Intrusion detection using machine learning

MD. ANAS

STD.NO. 150587

MD. SAYEDUZZAMAN
STD. NO. 0405035

Presentation Outline

- ❑ Introduction
- ❑ Previous works
- ❑ Motivation
- ❑ Problem statement
- ❑ Program
- ❑ Future direction and conclusion

Introduction

- Intrusion detection system (IDS)
  - monitors a network or systems for malicious activity or policy violations.
  - reports intrusion activities to an administrator.
  - has scope from single computers to large networks.
- Classifications
  - Network intrusion detection systems (NIDS)
    - analyzes incoming network traffic.
  - Host based intrusion detection systems (HIDS)
    - e.g. monitors important operating system files

Introduction(cont'd)

- Signature-based detection
  - Looks for "known patterns" in database
  - Low false alarm rate , accurate and fast
  - Unable to detect new type of attack
  - "Only strong as its rule set"

- Anomaly-based detection
  - Tracks unknown unique behavior pattern.
  - Uses machine learning techniques.
  - Helps to reduce the "*limitations problem*"
  - High false alarm rate..

Previous Works

Intrusion Detection Using Error Correcting Output Code Based Ensemble

(S. M. AbdElrahman, Ajith Abraham, 2014)

- Used methods
  - *Meta learning ensemble Methods*
  - *One-against-all (OAA)*
  - *One-against-one (OAO)*
  - *Error correcting code (ECOC).*

- **Bottleneck** : No feature selection method

Previous Works(cont'd)

Critical study of neural networks in detecting intrusions

(Rachid Beghdad,2008)

- Used methods
  - Multilayer perceptron (MLP),
  - Generalized feed forward (GFF),
  - Self-organizing feature maps (SOFMs), *Error correcting code (ECOC).*
  - Principal component analysis networks (PCAs)
  - false alarm rate (8.16%)

- **Bottleneck** :High false alarm rate

Previous Works(cont’d)

Wrapper Model

(Ron Kohavi, George H. John, 1997)

- Finds subset of features from the feature space
- Uses search techniques such as Sequential, Complete or Random Search

- Measures accuracy of that generated feature subset using learning algorithm.
  - Support Vector Machine (SVM)
  - Artificial Neural Network (ANN)
  - Nearest Neighbor

- **Bottleneck** : Computationally expensive

Previous Works(cont’d)

Hybrid Approach

(Manoranjan Dash, Huan Liu,1997.)

- Combination of both filter and wrapper approach.
- Uses intrinsic property of the dataset along with ML algorithm
- Removes extremely redundant features through filter approach
- Remaining features are applied in wrapper approach.

Motivation

Data quality affects the accuracy of data mining algorithms.

- Two important aspects
  - data relevance
  - data redundancy

- Concerns
  - allow algorithms to operate faster
  - improvement of accuracy of data mining algorithm
  - relevance of features
  - pairwise features correlation
  - redundant , irrelevant features affect accuracy of learning algorithms

Problem Statement

Improvement of detection rate of network attacks such as DOS, U2R, R2L and Probe by extracting appropriate features set.

Correlation metrics for feature extraction

- Two correlation metrics
  - Decision independent correlation (DIC)
  - Decision dependent correlation (DDC)

- Considerations
  - Dependency among the features,
  - Dependency with respect to a given data mining task
  - Removing data redundancy.

DIC metric

Quantifies the relevance and the correlation among features .

I(Y;X) = mutual information between decision Y and features X

H(X) = uncertainty of feature X

DIC : decision independent correlation

- 0 <= DIC(X_i, X_j) <= 1
- DIC(X_i, X_j) = 0
  - features Xi and Xj are uncorrelated.
- DIC(X_i, X_j) = 1
  - Full prediction between the features

$$DIC_{X_j}(X_i, X_j) = \frac{I(X_i; X_j)}{H(X_j)},$$

$$DIC_{X_i}(X_i, X_j) = \frac{I(X_i; X_j)}{H(X_i)}.$$

DDC metric

DDC : decision dependent correlation

- decision Y associated with the Xi , Xj features
- improves the accuracy of the decision variables.
- $Q_y(X_i, X_j)$= Correlation measure to quantify the information redundancy between Xi and Xj with respect to Y.
- $Q_y(X_i, X_j)$ = 1 when Xi , Xj fully correlated with respect to Y

$$Q_Y(X_i, X_j) = \frac{I(Y;X_i) + I(Y;X_j) - I(Y;X_i,X_j)}{H(Y)}$$

$$\begin{aligned} I(Y;X_i,X_j) &= I(Y;X_i) + I(Y;X_j|X_i) \\ &= I(Y;X_j) + I(Y;X_i|X_j). \end{aligned}$$

Features set evaluation

S : Feature subset

e(S) = subset evaluation measure

- an evaluation heuristic
- specifies a subset of features with regard to the decision functions
- DDC regarded as the penalty.
- the bigger e(S), the better the feature subset

$$e(S) = \frac{\sum_{\forall j \in I_m} I(Y; X_j)}{H(Y)} - \sum_{\forall i,j \; i \neq j \; i,j \in I_m} Q_Y(X_i, X_j).$$

Feature Extraction Algorithm (FEA)

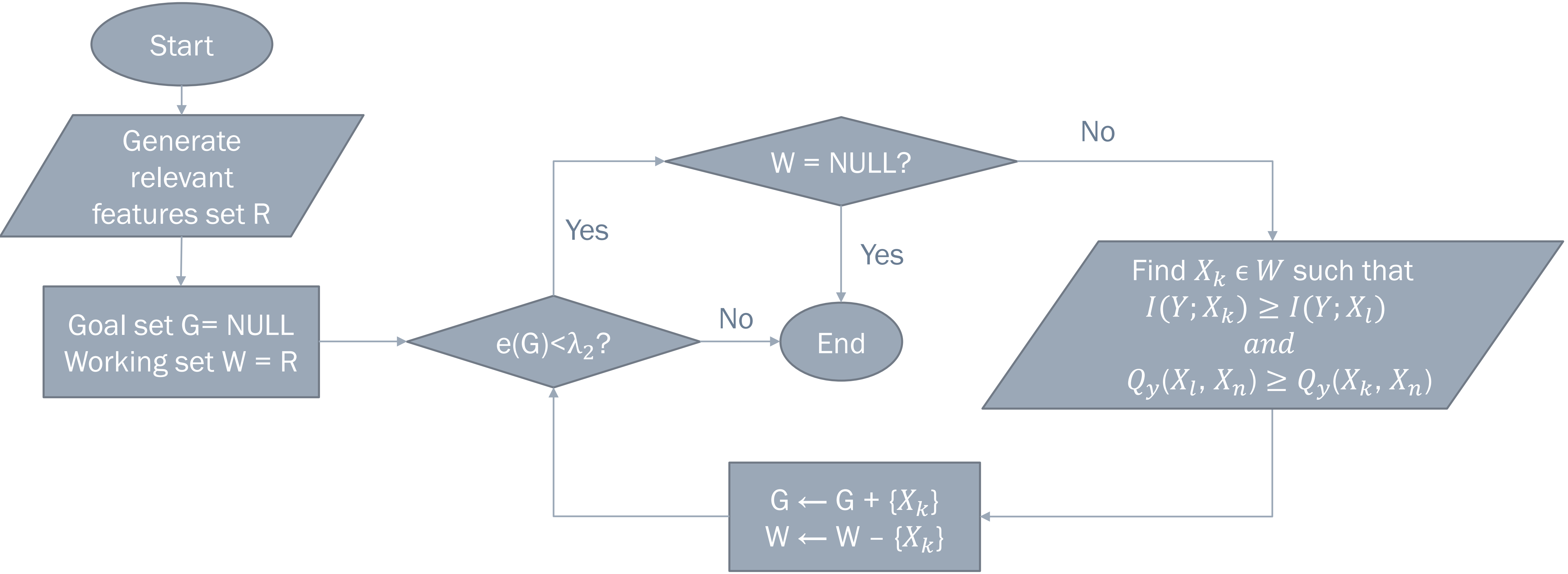
Goal

- Select the minimum set of features
- Select strongly related features to the desired decision variable
- Decrease redundancy among features.

Two functional modules.

- Focusing on removing irrelevance.
- Focusing on eliminating redundancy

Feature Extraction Algorithm (FEA)



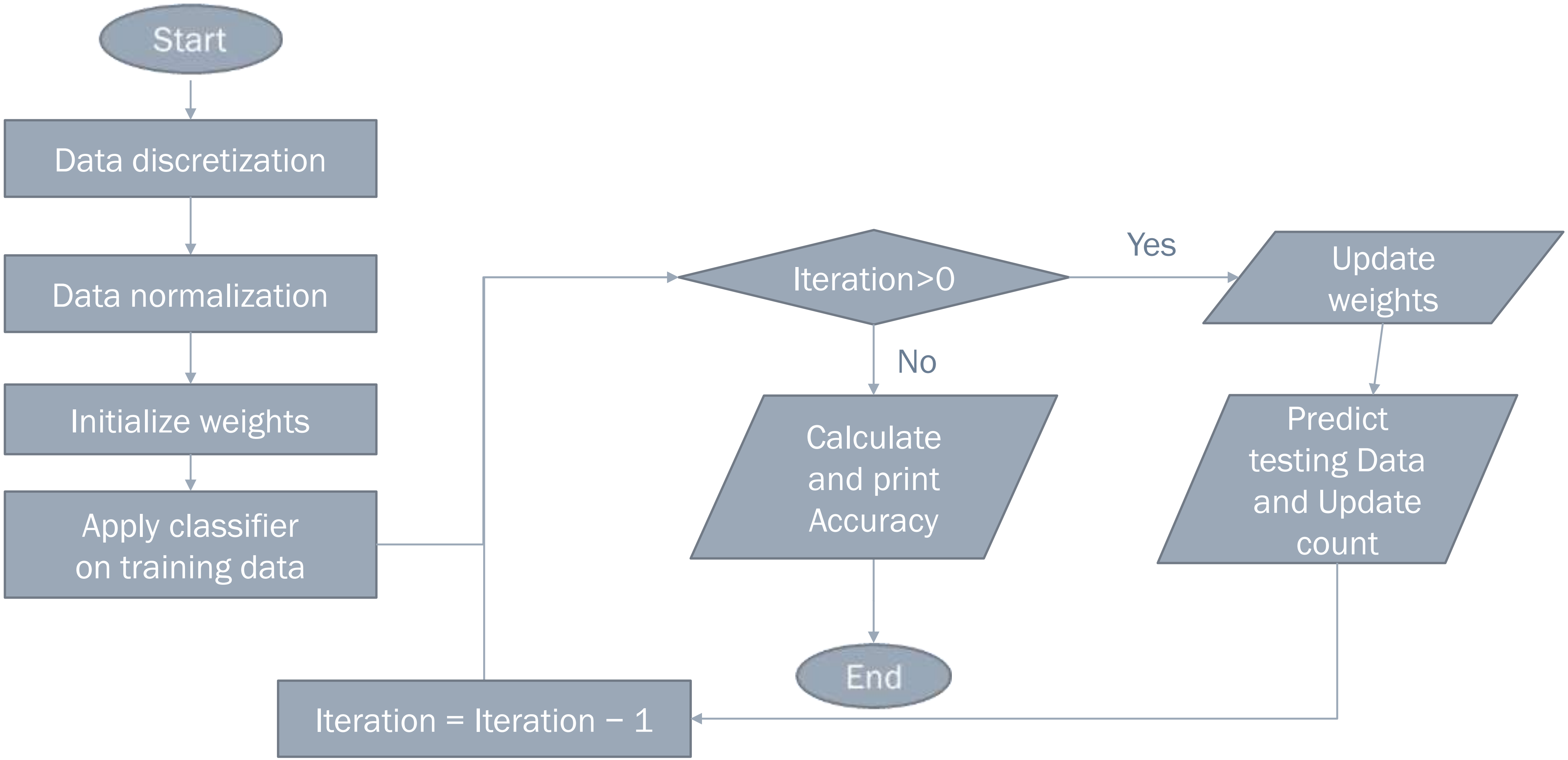
Classification Algorithm (CA)

- A machine learning approach to learn a classification function.
- The classifier has a linear function of weighted features

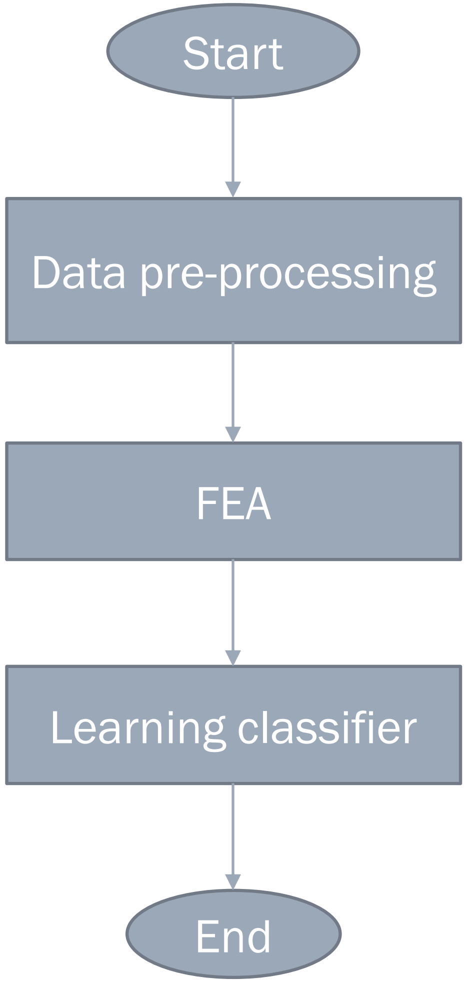
$$f(X) = \sum_{i=1}^{|X|} (w_i \,.\, x_i)$$

- Weights are generated randomly initially
  - Adjusted by back propagation later
- Stopping criteria set by number of iteration set initially.

Classification Algorithm (CA)



Final Intrusion Detection System (IDS)



Future direction and conclusion:

- DDC and the subset evaluation heuristic metric can be used to select the proper feature subset.
- Based on these features, the Learning algorithm can be a better classifier than sequential selection strategy.
- In future we will try to improve the accuracy, currently it is around 60%.

THANK YOU