大規模言語モデルを用いた侵入検知手法の性能分析

小松翔[1]　森康祐[1]　廣津登志夫[1]

概要：近年，大規模言語モデル（LLM）を用いた侵入検知が研究されている．しかし，LLM が従来手法と同等以上の分類精度を示した場合でも，従来手法の代替となり得るかは明らかではない．同程度の分類精度であっても，個々の通信に対する判定結果が異なる可能性があるためである．本稿では，3 種類の DDoS 検知データセットに対し，ロジスティック回帰，ランダムフォレスト，多層パーセプトロン，Qwen を用いて二値分類を行った．元の特徴量を用いる条件と，偏りや重複のある特徴量を除いた条件を設け，計 24 条件の性能を比較した．さらに，同一通信に対する特徴量条件間の正誤の変化と，特徴量削除後の各従来手法と Qwen の正誤を 1 対 1 で比較した．その結果，特徴量条件の変更に伴う判定の変化には手法間で違いが見られた．今後は通信途中に実際に観測できる情報を利用し，実環境に近い検知条件で判定挙動を検証する．

Performance Analysis of Intrusion Detection Methods Using Large Language Models

1. はじめに

DDoS 攻撃の検知には通信特徴量を用いた機械学習手法が広く利用されており，近年は大規模言語モデル（LLM）を用いた侵入検知も研究されている [1], [2]．しかし，LLM が正しく検知できる通信や誤判定する通信が従来手法と一致しているかは不透明であり，従来手法と同等以上の分類精度を示したとしても従来手法をそのまま置き換えられるとは限らない．

本研究では，LLM が従来手法の代替となり得るかを検討するため，両者の判定結果の差異に着目する．本稿では，3 種類の DDoS 検知データセットを用い，従来手法と LLM の分類性能を比較するとともに，特徴量条件を変更した際の各手法の正誤の変化と，同一通信に対する従来手法と LLM の判定結果の違いを分析する．

2. 関連研究

Mahmoodi ら [1] は，通信特徴量をテキスト化し，LLM を用いた DDoS 検知を評価した．Mehavilla ら [2] は，フローベースの侵入検知について，LLM と機械学習・深層学習モデルの性能を比較した．本稿では性能指標に加え，同一通信に対する正誤を特徴量条件間で照合し，従来手法を Qwen に置き換えた場合の判定上の差異を各手法との 1 対 1 比較から調べる．

3. 実験方法

CIC-IDS2017，CSE-CIC-IDS2018，CICIoT2023 の 3 種類のデータセットを使用し，正常通信（Benign）と DDoS 通信の二値分類を行った．実験にはデータセットを Benign と DDoS を 1 対 1 に調整し 5 万件にした `50k-balanced` 版を用い，そのうち 4 万件を学習用，5 千件を過学習の検知用，5 千件を評価用データとした．`full` は元の特徴量を用い，`clean-1` は予備調査で値の偏りや重複を確認した列を除いた比較用条件とする．CIC-IDS2017 では DDoS でポート 80 に集中した `Destination Port` の 1 列と，ほかの列と重複する 6 列の計 7 列，CSE-CIC-IDS2018 では `Dst Port` と `Init Fwd Win Byts` の 2 列，CICIoT2023 ではクラス間で値が偏る `Number` の 1 列を除いた．

従来手法にはロジスティック回帰（LR），ランダムフォレスト（RF），多層パーセプトロン（MLP）を，LLM には Qwen3-8B-Base（Qwen）を使用した．合計 24 条件について，以下の 3 点を調べる．（1）`full` と `clean-1` の分類性能，（2）同一手法で特徴量条件を変更した際の正誤の遷移，（3）`clean-1` で LR，RF，MLP のそれぞれと Qwen の正

[1] 法政大学
Hosei University

© 2026 Information Processing Society of Japan

表 1 50k-balanced 条件における分類精度（Accuracy，%）
Table 1 Classification accuracy (%) in the 50k-balanced setting.

データセット	特徴量条件	LR	RF	MLP	Qwen
CIC-IDS2017	full	99.90	100.00	99.84	99.98
	clean-1	99.90	99.98	99.76	99.94
CSE-CIC-IDS2018	full	99.62	99.88	99.02	99.98
	clean-1	98.68	99.44	97.72	99.96
CICIoT2023	full	100.00	100.00	100.00	100.00
	clean-1	100.00	100.00	100.00	99.98

表 2 2018 データセットにおける正誤の遷移（件）
Table 2 Correctness transitions on CSE-CIC-IDS2018.

手法	正 → 正	正 → 誤	誤 → 正	誤 → 誤
LR	4928	53	6	13
RF	4972	22	0	6
MLP	4856	95	30	19
Qwen	4997	2	1	0

誤を 1 対 1 で比較し，従来手法を置き換えた場合の判定結果の違いを調べる．正誤は正解ラベルと予測ラベルの一致によって定義する．

4. 実験結果と考察

4.1 分類性能の比較

表 1 に 24 条件の Accuracy を示す．全条件で 97.72% 以上を示した．特に差のあった CSE-CIC-IDS2018 では `clean-1` への変更に伴い，従来 3 手法の Accuracy は 0.44〜1.30 ポイント低下したが，Qwen の低下は 0.02 ポイントであった．しかしながら，この実験では全手法を通して 100% に近い精度が出ているため，この結果が各手法の差を誤差なく示していると決めることはできない．

4.2 特徴量条件を変更した際の判定変化

同一通信について，`full` と `clean-1` の正誤の遷移を集計した．**表 2** には変化の大きかった CSE-CIC-IDS2018 を示す．MLP は正解から誤判定へ 95 件，誤判定から正解に 30 件変化したが，Qwen ではそれぞれ 2 件，1 件であった．2017 の予測変更数は LR，RF，MLP，Qwen の順に 4，1，4，2 件，2023 では 0，0，0，1 件であった．とくに 2018 の MLP では，誤判定が 65 件増加した一方，個別には 125 件で正誤が入れ替わった．したがって，手法ごとに判定する際の頼る情報が異なり，Accuracy の差のみではその挙動を十分に捉えられない．

4.3 clean-1 における Qwen と従来手法の比較

`clean-1` で各従来手法と Qwen の正誤を同一通信ごとに比較した．**表 3** には，従来手法のみ正解（従来正・Q 誤）と，Qwen のみ正解（従来誤・Q 正）の件数を示す．2018 では後者が LR で 65 件，RF で 26 件，MLP で 113 件であった．うち LR の 62 件，MLP の 107 件は Benign であり，従来手法の誤検知を Qwen が正しく判定した例である．一方，2017 では RF が正解した DDoS を Qwen が 2 件見逃し，2023 でも従来 3 手法が正解した DDoS を Qwen が 1 件見逃した．したがって，置換時には誤検知と見逃しの違いを考慮する必要がある．

表 3 clean-1 における各従来手法と Qwen の 1 対 1 比較（件）
Table 3 Pairwise correctness comparison with Qwen (clean-1).

データ	手法	従来正・Q 誤	従来誤・Q 正
2017	LR	2	4
	RF	2	0
	MLP	1	10
2018	LR	1	65
	RF	0	26
	MLP	1	113
2023	LR	1	0
	RF	1	0
	MLP	1	0

5. おわりに

本稿では，従来 3 手法と Qwen について，特徴量条件による検知性能と正誤の遷移，および同一通信に対する手法間の判定結果を比較した．検知性能が近い場合でも判定が異なる通信が存在し，Qwen への置換によって正しく判定できるようになる通信と，反対に誤判定となる通信の双方を確認した．一方，本実験で用いた特徴量には，通信が終了した後に確定する統計量も含まれているため，本来の検知環境とは結果が異なる可能性がある．そこで今後は，フローを分割し順に読み込ませることで各手法の判定挙動を比較する．

参考文献

[1] Mahmoodi, M. and Jameii, S. M.: Utilizing Large Language Models for DDoS Attack Detection, OTCON (2024). DOI:10.1109/OTCON60325.2024.10688345.
[2] Mehavilla, L. et al.: Evaluating Large Language Models Effectiveness for Flow-Based Intrusion Detection: A Comparative Study with ML and DL Baselines, Artificial Intelligence Review, Vol. 59, Article 50 (2026). DOI:10.1007/s10462-025-11432-2.

© 2026 Information Processing Society of Japan