conceptual hierarchy: This represents the hierarchical (a.k.a. IS-A) structure of concepts. For example, the knowledge "a taxi is one kind of car" is represented by the logical form $\forall x. Taxi(x) \Rightarrow Car(x)$. This knowledge allows us to perform reasoning on various levels of abstraction.

Intention-situation axiom: The axiom describes the causal relation between an action and the situation in which the action is likely to be taken. For example, the knowledge "a vehicle v is likely to overtake a large vehicle c_l which is traveling in the same lane l as v and is in front of v" is represented by the logical form $\forall v,c_l,l. Vehicle(v) \cup LargeCar(c_l) \cup InFrontOf(v,c_l) \cup On(v,l) \cup On(c_l,l) \Rightarrow RiskyAction(Overtake,v,c_l)$.

As described in Section IV-B, a simulator of physical data requires an entity-action-object tuple as an input. The knowledge base in our previous work [8] is tailored for predicting a risky entity-action tuple, which is insufficient to integrate a physics simulator with symbolic inference. Popular machine-learning approaches for classification or ranking are also adversely affected by predicting this kind of richer information. Thus, the use of first-order logic as a representation enables us to easily accommodate a richer information structure.

2) Cost function

We employ the cost function of Weighted Abduction [19] as the abductive cost function. In weighted abduction, observation O and hypothesis H are represented by the conjunction of existentially quantified literals. Each literal has a positive real-valued *cost* (henceforth referred to as $l^{\$100}$). The cost of observation manages the uncertainty of observations. Background knowledge B is a set of Horn clauses. Each literal in the body of Horn clauses is assigned a positive real-valued *weight* (referred to as $l_1^{0.6} \wedge l_2^{0.6} \Rightarrow l_3$).

We now describe the cost function. Let $nonexp(H)$ be a set of non-explained literals in H. In the weighted abduction, the cost of a hypothesis H is defined by the sum of non-explained literals:

$$Cost_{WA}(H) = \sum_{h \in nonexp(H)} cost(h), \quad (1)$$

where $cost(h)$ is the cost of literal h. If h is a non-observed literal, $cost(h)$ is calculated by $cost(obs(h)) \times \prod_{a \in axioms(h)} weight(a)$, where $axioms(h)$ is the set of axioms used for deriving h, and $obs(h)$ is the observed literal backchained on to hypothesize h. If h is an observed literal, $cost(h)$ is simply a real-valued cost attached to h in the input. See Hobbs *et al.* [20] for further details. In this study, we extend (1) in two ways.

a) Learning reliability of axioms

Hobbs *et al.* [19] did not provide a method to learn the parameters of the cost function. Following our previous work [8], we learn the reliability of axioms by extending (1) as follows:

$$Cost(H) = Cost_{WA}(H) + \mathbf{w}_a \times F_a(H), \quad (2)$$

where $F_a(H)$ is a feature vector that is constructed from the set of axioms used for deriving H and $\mathbf{w}_a$ is a real-valued weight vector.

b) Context-dependency

According to the cost function, the goodness of the hypothesis depends on the plausibility of axioms used for deriving H and the uncertainty of observations, but *not* on the observed context. However, this is not suitable for traffic risk prediction. To incorporate the context-dependency, we further extend (2) as follows:

$$Cost(H) = Cost_{WA}(H) + \mathbf{w}_a \cdot F_a(H) + \mathbf{w}_c \cdot F_c(H,O), \quad (3)$$

where $F_c(H,O)$ is a feature function that returns a d-dimensional vector that is determined by a given hypothesis H and an observation O, and $\mathbf{w}_c$ is a real-valued weight vector.

In this study, we take a two-step supervised learning approach to learn $\mathbf{w}_c$ and $\mathbf{w}_a$. We first learn $\mathbf{w}_c$ by using a ranking SVM [20], where all the features are binary features encoding (i) literals describing a ranked object and action (prefixed with "obj" and "action_") and (ii) literals describing the other traffic objects in a traffic scene (prefixed with "context_"). Since the combinations of features are considered important for risk prediction, we use a polynomial kernel of degree 2. We then use a latent structured perceptron approach [21] to learn $\mathbf{w}_a$, fixing $\mathbf{w}_c$ and using the binary feature function that returns a 0-1 vector, where the value of the i-th element is 1 if the i-th axiom is used in H; 0 otherwise as F_a. In our experiment, we use Phillip [22] as an abductive inference engine[2].

We render the inference tractable by extending the cost function of Weighted Abduction as follows: (1) the cost of unification between hypothesized literals is ¥, (2) the cost of backward inference on observed literals except action/3 is ¥. This amounts to performing a best-proof search for a literal action/3 using $O \in B$ as a background knowledge base.

B. Action-Based Physical Simulations

In addition to the qualitative inference described thus far, we use a simulation of physical data to re-rank the risk prediction results based on physical information such as the location, distance, and velocity. We assume the following input and output of the simulator:

Input: (i) the road structure, (ii) the quantitative information of traffic agents (i.e., position, direction, velocity), and (iii) the intention of each object represented by a first-order logic literal (e.g., $RiskyAction(Stop, Car, YellowSignal)$) for "*Car* will stop at *YellowSignal*.");

Output: (i) the predicted future trajectories of each traffic agent, and (ii) information about collision between traffic agents (e.g., time-to-collision (TTC)).

As illustrated in Fig. 2, the physical data simulator receives an output from the abductive reasoner, which contains a risky entity-action-object tuple, which is required by the simulator input. The need to use this varied information suggests that there could possibly be ways to combine a physical simulation with symbolic inference. In this paper, we introduce a simple, pipeline re-ranking model as a preliminary study for this new challenge. More advanced and complicated combination methods will be explored in future work.

After running the physical simulation, we simply re-rank risky entity-action-object tuples based on the TTC with the ego-vehicle because our aim is to detect an entity-action-object tuple that poses a risk *to the ego-vehicle*.

The re-ranking score function for a risk $r = (e, a)$ after n seconds is then defined as follows:

$$Score_{TTC} = Cost(H_r) + \frac{w}{Cost(H_r)} \cdot \begin{cases} |n - TTC| & \text{if } e \text{ collided with ego-vehicle} \\ a & \text{otherwise,} \end{cases} \quad (4)$$

where H_r is a hypothesis associated with r.

We empirically set $w = 10$ and $a = 10$ as a result of manual adjustment on a development set.

Then, we implement the physical data simulator by using an action-based motion model using prototype trajectories [3].

Given an input, the simulator generates a trajectory for each traffic object based on a predefined set of prototype trajectories, where a prototype trajectory contains landmarks and a parametric trajectory represented by a set of points and acceleration. We combine several prototype trajectories to generate a final trajectory, transforming these prototype trajectories by scale and an angle. We use 14 of these trajectories as a preliminary study.

TABLE I: CLASSIFICATION OF CAUSE OF BRAKING.

.Label	Cause	#	Freq. (%)
Rule	Traffic rules (e.g., red light traffic signals)	209	20.9
Avoidance	To avoid imminent collisions (e.g., a leading vehicle brakes suddenly)	544	54.5
Prediction	To be on the safe side (e.g., overtaking a bicycle)	166	16.6
Other	Changing lane, entering a road, etc.	49	4.9
Unknown	Unable to judge the cause	31	3.1

TABLE II: CLASSIFICATION OF BEHAVIOR OF THE EGO-VEHICLE.

Label	#	Freq. (%)
Direct	373	53.4
Change	281	39.6
Other	50	7.0

V. EVALUATION

The purpose of our evaluation is twofold. The first evaluation is to check whether the proposed model can enrich the prediction results without disrupting the simple statistical ranking model. The second evaluation aims to determine the effectiveness of the integration of the physical simulation. Our previous work [8] evaluated the model on a small dataset

[2] https://github.com/naoya-i/phillip

that was not realistic (i.e., illustrations from a textbook), but in this work we evaluate our model on a large database of near-miss recordings that were collected by drive recorders mounted in taxis.

A. Task Setting

Given the traffic scene two seconds before an actual near-miss event, our task is to identify a risky action-object tuple that causes the near-miss incident. In this experiment, we assume the input to our task to be a two-dimensional bird-view map that represents the traffic scene two seconds before the actual near-miss event.

TABLE III: ACCURACY OF RISK PREDICTION MODELS

	Validation			Test		
Model	Acc@1	Acc@3	Acc@5	Acc@1	Acc@3	Acc@5
BASELINE	**52.8 (38/72)**	80.6 (58/72)	**90.3 (65/72)**	55.6 (40/72)	**77.8 (56/72)**	**93.1 (67/72)**
INFERENCE	51.4 (37/72)	80.6 (58/72)	**90.3 (65/72)**	**58.3 (42/72)**	**77.8 (56/72)**	91.7 (66/72)
INF+PHYSIM	51.4 (37/72)	**81.9 (59/72)**	**90.3 (65/72)**	**58.3 (42/72)**	**77.8 (56/72)**	91.7 (66/72)

The two-dimensional map contains information about the road structure and traffic objects (position and direction of movement). The physical simulator uses this information to simulate the physical data. The logical representation of each traffic scene is automatically generated from the map according to Section IV-A1. In this experiment, we assume the accuracy of sensor technologies to be perfect, in order to focus on exploring the methodology of risk prediction.

We evaluate our results by using a task-ranking framework. We use $Acc@k$ (Accuracy at k) as a metric, which is the fraction of problems for which the correct prediction is made within rank k.

B. Dataset

We use a database of near-miss accidents published by the Tokyo University of Agriculture and Technology. The dataset consists of over 100,000 video recordings of near-miss accidents recorded by the drive recorders of a taxi[3]. Our evaluation involved the use of over 3,000 videos recorded in 2014.

The dataset includes traffic scenes that have no potential risks indeed, since they are collected based on the occurrence of sudden braking. We conducted a corpus study on the dataset to classify scenes that are needed to predict those risks that truly exist because of braking. The classification was entrusted to a person who was not involved in the development of the dataset. We classified about 1,000 scenes in which the object posing a risk enters the view of the ego-vehicle two seconds before the actual near-miss incident. Table I presents the classification result. The scenes labeled Avoidance and Prediction are considered to be potentially risky; thus, we subjected them to further classification.

The subsequent classification is based on whether the ego-vehicle changed its trajectory during the two seconds preceding the actual near-miss event. We collected those scenes in which the ego-vehicle did not change its trajectory in order to create the task setting "What potential risks are there when the ego-vehicle maintains its current speed?" Table II contains the classification result. Direct label denotes scenes in which the ego-vehicle did not change its trajectory, and Change label denotes scenes in which it did. Finally, we use 379 scenes labeled as Direct as evaluation data, and divided the dataset into a training set, test set, validation set in the ratio of 3:1:1. The data corresponding to the above labels and ID on the database is publicly available[4]. We created a benchmark dataset by manually creating a two-dimensional bird-view map for each video.

C. Models

The evaluation entails a comparison of the following three models.

BASELINE: predicts a risk by an abductive reasoner with the cost function $Cost(H) = \mathbf{w}_c \cdot \mathbf{f}_c(H,O)$. This baseline directly models a mapping between observed information and a risky entity-action tuple. As mentioned in Section IV-A2, the weight vector $\mathbf{w}_c$ is trained by using a ranking SVM [20]; therefore, the result of this baseline indicates the basic performance of a statistical machine-learning-based ranker.

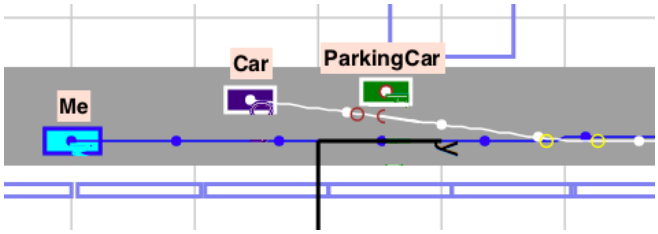


Fig. 4. Example trajectories produced as output by our physics simulator. Scene ID is 1397 on the database. The boxes represent vehicles, and the lines drawn from vehicles represent trajectories.

INFERENCE: predicts a risk by an abductive reasoner with the full cost function described in Section IV-A2. Rather than simulating the physical data, the model uses only symbolic information for the risk prediction.

INF+PHYSIM: the proposed model, which predicts a risk by using an abductive reasoner combined with the physical simulator described in Section IV-B.

D. Evaluation Results

The results are provided in Table III. By comparing the BASELINE and INFERENCE models, we observe that adding abductive reasoning does not affect the performance negatively: the output is successfully enriched such that a simulation of physical data can be performed. Furthermore, we observed the performance improvement on the test set. Manual inspection reveals that some mistakes made by the baseline model were corrected by an inference-based prediction.

Fig. 4 shows an example of the improvement. The risk is that *Car* will change lanes to avoid *ParkingCar*, which might lead to a collision between the ego-vehicle (i.e., *Me*) and *Car*.

[3] http://web.tuat.ac.jp/~smrc/drcenter.html (in Japanese)
[4] https://github.com/reiyw/traffic-scene-understanding

Although the BASELINE model predicted that *Car* will stop, the INFERENCE model predicted that *Car* will change lanes based on *the richer information* that the target lane is adjacent to the current lane.

Using the predicted rich information, the physical simulator predicted the future trajectories of each traffic agent. As illustrated in Fig. 4 the simulator correctly predicted the future trajectories and also inferred that *Me* will collide with *Car* after 3.6 seconds. This indicates that our logical inference framework successfully connects the world of symbolic inference to the physical world. A simple machine-learning-based ranker or classifier would find it relatively harder to predict these kinds of richer explanations.

Finally, we compare the INFERENCE model with the INF+PHYSIM model. The results indicate that the simulation did not improve the qualitative inference outcome. In future work, we plan to conduct a more in-depth analysis of the results of the simulation of physical data with the aim of refining the entire framework to improve the results.

VI. CONCLUSIONS

We have developed an Advanced Driver Assistance System (ADAS) with the ability to recognize potential risks in traffic scenes and provide the reasoning for its prediction. This involved extending our previous qualitative risk prediction model by adding the simulation of physical data to overcome the weakness of qualitative inference. Our evaluation of a real-life traffic incident database demonstrates the potentiality of our approach.

In future, we plan to refine the task setting to allow for a more practical evaluation. Currently, the task setting requires us to predict a risk *exactly two seconds after* the input scene; however, in practice, the ability to predict *any* risks after the input scene is expected to be beneficial. Another future task would include evaluating the quality of the produced explanations.

REFERENCES

[1] E. Rendon-Velez, I. Horváth, and E. Z. Opiyo, "Progress with situation assessment and risk prediction in advanced driver assistance systems: A survey," in *Proc. 16th ITS World Congress*, Dec. 2009.

[2] K. Bengler, K. Dietmayer, B. Färber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: Review and future perspectives," *IEEE Intell. Transport. Syst. Mag*., vol. 6, no. 4, pp. 6–22, 2014.

[3] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *Robomech Journal*, vol. 1, no. 1, p. 1, 2014.

[4] *Kiken-Yosoku-Master*, Chubu Nippon Driver School, 1999.

[5] A. Armand, D. Filliat, and J. I. Guzman, "Ontology based context awareness for driving assistance systems," in *Proc. 2014 IEEE Intelligent Vehicles Symp. Proc*., Dearborn, MI, USA, June 8-11, 2014, pp. 227–233, 2014.

[6] M. A. Mohammad, I. Kaloskampis, Y. Hicks, and R. Setchi, "Ontology-based framework for risk assessment in road scenes using videos," in *Proc. 19th Int. Conf. in Knowledge Based and Intelligent Information and Engineering Systems*, KES 2015, Singapore, 7-9 September, 2015, pp. 1532–1541, 2015.

[7] L. Zhao, R. Ichise, T. Yoshikawa, T. Naito, T. Kakinami, and Y. Sasaki, "Ontology-based decision making on uncontrolled intersections and narrow roads," in *Proc.* 2015 *IEEE Intelligent Vehicles Symp*., IV 2015, Seoul, South Korea, June 28 - July 1, 2015, pp. 83–88, 2015.

[8] N. Inoue, Y. Kuriya, S. Kobayashi, and K. Inui, "Recognizing potential traffic risks through logic-based deep scene understanding," in *Proc. 22nd ITS World Congress*, 2015.

[9] A. Broadhurst, S. Baker, and T. Kanade, "Monte Carlo road safety reasoning," in *Proc. IEEE Intelligent Vehicles Symp*., 2005, pp. 319–324, June 2005.

[10] M. Althoff, O. Stursberg, and M. Buss, "Model-based probabilistic collision detection in autonomous driving," *IEEE Trans. Intell. Transport. Syst.*, vol. 10, no. 2, pp. 299–310, 2009.

[11] A. Ambardekar, M. Nicolescu, G. Bebis, and M. N. Nicolescu, "Vehicle classification framework: a comparative study," *EURASIP J. Image and Video Processing*, vol. 2014, no. 29, 2014.

[12] A. Ess, B. Leibe, K. Schindler, and L. J. Van Gool, "A mobile vision system for robust multi-person tracking," in *Proc.* 2008 *IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition (CVPR 2008)*, 24-26 June 2008, Anchorage, Alaska, USA, 2008.

[13] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: Survey and experiments," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 31, no. 12, pp. 2179–2195, 2009.

[14] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in *Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition (CVPR 2009)*, 20-25 June, 2009, Miami, Florida, USA, pp. 304–311, 2009.

[15] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, June 16-21, 2012, pp. 3354–3361, 2012.

[16] S. Zhang, R. Benenson, M. Omran, J. H. Hosang, and B. Schiele, "How far are we from solving pedestrian detection?" CoRR, abs/1602.01237, 2016.

[17] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 34, no. 4, pp. 743–761, 2012.

[18] C. R. C. Souza and P. E. Santos, "Probabilistic logic reasoning about traffic scenes," in *Proc. 12th Annual Conference towards Autonomous Robotic Systems*, TAROS 2011, Sheffield, UK, August 31-September 2, 2011. pp. 219–230, 2011.

[19] J. R. Hobbs, M. E. Stickel, D. E. Appelt, and P. A. Martin, "Interpretation as abduction," *Artif. Intell*., vol. 63, no. 1-2, pp. 69–142, 1993.

[20] T. Joachims, "Optimizing search engines using clickthrough data," in *Proc. Eighth ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, July 23-26, 2002, Edmonton, Alberta, Canada, pp. 133–142, 2002.

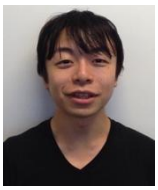
[21] X. Sun, T. Matsuzaki, D. Okanohara, and J. Tsujii, "Latent variable perceptron algorithm for structured classification," In *Proc. 21st Int. Joint Conf. Artif. Intell. (IJCAI)*, Pasadena, CA, vol. 9, pp. 1236–1242, 2009.

[22] K. Yamamoto, N. Inoue, K. Inui, Y. Arase, and J. Tsujii, "Boosting the efficiency of first-order abductive reasoning using pre-estimated relatedness between predicates," *Int. J. Mach. Learning and Computing*, vol. 5, no. 2, pp. 114, 2015.

Ryo Takahashi received the B.E. degree in 2016 from Tohoku University. He is a graduate student in the Department of System Information Sciences, Tohoku University. His research interests include logical inference, machine learning, and natural language processing.

Naoya Inoue received the M.S. degree in engineering from the Nara Institute of Science and Technology in 2010 and the Ph.D. degree in information science from Tohoku University in 2013. He joined DENSO Corporation as a researcher in 2013. He has been an assistant professor at Tohoku University since 2015. His research interests include inference-based discourse processing and language grounding problems.

Yasutaka Kuriya received the M.Eng. degree in 2009 from the Kyushu Institute of Technology. He is working with DENSO CORPORATION as a staff researcher. His research interests include image processing and software engineering.



Sosuke Kobayashi received the B.E. degree in 2014 from Tohoku University. He is a graduate student in the Department of Information and Intelligent Systems, Tohoku University. He has authored and co-authored publications in the area of natural language processing. His current research interests include machine learning and natural language processing.

Kentaro Inui received his doctorate degree of engineering from Tokyo Institute of Technology in 1995. He has experience as an assistant professor at Tokyo Institute of Technology and an associate professor at Kyushu Institute of Technology and Nara Institute of Science and Technology, he has been a professor of Graduate School of Information Sciences at Tohoku University since 2010. His research interests include natural language understanding and knowledge processing. He currently serves as the IPSJ director and ANLP director.