Isolation Forestを用いた IoT向け異常検知手法に関する考察

菅田 大輔[1,a)]　安全 花子[1,2,†1]

概要：概要

キーワード：Isolation Forest，IoT，IDS，異常検知

A Study on Anomaly Detection Method for IoT using Isolation Forest

Daisuke Sugata[1,a)]　Hanako Anzen[1,2,†1]

Abstract: abstract

Keywords: Isolation Forest, IoT, IDS, Anomaly Detection

1. はじめに

以下のことを書く.

- 大目標：小規模な環境向けのIoT向け異常検知手法の提案・「IoT環境を考えたとき、iforestが軽量でうまくいきそうなので、〜うまくいかないことも多かった。そこで、データセットの〜を工夫して、そのやり方を報告する」→本研究の目的を最初に述べちゃう
- 大目標を実現する必要性：卒論と同じ
- 問題提起（大目標を達成するために必要なことを述べ，そのためにどのような問題があるのかを述べる (小目標に分割する).）関連研究をリサーチしてみると参考にした論文が良さそうだ。しかし、いくつか課題もあるので改善したい 1. 特徴量エンジニアリング　2. 判定の組み合わせアルゴリズム
- （関連研究：問題を解決するための従来研究を紹介）
- 本研究の目的：1. 特徴量エンジニアリングの比較 2. 判定の組み合わせアルゴリズムの改善

2. 研究方法

2.1 Isolation Forestの説明

Isolation Forest（以下，iForest）は，外れ値検出のためのアルゴリズムである。iForestは異常データが少数であり、離れているという前提に基づいている。ランダムにデータを分割していくと、異常データは相対的に早く分離される。iForestは以下のステップで実行される。

1. データの分割

ランダムに選んだ特徴量から、ランダムに選んだ値をもとにデータを分割する。これを一定回数繰り返し、複数のツリーを作成する。

2. 異常スコアの算出

データがツリーの枝に到達するまでの平均パスをもとに、異常スコアを算出する。具体的な計算式は以下の通りである。

[1] 東京工業大学 情報理工学院 数理・計算科学系
Department of Mathematical and Computing Sciences, School of Computing, Tokyo Institute of Technology
[2] 株式会社 YY セキュリティ研究所
Security Laboratories, YY Corporation
[†1] 現在，国立研究開発法人 ZZ 研究所
Presently with National Institute of ZZ
[a)] sugata.d.aa@m.titech.ac.jp

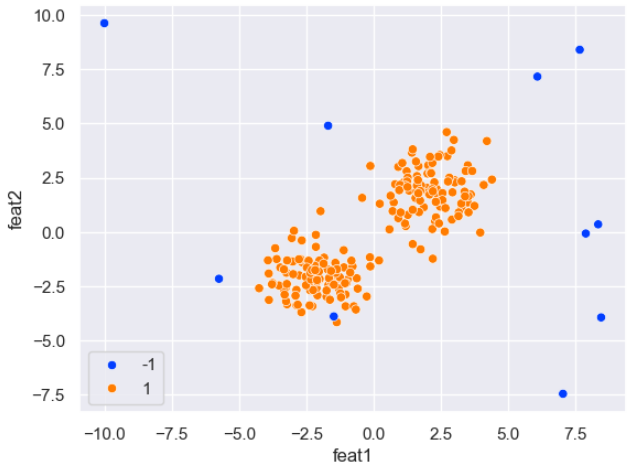

図 1　デモデータに関する図の説明（和文）
Fig. 1　Description of the dim_vs_accu figure (English).

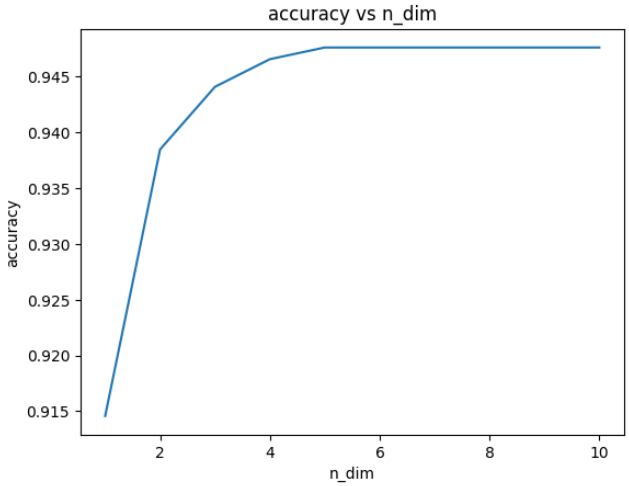

図 2　dim_vs_accu に関する図の説明（和文）
Fig. 2　Description of the dim_vs_accu figure (English).

ここで、$E(h(x))$ はデータ点 x の平均パス長、$c(n)$ はデータセットのサイズ n に依存する定数である。

3. 異常判定

先ほど計算した異常スコアをもとに異常検知を行う。通常、トレーニングデータの異常スコアの上位 10%を閾値として設定し、それを超えたデータを異常と判定する。

2.2　IDS の概要

IDS の概要をかく．採用した理由を述べる．

2.2.1　全体像

今回の研究では，以下のような IDS で異常検知を行った．

2.2.2

3.　事前実験

Isolation Forest を異常検知手法として使用する際，どのような問題があるのかを明らかにするため，事前実験を行った．はじめにデモデータを使用して，iForest の挙動を確認した．その考察をもとに、特徴量選択手法の提案を行なった

使用したデモデータは、(2.5, 2.5) と (-2.5, -2.5) を中心とした正常データ群と、10 から-10 の範囲に一様に分布した異常データ群からなる．二次元の場合のデモデータを図 1 に示す．

はじめに、デモデータと特徴量数の関係を調査した．図 2 に示すように、特徴量数が増えるにつれて、異常検知の精度が単調に向上することがわかった．また、精度ののびは増加に比例して緩やかになっていることもわかる。ゆえに、iForest は十分な特徴量数が存在すると言える。

パケット通信を監視して得られたデータセットの全ての特徴が、異常検知に有効であるわけではない。iForest は特徴量同士の重みづけを行わないため、判定に有効でない特徴量が混ざると精度が低下すると考えられる。そこで、図

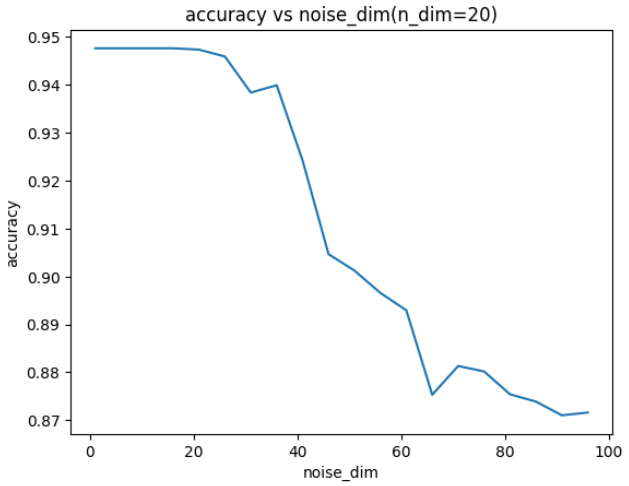

図 3　noise_accu に関する図の説明（和文）
Fig. 3　Description of the noise_accu figure (English).

3 に示すように、ノイズとなる特徴量を混ぜた時の精度を調査した。この結果から、ノイズとなる特徴量が混ざると精度が低下することがわかった。また、今回の実験の場合だと、ノイズとなる特徴量が判定に有効な特徴量数の 2 倍以上になると、精度が急激に低下することがわかった。

前の実験から、データセットからノイズとなる特徴量を取り除くことが重要であることがわかった。ところで、iForest はツリーベースの異常検知手法である。そこで、同じくツリーベースの Randam Forest から特徴の重要度を算出すれば、ノイズとなる特徴量を取り除けるのではないかと考えた。図 4 は、Random Forest で算出した特徴量の重要度を表している。このグラフは、ノイズ特徴量を判別できていることがわかる。そして、実際にノイズ特徴量を取り除いた場合の精度を調査したところ、精度は???%から 0.945%まで向上した。この結果から、Feature Importance による特徴量選択手法は精度の向上に有効ではないかと考

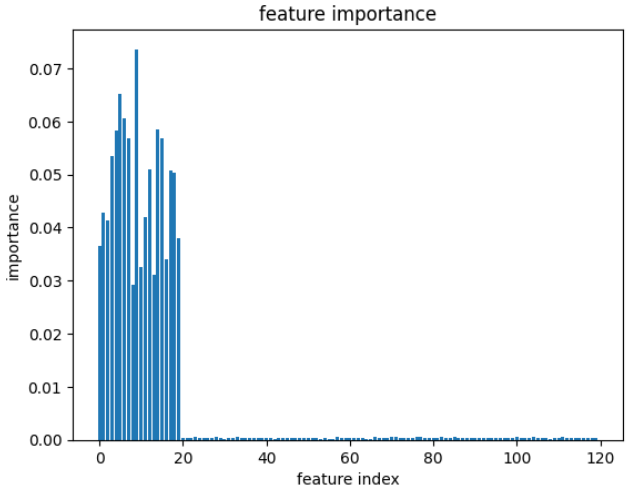

図 4　select_noise に関する図の説明（和文）
Fig. 4　Description of the select_noise figure (English).

えた。

4.　結果と考察

（本論文のメイン）

4.1　実験方法

4.1.1　実験環境

他の人が再現できるように実験環境を書く. 1. 特徴量エンジニアリングの比較 2. 判定の組み合わせアルゴリズムの比較

4.1.2　データセット

使用したデータセットの概要と, その妥当性について述べる.

4.1.3　評価指標

使用した評価指標と, その妥当性について述べる.

4.2　結果

実験の結果，得られるデータから読み取れる客観的事実を書く．この時，論文の目的を達成するためにどのような主張をどのような結果（データ）に基づいて説明すべきかを考える.

4.3　考察

4.3.1　本論文における目的に即した結論を導く

- 本結果を一般化したどのような結論を導き出せるかを，論文の目的に即して述べる.　1. このくらい有効特徴量あればいける　2.iforest ではうまくいく
- 実験結果の妥当性を説明する.

4.3.2　結果から予測される問題を提起する.

- 結果が生じた理由について考察する.　1. 2. グラフの分布を見ると、縦横で切るより斜めで切ったほうがいい

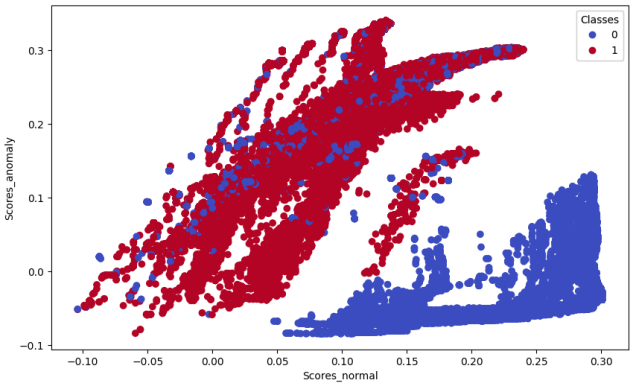

図 5　FI_UNSW に関する図の説明（和文）
Fig. 5　Description of the FI_UNSW figure (English).

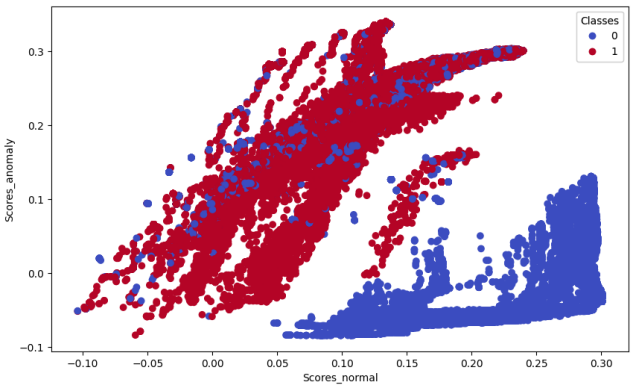

図 6　classes_UNSW に関する図の説明（和文）
Fig. 6　Description of the classes_UNSW figure (English).

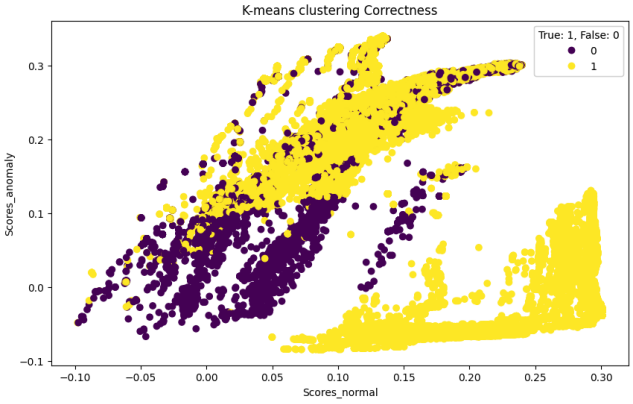

図 7　collectness_UNSW に関する図の説明（和文）
Fig. 7　Description of the collectness_UNSW figure (English).

- 本実験結果を認めると，どのような現象の予測や応用可能性があるかを述べる. 1 . 2 . より良い局面、より良いアルゴリズムがあるかも

5.　おわりに

おわりにを書く.

謝辞　謝辞を書く.

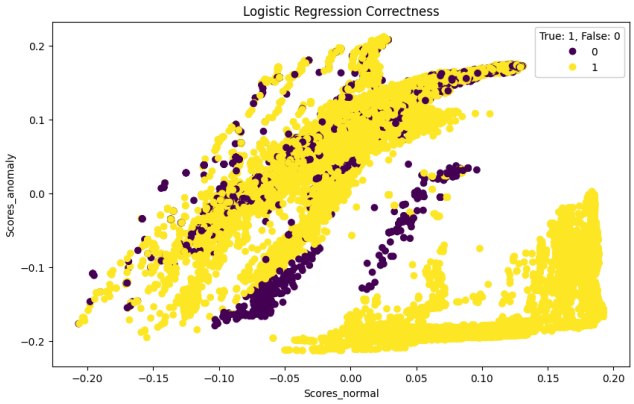

図 8　collectness_UNSW2 に関する図の説明（和文）

Fig. 8　Description of the collectness_UNSW2 figure (English).

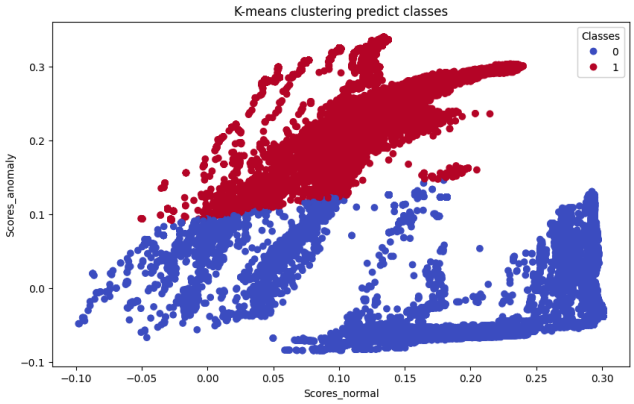

図 9　predict_classes_UNSW に関する図の説明（和文）

Fig. 9　Description of the predict_classes_UNSW figure (English).

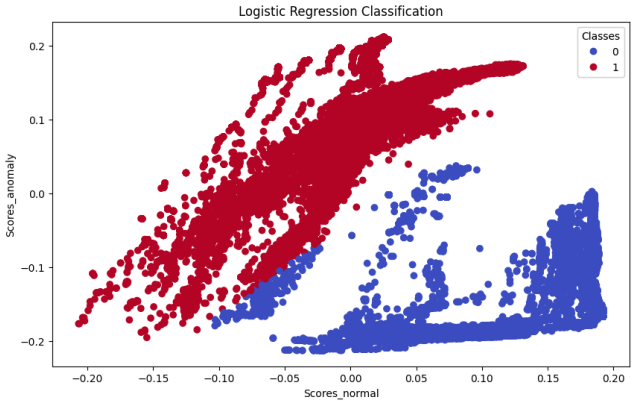

図 10　predict_classes_UNSW2 に関する図の説明（和文）

Fig. 10　Description of the predict_classes_UNSW2 figure (English).

参考文献