Multilevel Feature Extraction and Selection for Network Intrusion Detection

Introduction
- Using the recently collected Aegean Wi-Fi Intrusion Dataset (AWID) of wireless network server logs which contain real traces of both normal and intrusive network traffic, this study explores methodologies to detect different wireless attacks using clustering methods over the high-dimensional dataset.

Research goal:
- Investigate if the combined feature extraction and feature selection methodology can improve:
  - Prediction accuracy of the different classification methods as the intermediate results;
  - Prediction accuracy of the different clustering methods as the final results.

Contribution:
- Adopt an unsupervised feature extraction and a supervised feature selection methodology to increase accuracy of clustering to detect all attack classes.
- Construct an unsupervised multilevel Auto-Encoder with two consecutive hidden layers to extract features.
- Use four Machine Learning Classifiers with Decision Tree (CART), Random Forests (RF), Artificial Neural Network (ANN), and Support Vector Machine (SVM) to select features from the combined feature pool.
- Build a framework named Deep Learning Feature Extraction and Selection Intrusion Detection System (DLFESIDS) to detect and predict intrusion attempts over the 802.11 wireless networks.
- Combine multilevel auto-encoders for unsupervised feature extraction with a feature selection methodology to reduce dimensionality.
- Apply Clustering Algorithms in the transformed subspace.

Framework of FESACIDS

FESACIDS Data Preprocessing
- Transform the attributes with "object" datatype into numerical.
- Map the string to integer classes: 1 for normal instances, 2 for an impersonation, 3 for a flooding, and 4 for an injection attack.
- Map the symbolic attributes to integer values.
- Transform the hexadecimal data into integer values.
- Transform the attributes with "object" datatype into numerical;
- Transform the classification labels into 4 binary features ('is Normal', 'is Impersonation', 'is Flooding', 'is Injection');
- Map the symbolic attributes to integer values;
- Train the linear SVM with recursive feature elimination.
- Select the features on the top five levels of the tree according to the stored feature weights.
- Select the features whose weights are greater than a threshold.
- Select the features by the four feature selection methods over the 154 features and the combined 204 features.
- Use only one hidden layer. I use the rectifier activation function. The softmax layer minimizes the cross-entropy loss function.
- The grid-best tree is trained on the training set and then used to predict classification over the testing set.
- The feature extraction step makes the feature set more sensitive to the correlation among features.
- The Hopkins Statistic is 0.9997. The data is highly clustered.
- PCA produces the three principal components (PCs) from the 48 input features selected by the SVM-RFE method.
- Accuracy scores are low due to multiclass classification and separate generating process for the training set and testing set.
- The feature extraction step makes the feature set more sensitive to the feature selection methods. But it also introduces high correlation among features.
- The SVM with the recursive feature elimination procedure (RFE) has the highest accuracy based on the its selected features. RFE is better than the ranked feature elimination because it takes into account of the combined effect of features on the classification.

FESACIDS Principal Component Analysis
- The Hopkins Statistic is 0.9997. The data is highly clustered.
- PCA produces the three principal components (PCs) from the 48 input features selected by the SVM-RFE method.

FESACIDS Model Comparison
- The feature extraction and selection procedure improves prediction accuracy for Classification and Clustering.
- The hierarchical clustering approach performs better than the flat partitioning approach in clustering this dataset.

Conclusion
- The feature extraction and selection procedure improves prediction accuracy for Classification and Clustering.
- The hierarchical clustering approach performs better than the flat partitioning approach in clustering this dataset.

References