

NII



統計的機械学習と 演繹的形式推論

システムの信頼性と説明可能性へのアプローチ

蓮尾 一郎

国立情報学研究所 (NII)

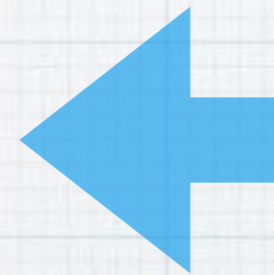
JST ERATO 蓮尾メタ数理システムデザインプロジェクト

- 東大 (BSc 数学), 東工大 (MSc 論理学), U Nijmegen (PhD 計算機科学)
- 京大数理研 → 東大情報理工 → NII
- さきがけ西浦領域OB
- 専門: 理論計算機科学, 論理学, 計算機科学における代数的・圏論的構造



https://commons.wikimedia.org/wiki/File:Waymo_self-driving_car_front_view.gk.jpg

高機能・高効率のシステムの
設計・実装
および
安全性保証
説明責任



統計的機械学習
演繹的形式推論

識別ルール（犬っぽさ，猫っぽさ）を，
パラメータ（重み）として自動で発見

統計的機械学習

例：ニューラルネット（NN）で犬と猫を判別

データ

（ラベル付き画像）
を準備

$$\left\{ \begin{array}{l} (\text{pic}_1, \text{label}_1), \\ \dots, \\ (\text{pic}_K, \text{label}_K) \end{array} \right\}$$

where $\text{pic}_i \in \mathbb{R}^{640 \times 480}$,
 $\text{label}_i \in \{\text{dog}, \text{cat}\}$

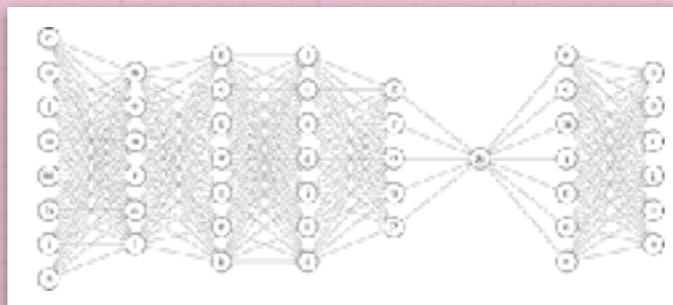


<https://an.wikimedia.org/wiki/Date:Media/>



ニューラルネット
を準備

$$\mathbb{R}^{640 \times 480} \xrightarrow{f_p} \{\text{dog}, \text{cat}\}$$

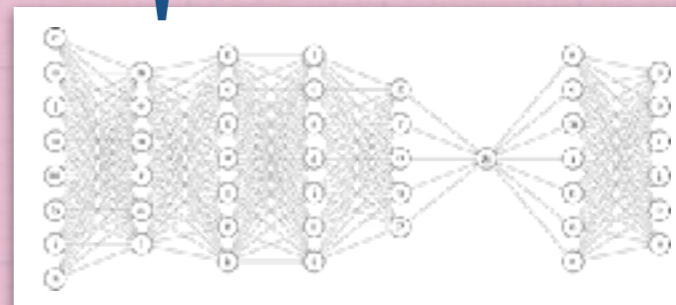


ニューラルネット
を訓練

Find p such that

$$\text{label}_i \approx f_p(\text{pic}_i)$$

Gradient descent,
backpropagationで
パラメータ（重み）を
更新



数学における証明と同じ

記号化 → 計算機による支援・自動化

演繹的形式証明

公理を準備

ワン！と鳴けば犬である

うちのポチはワン！と鳴く
(観察)

公理と推論規則から命題を演繹

* 推論規則の例：
三段論法

$$\frac{A \quad A \supset B}{B}$$

* 演繹 (証明) の例

1. ポチはワンと鳴く
(公理より)
2. ワンと鳴けば犬である
(公理より)
3. ポチは犬である
(1, 2, 三段論法より)

* 応用上は、すでに与えられた結論に対し、証明を探索することも多い



統計的機械学習 vs 演繹的形式推論



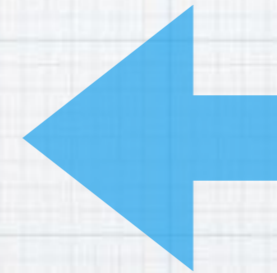
$$\frac{A \quad A \supset B}{B}$$

統計的機械学習		演繹的形式推論
データのノイズを許容	入力の誤り	公理は絶対 誤りは想定せず
保証されない	結論の正しさ	論理的に保証 (cf. 数学的証明)
高い データから自動で特徴量発見	スケーラビリティ	低い 公理の準備は人力 (cf. エキスパートシステム)
低い 判断の理由はパラメータ (重み)	説明可能性	高い 推論過程が証明として明示的



https://commons.wikimedia.org/wiki/File:Waymo_self-driving_car_front_view.gk.jpg

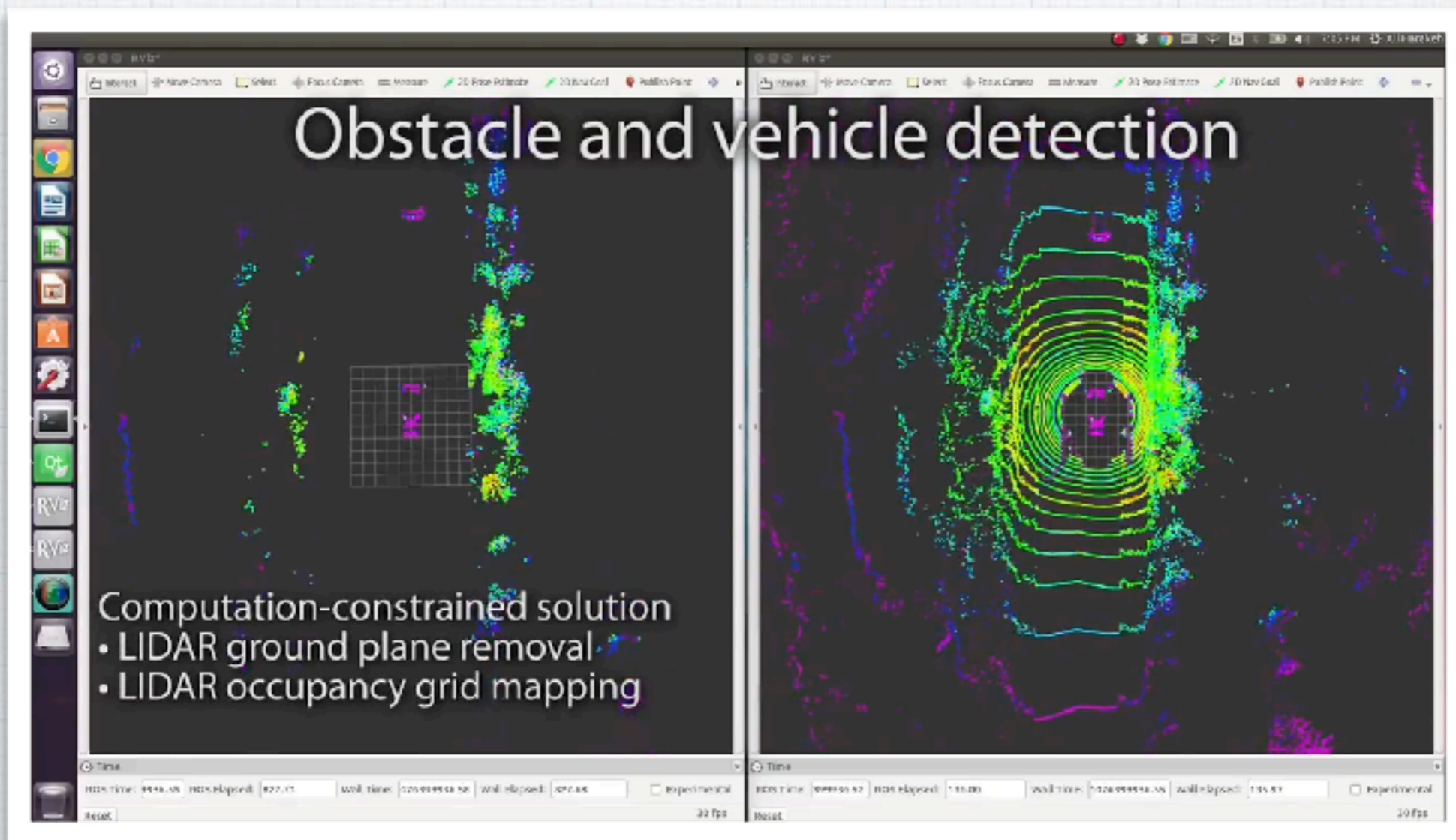
高機能・高効率のシステムの
設計・実装
および
安全性保証
説明責任



統計的機械学習
演繹的形式推論

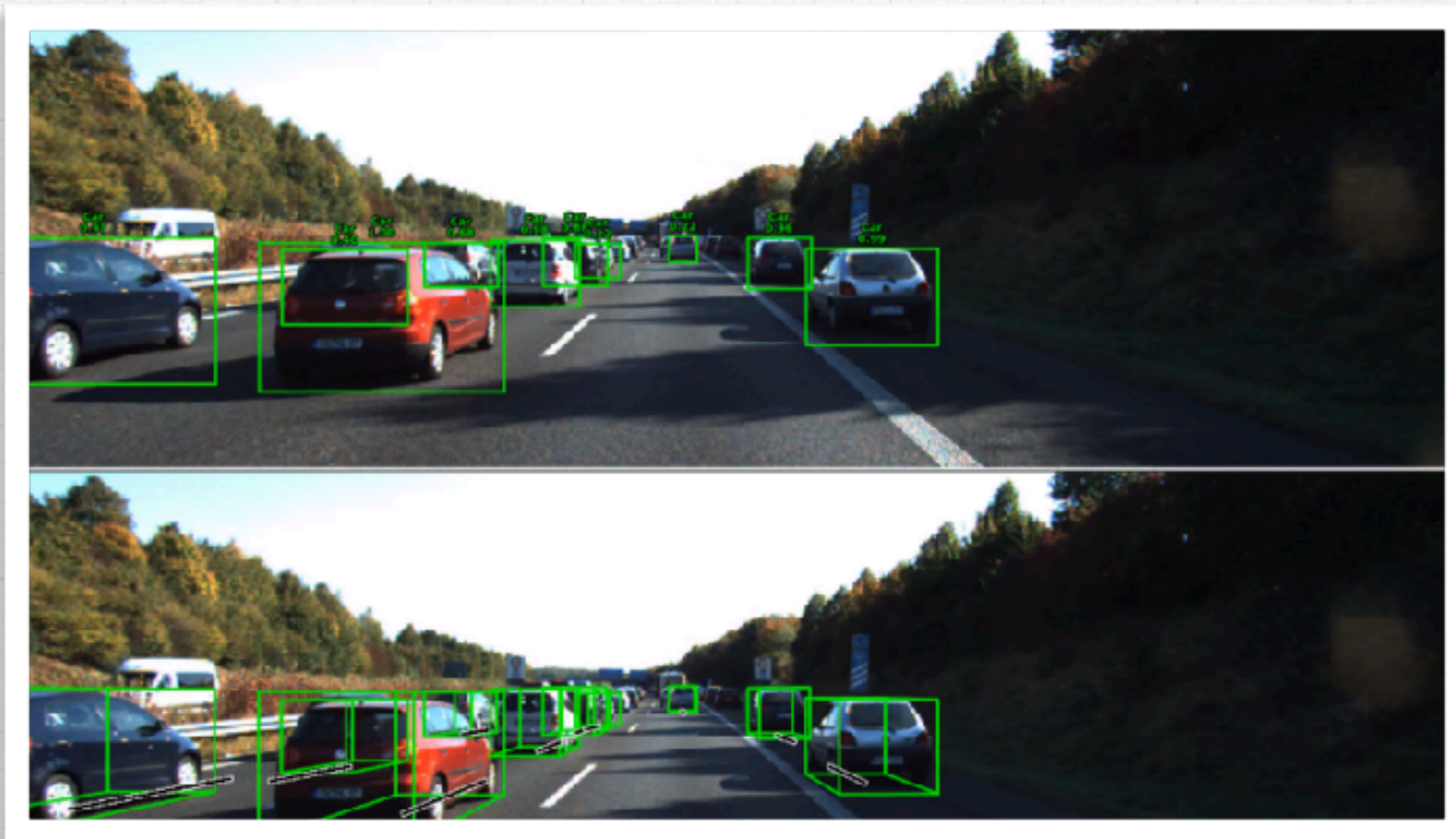
システムの例：自動運転

* センシング → 物体認識 → 経路計画 → 経路トレース



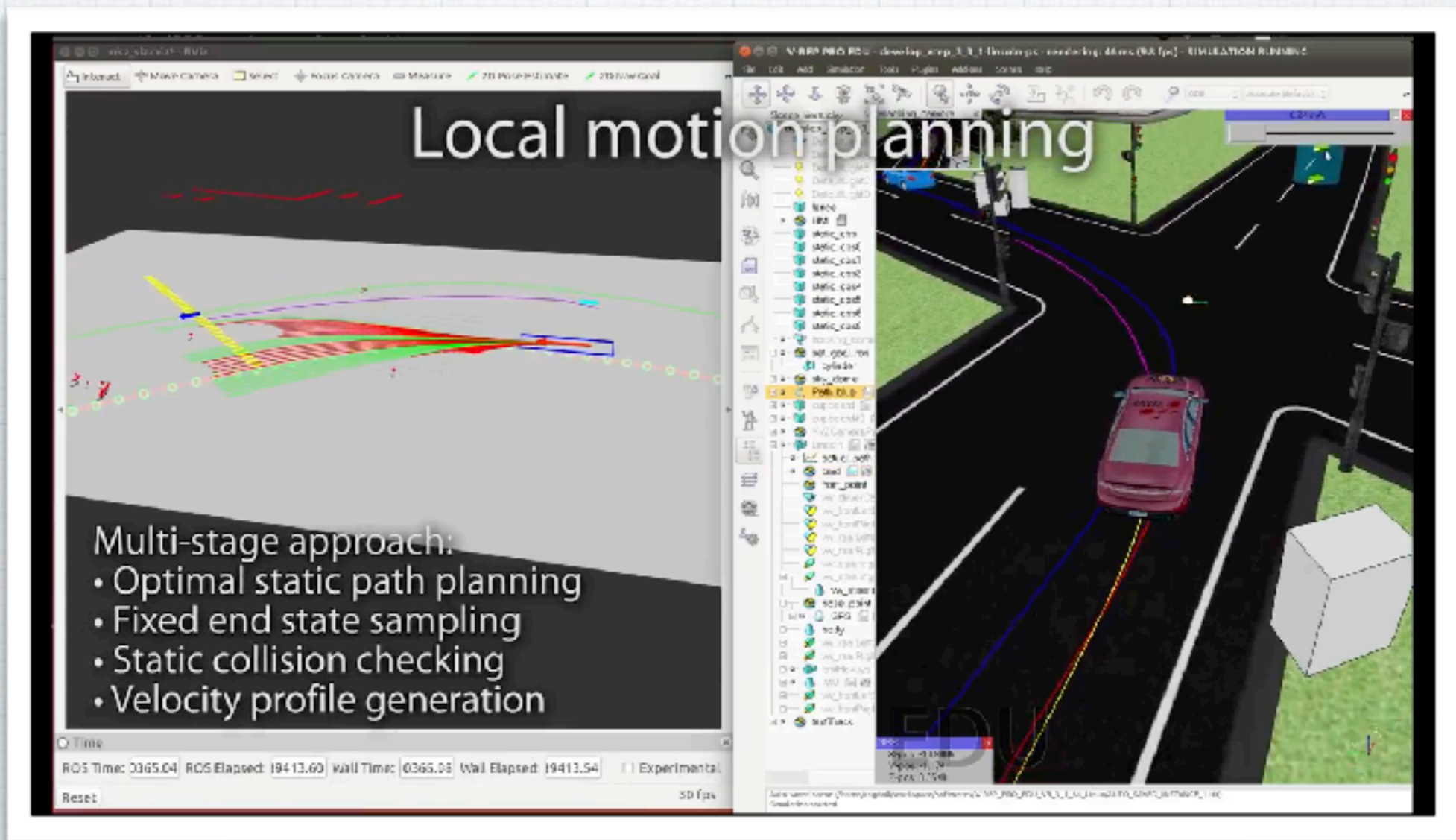
システムの例：自動運転

* センシング → 物体認識 → 経路計画 → 経路トレース



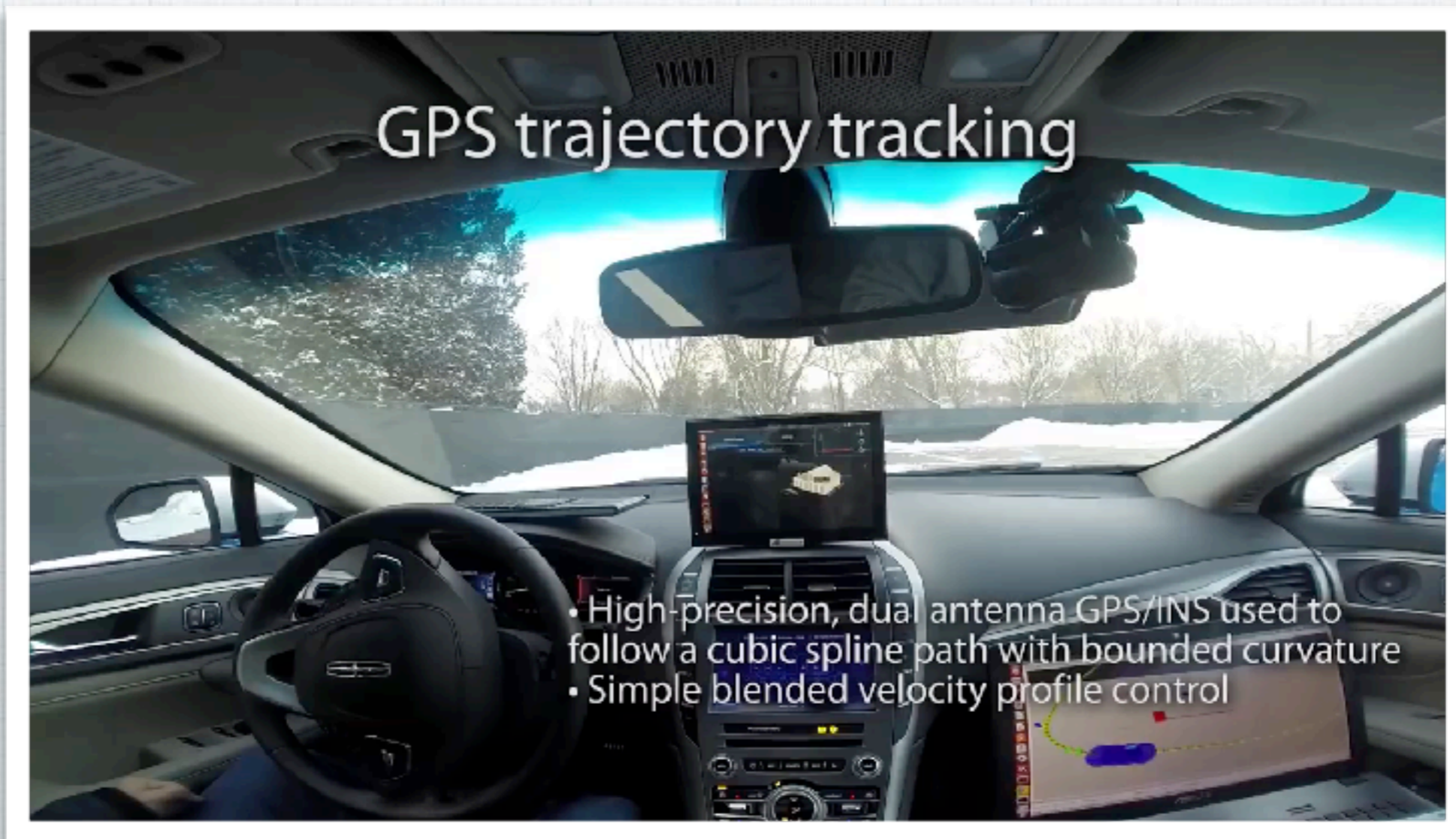
システムの例：自動運転

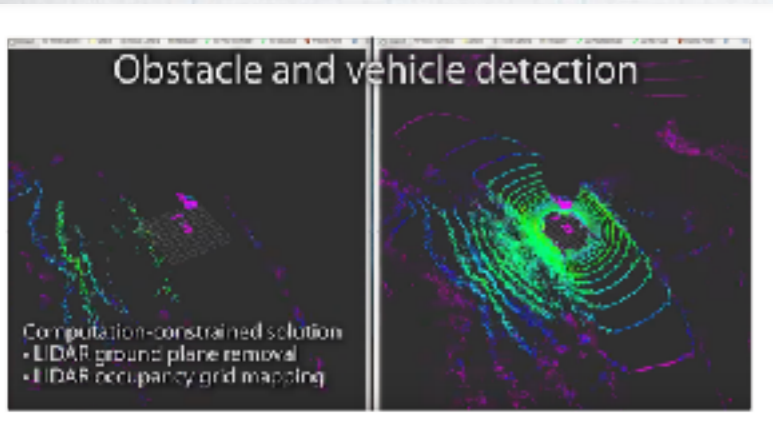
* センシング → 物体認識 → 経路計画 → 経路トレース



システムの例：自動運転

* センシング → 物体認識 → 経路計画 → 経路トレース

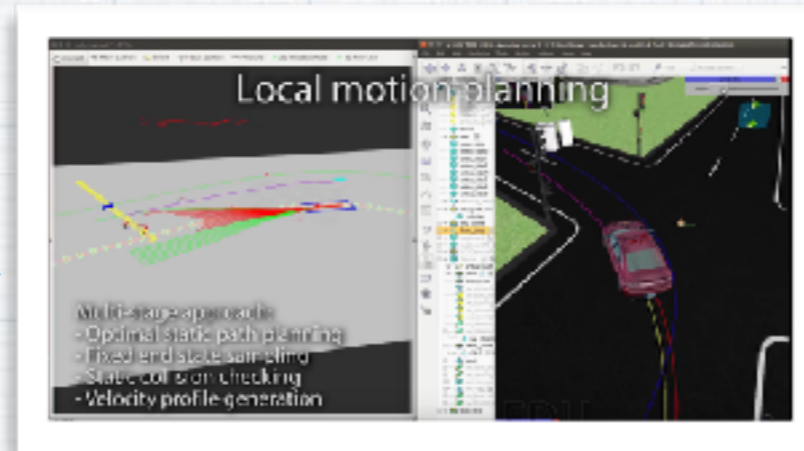




センシング



物体認識



経路計画



経路トレース

- * システムの実装
- * 統計的機械学習が各ステージで用いられる（特に物体認識）
- * システムの解析は別の話
- * 安全性・信頼性・品質の保証
- * テスト（統計的保証）
- * 演繹的形式推論が用いられる → 形式手法



https://commons.wikimedia.org/wiki/File:Waymo_self-driving_car_front_view.gk.jpg

高機能・高効率のシステムの

設計・実装

および

安全性保証

説明責任

統計的機械学習

演繹的形式推論

- * システムの実装
- * 統計的機械学習が各ステージで用いられる（特に物体認識）
- * システムの解析は別の話
- * 安全性・信頼性・品質の保証
- * テスト（統計的保証）
- * 演繹的形式推論が用いられる → 形式手法

統計的機械学習と演繹的形式推論

* 協働が hot topic

1. Safe Learning

[Akametalu, Kaynama, Fisac, Zeilinger, Gillula & Tomlin, CDC'14]

[Zhou & Li, CAV'18] 他

2. ニューラルネットから 有限オートマトンを抽出

[Weiss, Goldberg & Yahav, ICML'18] 他

3. 統計的機械学習による証明

[Faeber, Kaliszyk & Urban, CADE'17]

[Okudono, Nishida, Kojima, Suenaga, Kido & Hasuo, APLAS'17] 他

Safe Learning



- * 統計的機械学習：出力は統計的
- * 自動運転車が他車に衝突しないことの保証？
最善を尽くしたことの説明責任？
- * 「安全例を使って学習 → 安全」ではない
- * 主なアプローチ：システムレベルでの安全性保証

論理的 “safety envelope”

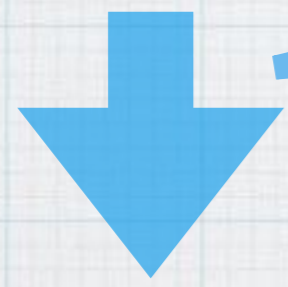
機械学習

コンポーネント

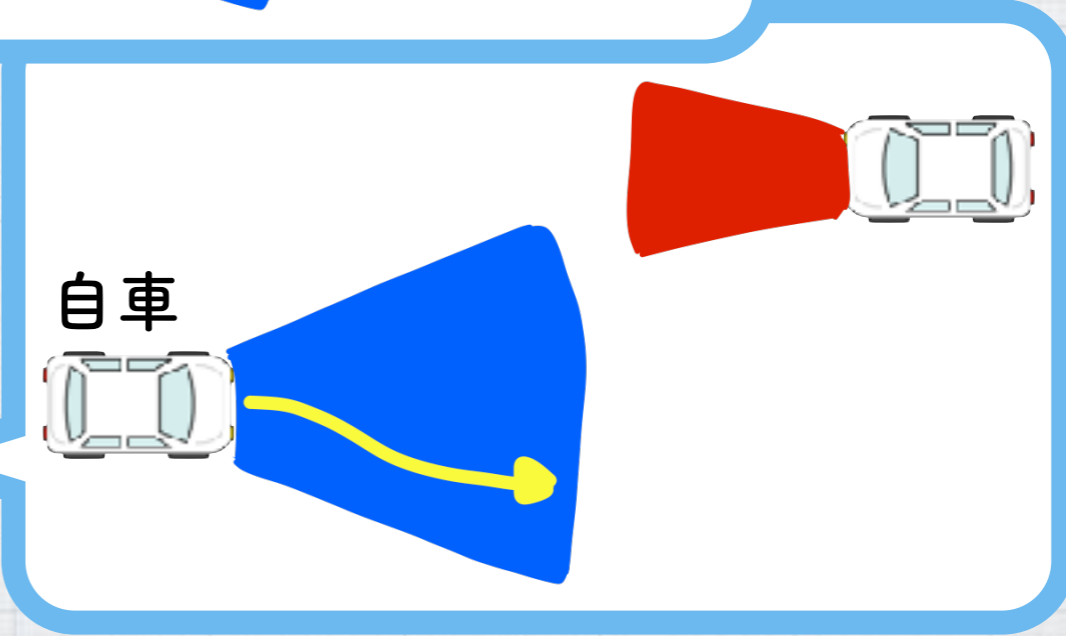
Safe Learning: 経路計画における例

[Akametalu, Kaynama, Fisac, Zeilinger, Gillula & Tomlin, CDC'14] 他

演繹的
可達性解析



機械学習による
経路計画



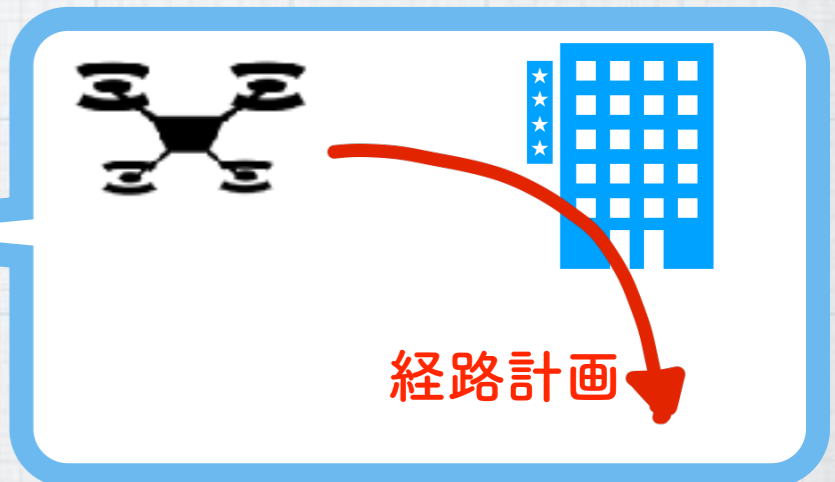
探索範囲を safety envelope に限定

Safe Learning: 反例に基づく段階的改良

[Zhou & Li, CAV'18]

反例を
フィードバック
(負のバイアス)

機械学習による
意思決定



安全性検証は expensive
反例フィードバックのやり方がキモ。学習アルゴリズムに依存
[Zhou & Li] では apprenticeship learning に適用

安全でない

演繹的
安全性検証
(モデル検査)

安全

計画を
実行

統計的機械学習と演繹的形式推論

* 協働が hot topic

統計的機械学習は
black/gray box

1. Safe Learning

[Akametalu, Kaynama, Fisac, Zeilinger, Gillula & Tomlin, CDC'14]

[Zhou & Li, CAV'18] 他

2. ニューラルネットから 有限オートマトンを抽出

統計的機械学習の
white box 化を目指す

[Weiss, Goldberg & Yahav, ICML'18] 他

3. 統計的機械学習による証明

[Faeber, Kaliszyk & Urban, CADE'17]

[Okudono, Nishida, Kojima, Suenaga, Kido & Hasuo, APLAS'17] 他

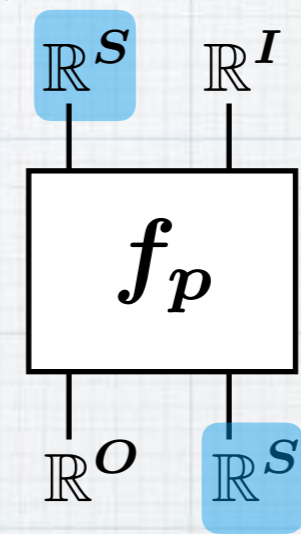


ニューラルネットから有限オートマトンを抽出

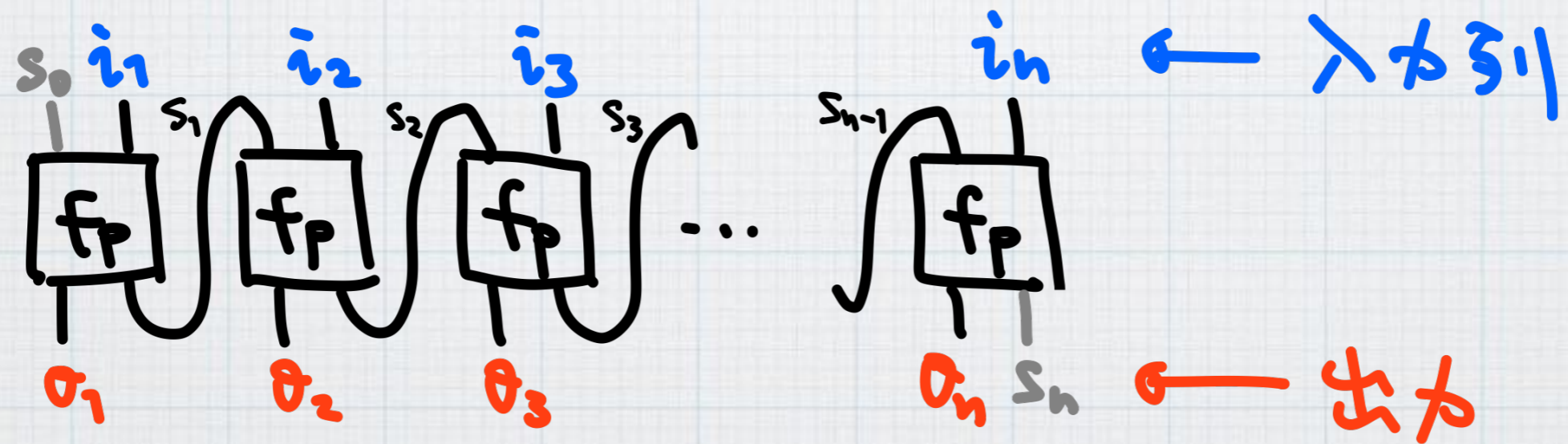
[Weiss, Goldberg & Yahav, ICML'18]

* Recurrent Neural Network (RNN)

$$\mathbb{R}^S \times \mathbb{R}^I \xrightarrow{f_p} \mathbb{R}^O \times \mathbb{R}^S$$



* 動作モード



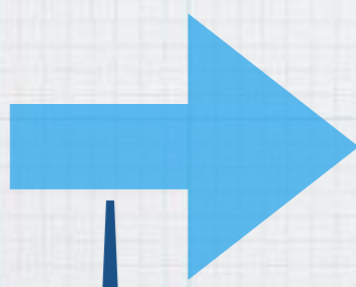
* 自動翻訳などに応用

ニューラルネットから有限オートマトンを抽出

[Weiss, Goldberg & Yahav, ICML'18]

* Recurrent Neural Network (RNN) から有限オートマトンを抽出

$$\mathbb{R}^S \times \mathbb{R}^I \xrightarrow{f_p} \mathbb{R}^O \times \mathbb{R}^S$$

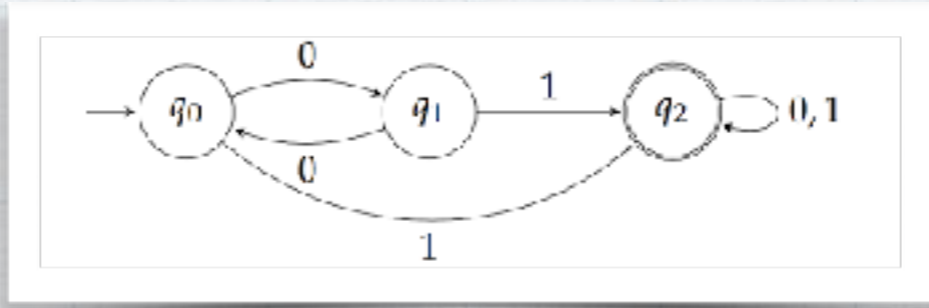


$$Q \times I' \xrightarrow{g} O' \times Q$$

(Q, I', O' are finite set)

モデル検査など，形式手法を用いて安全性検証可能，「理解」可能

- * Angluin のオートマトン学習アルゴリズム
- * ユークリッド空間のクラスタリング



統計的機械学習と演繹的形式推論

* 協働が hot topic

1. Safe Learning

[Akametalu, Kaynama, Fisac, Zeilinger, Gillula & Tomlin, CDC'14]

[Zhou & Li, CAV'18] 他

2. ニューラルネットから 有限オートマトンを抽出

[Weiss, Goldberg & Yahav, ICML'18] 他

3. 統計的機械学習による証明

[Faeber, Kaliszyk & Urban, CADE'17]

[Okudono, Nishida, Kojima, Suenaga, Kido & Hasuo, APLAS'17] 他

統計的機械学習による証明

* 統計的機械学習の判断の不確かさ

→ 証明探索であれば話は別

* 「証明の候補」を統計的機械学習で発見

* その正当性を演繹的に確認 → 正しさの論理的保証

* cf. proof search vs proof check.

証明問題の採点

* 証明探索を強化学習により効率化

[Faeber, Kaliszyk & Urban, CADE'17]

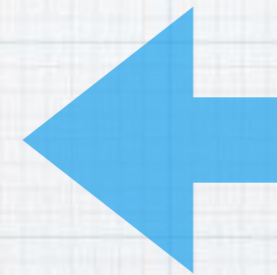
* 数値最適化による補題（数学的帰納法の仮定）の発見

[Okudono, Nishida, Kojima, Suenaga, Kido & Hasuo, APLAS'17]



https://commons.wikimedia.org/wiki/File:Waymo_self-driving_car_front_view.gk.jpg

高機能・高効率のシステムの
設計・実装
および
安全性保証
説明責任



統計的機械学習
演繹的形式推論

機械学習 + ソフトウェア工学

- * 情報システムの安全性・信頼性・品質保証
 - * 形式手法：演繹的形式推論，
モデル検査・定理証明，論理的保証
 - * テスト：統計的保証，形式手法を用いて効率化
→ ソフトウェア工学
- * 機械学習システムに対してもソフトウェア工学的手法は有効

機械学習システムのための数学



統計学
数理最適化

統計的機械学習
の基礎

論理学
(数学基礎論)

理論計算機科学

制御理論

力学系理論

システムの設計
と理解の基礎

幾何学

圏論

代数学

多様な理論の統合の
ためのメタ理論



プロジェクト紹介

- * プロジェクト目標：工業製品の設計サポート
 - * 形式手法をソフトウェアから物理情報システムへ
 - * 工業製品の品質保証をサポート，省力化
- * Challenge: 関連分野の多様性
 - * 計算機科学，制御理論，ソフトウェア工学，機械学習， etc.
- * 方法論：圏論・代数学的メタ理論による統合
- * > 10 名の特任教員・研究員を含む 50 名体制で理論研究を推進.
- * 国内製造業企業（自動車関連），U Waterloo 自動運転プロジェクトと協働，社会応用
- * 特任教員・特任研究員募集中（-2022）. <https://group-mmm.org/eratommsd/openpositions/>

論理学,
 代数学, 圏論,
 ...



Abstract Technique
 $T[\]$

$$A ::= \text{true} \mid \text{false} \mid A_1 \wedge A_2 \mid \neg A \mid a_1 < a_2 \mid \forall x \in *N. A \mid \forall x \in *R. A$$

$$\begin{array}{ccc} FX \xrightarrow{F\text{beh}_c} FZ & & FX \xrightarrow{Ff} FY \\ c \uparrow & \uparrow \text{final} & c \uparrow \quad \exists \quad \uparrow d \\ X \xrightarrow{\text{beh}_c} Z & & X \xrightarrow{f} Y \end{array}$$

system behavior simulation

Identify
 "mathematical
 essence"

Choose
 parameter e_1

Choose
 parameter e_2

Existing Technique
 $T_1 = T[e_1]$

Novel Technique
 $T[e_2]$

```

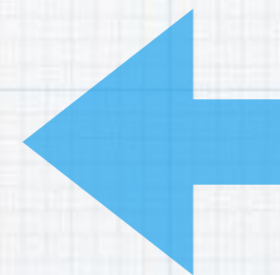
    'replace_interests' => false,
    'send_welcome' => false,
  })
  success('error', {result}) {
    'result' = 'error' ('response' => 'error', 'message' => 'error');
    'result' = 'success' ('response' => 'success');
  }
  success('errResult');
  
```

- * 物理情報システムの多様な応用に即応
- * 産業界（製造業）のニーズに根ざした理論展開
- * 現代数学の抽象性から生まれる応用



まとめ

高機能・高効率のシステムの
設計・実装
および
安全性保証
説明責任

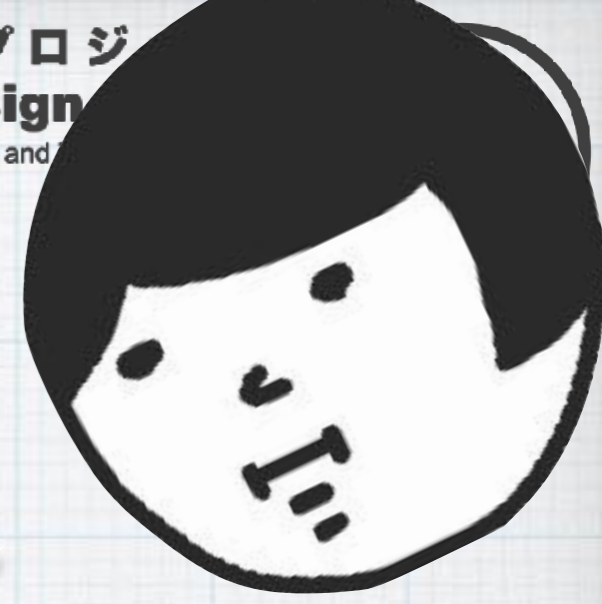


統計的機械学習
演繹的形式推論

- * 両者の協働は不可欠。協働の仕方は多様かつ発展途上
- * 論理学（数学基礎論）の役割，統計的手法への歩み寄り
- * 自動運転など，社会からの要請は大
→ 応用を軸にした数学諸分野（および周辺分野）の協働の
必要性と可能性
- * 圏論・代数学は協働のためのメタ理論



予備スライド

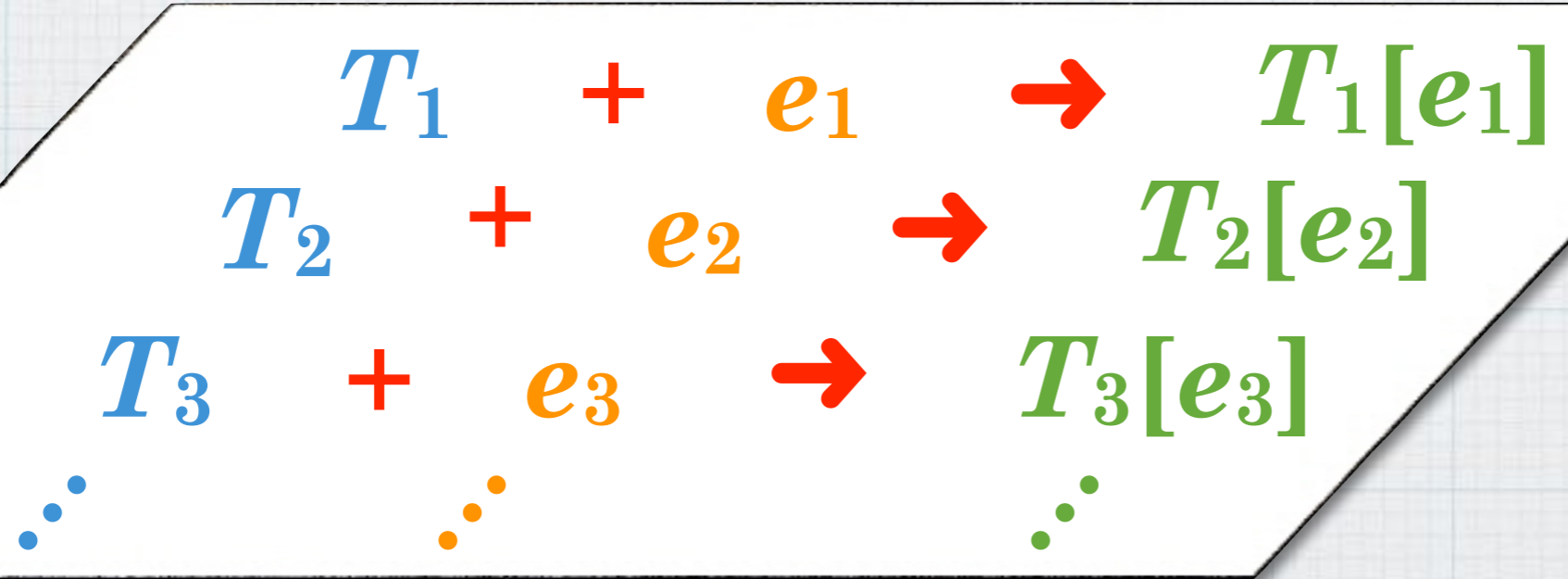
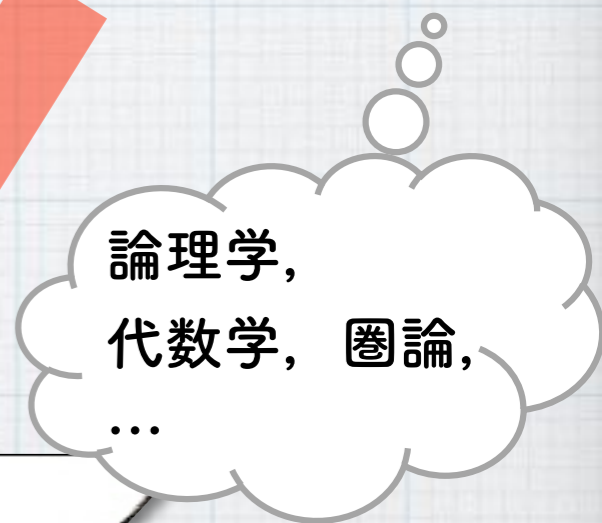


独自の方法論： メタ理論による移転

理論拡張のプロセス
自体を理論的に解析

... $T + e \rightarrow T[e]$
の包括的な一般論を構築
(さまざまな T と e に一挙に適用可能)

メタ理論家



ソフトウェアの
ための形式手法

新たな関心事
連続量, 不確かさ, ...

物理情報システム
のための形式手法

論理学,
 代数学, 圏論,
 ...



Abstract Technique
 $T[\]$

$$A ::= \text{true} \mid \text{false} \mid A_1 \wedge A_2 \mid \neg A \mid a_1 < a_2 \mid \forall x \in {}^*\mathbb{N}. A \mid \forall x \in {}^*\mathbb{R}. A$$

$$\begin{array}{ccc} FX \xrightarrow{F\text{beh}_c} FZ & & FX \xrightarrow{Ff} FY \\ c \uparrow & \uparrow \text{final} & c \uparrow \quad \exists \quad \uparrow d \\ X \xrightarrow{\text{beh}_c} Z & & X \xrightarrow{f} Y \end{array}$$

system behavior simulation

Identify
 "mathematical
 essence"

Choose
 parameter e_1

Choose
 parameter e_2

Existing Technique
 $T_1 = T[e_1]$

Novel Technique
 $T[e_2]$

```

    'replace_interests' => false,
    'send_welcome' => false,
  })
  success('err', {result}) {
    'result' = 'err' ('response' => 'error', 'message')
    'result' = 'err' ('response' => 'success');
  }
  console.log('errResult');
  
```

- * 物理情報システムの多様な応用に即応
- * 産業界（製造業）のニーズに根ざした理論展開
- * 現代数学の抽象性から生まれる応用

