

機械学習を用いた危険行動検知手法の開発

Development of Dangerous Behavior Detection Method using Machine Learning

東京大学	川崎 祐典	Yusuke Kawasaki	No-Member
東京大学	出町 和之	Kazuyuki Demachi	Member
東京大学	笠原 直人	Naoto Kasahara	Member

Abstract

After Fukushima accident, threats of terrorism are increasing. Especially, the threat of sabotage by malicious insider is significant problem. But there is not a concrete measure to detect insider. So we develop fundamental technique of abnormal behavior detection.

Keywords: Deep Learning, Deep Belief Nets, Insider Detection

1. 緒言

福島事故以降、原子力施設における核セキュリティの重要性は増大している。原子力発電所における核セキュリティの最大の目的は悪意のある行為者による原子力施設の安全機能の喪失のリスクを可能な限り下げることである。その中でも特に重要視すべきものが妨害破壊行為である。妨害破壊行為は一般的にアウトサイダーによるものとインサイダーによるものの2つに分類される。現状の核セキュリティはアウトサイダーへの対応に比べインサイダーへの対応は十分とは言えない可能性がある。そのため原子力施設内でインサイダーが行う妨害破壊行為を検知できる技術の開発が強く求められている。本研究では、機械学習法のひとつであるDeep Belief Network(以下DBN)に基づいて画像解析によるインサイダー検知のための基礎技術を開発し、その有効性を検証することを目的とする。

2. 理論

2.1 背景差分を用いた画像内の人物の検出

本研究では、カメラで取得した動画の中から人及び不審物を抽出する方法として、背景差分法を用いた。背景差分法とは、事前に用意した背景画像とカメラから得たリアルタイム画像との差分をとり閾値判定を行うこと

で背景画像にない物体を検出する方法のことである。検出後それぞれの対象に対して重心の計算を行うことで位置情報を取得する。Fig.1に背景差分法により取得された画像を示す。



Fig.1 Image using background subtraction

2.2 機械学習を用いた行動の分類

次に、抽出された人物画像から人の行動を検知して分類する必要がある。人の行動を複数に分類する際の手法には機械学習がよく用いられる。画像を対象として機械学習を行う場合別途画像から特徴抽出を行うのが一般的であるが、特徴抽出の方法は画像ごとに適した方法が異なるため今回のような様々な行動が想定される場合は効率的でない。よって本研究では特徴抽出のプロセスを自動で行えるDeep Learning[1]を用いる。

2.2.1 Deep Learning

Deep Learning[1]とは、多層のニューラルネットを用いた機械学習法であり、その特徴として従来人の手で行わ

連絡先: 〒113-0033 東京都文京区本郷7丁目3-1、東京大学大学院工学系研究科原子力国際専攻
E-mail : yu_ten@live.jp

れていた特徴抽出を自動で行えることが挙げられる。
Deep Learning の種類は様々存在するが本研究ではそのうちの一つである Deep Belief Nets(DBN)を用いる。

2.2.2 Deep Belief Nets

Deep Belief Nets(以下 DBN)[2]は Hinton らによって発表された Deep Learning の手法の一つで、学習は各層ごとの特徴抽出のための教師無し学習とクラス分類器を含めた全体で多クラス分類のための教師有り学習の二つの段階に分かれている。

2.2.3 教師無し学習

教師無し学習では各層の Restricted Boltzmann Machine のパラメータの更新を行う。Restricted Boltzmann Machine[3]は各素子が{0,1}の値をとり入力層同士及び隠れ層同士に結合がないネットワークである。学習の手順は以下の通り。

- (1) 適当な v を入力
- (2) 確率分布 $p(h|v)$ に従って h を算出
- (3) 確率分布 $p(v|h)$ に従って v' を算出
- (4) v' を v に近づけるようにパラメータを更新

h :隠れ素子, v :可視素子

w :各層間の重み, b,c :バイアス

$$p(h_i = 1|v) = \sigma(\sum_{j=1}^m w_{ij}v_j + c_i) \quad (1)$$

$$p(v_j = 1|h) = \sigma(\sum_{i=1}^n w_{ij}h_i + b_j) \quad (2)$$

パラメータの更新には対数尤度による最尤法を用い途中計算量を削減するためにギブズサンプリングと Contrastive Divergence 法[4]を用いた。

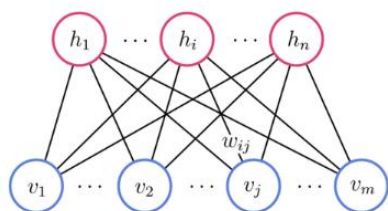


Fig.2 Network of Restricted Boltzmann Machine

2.2.4 教師有り学習

教師有り学習では、教師データ(学習用画像及び学習用ラベル)を用いてネットワーク全体の結合荷重及びバイアスを誤差逆伝播法及び最急降下法を用いて更新する。

3. 検証と結果

今回検証用の動画データとして、人が画面左から歩いて

きて床に物を置いて去る、という爆弾設置を模擬した動画を用意し、画面内の右側を進入禁止エリアと仮定した上で、「侵入禁止エリアへの侵入及び不審物の検知」と「物を床に置く行動の検知」の二つを行い、これらの組み合わせによって「人が侵入禁止エリアに物を置いたことの検知」の実現ができるかの検証を行った。

3.1 侵入禁止エリアへの侵入及び不審物の検知

まず背景差分法を用いて動画像から人の侵入及び不審物の検出ができるかを検証した。なお、今回は検知した対象に対して機械学習を用いた物体認識を行い人と物体の判定を行った。なお、緑点が人、赤点が物体、赤線より右側が侵入禁止エリアを意味している。Fig. 3, 4 に示す結果から検知が精度よく行えていることが分かる。



Fig.3 Detected object from an image

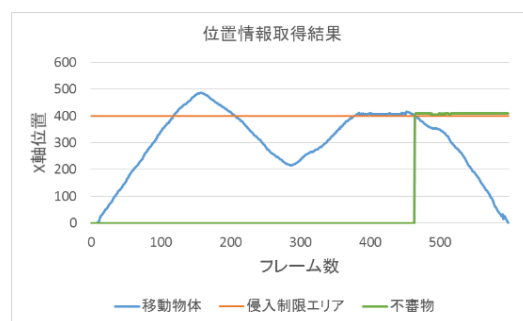


Fig.4 Time series data of position of object and human

3.2 物を床に置く行動の検知

次に、物を床に置く行動の検知の準備として、機械学習を用いて人の姿勢の判定が行えるかを検証した。この検証では、別途人が画面内で立ち及びすわりを繰り返す動画を撮影し、「立ち」と「しゃがみ」の分類を行った。

3.2.1 人の姿勢の判別

「立ち」と「しゃがみ」の分類を行った。学習用画像を Fig. 5 に、これらの画像による学習後のネットワークによって得た動画像に対する分類結果を Fig. 6 に示す。図 5, 6 に示すように姿勢の分類が精度良く行えたことが分かる。

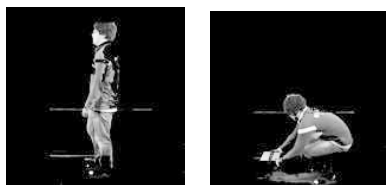


Fig.5 Images for pre-training and training DBN

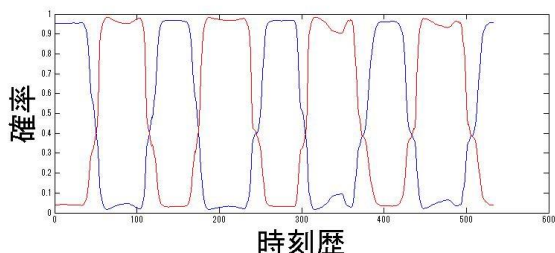


Fig.6 Probability time series data (blue : standing, red : sitting)

3.2.2 物を床に置く行動の検知

次に、分類を「立ち及び歩き」と「物を上に置く動作」と「物を下に置く動作」の3つに増やし元の検証用動画に対して行動分類を行った。学習用画像を Fig. 7 に、これらによる学習後のネットワークによって得られた分類結果を Fig. 8, 9 に示す。なお、Fig. 8 における棒グラフはネットワークによる各行動の確率を示している。Fig. 8, 9 に示すように、検知が精度よく行えていることが分かる。



Fig.7 Images for pre-training and training DBN

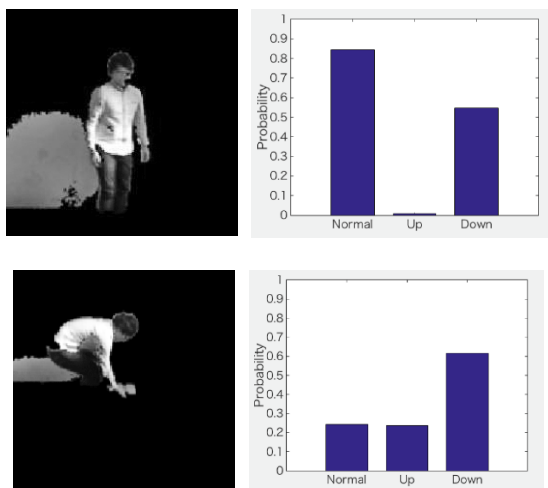


Fig.8 Probability of image

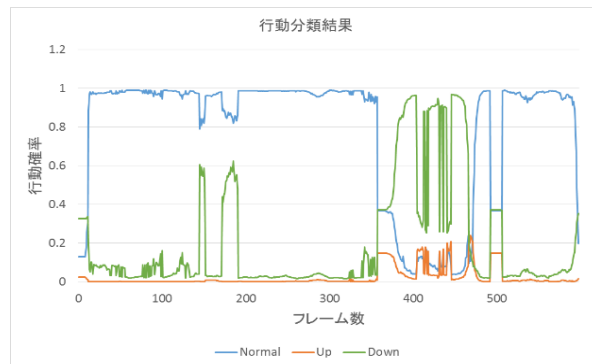


Fig.9 Time series data of probability

3.3 人が侵入禁止エリアにものを置いたことの検知

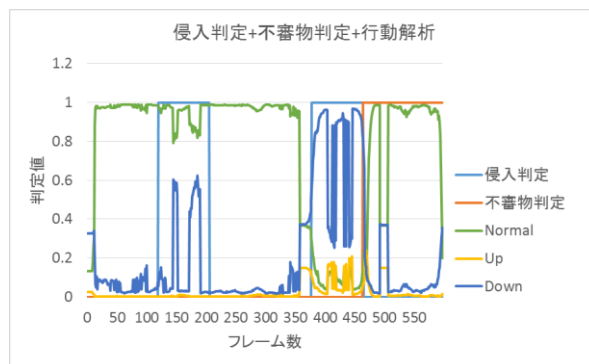


Fig.10 Total graph of result of each detection

Fig. 10 より、358 フレーム以降で侵入の検知及び「下に物を置く動作」の行動確率が大きくなり、その後不審物が出現している。これは侵入禁止エリアで不審物を設置するための異常行動を行った可能性が高いと判断できる。これに対し142~193フレームでは侵入の検知がされているが他の異常は検知されていないため異常度が低いと判断できる。このように複数の検知結果を組み合わせることで精度の高い異常判定が行えることがわかる。

以下ではこの組み合わせによる判定を定量的に行うために評価式を用いた異常値算出法とネットワークを用いた異常値算出法の2種類を適用した。

3.3.1 評価式を用いた異常値算出

まず評価式を用いた異常値の算出を行った。位置情報、行為情報、不審物情報の三つに対して個別に異常値を算出し、足し合わせることで異常度を算出した。今回用いた評価式は以下の通りである。ある時刻で、 V_1 は位置情報に対する異常値を過去 r ステップ中何%の割合で侵入禁止エリアにいたかで表し、 V_2 は行為情報に対する異常値を

異常行動を行っているかで表し (α は行動ごとの係数、 P は行動ごとの確率を表す)、 V_3 は不審物情報に対する異常値をその物体が侵入禁止エリアにあるかで表し、 V_1 はこれら三つを足し合わせて正規化したものである。

$$V_1(t) = \frac{\sum_{i=t-r}^t X_i(x_i)}{r}$$

$$X_i(x) = 1 (x > threshold) \quad X_i(x) = 0 (x \leq threshold)$$

$$V_2(t) = \frac{1}{1 + \exp(-P(t))}$$

$$P(t) = \alpha_{ac1}p_{ac1}(t) + \alpha_{ac2}p_{ac2}(t)$$

$$V_3(t) = I(x_t)$$

$$I(x_t) = 1 (x > threshold) \quad I(x_t) = 0 (x \leq threshold)$$

$$V(t) = \frac{V_1(t) + V_2(t) + V_3(t)}{3}$$

この評価式による異常値算出を単体の検知と比較するために位置情報だけによる異常値算出と比較した図を以下に示す。Fig.11 からわかるように、一つの情報による検知よりも正確に異常を検出することが可能である。

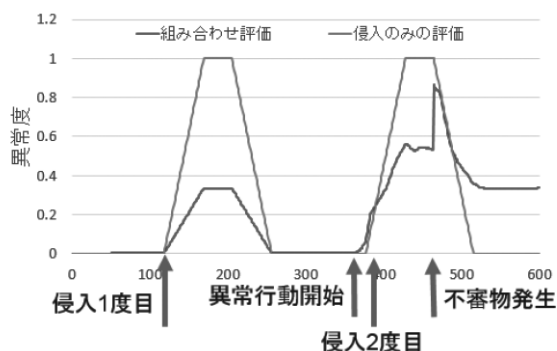


Fig.11 Comparison of rates of abnormality

3.3.2 ネットワークによる異常値算出

次にネットワークによる異常値算出を行った。検知した情報を統合してネットワークに入力し異常値を出力として得る、という方法である。今回は任意の時刻に対して過去 50 ステップ分の情報を与えて異常値を算出した。これによる結果と 3.3.1 節による結果を比較したものを Fig.12 に示す。Fig.12 からわかるように 3.3.1 節よりも敏感に以上の検出が行えることがわかる。

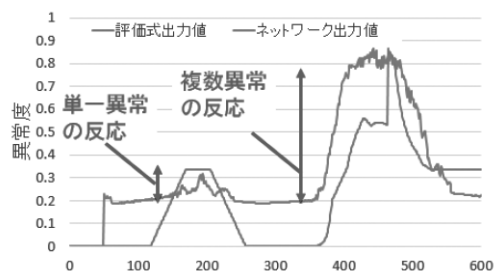


Fig.12 Comparison of rates of abnormality(3.3.1&3.3.2)

4. 結論及び課題

機械学習法のひとつである DBN を用いてインサイダーの危険行動検知の手法を作製し、その有効性を示した。そのための基礎技術として画像からの物体検知及び行動検知の技術を開発した。

今後はリアルタイム処理のための計算の軽量化及び照明の変化等も考慮した物体検知の精度の向上が必要であり、また、建物全体の監視による行動ルート等の位置情報による異常検出を取り入れるなどによる更なる拡張も考えられる。

参考文献

- [1] Quoc V. Le, Marc'Aurelio Ranzato et al. "Building High-level Features Using Large Scale Unsupervised Learning", Proceedings of the 29th International Conference on Machine Learning, Edinburgh, Scotland, UK, 2012.
- [2] Geoffrey E. Hinton et al. "A Fast Learning Algorithm for Deep Belief Nets", Neural Computation 18, 1527-1554(2006)
- [3] Asja Fischer, Christian Igel "An Introduction to Restricted Boltzmann Machines" Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications Lecture Notes in Computer Science Volume 7441, 2012, pp 14-36
- [4] Carreira-Perpinan, M. A., & Hinton, G. E. (2005). "On contrastive divergence learning." In R. G. Cowell & Z. Ghahramani (Eds.), Artificial Intelligence and Statistics, 2005. (pp. 33-41). Fort Lauderdale, FL: Society for Artificial Intelligence and Statistics.

