22nd ITS World Congress, Bordeaux, France, 5–9 October 2015

Paper number ITS-1432

Recognizing Potential Traffic Risks through Logic-based Deep Scene Understanding

Naoya Inoue[1*], Yasutaka Kuriya[2], Sosuke Kobayashi[1], Kentaro Inui[1]
1. Tohoku University, Japan
6-6-05 Aramaki Aza Aoba, Aobaku, Sendai, Miyagi 980-8579, Japan
Tel: +81-22-795-7091 / Email: naoya-i@ecei.tohoku.ac.jp
2. DENSO CORPORATION, Japan

Abstract
Automatic recognition of risks in traffic scenes is a core technology of Intelligent Transportation Systems. While most of the existing work on traffic risk recognition has been done in the context of motion prediction of vehicles relying on directly observed information (e.g., velocity), exploiting implicit information inferable from observed information (e.g., the intention of pedestrians) has rarely been explored. This paper proposes a novel risk prediction model that uses abductive reasoning to infer implicit information from observations and jointly identifies the most-likely risks in traffic scenes. To evaluate our model, we create a novel benchmark dataset that contains a wide variety of risk prediction problems. Our experiments indicate that the abduction-based framework has a great potential for solving risk prediction problems. The developed dataset is made publicly available for research purposes.

Keywords:
Abductive Reasoning, Deep Scene Understanding, Potential Traffic Risks

1 Introduction

There has been a growing interest in Advanced Driver Assistance Systems (ADASs), which provide drivers with safe driving [9, 14, 20]. Recent advances in the technology of car sensor devices (e.g., radar sensors, vision systems) and image recognition, the fundamental technology of ADASs, accelerate the research and development of ADASs. ADASs are one of the core technologies of Intelligent Transportation Systems, which have received increasing attention through competitions such as DARPA Grand and Urban Challenges.

In this paper, we develop an ADAS to automatically identify possible traffic accidents

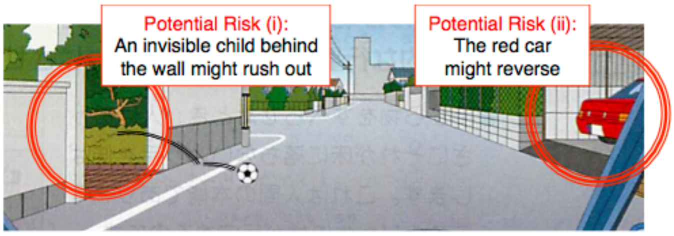

Fig 1. Example potential risks in traffic scenes.

several seconds before they may actually occur. We refer to such kind of dangerous events, implicit events that will occur within several seconds as *potential risks*. For example, the traffic scene shown in Figure 1 include the potential risk that there might be a child behind the left-hand-side wall and the child might rush out to follow the soccer ball. We expect to drastically decrease car accidents by notifying such potential risks to car drivers as early as possible.

Most of the existing work on traffic risk recognition has been done in the context of motion prediction of vehicles [1, 2, 3, 11]. However, the existing prediction models do not fully take into account implicit contextual information, which is not directly observed but can be inferred from observed information, such as the predicted future behavior of other traffic objects (e.g., backing up from a parking spot) and the existence of invisible objects (e.g., the invisible child behind the wall); see [9] for a detailed survey. Because traffic objects may suddenly change their future behavior in response to changes in the state of the surrounding context (e.g., the vehicle in front changes lanes, nearby traffic lights change), ignorance of such contextual information makes it unreliable to predict long-range risk, which is important for potential risk prediction. Additionally, the existing work does not provide drivers with explanations as to why the traffic scene is risky (e.g., who takes what action, why she/he takes that action).

In this paper, we propose a context-aware risk prediction model that exploits first-order logic-based abductive reasoning. *Abduction* is inference to the best explanation. In the field of artificial intelligence research, abduction is widely used for knowledge-based systems, such as diagnostic systems and plan recognition systems [5, 10]. An abductive framework allows us to predict long-range movements of traffic objects by using implicit contextual information, and simultaneously provides deeper explanations as to why the traffic scene has a risk. The declarative nature of abduction allows us to abstract away from the procedural process of inferences, concentrating on describing relevant knowledge in a declarative fashion (see Sec. 2.1).

2 Background

2.1 Related Work

Much research on risk prediction in traffic scenes has been done in the context of motion/path planning of vehicles [1, 2, 3, 11]. The main interest of the community is in how to correctly predict the trajectory of vehicles based on directly observed information such as position, speed and the state of traffic lights. For example, Broadhurst et al. [1] predict the potential future trajectories of an ego-vehicle and other traffic objects based on kinematic models to detect a collision between traffic objects. Ortiz et al. [11] estimate the sequence of future driver maneuvers of an ego-vehicle with a Multi-Layer Perceptron, where the features include the physical state of the vehicle (e.g., position, speed), driver behavior (e.g., head movement), and surrounding contextual information (e.g., the state of nearby traffic lights).

On the other hand, *implicit information* inferable from observed information such as the intention of traffic objects and the existence of invisible objects is also an important clue for risk prediction, but has rarely been explored. One notable exception can be seen in [7]. Lattner et al. [7] propose a theorem proving-based approach. Similar to our work, Lattner et al. [7] use a logical knowledge base to prove the risk in traffic scenes from observed information, inferring implicit information on qualitative knowledge representation such as the type of traffic objects and the discretized speed of pedestrians in a traffic scene.

However, since a theorem prover only checks if the risk is provable or not, the quality of proof is not taken into account, unlike in our model. Moreover, they report the evaluation of their model on one classic example of a traffic scene. In contrast, our model evaluates the goodness of risk proofs in terms of evidentiality seen in a traffic scene and is tested on a much wider variety of risk prediction problems; see Sec. 3 and Sec. 4 for further details.

2.2 Grounding technology

Grounding technologies, including image/motion recognizers and radars, are recently making significant advances. For object recognizers, a number of benchmark datasets are publicly available [17, 18, 19], and they have been extensively studied over the years. Benenson et al. [17] compare around 40 pedestrian detectors on the Caltech pedestrian detection benchmark, and report that the best method, Katamari-v1, achieves a 22.5% miss rate. In fact, these technologies have already been applied to traffic scene understanding [16]. Regarding other grounding technologies, such as radar and vision cameras, extensive research has also been done; see [20] for a good overview.

Obviously, in real-life applications, the whole system needs to be grounded to the real world by means of high-resolution cameras, accurate image recognizers, physics simulators, etc. In addition, we clearly need additional effort to calibrate our risk prediction system with possible noisy inputs. However, exploring the use of implicit information in risk prediction is a big research issue in itself. Therefore, we leave the integration of these technologies into our system for future work.

2.3 Abduction

Abduction is inference to the best explanation. Abduction is widely used in knowledge-based systems, such as diagnostic systems and plan recognition systems [5, 10]. 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: An explanation H such that $H \cup B \vDash O, H \cup B \nvDash \bot$, where H is a set of literals or substitutions. Each element in H is called an elemental explanation. Throughout the paper, $\vDash, \bot$ means logical entailment and logical contradiction, respectively.

In this paper, we assume that all variables occurring in a logical form of background knowledge are universally quantified with the widest possible scope, unless it is explicitly stated as existentially quantified. On the other hand, we assume that variables occurring in an explanation and observation are existentially quantified implicitly.

Typically, given observations O, we have more than one explanation H that explains O. We call each of them a *candidate explanation*, and denote a set of candidate explanations of O given B as $C(O, B)$. The goal of abduction is to find the best explanation among candidate explanations by a specific evaluation measure. In this paper, we formulate abduction as the task of finding the maximum-score explanation H^* among $C(O, B)$. Formally, we find $H^* = \arg\max_{H \in C(O,B)} score(H)$, where *score* is a function that maps each H in $C(O, B)$ to a real number, which is called the abductive score function. In the literature, several kinds of score functions have been proposed, including cost-based and probability-based [2, 5, 13]. We elaborate our score function in Sec. 3.5.

3 Recognizing Potential Risks with Logic-based Deep Scene Understanding

3.1 Task definition

We define the problem of traffic risk prediction with respect to a driver d at time t as follows:

Given: An image description $s_{d,t}$ of a traffic scene from a driver d's viewpoint at time t, where $s_{d,t}$ is a set of literals in first-order predicate logic,

Find: A set R of potential risks, where each potential risk r_i consists of an object-action tuple (o_i, a_i).

We elaborate on the representation language of an image description in Sec. 3.3. Roughly, the image description includes a set of qualitatively-represented first-order logical atoms that describe both physical information and symbolic information (e.g., {*wall(Wall), ball(Ball), car(RedCar), left-front-of(Now, Wall, Me)*, ...}). In this paper, we simply assume that the logical forms are correctly obtained from sensors such as LIDAR or vision cameras. (See Sec. 2.2 for the validity of this assumption.)

3.2 Risk prediction as abduction

One simple approach to solving this task is to directly model a mapping between the combinations of observed ground-level clues and potential risks. For example, one could create a straightforward rule like: “if (i) there is a child x, (ii) x is in left front of the ego-vehicle and (iii) x has an umbrella, then x might rush out.” This kind of rule could be implemented in a production system or a knowledge-based system (a.k.a. expert systems), or could be learned by feature-based machine learning algorithms. However, such approaches to directly mapping observable clues to risks may not be robust enough to deal with a large variety of unseen

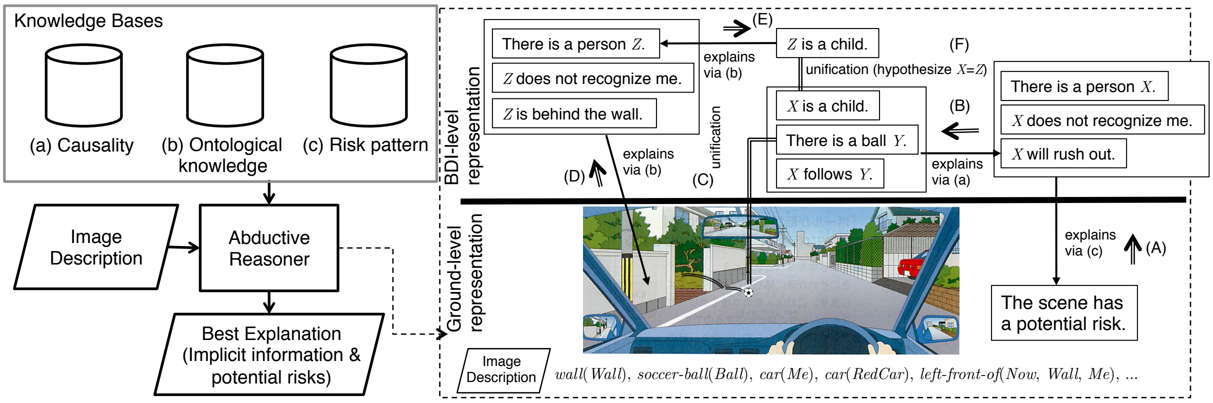

Fig 2. Overall framework. Arrows with alphabets represent information flow.

cases in real-life applications.

The main idea of our approach is as follows. We keep the prediction rules as general as possible, and try to find the best explanation for the condition of general prediction rules based on multiple pieces of evidence at different levels of abstraction. One can view the process of finding the best explanations as the projection of observed information onto a latent space of features useful for risk prediction. Introducing a latent feature space has been proved effective for a wide range of AI tasks [4, 8].

In our task, the latent information we particularly consider useful includes the beliefs, desires, and intentions (BDI) of each traffic object appearing in a given traffic scene. If the BDI information of objects is available, it can then be used to predict their next actions, which leads to a better prediction of potential risks. BDI information is not observable, but may sometimes be inferable from the observable behavior of the object and the observable information on the traffic environment.

Our overall framework is shown in Figure 2. To formulate the above idea, we regard the task of potential risk prediction as the task of abductive theorem proving. This is analogous to the spirit of abductive text understanding [5, 12]; however, our work is the first work to apply abduction to the task of potential risk prediction. That is, given an image description s, we try to prove a proposition “there exist some risks” from s, using a background knowledge base. More specifically, we perform abductive reasoning, regarding an input image description and the “risk’" proposition as observation O. For background knowledge B, we use three kinds of knowledge bases: (a) causality (e.g., “a pedestrian must stop when the traffic signal is red,” “a pedestrian who does not recognize the ego-vehicle rushes out”), (b) ontological knowledge (e.g., “trucks and buses are both large vehicles”), and (c) risk patterns (e.g., “a person can rush out”).

Let us illustrate the reasoning process with one possible explanation shown in Figure 2. The explanation is generated with the following reasoning processes: (A) given the input scene, we hypothesize that there is a person X who does not recognize me and who would rush out, using (c) the knowledge about potential risks; (B) hypothesize X to be the child who follows a ball Y, using (a) the knowledge about causality; (C) hypothesize the ball Y to be identical to the ball in the observed ground-level information; (D) hypothesize there is a person Z behind the wall who does not recognize me; (E) hypothesize Z to be a child; (F) hypothesize Z and X to be identical. Based on this explanation, we identify "a child who follows the observed soccer ball behind the wall and who will rush out." as a risky object-action tuple.

The main advantage of using abduction is characterized as its declarative nature. We can abstract away from the process of inferences, concentrating on creating a sophisticated knowledge base in the declarative fashion. For example, we do not specify the order of reasoning processes (A) through (F) in Figure 2 beforehand; an abductive inference engine finds the best way to apply the inference rules in a knowledge base. In contrast, procedural modeling additionally requires us to specify when, where, and how to use the knowledge base. This may not be a promising approach in cases where we want to combine several kinds of inferences. For those who are not familiar with the notion of declarative and procedural modeling, see Sec. 2.1 for further descriptions.

We have another important advantage in using abduction for potential risk prediction, which is that the latent information inferred by abduction is interpretable by a human. As described in Sec. 1, it is important for potential risk prediction systems to provide car drivers with deeper explanations as to why the system derives a potential risk. To fulfill this requirement, we can exploit the inferred latent information as the explanation for a potential risk. For example, in the above example, we can provide "children are playing with the soccer ball behind the wall" as the explanation for the potential risk "an invisible child will rush out." In contrast, other machine-learning techniques using latent information such as deep neural networks do not produce interpretable latent information.

3.3 Knowledge representation

The first-order language we use for representing our background knowledge is the following. One may think of some other alternative ways of knowledge representation. We leave the further exploration of the issue of knowledge representation to future work.

Constants. Throughout all traffic scenes, the following constants are used: *Me*: the ego-vehicle, *MyLane*: the current own lane, *OppositeLane*: the opposite lane, *MyWalkingLane*: the walking lane on the ego-vehicle side, *OppositeWalkingLane*: the walking lane on the opposite side of the ego-vehicle. *Now*: the current time, *Future*: sometime in the future. Also, all the objects in a traffic scene are represented by constants (e.g., *RedCar*, *MotorCycle*).

Predicates. To describe properties of objects (e.g., type) and the relation between objects (e.g., relative position) in a traffic scene, our language includes the following predicates:

- Type of object: one-place predicate representing a concept type of object. We enumerate about 50 typical types of traffic objects from material for driving lessons. For example, *signal(x)* means that x is a signal.
- State of object: one-place predicate. We define the following predicates: *wet(x), icy(x),*

muddy(x), snowy(x).

- Time-sensitive state of object: two-place predicate representing the state of an object at a certain time. We define about 20 types of predicates to represent states. For example, *left-head-lamp-on(t,x)* means that a head lamp of *x* is turned on at time *t*. The time is represented by either *Now* or *Future* as defined above.
- Intention of object: one-place predicate representing the intention of an intentional object. The predicates include *will-rush-out(x), will-avoid(x), will-go-front(x)*, etc. The main task of potential risk prediction is to infer the intentions of objects.
- Relative position between objects: more than one-place predicate representing the relative position between objects at a certain time. We define about 15 predicates. For example, *in-front-of(t, x, y)* means that *x* is in front of *y* at time *t*.
- Potential risk: *risk(r,p)* representing that there exists risk *r* for traffic object *p*.

3.4 Background knowledge

Using the knowledge representation above, we encode our commonsense knowledge as logical axioms. Our background knowledge consists of three kinds of knowledge.

Causality. To infer intentions of objects, we encode a relation between states and intentions in background knowledge. For instance, we write for all *x,y*: *large-vehicle(x) & in-front-of(Now, x, y)* → *will-avoid(y)* to represent the knowledge that if large vehicle *x* (e.g., truck) is in front of vehicle *x, y* is likely to avoid *x*. Note that the knowledge is not intended to express a 100% rigid logical implication.

Ontological knowledge. We encode a conceptual hierarchy and disjointedness between concepts in background knowledge. For instance, we write for all *x*: *bicycle(x)* → *vehicle(x)* to represent the fact that a bicycle is one kind of vehicle. We write for all *x*: *car(x) & bicycle(x)* → ⊥ to represent the fact that something cannot be a car and a bicycle simultaneously.

Risk pattern. Based on observed information and implicit information inferred by the knowledge above, we define a mapping between a traffic scene and potential risks. For example, we write for all *x, y*: *in-front-of(Now, x, y) & will-stop(y)* → *risk(r, y)* to represent that if something *x* would stop in front of vehicle *y*, that will be a potential risk to *y*.

In our experiment, we manually construct a knowledge base, which consists of 32 causal relations, 2,077 entries for ontological knowledge (51 conceptual hierarchy, 2,026 conceptual disjointedness) and 11 risk patterns. The knowledge construction process is all done on the training data mentioned in Sec. 4.2.

3.5 Score function

As described in Sec. 2.3, we model the abductive score function as the weighted linear feature function $\mathbf{w}\cdot\mathbf{f}(H)$, where $\mathbf{w}$ is a real-valued weight vector and $\mathbf{f}$ is a feature function of a hypothesis. In this paper, we design the feature function based on the following intuition: a hypothesis is (un)reliable if the axioms used to derive it are (un)reliable. More specifically, we define the feature function as follows: $\mathbf{f}(H) = \mathbf{f_e}(h_1) + \mathbf{f_e}(h_2) + ... + \mathbf{f_e}(h_n)$, where h_i is an element of *assumptions*(H), and *assumptions*(H) is a set of elemental hypotheses that are not explained by H, B, and O. That is, we characterize the hypothesis H with a set of unexplained literals. Moreover, we decompose the feature function of unexplained elemental hypothesis $\mathbf{f}_e(h)$ as follows: $\mathbf{f}_e(h) = \mathbf{f}_a(a_1) + \mathbf{f}_a(a_2) + ... + \mathbf{f}_a(a_n)$, where a_i is an element of *axioms*(h), and *axioms*(h) is a set of axioms used for hypothesizing h.

For $\mathbf{f}_a(a)$, we simply introduce binary functions for each axiom, namely $\mathbf{f}_a(a)$ = {1 if a is "for all x, y: *in-front-of(Now, x, y) & will-stop(y)* → *risk(r, y)*"; 0 otherwise, 1 if a is "all x: *bicycle(x)* → *vehicle(x)*"; 0 otherwise, ...}. One can design another type of feature function; for example, $\mathbf{f}_a(a)$ = {1 if a is an axiom about risk pattern; 0 otherwise, 1 if a is an axiom about causality; 0 otherwise, ...}, which shares the weights across axioms. Because the performance of risk prediction is sensitive to the design choice of the feature function, we will explore the proper design of the feature function in our future work.

In order to tune the weight vector $\mathbf{w}$, we adopt the machine-learning framework for abduction from [21], replacing the weight update algorithm with Soft Exact Confidence-Weighted Learning [22], the state-of-the-art online machine-learning algorithm. Intuitively, the weight learner finds the weight vector that minimizes the risk prediction error on a training dataset. In our feature design, it learns the importance of the reliability of axioms for representing the reliability of an unexplained hypothesis. Through the learning process, the axioms useful for discriminating risky events from non-risky events are given higher weights, while the axioms not contributing to the risky/non-risky discrimination are given lower weights. This enables the prediction model to be more robust for redundancy or inconsistency between inference rules in the knowledge base.

4 Evaluation

In our experiments, we compared our abduction-based model against a simple machine learning-based baseline model that relies only on directly observed ground-level information, using the dataset described in Sec. 4.1. The goal is to estimate how useful the inference of implicit information is for potential risk prediction.

4.1 Dataset

An important issue for ADASs is how to deal robustly with a wide variety of traffic scenes. However, to the best of our knowledge, there is no existing standard benchmark dataset that includes a wide variety of traffic scenes. Prior researchers on risk prediction evaluate their models only on a severely limited variety of scenes.

To create such a benchmark dataset, we collected potential risk prediction problems from the textbook published by Chubu-Nippon-Driver-School [24] and Web materials for driving lessons (http://www.honda.co.jp/safetyinfo/kyt/training; http://www.jaf.or.jp/eco-safety/safe ty/danger; http://www.bridgestone.co.jp/csr/tiresafety/training; http://www.nasva.go.jp/fusegu/). The materials cover a wide range of risky situations, from pedestrians rushing out at a crossing to slipping on icy roads. Each problem consists of an illustrated traffic scene from the driver's viewpoint and one or a few typical risks in the traffic scene annotated by a human.

For each problem, we manually convert the traffic scene into a logical form and specify the object-action tuples that are described as a potential risk in the problem description. As a result, we extract 93 problems from Chubu-Nippon-Driver- School [24] as training data, and 100 unseen problems randomly sampled from the web materials as test data. The dataset contains 14 traffic objects, 1.6 risky object-action tuples per scene on average, and 4 types of risky actions in total: rushing out, sudden slowdown, taking control of someone, and passing by. To the best of our knowledge, no previous work has evaluated the proposed systems on such a wide variety of risk prediction problems. We plan to make our dataset publicly available so that it can be used as a shared benchmark in the research community to compare across different ADASs.

4.2 Setting

As an evaluation measure, we use Precision@k and Recall@k which are defined as follows: Precision@k = NCPs(k)/NOs(k); Recall@k = NCPs(k)/NOLs, where NCPs(k) is the number of correct predictions in k-best outputs, NOs(k) is the number of output object-action tuples in k-best outputs, and NOLs is the number of labels in the test set. We do not require each system to infer the types of invisible objects (e.g., it suffices to infer that something behind the wall might rush out in Figure 1). We also use F-measure@k, which is the harmonic mean of Precision@k and Recall@k.

To obtain k-best explanations, we follow [15]. We perform abductive reasoning k times, imposing a constraint that the best explanation found in the n-th turn should not include the potential risks found in 1, 2, ..., (n-1)-th turns. For example, if the best explanation contains the potential risk (*Car1*, *will-rush-out*), we find the second-best explanation under the constraint that (*Car1*, *will-rush-out*) is not included in the best explanation.

For learning and inference, we have extended Phillip (https://github.com/kazeto/phillip), an efficient open-source abductive inference engine that runs on first-order logic [23]. By default, we initialized the weights of axioms with -1 to penalize backward inference. This enables the abductive score function to prefer minimal explanations that can infer annotated potential risks. We also tried the zero-initialization of weights, but it gave worse results.

To evaluate our model, we performed 10-fold cross validation on the test data, where we also used the training data for learning the weights. We emphasize that the training data used for knowledge construction is not included in the test partition of the 10-fold cross validation.

4.3 Baseline model

In order to understand the difficulty of potential risk prediction, we created (i) a baseline

Model	Precision@k	Recall@k	F
Baseline (random)	1.0 (1/100)	0.6 (1/161)	1.1
Baseline (majority)	22.6 (95/420)	**59.0 (95/161)**	32.7
Baseline (SVM, k=1)	30.0 (30/100)	18.6 (30/161)	23.1
Baseline (SVM, k=2)	30.5 (61/200)	37.9 (61/161)	33.9
Baseline (SVM, k=3)	28.0 (84/300)	52.2 (84/161)	**36.5**
Abduction (k=1)	**31.5 (39/124)**	24.2 (36/161)	27.4
Abduction (k=2)	30.3 (59/195)	36.6 (69/161)	33.1
Abduction (k=3)	22.5 (62/276)	38.5 (78/161)	28.4

Table 1. Results of potential risk prediction.

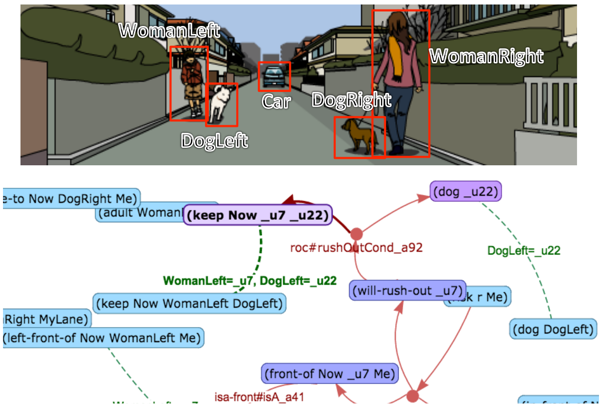

Fig 4. Example inference results.

model that randomly chooses a risky object from a scene and then randomly outputs a risky action from the 4 types of risky action candidates, and (ii) one that simply predicts that all the people and cars in a scene will rush out, which is the major source of potential risks in the training data.

Secondly, we built a baseline model that directly models a mapping between observed information and potential risks. Given an image description, the baseline model enumerates all the possible candidate potential risks (object-action tuples) and ranks them to output k-best risks. We use Ranking SVM [6] to train the ranking model, 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 used a polynomial kernel of degree 2.

4.4 Results

The results of our experiment are shown in Table 1. The random baseline models indicate poor performance because we have 14 traffic objects per scene on average and 4 risky action candidates (i.e., the probability of the output being correct is 1/56=1.8%). The simple majority baseline achieved F measures of 32.7, although the prediction rule is very simple.

Both SVM and abduction models performed better than the simple baseline models in k=3 and k=2, respectively. Contrary to what we expected, the abduction-based model did not outperform the SVM baseline model that relies on only directly observed information.

To understand the potentiality of our abduction-based model more deeply, we analyzed the improved examples. An example is shown in Figure 4. The correct answer for this problem is that the left-hand-side woman might rush out because her dog suddenly starts to run. While the baseline model ranked the correct risk in 2nd place, our abduction-based model ranks it in 1st place. The best explanation for this problem, an output from the visualization module provided by Phillip, is shown in the bottom of Figure 4. Although our abduction-based model could not obtain a significant gain in the overall performance, we believe that abduction-based modeling has great potential for potential risk-oriented ADASs because interpretable explanations are produced.

4.5 Error analysis

We analyzed 30 randomly-sampled error instances and found that in 29 instances the correct risks are at least included in candidate explanations. To understand why these 29 correct risks are ranked lower, we further manually checked whether top-ranked wrong risks are inferred via reasonable inference rules. We found that the majority of the erroneous potential risks are derived via unreasonable inference rules, which is caused by a lack of physical information such as the precise positions and directions of traffic objects.

For example, the system needed to understand that a man is facing a bus stop in order to board a bus and therefore the man is more likely to rush out than other people in the same traffic scene. Similarly, the system needed to understand that the woman is leaving the bus stop, and she is not facing the road. This inference requires the system to perform reasoning about physical information such as the direction of the pedestrian, the position of the bus stop, etc. The current qualitative representation loses such precise information. In future work, we plan to integrate a physics simulator with the current risk prediction system to cope with such quantitative information properly.

5 Conclusions

While previous work on traffic risk recognition has mostly relied on directly observed information, we have explored a context-aware logical abduction-based risk prediction model that exploits implicit information inferable from observed information. Our experiments have shown that the abduction-based model is still not mature compared with a simple machine learning-based model, but indicates that the abduction-based framework has a good potential for solving risk prediction problems. The benchmark dataset and axiomatized knowledge base are to be made publicly available for research purposes.

Clearly, as revealed by the error analysis, our abduction-based model needs non-discretized quantitative information representing physical information of traffic scenes, such as the positions and shapes of traffic objects. Our future directions include exploring the integration of symbolic inference into physics-based simulation, using first-order logic as the interface. At the same time, we also explore the integration of our prediction system with the grounding technologies such as sensor devices and vision cameras through a road test in a more practical situation. To extend our problem setting, we will also continue to develop our benchmark dataset so as to evaluate the discriminative ability of risky vs. non-risky situations of risk prediction systems. For a more practical evaluation, we also plan to use the video traffic scenes recorded via drive recorders as the benchmark dataset.

References

1. Broadhurst A., Baker S., and Kanade T. 2005. Monte carlo road safety reasoning. In Proceedings of the IEEE Intelligent Vehicles Symposium, 319–324. IEEE.
2. Charniak E., and Goldman R.P. 1991. A Probabilistic Model of Plan Recognition. In AAAI, 160–165.
3. Greene D., Liu J., Reich J., Hirokawa Y., Shinagawa A., Ito H., and Mikami T. 2011. An efficient computational architecture for a collision early-warning system for vehicles, pedestrians, and bicyclists. Intelligent Transportation Systems, IEEE Transactions on 12(4): 942–953.
4. Hinton G., Deng, L., Yu D., Dahl G.E., Mohamed A.-r., Jaitly N., Senior A., Vanhoucke V., Nguyen P., Sainath T.N., et al. 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Processing Magazine, IEEE 29(6): 82–97.
5. Hobbs J.R., Stickel M., Martin P., and Edwards D. 1993. Interpretation as abduction. Artificial Intelligence 63:69–142.
6. Joachims T. 2002. Optimizing search engines using click- through data. In Proc. Of ACM SIGKDD, 133–142.
7. Lattner, A.D., Timm I.J., Lorenz M., and Herzog O. 2005. Knowledge-based risk assessment for intelligent vehicles. In Proc. of KI- MAS '05, Waltham, Massachusetts, USE, 191–196.
8. Lee H., Grosse R., Ranganath R., and Ng A.Y. 2009. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In Proceedings of the 26th Annual International Conference on Machine Learning, 609–616. ACM.
9. Lefe`vre S., Vasquez D., and Laugier C. 2014. A survey on motion prediction and risk assessment for intelligent vehicles. ROBOMECH Journal 1(1): 1–14.
10. Ng H.T., and Mooney R.J. 1992. Abductive Plan Recognition and Diagnosis: A Comprehensive Empirical Evaluation. In KR, 499–508.
11. Ortiz M.G., Fritsch J., Kummert F., and Gepperth A. 2011. Behavior prediction at multiple time-scales in inner-city scenarios. In Intelligent Vehicles Symposium (IV), 2011 IEEE, 1068–1073. IEEE.
12. Ovchinnikova E., Montazeri N., Alexandrov T., Hobbs J.R., McCord M.C., and Mulkar-Mehta R. 2011. Abductive Reasoning with a Large Knowledge Base for Discourse Processing. In IWCS.
13. Raghavan S., and Mooney R. J. 2010. Bayesian Abductive Logic Programs. In Star-AI 10, 82–87. Rajamani, R. 2011. Vehicle dynamics and control. Springer.
14. Rendon-Velez E., Horvath I., and Opiyo E.Z. 2009. Progress with situation assessment and risk prediction in advanced driver assistance systems: A survey. In Proceedings of the 16th ITS World Congress, 21–25.
15. Santos Jr. E. 1991. On the generation of alternative explanations with implications for belief revision. In Proc. of UAI, pp. 339–347.
16. Souza C.R., and Santos P.E. 2011. Probabilistic logic reasoning about traffic scenes. In Towards Autonomous Robotic Systems. Springer. 219–230.
17. Benenson R., Omran M., Hosang J., Schiele B. 2014. Ten years of pedestrian detection, what have we learned? In ECCV 2014, CVRSUAD workshop.
18. Ess A., Leibe B., Schindler K., Van Gool L. 2008. A mobile vision system for robust multi-person tracking. In CVPR, IEEE Press (June 2008).
19. M. Enzweiler and D.M. Gavrila. 2009. "Monocular Pedestrian Detection: Survey and Experiments." In IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 31, No. 12, pp. 2179-2195.
20. Bengler K., Dietmayer K., Farber B., Maurer M., Stiller, C and Winner H. 2014. Three Decades of Driver Assistance Systems: Review and Future Perspectives. In Intelligent Transportation Systems Magazine, IEEE (Volume: 6, Issue: 4).
21. Inoue N., Yamamoto K., Watanabe Y., Okazaki N. and Inui K. 2012. Online Large-margin Weight Learning for First-order Logic-based Abduction. In Proceedings of the 15th Information-Based Induction Sciences Workshop, pp. 143-150.
22. Wang J., Zhao P., and Hoi S. 2012. Exact Soft Confidence-Weighted Learning. In Proceedings of 29th International Conference on Machine Learning.
23. Yamamoto K., Inoue N., Inui K., Arase Y., and Tsuji J. 2014. Boosting the Efficiency of First-order Abductive Reasoning Using Pre-estimated Relatedness between Predicates. International Journal of Machine Learning and Computing, Vol. 5, No. 2, pp. 114-120.
24. Chubu-Nippon-Driver-School. 1999. Master of your driving (in Japanese). TOYOTA Nagoya Education Center.