

概要

1. NEologd に含まれていない単語の収集
2. NEologd に含まれる単語のベクトルの学習

Github: <https://github.com/reiyw/intern-line>

1. NEologd に含まれていない単語の収集

- 収集された単語: new_words.txt
 - 例:
 - 言語批判論集 ゲンゴヒハンロンシュウ
 - 言語技術 ゲンゴギジュツ
 - 言語の数理 ゲンゴノスウリ
- リソース: 日本語版 Wikipedia ダンプデータ (20160601)
 - <https://dumps.wikimedia.org/jawiki/20160601/>
 - jawiki-20160601-pages-articles.xml.bz2 2.2 GB
- 方針:
 - できるだけ固有表現のみを抽出
 - 特に, NEologd には作品名が足りてなさそう
 - 正規表現だけでラクに集めてくる
 - 数が必要である場合に API は叩きたくない

手順

1. WikiExtractor (<https://github.com/attardi/wikiextractor>) でダンプされた xml をプレーンテキスト化
 - マークアップの削除など
2. 括弧 “『』” で囲まれている単語を抽出
 - 『』は作品名, 書籍名を表すために用いられる
3. 英数字だけからなる文字列を削除
 - 日本語形態素解析のための辞書なので, 英単語は念のため除外
4. 文字数が 2 以下または 31 以上である文字列を削除
 - 2 以下: 作品名である場合もあるが, ほとんどはノイズ
 - 31 以上: 『』が引用文のために用いられる場合がある
5. 半角の記号類を含む文字列を削除
6. NEologd との重複単語削除
 - LevelDB 使用
7. NEologd を用いた MeCab で読みの付与

作成したプログラムなど

- extract_name_of_work.sh: 前頁の 2-5 を処理

2. NEologd に含まれる単語ベクトルの学習

- リソース: 1 と同様

手順

1. WikiExtractor (<https://github.com/attardi/wikiextractor>) でダンプされた xml をプレーンテキスト化
2. 1行1文に変換
3. 行をランダムに並び替える
4. NEologd を用いた MeCab で形態素解析
5. 学習したい単語の基本形を抽出
 - 名詞・動詞・形容詞だけを残す
 - ただし、非自立な動詞・名詞、トップワードは削除する
6. word2vec (<https://github.com/dav/word2vec>) で学習
 - 300次元とした以外はデフォルトのパラメタを使用

作成したプログラム

- make_corpus.sh: 前頁の 2-3 を処理
- mecab2words.py: 前頁の 5 を処理

自己 PR

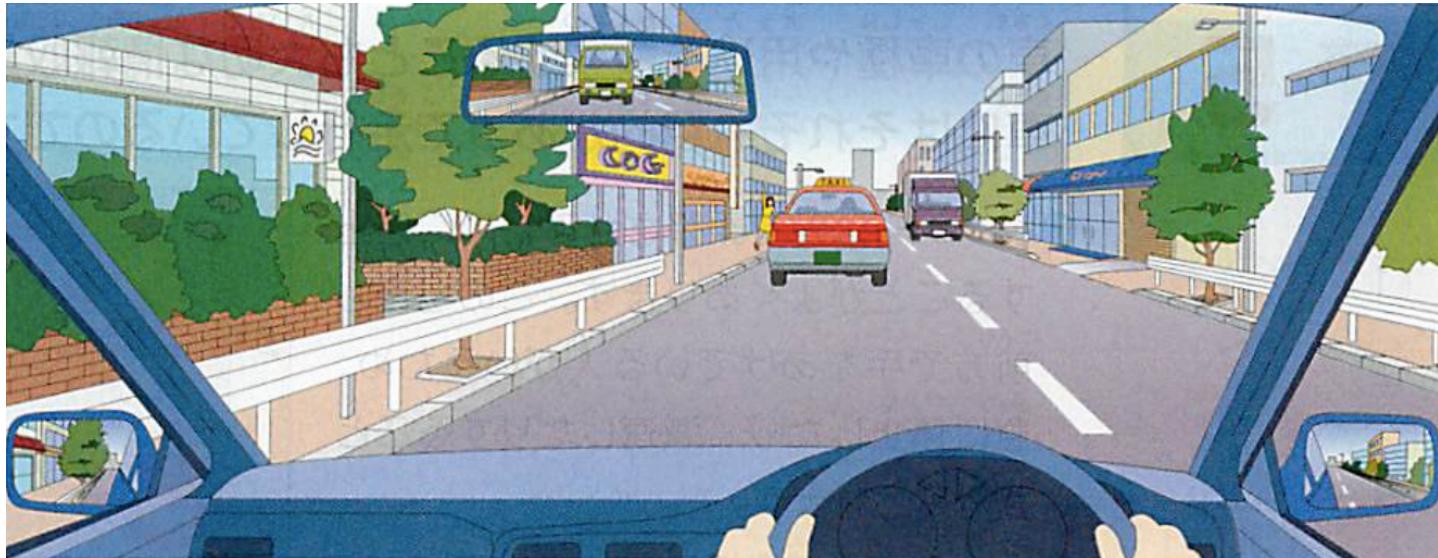
- 研究についてはエントリーシートに書いた通りです。最近人工知能学会で発表しましたので、その時のスライドを参考に添付しておきます。

交通オントロジーと 説明生成に基づく交通危険予測

高橋 諒 (東北大),
井之上 直也 (東北大),
栗谷 康隆 (DENSO CORPORATION),
山本 風人 (東北大),
乾 健太郎 (東北大)

本研究のタスクと狙い

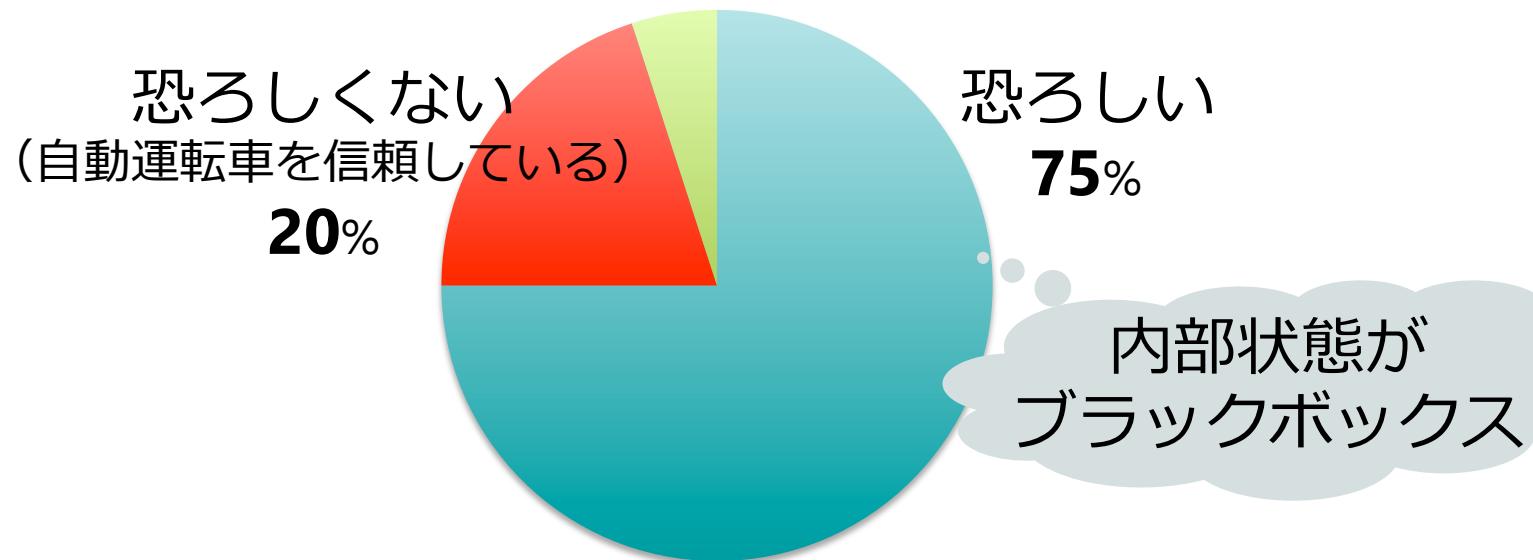
- 都市交通シーンにおける潜在的な危険を予測
 - 特に、衝突まで 2 秒以上かかるような危険



- 何が：タクシーが急停止して危険
- なぜ：タクシーは人を乗せるために急停止するものであるから

「なぜ」を提示する必要性が高まっている

自動運転車に運転を任せることが

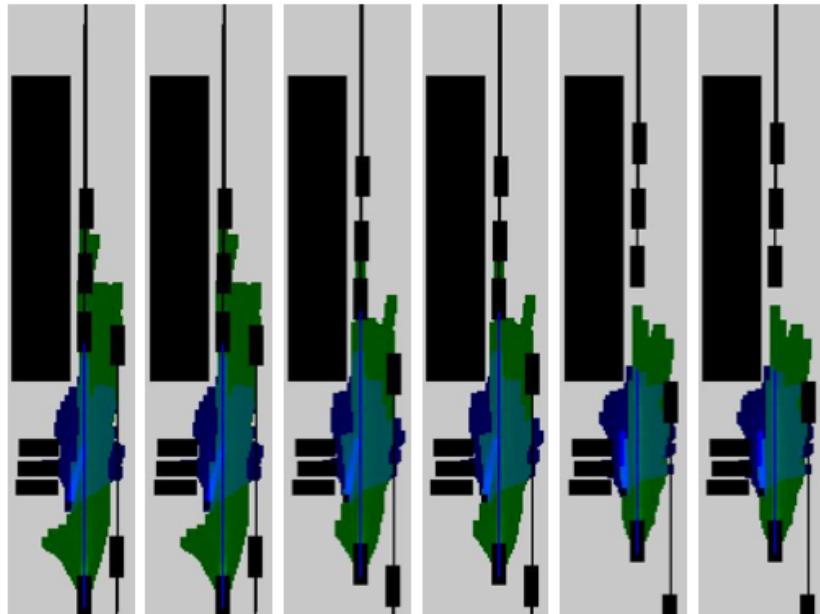


出典：全米自動車協会 (2016)

自動運転車の判断の根拠（なぜ）を
可視化することでホワイトボックス化の一助に

直接軌道を予測して危険予測する先行研究

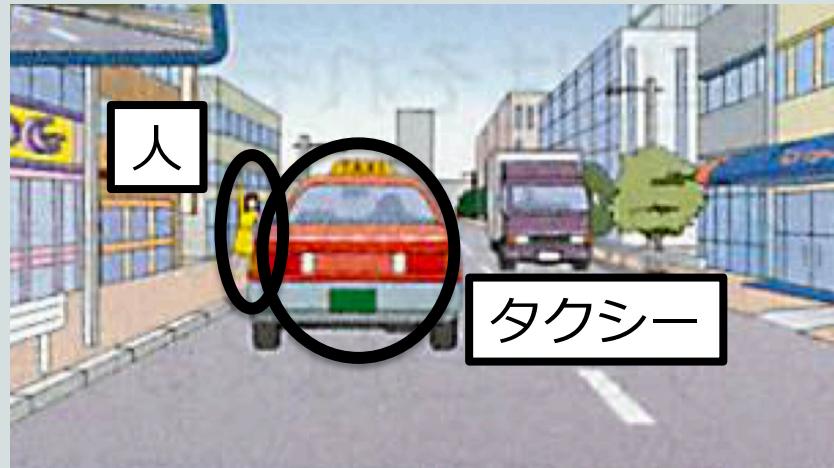
- 物理シミュレーションに基づく危険予測モデル



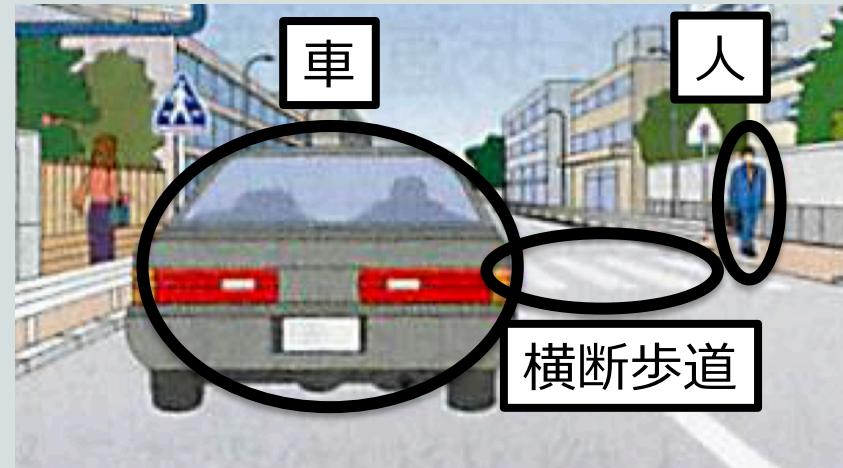
[Broadhurst+ 05]

- ✓ 衝突するまでの時間 (TTC) を含む精緻な予測が可能
- ✗ 何らかの目的を達成するための将来の行動変化を捉えられない

何らかの目的のために行動変化する例



タクシー \wedge 人
 \Downarrow
タクシーが人を
乗せるために止まる



車 \wedge 横断歩道 \wedge 人
 \Downarrow
横断歩道を渡る人の
ために車が止まる

何らかの目的達成のための行動変化を捉えるには
物体同士の関係のモデルが必要 [Rendon-Velez+ 09]

物体同士の関係のモデルを含む先行研究

- オントロジーを用いて関係をモデル化し、論理推論のためのルールを記述してそれぞれのタスクに適用 [Armand+ 14, Mohammad+ 15, Zhao+ 15]



歩行者(p) \wedge 車道(r) \wedge 車道上(p, r)
 \Downarrow
高リスク(p)

- ✓ 説明的：ホワイトボックス
- ✗ TTC を含む精緻な危険予測が苦手
- ✗ 観測の不確実性への対応ができない

アイデア | 定性推論と定量推論を組み合わせる

アプローチ	利点	欠点
軌道計算 (定量推論)	<ul style="list-style-type: none">精緻な危険予測	<ul style="list-style-type: none">物体同士の関係を捉えるのが難しい説明的でない (論理推論と比べて)
論理推論 (定性推論)	<ul style="list-style-type: none">物体同士の関係を捉える説明的	<ul style="list-style-type: none">精緻でない<u>観測の不確実性への対応不可</u>

仮説推論を応用

二つのアプローチを統合して
それぞれの利点を活かす

枠組み

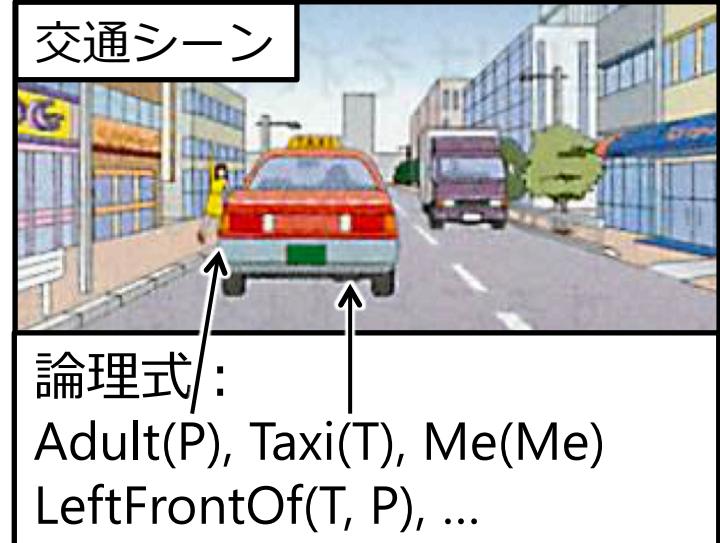
危険予測問題の定義

入力

センサー（カメラ, LIDAR など）
より得られる情報

- 定量情報
(物体の形状, 位置, 速度)
- 交通シーンを記述する論理式
※論文参照

交通シーン



出力

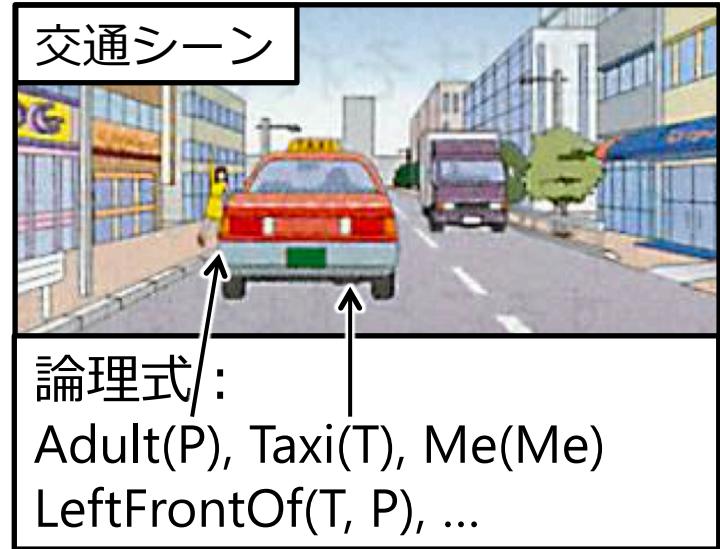
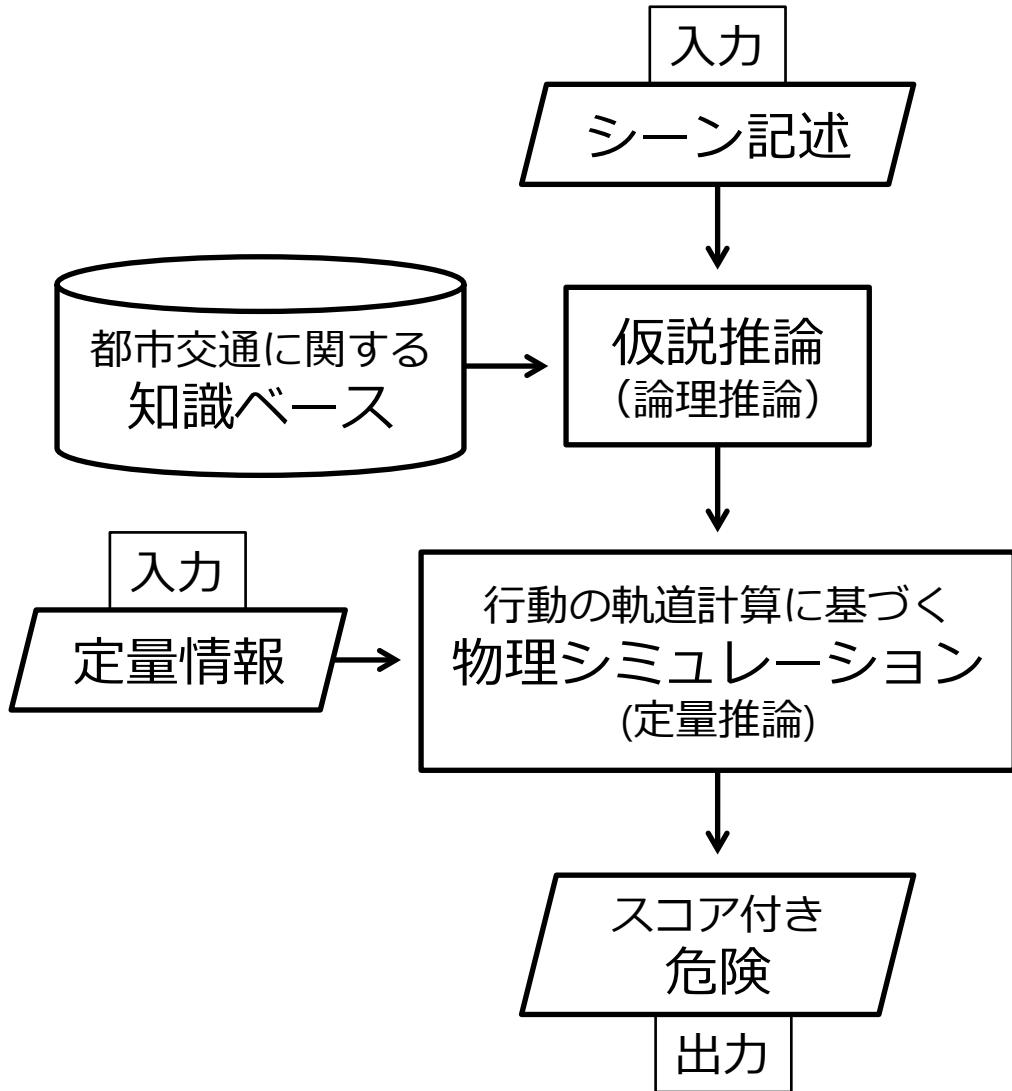
低いほど
危険

自車の観点でのスコア付き危険

- (移動体, 行動) のペア
- その説明

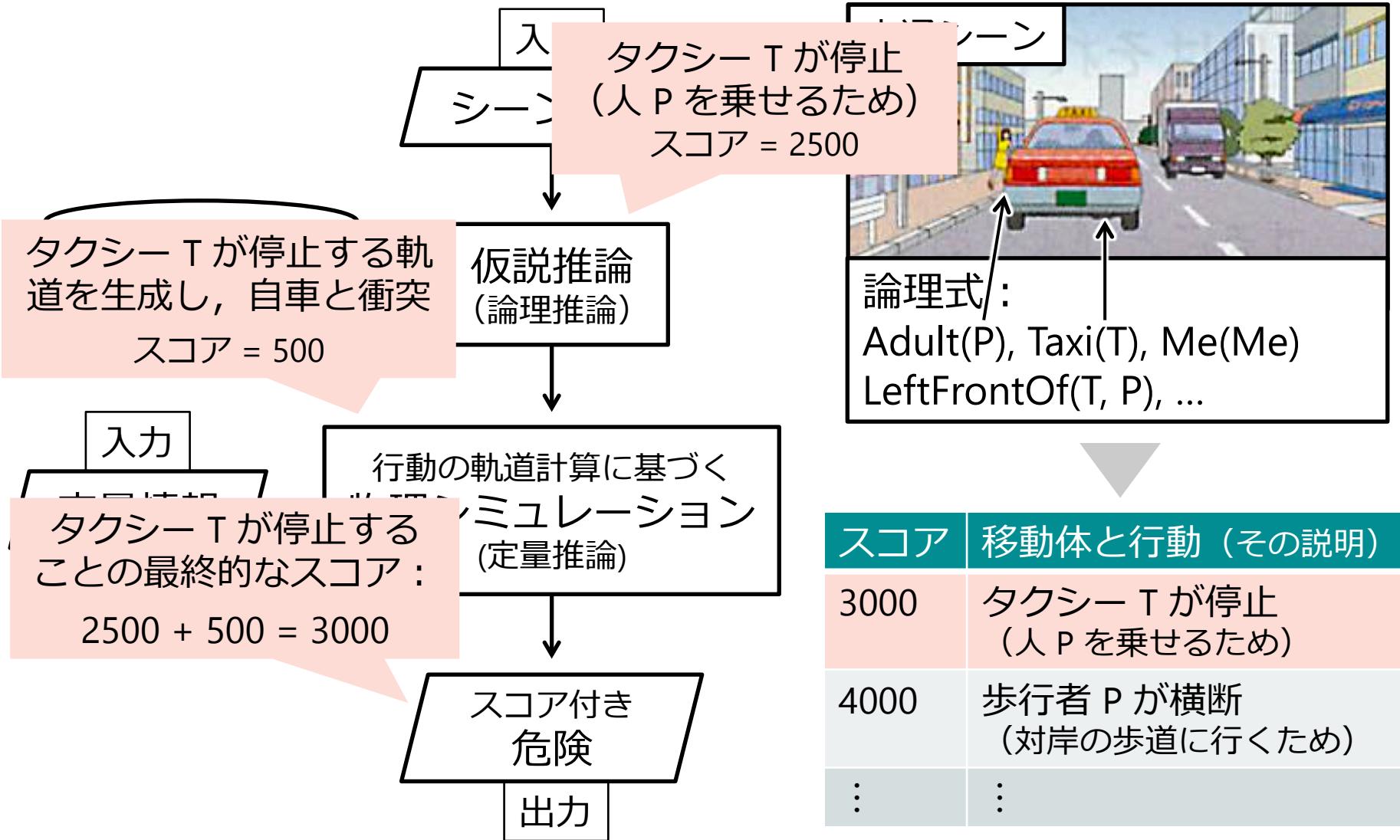
スコア	移動体と行動 (その説明)
3000	タクシー T が停止 (人 P を乗せるため)
4000	歩行者 P が横断 (対岸の歩道に行くため)
:	:

危険予測の枠組み

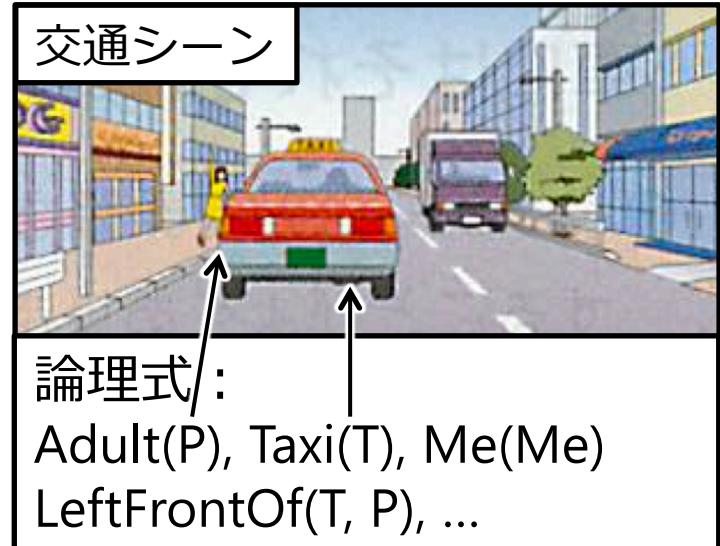
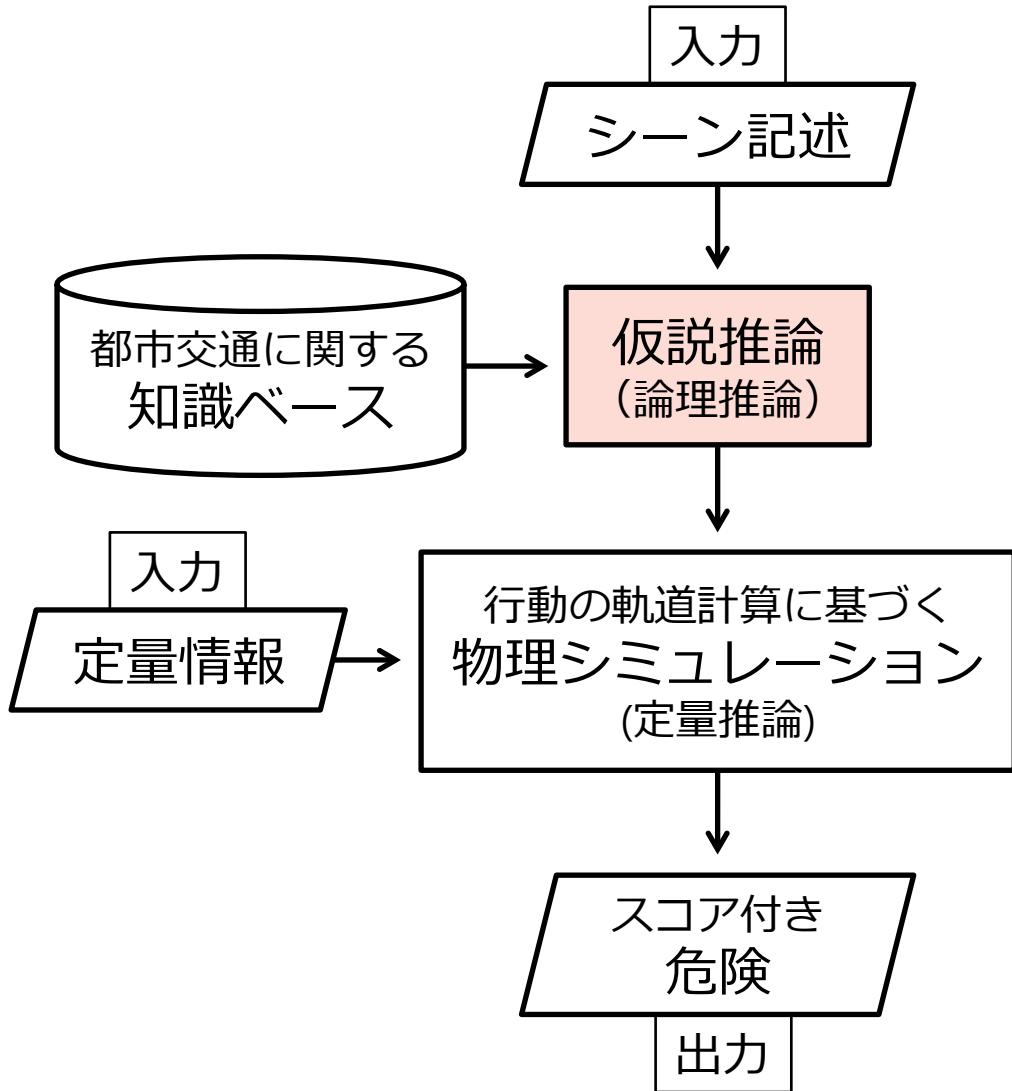


スコア	移動体と行動（その説明）
3000	タクシー T が停止 (人 P を乗せるため)
4000	歩行者 P が横断 (対岸の歩道に行くため)
⋮	⋮

危険予測の枠組み



危険予測の枠組み



スコア	移動体と行動（その説明）
3000	タクシー T が停止 (人 P を乗せるため)
4000	歩行者 P が横断 (対岸の歩道に行くため)
⋮	⋮

仮説推論とは

観測に対する最良の仮説（説明）を求める推論

仮説



スプリンクラー消し忘れ？ 夜のうちに雨が降った？



観測



自宅の庭が濡れている



隣りの庭も濡れている

仮説推論を行動予測に用いる

※スコア関数: [Hobbs 93]

背景知識

「大人は歩行者である」

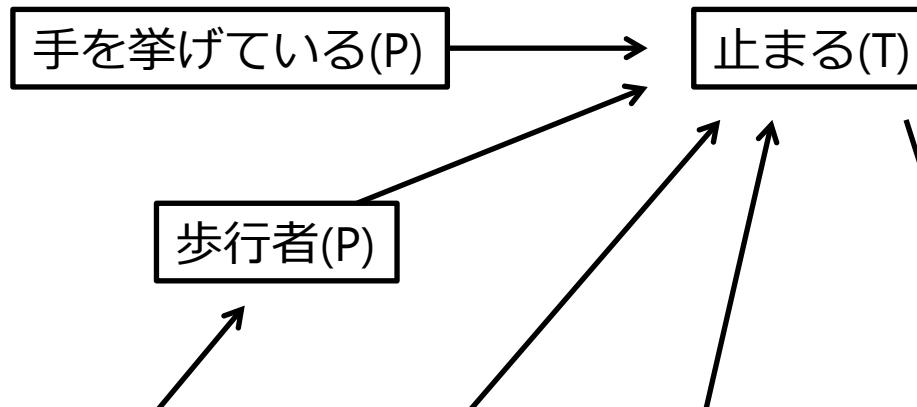
大人(x) → 歩行者(x)

「タクシーは客を乗せるために止まる」

タクシー(t) ∧ 歩行者(p) ∧ 左前方(t, p) ∧ 手を挙げている(p) → 止まる(t)



仮説



観測

自車(M) ∧ 大人(P) ∧ タクシー(T) ∧ 左前方(T, P) ∧ 何らかの行動(T)

仮説推論を行動予測に用いる

※スコア関数: [Hobbs 93]

背景知識

「大人は歩行者である」

処理の手順を捨象した宣言的知識

「タクシーは客を乗せるために止まる」



タクシー(t) \wedge 歩行者(p) \wedge 左前方(t, p) \wedge 手を挙げている(p) \rightarrow 止まる(t)

仮説

手を挙げている(P)

止まる(T)

観測の不確実性に自然に対応

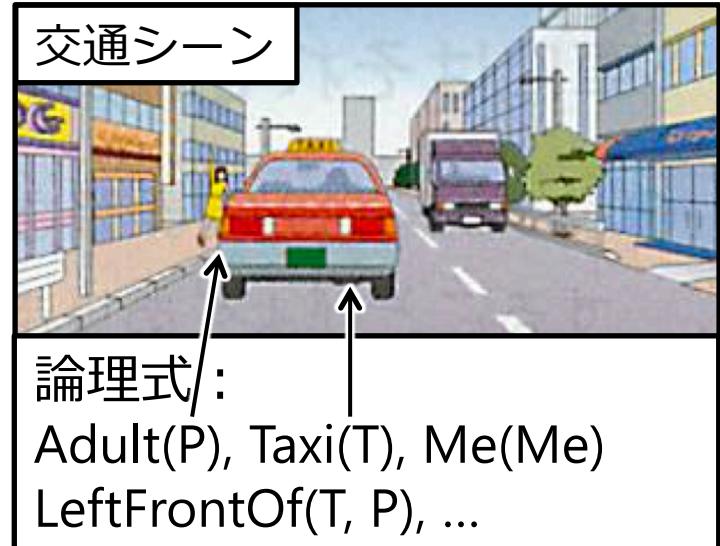
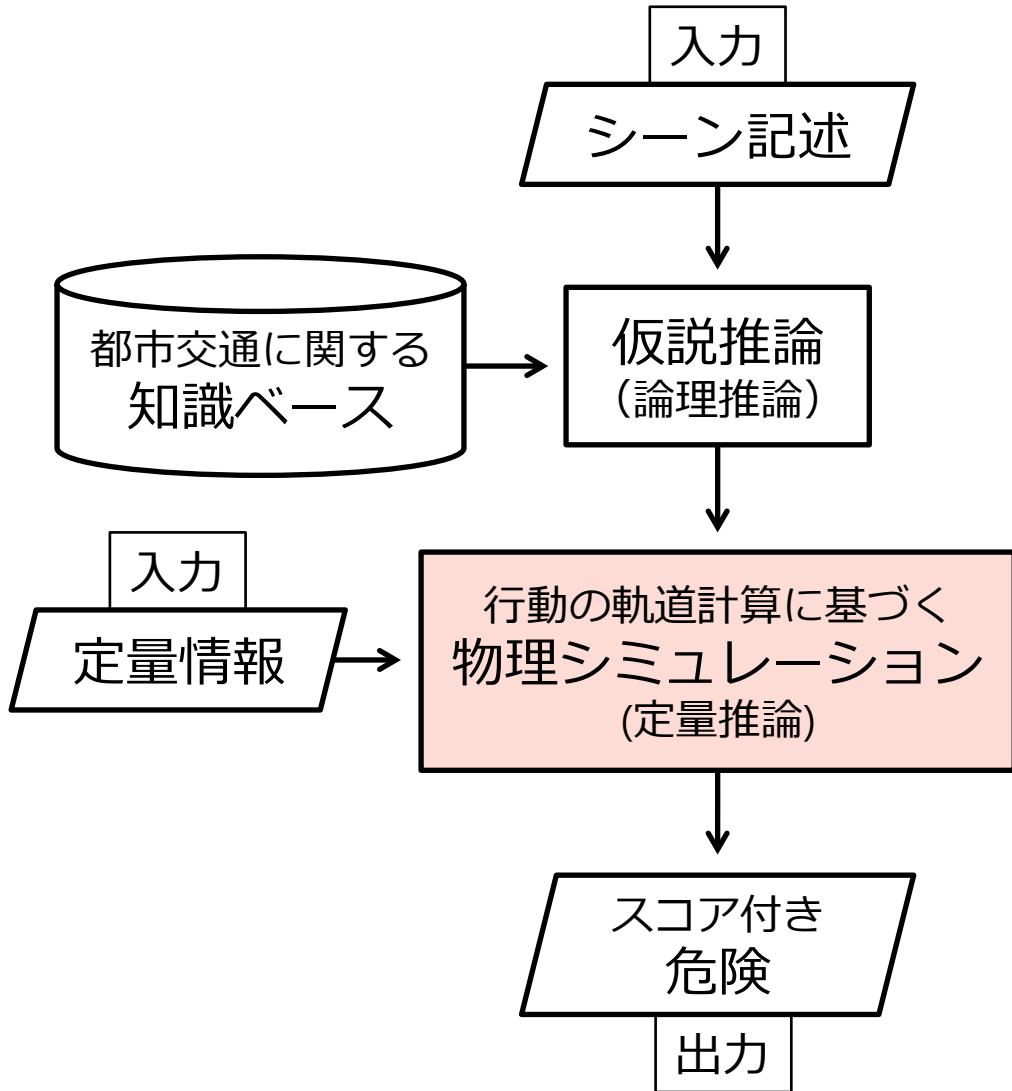
歩行者(P)

仮説 = 予測に対する説明

観測

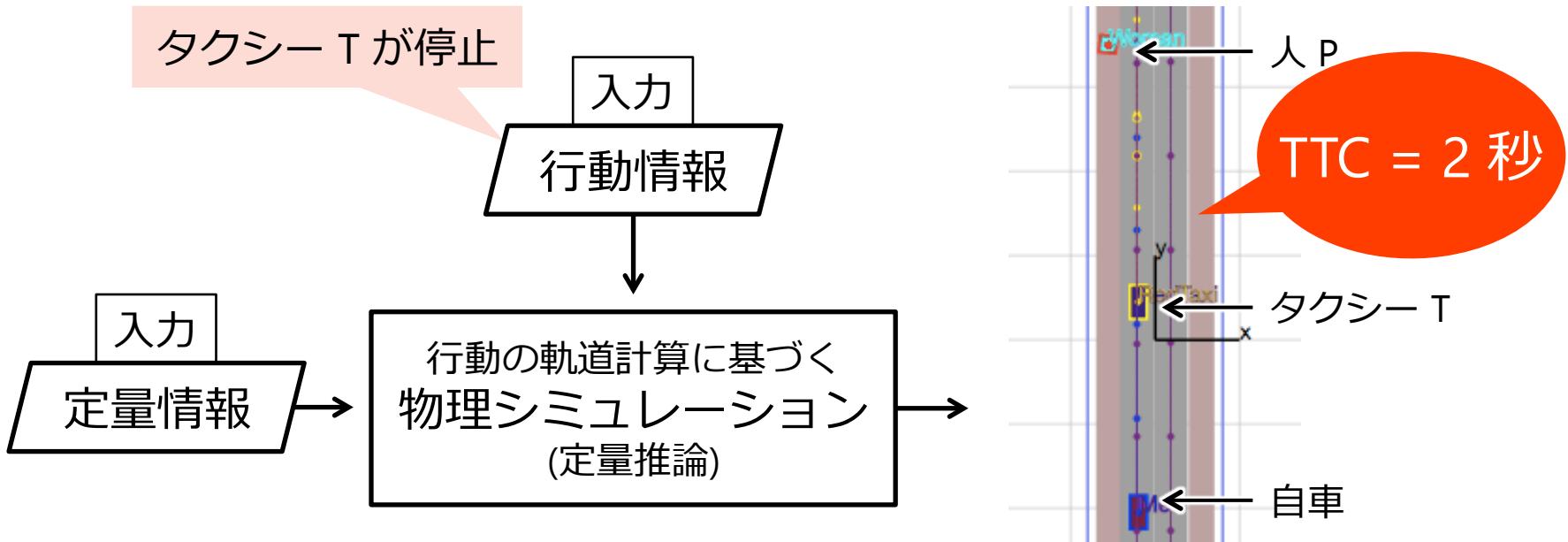
自車(M) \wedge 大人(P) \wedge タクシー(T) \wedge 左前方(T, P) \wedge 何らかの行動(T)

危険予測の枠組み



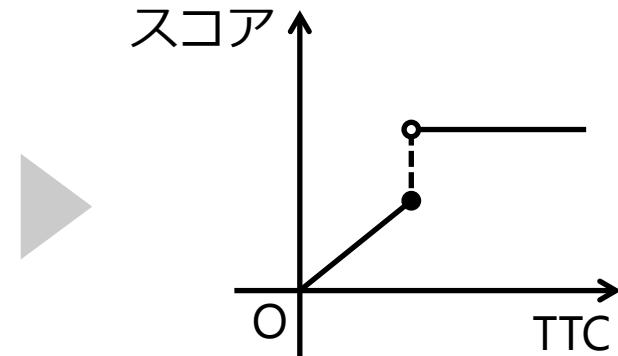
スコア	移動体と行動（その説明）
3000	タクシー T が停止 (人 P を乗せるため)
4000	歩行者 P が横断 (対岸の歩道に行くため)
⋮	⋮

行動の軌道計算に基づく物理シミュレーション



スコア関数

- TTC が小さいほうがより危険
- TTC が大きいときは仮説推論を信用



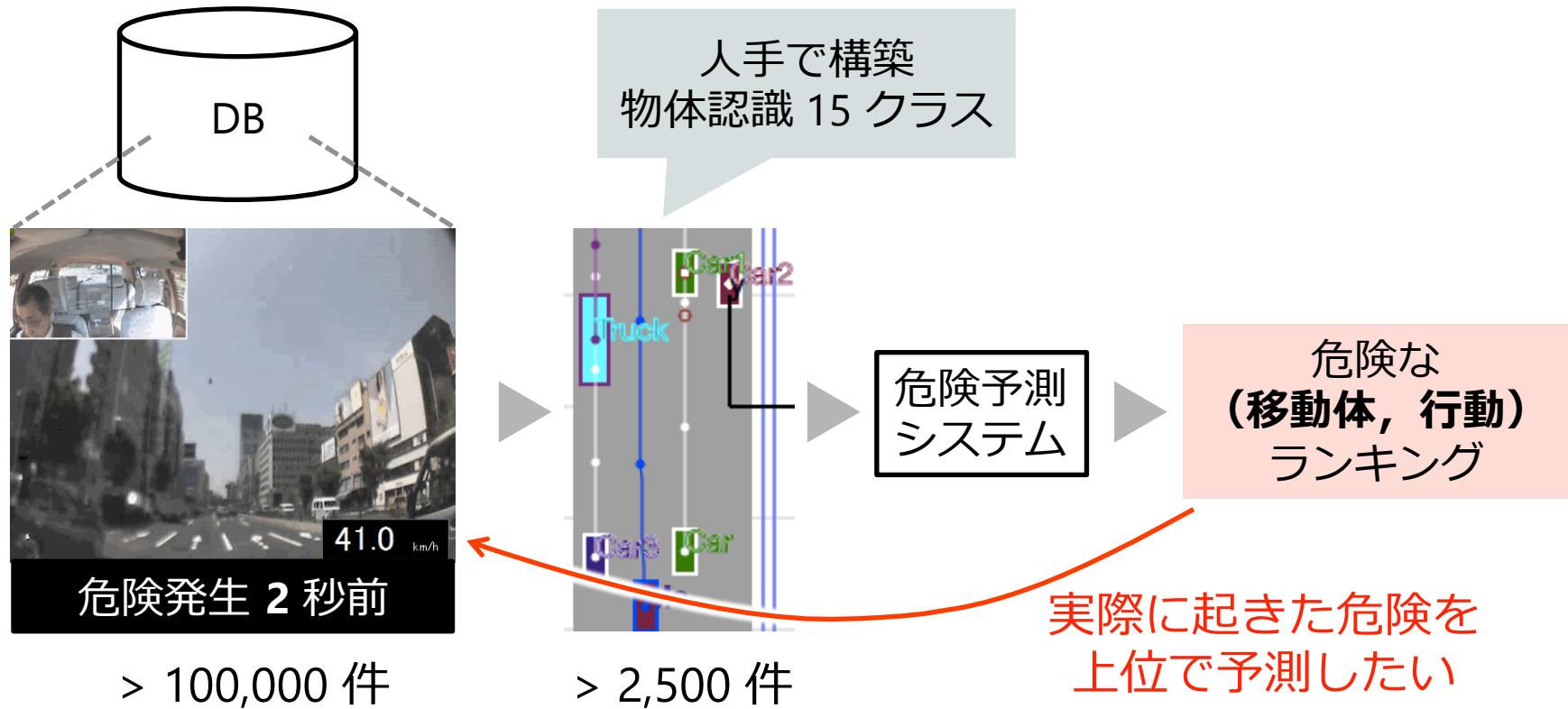
実事例での評価実験

実験設定

評価用データセット

東京農工大学大学院 ドラレコデータベース

<http://web.tuat.ac.jp/~smrc/drcenter.html>



論理推論部 | one-best での評価結果

表 1: 行動予測の性能

モデル	Precision	Recall	F
MAJORITY	24.0 (6/25)	24.0 (6/25)	24.0
DEDUCTIVE	60.0 (15/25)	60.0 (15/25)	60.0
ABDUCTIVE	64.0 (16/25)	64.0 (16/25)	64.0

仮説不可能な
定理証明器

仮説推論

期待に反して性能向上はわずか 😞

- ▷ 観測が十分なため、仮説の必要がある問題が少ない
- ▷ DEDUCTIVE でも擬似的に “仮説可能な” 定理証明に

期待通りに動作した例



スコア	移動体と行動
4220.5	車1 が 停止
4220.5	車1 が 右車線変更
4220.5	車1 が 左車線変更
4220.5	車2 が 停止
4220.5	車2 が 右車線変更
4220.5	トラック が 停止
:	:



スコア	移動体と行動
14220.5	車1 が 停止
14220.5	車1 が 右車線変更
4310.5	車1 が 左車線変更
14220.5	車2 が 停止
14220.5	車2 が 右車線変更
14220.5	トラック が 停止
:	:

なぜ
前方に停止
車両があるので
左車線変更

問題点・限界

問題点

- 「衝突（しないけど）しそう」を考慮できない
 - ▶ Time to Closest Encounter (TCE)
 - ▶ Distance of Closest Encounter (DCE)

原理的にできない

- 行動の連鎖 (chain reaction) を捉えられない
 - 前方の車が
「何らかの理由のために急停止する前方の車」
のために急停止

まとめと今後の課題

まとめ

- 交通危険予測のために、仮説推論と物理シミュレーションを統合
- 説明生成に基づくので、なぜ危険かまで含めた予測が可能
- 交通危険予測の文脈では、これまでで最も大規模な実世界を元にしたデータセットを作成

今後の課題

- 説明（なぜ）の定量評価
- 実際のセンサー機器と統合