テーマ番号	EP076			
プロジェクトテーマ	和文	深層学習を用いた侵入検知システムに関する研究	指導教員	長田 茂美 教授
	英文	Research on the Intrusion Detection System Based on Deep Learning		
プロジェクトメンバー		4EP3-32　仲地　駿人 (Takato Nakachi)		

Abstract　近年，さまざまなモノがインターネットに接続される IoT(Internet of Things)の時代において，モノを標的としたサイバー攻撃を検知することは必須の機能となっている．サイバー攻撃を検知するシステムとして，侵入検知システム(IDS, Intrusion Detection System)があるが，IDS では，必ずしも新種サイバー攻撃に対応できるとは限らない．また，セキュリティ人材不足から，攻撃種類の分類を自動化したいという要求もある．本研究では，畳み込みニューラルネットワーク(CNN, Convolution Neural Network)を用いて，既知攻撃を検知および分類するシステムを開発し，評価実験によりその有用性を確認した．また，CNN を用いた IDS によって，未知攻撃を検知および分類できる可能性も示した．

Keywords　intrusion detection system, deep learning, convolutional neural network, multilayer perceptron, KDD CUP 1999.

1. まえがき

近年，ICT 技術の発展により，さまざまなモノがインターネットに接続される IoT 時代に突入し，新種のマルウェアやサイバー攻撃は増加の一途を辿っている．それに伴い，サイバー攻撃を検知する侵入検知システム(IDS, Intrusion Detection System)が注目を集めている．IDS は，ネットワークやサーバへの通信データを監視し，異常がある通信データを検知して，管理者へ通知する．IDS が攻撃を検知する方式として，シグネチャ型 IDS とアノマリ検知型 IDS がある．前者のシグネチャ型 IDS は，既知の攻撃パターンをデータベースに登録しておき，監視している通信データとのパターンマッチングにより攻撃を検知するため，未知攻撃を検知できないという問題がある．一方，後者のアノマリ検知型 IDS は，通信データと IDS に登録されている特徴との類似性を使って攻撃を検知するため，未知攻撃を検知できるが，誤検知が多いという問題がある．

セキュリティ人材不足[1]が問題となっている今日，攻撃の種類によって対策法が異なるため，攻撃の種類別に分類する必要があり，攻撃の検知および分類を自動的に実行できるシステムが求められている．

このような IDS の研究では，KDD CUP 1999 Dataset[2]が用いられている．これは，1998 DARPA Intrusion Detection Evaluation Data Set の通信データをセッション単位に加工したデータセットである．

本研究では，深層学習の一種である畳み込みニューラルネットワークを，この KDD CUP 1999 Dataset に適用し，既知攻撃の検知および分類，さらには，未知攻撃の検知および分類の自動化を目指す．

2. システム概要

2.1 通信データセット

KDD CUP 1999 Dataset には，4,893,980 件のフルデータ(kddcup.data)と，フルデータの約 10%を抽出した 494,021 件のデータ(kddcup.data_10_percent)がある．本研究では，kddcup.data_10_percent を学習用データ，テスト用データに分けて使用する．学習用データの全データ数は 345,791 件，テスト用データの全データ数は 148,230 件である．表 1 に，KDD CUP 1999 Dataset のクラス，ラベルとクラスごとのデータ件数を示す．

2.2 前処理

KDD CUP 1999 Dataset では，41 次元の特徴が用意されている．そのうち，3 つの特徴が文字列である．3 つの特徴はプロトコルタイプ，ネットワークサービス，接続状態である．これらの 3 つの特徴はいずれも重要な特徴であり，本研究では，文字列にダミー変数を適用した 122 次元の特徴を用いて検証する．また，学習を効率的に進めるため，全データに対して正規化を行う．

表1 クラス，ラベル，データ件数
Table 1 Classes, attack labels, and number of data.

class		label name	number of data	
			train	test
Normal		normal	68,946	29,232
Attack	Dos	back, land, neptune, pod, smurf, teadrop	274,017	117,441
	U2R	buffer_overflow, loadmodule, perl, rootkit	31	21
	R2L	ftp_write, guess_ passwd, imap, multihop, phf, spy, warezclient, warezmaster	797	329
	Probe	ipsweep, nmap, portsweep, satan	2900	1207

2.3 攻撃の検知および分類

攻撃の検知および分類には，畳み込みニューラルネットワーク(CNN, Convolutional Neural Network)を用いる．攻撃の検知は，通信データから攻撃通信データを検知する．すなわち，通信データを通常通信データと攻撃通信データに分類する．分類手法では，攻撃通信データを Dos 型，U2R 型，R2L 型，Probe 型の計 4 種類に分類する．

2.4 実験方法

図 1 に，未知攻撃の検知および分類の実験方法を示す．提案手法では，未学習の攻撃ラベルを未知攻撃として扱う．攻撃クラスの 1 つの攻撃ラベルをテスト用データとし，同じ攻撃クラスのテスト用データ以外の全攻撃ラベルを学習用データと扱い，Dos 型，U2R 型，R2L 型，Probe 型の計 4 種類に分類する．

テーマ番号	EP076

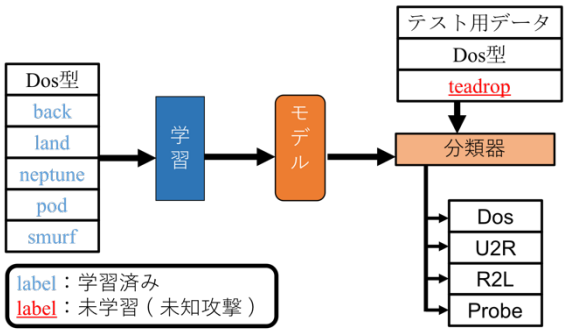

図1 実験方法

Fig. 1 Experimental method.

3. システム評価

CNN の有効性を検証するために，攻撃の検知および攻撃の分類に分けて検証を行い，多層パーセプトロン(MLP, Multilayer perceptron)との比較実験 I を行う．また，CNN による未知攻撃の検知および分類の検証も行う．また，学習用データ数の多いクラスと少ないクラスが混在しているため，学習用データ数の少ないクラスの特徴を無視して，学習用データ数の多いクラスの特徴を捉えてしまい，テスト用データが学習用データ数の多いクラスに分類されるという不均衡問題が発生する可能性がある．

そこで，本システムでは，学習用データ，テスト用データともに，1 クラス，最大でも 1000 インスタンスになるように調整を行った．また，本研究では，攻撃の検知および分類に重きを置くため，再現率(recall)を評価基準として扱う．

3.1 攻撃の検知

攻撃の検知では，通常通信データ(normal)と攻撃通信データ(attack)の 2 種類に分類する比較実験 I を行った．表 2 に，その結果を示す．比較実験 I の結果から，MLP と比べ，CNN の方が既知攻撃の検知に有効であるといえる．

表2 比較実験結果 I

Table 2 Result of comparative experiment I.

	MLP		CNN	
class	normal	attack	normal	attack
correct	983	2334	985	2349
mistake	17	16	15	1
recall	98.3%	99.319%	98.5%	99.8%

3.2 攻撃の分類

攻撃の分類では，Dos 型，U2R 型，R2L 型，Probe 型の計 4 種類のクラスに分類する比較実験 II を行った．表 3 に，その結果を示す．CNN による U2R 型，R2L 型の再現率は MLP と変わらないが，Dos 型，Probe 型の再現率は高くなっていることがわかる．したがって，CNN の方が既知攻撃の分類に有効であるといえる．

表3 比較実験結果 II

Table 3 Result of comparative experiment II.

	MLP			
class	Dos	U2R	R2L	Probe
correct	997	12	327	996
mistake	3	9	2	4
recall	99.7%	57.142%	99.392%	99.6%
	CNN			
class	Dos	U2R	R2L	Probe
correct	999	12	327	1000
mistake	1	9	2	0
recall	99.9%	57.142%	99.392%	100%

表4 評価実験結果

Table 4 Result of evaluation experiment.

class	label name	number of data	recall
Dos	back	2,203	0%
	land	21	100%
	neptune	107,201	0.8%
	pod	264	98.106%
	smurf	280,790	0%
	teadrop	979	90.194%
U2R	buffer_overflow	30	63.333%
	loadmodule	9	22.222%
	perl	3	100%
	rootkit	10	20%
R2L	ftp_write	8	75%
	guess_passwd	53	0%
	imap	12	0%
	multihop	7	71.428%
	phf	4	0%
	spy	2	0%
	warezclient	1,020	2.1%
	warezmaster	20	100%
Probe	ipsweep	1,247	98.4%
	nmap	231	98.701%
	portsweep	1,040	99.7%
	satan	1,589	99.2%

3.3 未知攻撃の検知および分類

CNN による未知攻撃の検知および分類に関する仮説を検証するために，各攻撃クラスの 1 つの攻撃ラベルをテスト用データとし，それ以外の攻撃ラベルを学習用データとして評価実験を行った．表 4 に，評価実験結果を示す．この結果から，CNN は類似した未知攻撃の検知および分類に有効であることがわかる．特に，Probe 型の再現率が，他の攻撃クラスの再現率と比べて高くなっている．その原因として，Probe 型はポートスキャンなどの攻撃対象に対する調査，偵察を目的する攻撃であり，攻撃通信データが類似している傾向にあることが考えられる．しかし，R2L 型の再現率は，他の攻撃クラスと比べ，低くなっている．R2L 型は外部から攻撃対象のサーバなどに不正ログインすることを目的とする攻撃であり，攻撃ラベルによってネットワークサービスなどの特徴が異なることに起因するものと考えられる．

今後は，学習用データ，テスト用データの組み合わせ方や新たな手法の検討をはじめ，未知攻撃の検知および分類の精度向上を図っていく予定である．

4. むすび

本研究では，CNN を用いて既知攻撃の検知および分類するシステムを開発し，MLP との比較実験によって，CNN を用いた侵入検知システムの有用性を確認した．また，評価実験により，類似した未知攻撃の検知および分類を実現できる可能性を示せた．今後も，評価および改良を継続し，実用的なシステムを目指す予定である．

参考文献

[1] 経済産業省：”IT 人材の最新動向と将来推計に関する報告書”， http://www.meti.go.jp/policy/it_policy/jinzai/27FY/ITjinzai_report_summary.pdf， 2016.6 2018 年 1 月 19 日参照

[2] Hettich S., and Bay S., “KDD Cup 1999 Data, “University of California The UCI KDD Archive, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html 1999.10.