Feature Selection for Anomaly–Based Network Intrusion Detection Using Cluster Validity Indices

Tyrone Naidoo_, Jules–Raymond Tapamoy, Andre McDonald_

Modelling and Digital Science, Council for Scientific and Industrial Research, South Africa

1tnaidoo2@csir.co.za

3amcdonald@csir.co.za

yCollege of Agriculture, Engineering and Science, University of Kwazulu–Natal, South Africa

2tapamoj@ukzn.ac.za

Abstract

A feature selection algorithm that is novel in the context of anomaly–based network intrusion detection is proposed in this paper. The distinguishing factor of the proposed feature selection algorithm is its complete lack of dependency on labelled data, which is rarely available in operational networks. It uses normalized cluster validity indices as an objective function that is optimized over the search space of candidate feature subsets via a genetic algorithm. Feature sets produced by the algorithm are shown to improve the classification performance of an anomaly–based network intrusion detection system over the NSL-KDD dataset. The system approaches the performance attained by using feature sets derived from labelled training data via existing wrapper and filter–based feature selection algorithms.