2018年度 優秀

敵対的生成ネットワークを用いた異常検知とその改良手法の評価

機械学習を活用した異常検知において、教師なし学習の手法を用いる有効性は高い。急速に増加するさまざまなデータに応じて検知精度を高めるための研究開発が行われている中で、敵対的生成ネットワーク（GAN：Generative Adversarial Network）による異常検知手法が提案されているが、先行研究におけるネットワーク構成ではマッピングが一つに定まらず再構成能力の低下につながることが指摘されている。本稿では、GANに基づく異常検知手法において、データのマッピングを一貫的に制約することによる改良手法を提案し、汎用データセットを用いた実験を通じて異常検知精度を改善できることを確認した。

1. はじめに

ディープ・ラーニングの活用シーンの一つに画像や時系列データの異常検知があるが、異常パターンは無数にあることが想定され、ラベル付けをして学習させる教師あり学習の適用が難しい場面がある。その場合、ラベルを使用せず正常例のみで学習データを作成する教師なし学習が効果的である。近年、教師なし学習の手法としてGANを適用し、従来の機械学習の手法に比べて精度改善を報告している研究がある[1][2]。本稿では、2018年に提案された3つのニューラル・ネットワーク構成のGANによる異常検知手法を改良する手法を提案し、既存手法との比較実験を行うことでその有効性を確認する。

2. 提案手法

図1の3つのニューラル・ネットワークからなるGANのアーキテクチャーを考える。Generatorは生成データの特徴量に相当するランダムノイズを入力とすることで偽のデータを生成し、Discriminatorはデータの真偽を判定するように学習を進めることで、Generatorは学習データに類似したデータを生成できるようになる。また、Encoderはデータに対応する特徴量を取り出す。異常検知を行う場合、与えられたデータからEncoderで特徴量を取り出し、Generatorで再構成を行う。正常データのみで学習を行った場合、学習データに含まれないデータは生成できないので、再構成できないデータを異常とみなし異常検知を行う。

2017年にSchleglら[1]は、GANを用いた異常検知としてAnoGANを提案した。AnoGANはGeneratorとDiscriminatorの2つのネットワークで構成され異常検知を実現している。続いてZenatiら[2]は、AnoGANを発展させ、データから特徴量へのマッピングの手法としてEncoderを用いることで、**図1**の構成に基づく方法を提案した。Encoderを用いることで、データとそれに対応する特徴量の間でより精度の高いマッピングが実現でき、異常検知精度の改善につながることが実験を通じて示された。本稿

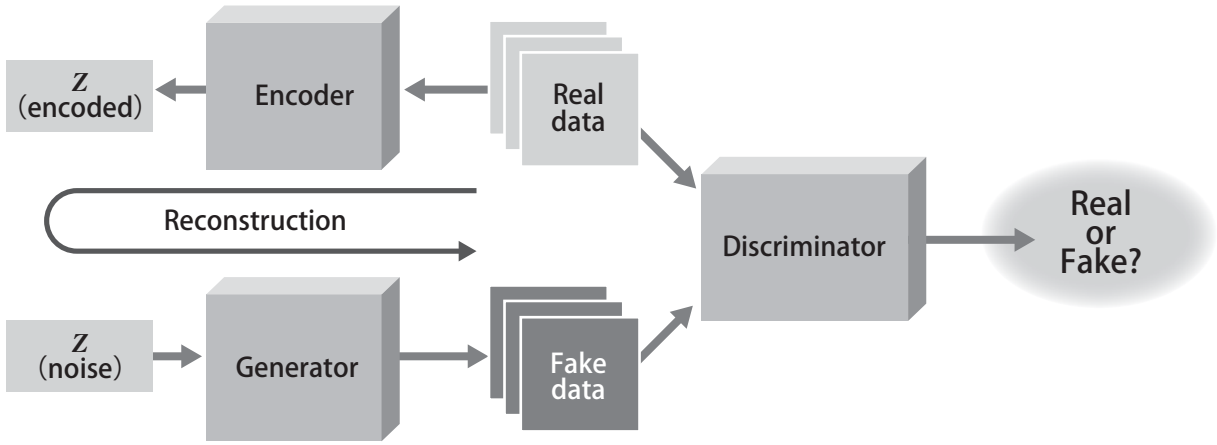

図1. GANのアーキテクチャー

ではZenatiら[2]の提案手法をEfficient-AnoGANと呼ぶ。

本稿ではこの先行研究に対する改善手法として、データと対応する特徴量の間のマッピングの改善方法を提案する。GANの構成において、マッピングが1つに定まらず再構成能力の低下につながることが指摘されている[3]。また、Liら[3]はGANのコスト関数にマッピングの制約項を設けることで、一貫的なマッピングに改善ができることを示している。このような制約は異常検知における場合においても有効であると考え、Efficient-AnoGANをベースにコスト関数にマッピングを制約する項を設け再構成能力を向上させることで、異常検知精度の改善につなげる手法を提案する。

3. 実験と評価

前章で提案した手法の有効性を確かめるため、一般に公開されているデータセットであるCIFAR-10(画像分類用データ)とKDD CUP99(ネットワーク侵入検知の数値データ)による実験を行った。CIFAR-10においては10カテゴリーのうち、1カテゴリーを正常、残り9カテゴリーを異常として扱う。また評価指標として、CIFAR-10はAUC(Area Under Curve)で評価し、KDD CUP99は適合率、再現率およびFスコアで評価する。比較対象として、One-Class Support Vector Machine、Efficient-AnoGANと比較する。

(1)CIFAR-10

実験結果を**表1**に示す。提案手法は、8カテゴリーにおいて比較対象手法のAUCを上回っている。カテゴリー全体の平均値では、従来手法より約5〜6ポイントの改善ができている。カテゴリーでその改善率には差があり、大きいものでは15ポイント超の改善となる。提案手法をEfficient-AnoGANと比較した場合、9カテゴリーにおいて改善されている。カテゴリーによって改善率に差があり、提案手法の精度が下がっているものも見られる。

(2)KDD CUP99

実験結果を**表2**に示す。提案手法は示したすべての指標において提案手法は他手法を上回っている。数値データにおいても、比較対象手法より良好な結果を示すことが確認できた。

4. おわりに

今後の課題として以下が挙げられる。

①ネットワーク・アーキテクチャーおよび学習プロセスの見直しによる異常検知精度の改善

②より実用的なデータセットを対象にした追加実験

※全文はPROVISIONのWebサイトに掲載しております。

[参考文献]
[1] T. Schlegl et al. : Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discover, arXiv, 2017.
[2] H. Zenati et al. : Efficient GAN-Based Anomaly Detection, arXiv, 2018.
[3] C. Li et al. : ALICE: Towards Understanding Adversarial Learning for Joint Distribution Matching, Neural Information Processing Systems(NIPS), 2017.

日本アイ・ビー・エム システムズ・エンジニアリング株式会社
ワトソン・ソリューション
ITスペシャリスト
平内 雅則
Masanori Hirauchi

2014年日本IBM入社。近年は日本IBM システムズ・エンジニアリングで、ディープ・ラーニングを中心に機械学習に関連するプロジェクトに参画。

表1. CIFAR-10の実験結果(AUC)

カテゴリー	OC-SVM	Efficient-AnoGAN	提案手法
Airplane	0.6623	0.7137	**0.7939**
Automobile	0.4941	**0.4977**	0.4711
Bird	0.6345	0.6865	**0.7545**
Cat	0.5327	0.5636	**0.6310**
Deer	0.6714	0.7459	**0.7509**
Dog	0.6030	0.5433	**0.6701**
Frog	**0.7227**	0.6885	0.7026
Horse	0.5979	0.5455	**0.6025**
Ship	0.6392	0.7183	**0.8090**
Truck	**0.6381**	0.4511	0.5206
全体平均	0.6195	0.6154	**0.6706**

表2. KDD CUP99の実験結果

手法	適合率	再現率	Fスコア
OC-SVM*	0.7457	0.8523	0.7954
Efficient-AnoGAN*	0.9200±0.00740	0.9582±0.0104	0.9372±0.0440
提案手法	**0.9524**	**0.9676**	**0.9599**

＊OC-SVM、Efficient-AnoGANの結果は、本実験と同条件の結果のためZetaniら[2]より引用