企画セッション | 部会・連絡会セッション：核不拡散・保障措置・核セキュリティ連絡会

2022年3月17日(木) 13:00 ~ 14:30 | G会場

[2G_PL] 核セキュリティ分野における人工知能技術の応用と課題

座長：宇根崎 博信 (京大)

[2G_PL01] 核鑑識及び核・放射線テロ現場初動対応関連技術における機械学習モデルの応用

*木村 祥紀[1] (1. JAEA)

[2G_PL02] 画像深層学習と自然言語処理を融合した検知手法の提案・開発・実装

*出町 和之[1] (1. 東大)

[2G_PL03] 重要インフラ管理のために我々はどのようにAIを活用できるのか？

*田中 淳裕[1] (1. NEC)

2022年春の年会

核不拡散・保障措置・核セキュリティ連絡会セッション

核セキュリティ分野における人工知能技術の応用と課題

Application of artificial intelligence in the field of nuclear security and its challenges

（1）核鑑識及び核・放射線テロ現場初動対応関連技術における機械学習モデルの応用

(1) Application of machine learning models on the technologies related to nuclear forensics and first response for nuclear and radiological terrorism

*木村　祥紀[1]

[1]JAEA

1. はじめに

日本原子力研究開発機構核不拡散・核セキュリティ総合支援センター（ISCN）では、不法移転やテロ事案などの現場から規制外の核物質及びその他放射性物質を押収し、押収されたそれらの物質の起源や履歴などを特定することで法執行機関による捜査活動を支援する技術的手段である「核鑑識」に関連した技術開発を進めている。核鑑識のプロセスは不法移転やテロ事案の現場における初動対応から、ラボでの分析及び分析データの解析を含み、それらに関連する技術的課題も多岐にわたる。ISCN では、核鑑識プロセスに関連する様々な技術的課題の解決を目的として、機械学習モデルを応用した新しい核鑑識技術の開発を行っている。本講演では、核セキュリティ分野における人工知能技術の応用例として、核・放射線テロ現場初動対応を含む核鑑識関連技術における機械学習モデルの応用に関する研究とその成果、今後の展望を紹介する。

2. 核鑑識における技術的課題と機械学習モデルの適用

近年、人工知能に関する中心的技術として様々な分野に応用されている機械学習は、サンプルデータから規則性などを学習し、解析の目的となるタスクを実現するモデルを構築する手法であるが、目的となるタスクやデータの種類・量などに応じて適切な機械学習モデルを選択することが非常に重要となる。ISCN では、機械学習モデルを応用した核鑑識技術開発に際し、核鑑識プロセス全体の評価と、それに基づいて機械学習モデルの潜在的なニーズを検討した。これらの評価に基づき、現在、核鑑識における技術的課題の解決に資する技術の実現を目標として、以下に関する技術開発を進めている。

① 深層ニューラルネットワークモデルによる放射性核種判定アルゴリズムの開発
② 深層距離学習モデルによる核物質表面パターン解析技術の開発
③ 核鑑識分析データの解釈における従来型機械学習モデルの応用に関する研究

3. 機械学習モデルを応用した核鑑識技術開発

3-1. 深層ニューラルネットワークモデルによる放射性核種判定アルゴリズムの開発

規制外の核物質及びその他放射性に関するテロ事案等の現場初動対応では、はじめに原因となる放射性核種の特定が必要不可欠となる。これらの現場初動対応は放射線測定の専門家ではない法執行機関を中心に実施されることが想定され、そのため持ち運びがしやすい携帯型の検出器で迅速かつ高精度な核種判定が自律的に可能となる技術が必要となる。本研究では、携帯型検出器で測定したガンマ線スペクトルの解析に深層ニューラルネットワーク（DNN）モデルを応用した核種判定アルゴリズムの基礎技術を開発した[1]。本アルゴリズムでは、DNN モデルによる解析で推定した測定ガンマ線スペクトルの解析対象エネルギー領域における計数寄与率に基づいて核種判定を行う。本アルゴリズムの性能を、高エネルギー分解能の HPGe 検出器及び低エネルギー分解能の CsI(Tl)検出器でそれぞれ測定したガンマ線スペクトルのテストデータで評価した結果、非常に高い核種判定性能を達成できることが確認された（図 1, 表 1)。本アルゴリズムにおいて、DNN

モデルの学習は検出器シミュレーションで構築した模擬スペクトルで行っており、同様の手法で様々なタイプの検出器に応用が可能な核種判定アルゴリズムを開発することができる。また、本アルゴリズムでは核種判定の結果だけでなく、DNN モデルによる計数寄与率の推定値をもとに初動対応者による結果の解釈を非常に容易に行うことができる。さらに、計数寄与率は核種判定だけでなく、現場における濃縮度に基づいたウランの分類推定といった用途にも応用可能である。

表 1: 核種判定アルゴリズムの性能評価結果

Detector	Model	Precision	Recall	F-score
HPGe	FC-ANN	1.000	0.880	93.62
HPGe	CNN-19	1.000	1.000	100.0
HPGe	Conventional (3σ*)	0.735	1.000	84.75
CsI(Tl)	FC-ANN	0.722	0.520	60.47
CsI(Tl)	CNN-19	0.958	0.920	93.88
CsI(Tl)	Conventional (3σ*)	0.882	0.600	71.43
CsI(Tl)	Conventional (2σ*)	0.667	0.800	72.73

* 従来法（ライブラリ比較法）におけるピーク検知閾値

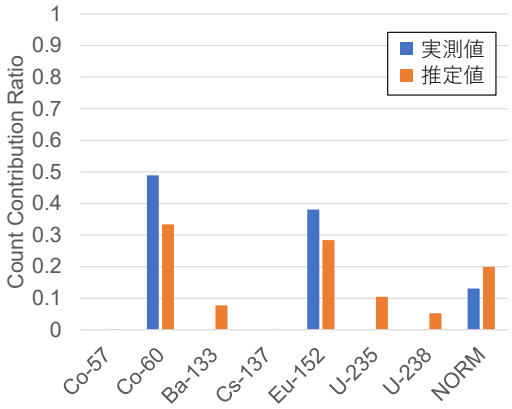


図 1 : DNN モデルによる計数寄与率の推定例
（CsI(Tl)検出器, ^{60}Co+^{152}Eu）

3-2. 深層距離学習モデルによる核物質表面パターン解析技術の開発

核鑑識分析では、規制外の核物質等の起源や履歴を特定するために、様々な側面から物質の特性を分析し、分析で得られたデータをもとに試料の異同識別解析を行う。本研究では、核鑑識分析における異同識別解析に関する技術的課題のひとつである顕微鏡画像データの解析に資する技術として、機械学習モデルを用いた核物質表面パターン解析技術の開発を進めている。核物質表面の微細パターンや粒子形状などの微細構造は電子顕微鏡による撮影で得ることができる物質特性であるが、定量的な分析をすることが難しく、特に表面パターンは客観的な解析が難しい。本研究では、画像解析分野で実績のある畳込ニューラルネットワーク（CNN）モデルを用いた深層距離学習により、表面パターンから試料を異同識別解析する技術を実現した [2]。本技術では、表面パターンを効果的に識別するための特徴量を抽出するモデルを距離学習により構築し、特徴量空間における距離に基づいて試料を分類する（図２）。電子顕微鏡で撮影したウラン精鉱試料の画像データにより本技術の分類性能を評価した結果、高い性能で試料の異同識別が可能であることを確認した。また核鑑識分析においては、既知試料への分類だけでなく、分析試料が未知試料であるかどうかの判断もまた非常に重要となる。本技術では、深層距離学習モデルを用いることで、特徴量空間における既知試料からの距離に基づいて、高い性能で未知試料を検知することも可能である。

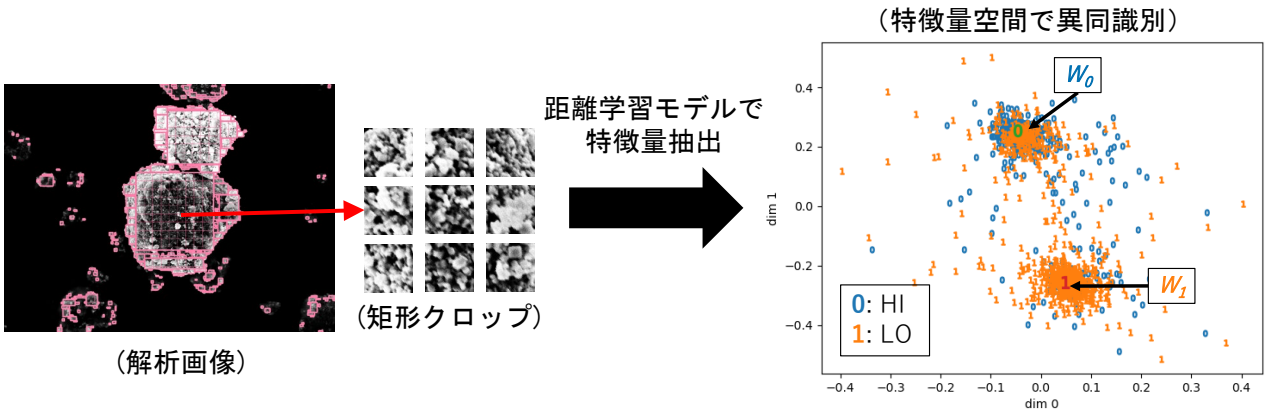


図２：深層距離学習モデルによる表面パターンに基づく異同識別解析（イメージ）

4. まとめと今後の課題

核鑑識においては、押収物質の起源や履歴といった分析結果が犯罪証拠やその後の裁判資料として利用される可能性があることから、機械学習モデルを用いて得られた分析結果の説明・解釈が重要となる。近年様々な分野で応用が進められている機械学習モデルのひとつであるニューラルネットワークモデルは、結果の解釈が非常に困難であることが知られており、解釈性の高い機械学習モデルを応用した技術や、機械学習モデルによる結果の解釈を支援する技術の開発が今後の課題として挙げられる。

サンプルデータから規則性などを学習することで解析の目的となるタスクを高い性能で実現する機械学習モデルを応用することで、核鑑識における様々な技術的課題を解決する新しい技術を実現することが可能であることを、具体的な研究開発の成果を交えながら紹介した。これらの技術は核鑑識プロセスと機械学習モデルの潜在的なニーズ評価に基づいて研究開発が進められているものであるが、核鑑識を始めとした核セキュリティ分野においては依然多くの技術的課題が存在しており、それらの解決方法として機械学習モデルの応用可能性は非常に高いものであると考えられる。

謝辞

本稿で示した成果の一部は文部科学省核セキュリティ強化等補助金事業及び科学研究費補助金（18K04646）.によるものです。

参考文献

[1] Y.Kimura, K.Tsuchiya., submitted, *Radioisotopes*.

[2] Y.Kimura, T.Murai, presentation in *Workshop: Application of AI to Nuclear Forensics* (2021).

---

*Yoshiki Kimura[1]

[1]JAEA.

2022年春の年会

核不拡散・保障措置・核セキュリティ連絡会セッション

核セキュリティ分野における人工知能技術の応用と課題
Application of artificial intelligence in the field of nuclear security and its challenges

（2）画像深層学習と自然言語処理を融合した検知手法の提案・開発・実装

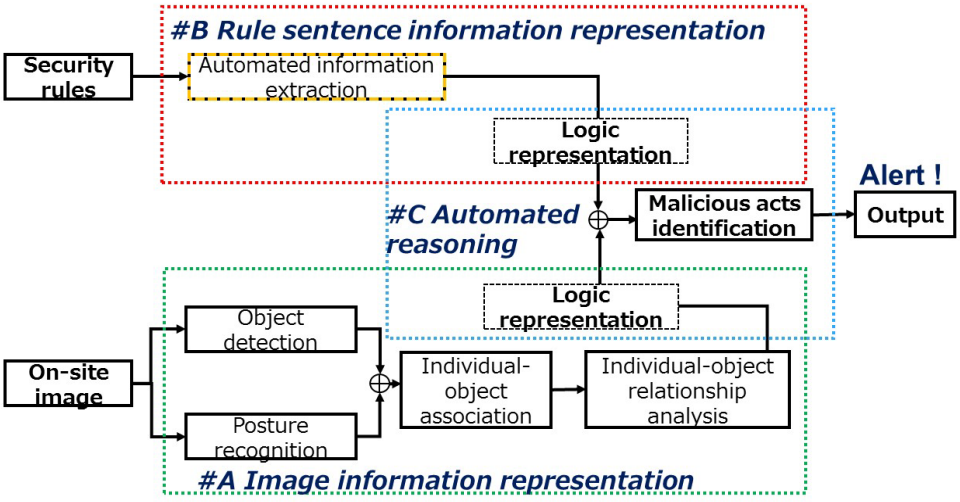
(2) Proposal, development, and implementation of detection methods by combining image deep learning and natural language processing

*出町 和之[1]，陳 実[1]
[1]東京大学

近年、窃盗や万引きなどの犯罪行為を画像深層学習を用いて検知する技術がさまざまに開発されてきている。しかしその多くは正常／正常からの逸脱に対する二値判定であり、複雑な状況（シーン）となり得る核物質の盗取や妨害破壊行為などの違反行為に対する検知へ適用するには複数種類の二値判定を複雑に組み合わせたプログラミングが必要となる。

一方、核物質の盗取や妨害破壊行為などの違反行為を禁じるルールは、文として定義することが可能である。もし、画像に映る複雑な状況をこれらルール文書と直接比較することで違反行為の有無を判定することができれば、複雑な判定プログラムを作成する必要が無くなり利便性が格段に向上すると期待できる。しかしそのためには、画像深層学習と自然言語処理という異なる２つの深層学習同士のインターフェイスが必要となる。東大・出町研では 3 種のアルゴリズムを提案・開発・実装し、このインターフェイスを実現した。

1 つめのアルゴリズムは「#A 画像情報の論理表現化」であり、監視カメラなどで撮影された状況を物と人の位置関係に基づき論理表現化するものである。ビッグデータに依存する AI 手法の主流とは異なり、非 AI 画像解析と AI 画像解析との融合により少ない教師データでの学習と時間短縮を実現した。２つめは「#B 安全文章情報の論理表現化」であり、形態素解析と関係性抽出により文意を論理表現へと変換した。３つめは「#C 自動判定」アルゴリズムであり、#A と#B の出力である論理表現同士の一致・不一致を判定して作違反行為の有無を自動検知する。以上の 3 つのアルゴリズムは自動計算プログラムとして実装済みである。この発表ではこれら 3 つのアルゴリズムについて詳細を解説する。



図：画像深層学習と自然言語処理の融合による違反行為検知アルゴリズム

*Kazuyuki Demachi[1] and Shi Chen[1]
[1]The University of Tokyo

2022年春の年会

核不拡散・保障措置・核セキュリティ連絡会企画セッション

核セキュリティ分野における人工知能技術の応用と課題

Application of artificial intelligence in the field of nuclear security and its challenges

（3）重要インフラ管理のために我々はどのように AI を活用できるのか？

(3) How can AI be used in the operation and management of important social infrastructure?

*田中 淳裕[1]

[1]NEC セキュアシステム研究所

1. はじめに

重要インフラの運用管理に AI 技術が使われ始めている[1, 2]。これはセンサーデータにより設備の稼働状況をモニタリングし大量のデータ処理を AI に行なわせることで、短時間で正確なモニタリングが可能となるという期待に基づいている。データ量の増大への対処はコンピュータプログラムが得意とすることであり、確立された手法（統計値の算出、外れ値による正常・異常の判断等）の利活用として導入が進められている。また最近では深層学習（DNN: Deep Neural Network）という手法を用いることで、従来技術では検出できなかったような微小な変化を初期段階で検出し、運用管理に活かす試みもなされている[3]。

このように AI 技術への期待感が高まる中で、重要インフラに AI を適用する際には、AI が認識する際の特徴（認識のクセ、バイアス）を理解し、認識バイアスへの対処が必要になると考える。以下では AI 利活用に潜むリスクについて、視覚情報処理における例を取り上げ今後の AI 活用における指針を議論する。

2. 視覚情報処理における誤認識

2-1. 錯視 --- 人間による誤認識

AI による認識のクセを議論する前に、人間が良く引き起こす誤解・誤認識について簡単に紹介する。最も有名な例がエッシャーやペンローズによるだまし絵であろう[4]。二次元の図として描かれた一見もっともらしい構造物がいくつも知られている。ただし実際に三次元の構造物を作ろうとしても物理制約を満たしていないために作ることができない（図 1: ペンローズの階段）。また視覚より得られた二次元情報をもとに三次元の物体を脳内で再構築するという人の認識のクセを活用することで、何度回転させても「右しか向かない矢印」を構成することが可能であることも知られている[5]。

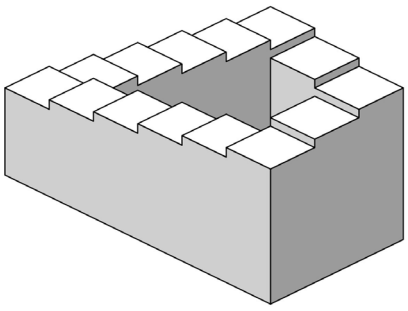


図 1: ペンローズの階段
（Wikipedia より）

人がどういう場合に誤解するのか、騙されるのかを事前に洗出しリスクを特定することは、重要な決定を行う際に不可欠であり、組織的な対処策を構築することが必要とされている。

2-2. 認識 AI による誤認識

画像認識 AI が騙される例は Goodfellow らの研究[6]により指摘された。人間には気が付かれないレベルのノイズを入力画像に加えることで、画像認識 AI が人とは全く異なる認識をしてしまう。図 2 はパンダの画像にノイズを乗せることで、高い信頼度でテナガザル（gibbon）と AI が誤認識した例である[7]。

この研究をきっかけとし、認識 AI に対する攻撃（Adversarial Examples; 敵対的入力）の研究が進められている。もし仮に敵対的入力が悪用されると、例えば道路の制限速度を誤認識させるように交通標識を汚す攻撃など、社会インフラに大きな影響を及ぼすような事態になることも予想される。

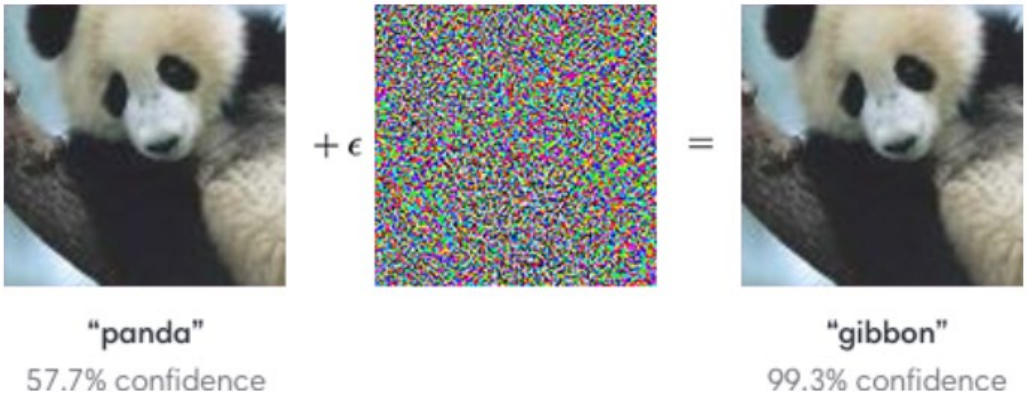


図 2: 敵対的入力による誤認識の例 [7]

3. AI 活用に向けて

人間の認識にはクセがあり誤認識をすることが知られている。同様に AI も誤認識をすることを、敵対的入力は我々に教えてくれた。深層学習 DNN はラベル付けされた正解データを大量に学習させることにより、従来不可能であった分類問題（例えば早期の異常検知等）への応用が期待されている。一方で詳細な動作原理に関しては未解決な部分も多く、信頼性を上げるための研究が盛んに進められている[8]。

AI を用いてシステムの信頼性を向上するためには、AI の認識の特徴を理解することで誤認識が起こり得るシーンを特定するとともに、他の手段で誤認識が起きたことを判別し、更に補間する手段を準備することが必要である。新しい技術の導入は一朝一夕に行うことはできない。多様な知見・経験を持つ有識者とともに、多方面からの検証活動を行いながら AI の導入を進めていきたい。

参考文献

1. 棗田昌尚、落合勝博、朝倉敬喜、林司、インバリアント分析技術の大規模物理システムへの適用－原子力発電所の監視への適用を例に－、情報処理学会 (IPSJ) デジタルプラクティス、Vol.6, No.3, pp. 207-214, 2015 年 7 月.
2. 相馬 知也、インバリアント分析による火力発電プラントの状態監視 : 見えなかった状態の可視化によるメンテナンスの効率化、保全学、2020 年 10 月, pp.14-17.
3. 吉永 直生、外川 遼介、網代 育大、時系列データ モデルフリー分析技術、NEC 技報、Vol.72, No.1, 2019 年 10 月.
4. 杉原厚吉、エッシャー・マジック --- だまし絵の世界を数理で読み解く、東京大学出版会、2011 年.
5. Kokichi Sugihara, 「立体錯視の世界」 第 1 回「右しか向かない矢印」、YouTube, <URL: https://youtu.be/qTFofXh9rE0>
6. I. J. Goodfellow, J. Shlens, and C. Szegedy, Explaining and harnessing adversarial examples. In 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015.
7. Attacking Machine Learning with Adversarial Examples, <URL: https://openai.com/blog/adversarial-example-research/>, 2017 年 2 月
8. 森川 郁也、機械学習セキュリティ研究のフロンティア、通信学会 基礎・境界ソサイエティ Fundamentals Review, Vol.2021,No.1, p. 37-46, 2021 年 7 月.

*Atsuhiro Tanaka[1]
[1]Secure Systems Research Labs., NEC Corporation