2020年度

AIPシンポジウム

成果報告会

RIKEN Center for Advanced Intelligence Project

理研AIP-NEC連携センターの取り組み

宮野　博義

理研AIP-NEC連携センター　副連携センター長 /

日本電気株式会社 バイオメトリクス研究所 所長代理

いままでの活動振り返りと2020年度活動テーマ

2019年度までの活動振り返り

- 3テーマ体制で、国内外学会への投稿・採録を通じ基盤技術を育成
- テーマ２・３の2テーマは産総研との三者連携体制で成果創出を推進

2020年度も実用化に向け3テーマの技術を強化

活動テーマ

	~2018	2019
1．少量の学習データで高精度を実現する学習技術の高度化	現場データ０という極端な仮定で**技術開発**（ゼロショットドメイン適応）	異常など一部に限定して現場データ無い場合に緩和 **ACML2019,WACV2020採択**
2．未知状況での意思決定を支援する学習／AI技術の高度化	膨大な候補から妥当な仮説を絞り込む高速論理推論 **FIT2018（ヤングリサーチャー賞）** **三者連携構築**	仮説のロバスト性を強化（観測との矛盾を高速検知） **NeurIPS2019採択（三者連携成果）**
3．複数AI間の調整にかかわる学習技術の理論解析	AI同士の交渉を仲介するコーディネータAIをモデル化 **三者連携構築**	交渉戦略向けの未知制約付きベイズ最適化高速手法を考案 **三者連携,SIP**

2020年度にテーマを増設

- ＡＩの技術に対する社会実装が進み、ＡＩに対するセキュリティの懸念が顕在化
  → **「テーマ4：実環境を想定したAXのリスク評価・対策」を2020年度に新設**

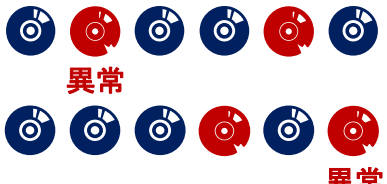
2020年度成果（テーマ1）

- 目的

従来AIが課題とする少量の学習データで高精度を実現する学習技術の高度化。
真のラベルよりも低コストで入手可能な弱ラベルを用いた学習手法の改善

知られている弱ラベル学習の例

PU学習



一部の正例データ（異常）のみラベル付けすれば学習できる

補ラベル学習

「～ではない」というラベルで学習できる

ノイズありラベル学習

ラベルに誤りが含まれていても学習できる

弱ラベル学習全般の課題：弱ラベルから真のラベルに対する損失を再構成すると、多くの場合、訓練損失が下に有界にならず、過学習が生じてしまう

- 2020年度成果
  - 以下の性質①②を満たす損失関数の十分条件を理論的に導出
    - 性質①：クラスの事後確率を正しく推定できること (properness)
    - 性質②：損失関数が下に有界であること (lower-boundedness)
  - 任意の弱ラベル損失に対して、性質①を失わずに性質②を持たせる正則化手法generalized logit squeezing(gLS)を考案
    過学習を回避し精度向上を実現

	性質①	性質②	CIFAR-10, WRN-28-2
BC	✓	✗	29.57 ± 1.58 %
BC + GA	✗	✓	36.87 ± 2.26 %
BC + gLS	✓	✓	**49.98** ± 2.59 %

BC: Backward Correction
GA: Gradient Ascent

提案手法

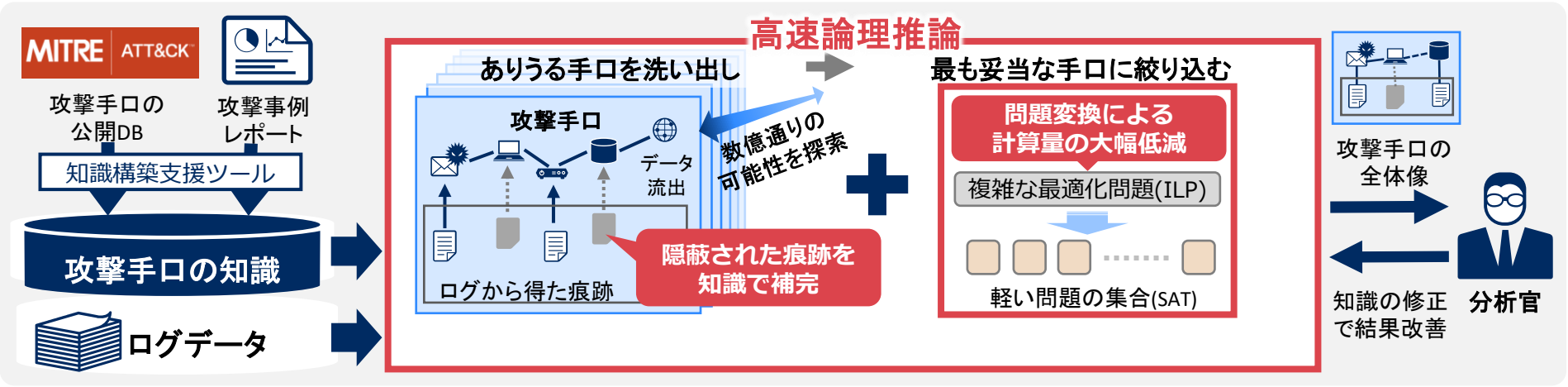
Lower-bounded proper losses for weakly supervised classification https://arxiv.org/abs/2103.02893

2020年度成果（テーマ2）

■ 目的

従来AIが課題とする未知状況での意思決定に対し、それを支援する高速論理推論を開発。高度化するサイバー攻撃手口の解明自動化を応用事例とし、有効性検証

高速論理推論によるサイバー攻撃手口推定



■ 2020年度成果

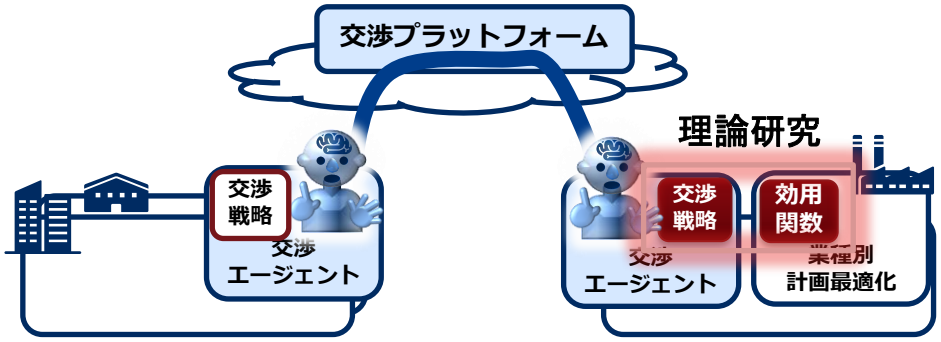
➢ 2019年度までに開発した高速論理推論技術をサイバー攻撃手口推定の問題に適用。隠蔽された痕跡を補完しつつ、実時間で手口を推定できることを確認

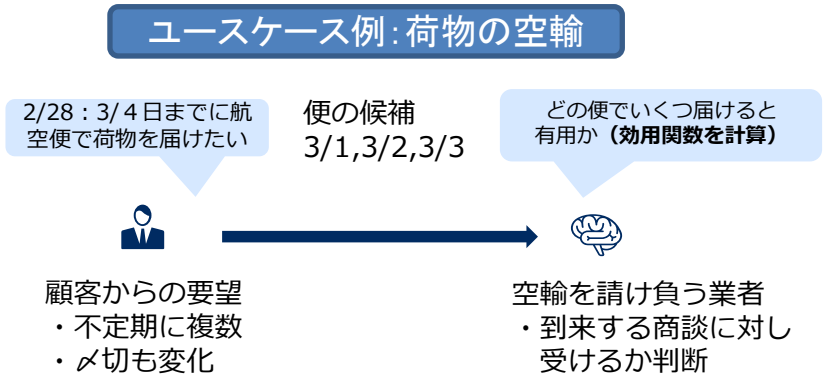
- ILPをSATに置き換えるという先に開発した汎用的な手法による高速化
- 知識記述の最適化や遅延評価が本タスクには有効であり、さらなる高速化を実現

2020年度成果（テーマ3）

■ 目的

AIが普及した社会を見据え、社会全体を最適化するAI間交渉・連携技術を開発

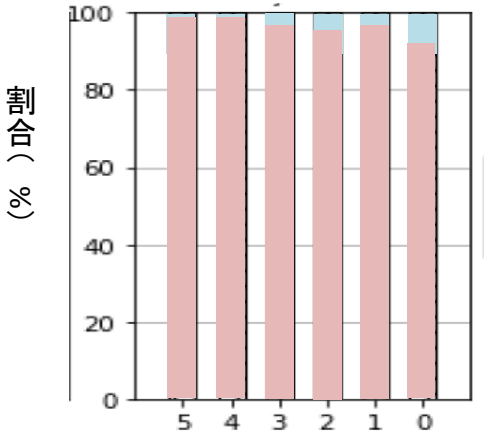




■ 2020年度成果

- 効用関数を確率的ナップザック問題の拡張として定式化
  - 到来する要望・種類が確率的・対応できる容量が可変な場合に対応
- 最適化手法の算出法、高速な近似解導出手法を考案
- ルールベース手法（空いている便に順に詰め込む）と比較し開発手法の優位性を確認

ルールベース（青）／開発手法（赤）
効用関数値が大きかったものの割合



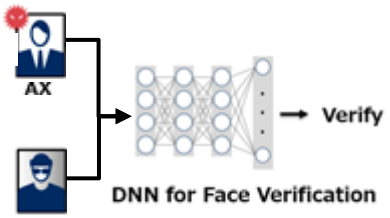
2020年度成果（テーマ4）

■ 目的

顔認証システムをターゲットに、実運用を想定したAdversarial Example(AX)の脅威を評価し、さらにAXに対するAIのロバスト学習技術を開発

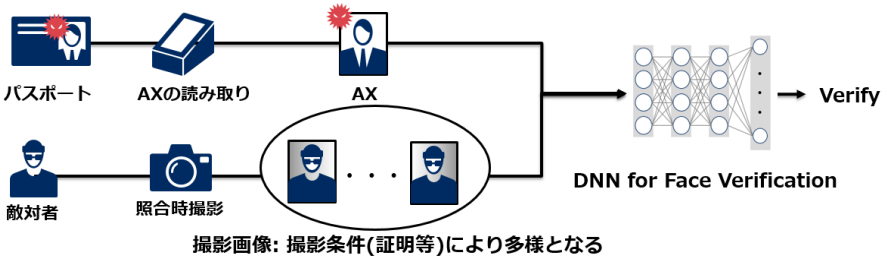
従来のAX研究

NNモデルに直接攻撃できる前提



本テーマの研究

実運用を想定、NNモデルには直接攻撃できない



顔認証は登録と照合の２プロセスあり、照合時のカメラ実施を想定

■ 2020年度成果

➢ 顔認証の運用を加味したAX攻撃手法の提案及びその対策手法を考案

- 撮影の変動（照明、姿勢）によらず攻撃できるAX手法を開発

➢ 従来よりも軽量で実現できる対策手法の考案

- 従来の複数のモデルを使い騙されにくくするアプローチに対し、同等の処理をモデル１つで実現可能に
- CIFAR10データセットを用いてパラメータ1/3で既存技術と同程度の堅牢性と精度が実現できることを実験的に確認