要 旨

不正侵入パケットに対する機械学習の適用

松尾　達郎

機械学習を用いた異常検知型侵入検知システムはシグネチャ型の問題点である未知の攻撃を検知できないという欠点を解消する手法として注目されている．そのため，異常検知型侵入検知システムの精度を向上する研究が行われている．しかし，近年のネットワークの高速化と取り扱うデータの増大により，異常パケットの判別に用いられる機械学習は，精度に加え学習の速さが重要になってきている．そこで，本研究では，異常検知型侵入検知システムの学習アルゴリズムとして，近年提案された高速な学習アルゴリズムである Extreme Learning Machine(ELM) を適用し，汎化性能が高く，多くの実問題で用いられているサポートベクターマシン (SVM) と，精度と学習速度を比較する．データセットには UCI Machine Learning Repository 提供のデータセット「KDD Cup 1999 Data」の訓練データサンプルを用いる．4,898,431 個のデータに対して，500〜1 行の一定行間隔のサンプリングを行うことでサブデータセットを生成し，それぞれのサブデータに対して判別を行う．実験の結果，ELM の学習時間はデータセットの大きさに対して線形的に増加し，データセット全体に対しては，SVM を用いた学習は精度が 99.91%，学習時間が 34,902 秒であるのに対し，ELM を用いた学習は精度が 99.84%，学習時間が 865 秒となることを示す．この結果は，ELM が精度を保ったまま，学習を高速化できることを示しており，高速通信路を短時間で通過する大きなサイズのデータに対する学習法として有用であると考える．

キーワード　異常検知型侵入検知システム，Extreme Learning Machine，サポートベクターマシン，KDD Cup 1999 Data

Abstract

A Study of Network Intrusion Detection System using Machine Learning Algorithm

Tatsuro MATSUO

Anomaly intrusion detection system(IDS) using machine learning has been studied by many researchers to overcome the weakness of signature-based IDS in detecting unknown attacks. Improvement of precision of classification is an important issue in the research area of anomaly IDS. High speed network produces tremendous amount of data to inspect by IDS. Then not only precision of classification, but also the speed of convergence is important for a real network application. In this study, the extreme learning machine(ELM), which is a novel neural-network learning algorithm proposed in 2006, is applied to IDS, and is compared with support vector machine(SVM), which is a popular learning algorithm applied in various area. The advantage of ELM is its convergence speed. The precision and convergence speed of ELM are compared with those of SVM. For abnormal packets classification, the dataset "KDD Cup 1999 Data" by University of California Irvine, Machine Learning Repository is used. Total number of the data is 4,898,431 and sub-datasets are generated by equal interval sampling. The discrimination is performed using each sub-dataset. The result shows linear increase of learning time using ELM for datasets, while the learning time of SVM increases in cubic to the number of data. SVM classifier achieves 99.84% of the precision and 34,902 seconds for learning. ELM classifier achieves 99.84% of the precision and 865 seconds for learning. These results show that learning speed of ELM improves while its

classification precision is preserved, and ELM has an advantage for IDS in a high speed communication network.

key words Intrusion Detection System, Extreme Learning Machine, Support Vector Machine, KDD Cup 1999 Data