Features Dimensionality Reduction Approaches for Machine Learning Based Network Intrusion Detection

Razan Abdulhammed¹, Hassan Musafer¹, Ali Alessa¹, Miad Faezipour¹², and Abdelshakour Abuzneid¹

¹Department of Computer Science and Engineering, University of Bridgeport, Bridgeport, CT, 06604, USA
²Department of Biomedical Engineering, University of Bridgeport, Bridgeport, CT, 06604, USA

Corresponding Author: Miad Faezipour (e-mail: mfaezipo@bridgeport.edu)

Summary of Modifications

We received valuable feedback from the respectful reviewers. They significantly helped us improve the quality of our manuscript. For that, we would like to thank the editor, associate/assistant editor and the reviewers. In summary, the changes applied to the manuscript based on the reviewers’ comments, are as follows:

• All the reviewers’ comments and suggestions have been carefully considered and addressed.
• A few references were added for better clarity.
• Explanations have been added/modified according to the reviewers’ comments and/or the authors’ observations.
• All the changes applied have been highlighted in the revised manuscript.

Now we explain the changes in detail per item. The items are categorized per reviewer. Our response to each item starts with ‘R:’ and is shown in boldface.

We are truly grateful to have been given the chance to revise the paper according to the reviewers’ comments. In what follows, we have provided the reviewers’ comments and our responses regarding how we have addressed each of the comments. We have made our best efforts to address all comments. Thanks so much.

Reviewer #1:

Comments to the Author:

Q. The authors analyzes two features dimensionality approaches (Auto-Encoder and PCA) for supporting Intrusion Detection Systems based on machine learning techniques. The paper is well written and presented in general but there are some points to be clarified.

R: Thank you very much. We are glad and grateful that the reviewer found our paper satisfying.

Q. In particular, the related work section should include a table in which the analyzed approaches are compared underlying pros and cons

R: We thank the reviewer for this useful suggestion.
• One common dimensionality reduction approach is the Missing Value Ratio (MVR) approach. The MVR approach is efficient when the number of missing values is high. For the CICIDS2017 dataset, the number of missing values is near zero. Therefore, we excluded the Missing Value Ratio approach. Other common approaches include the Forward Feature Construction (FFC) and Backward Feature Elimination (BFE) approaches. Both FFC and BFE are prohibitively slow on high dimensional datasets, which is the case for CICIDS2017 (>2,500,500 instances). As a result, we did not discuss these approaches. The PCA technique, on the other hand, is relatively computationally cost efficient, can deal with large datasets, and is widely used as a linear dimensionality reduction approach. The auto-encoder dimensionality approach is an instance of deep learning, which is also suitable for large datasets with high dimensional features and complex data representations.

• Three additional references regarding dimensionality reduction techniques were added and cited in the revised manuscript:

• We also added Table 1 in the revised manuscript, which shows the properties of AE and PCA for dimensionality reduction along with an explanation:

<table>
<thead>
<tr>
<th>Technique</th>
<th>Computational Complexity</th>
<th>Memory Complexity</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>$O(P^3)$</td>
<td>$O(P^2)$</td>
<td>Can deal with large datasets; fast run time</td>
<td>Hard to model nonlinear structures</td>
</tr>
<tr>
<td>AE</td>
<td>$O(n^2)$</td>
<td>$O(w)$</td>
<td>Can model linear and nonlinear structures</td>
<td>Slow run time; prone to overfitting</td>
</tr>
</tbody>
</table>

Q. In the fourth section, the authors should provide more details about parameters setting (i.e. lambda ,rho etc.)

R: We thank the reviewer for the valuable comment.

In the computations, the weights are multiplied by $\lambda$ to prevent the weights from growing too large.

The sparsity parameters and penalty are designed to restrict the activation of the hidden units, which reduces the dependency between the features.

We added the Design Principle Table below in the revised manuscript for the parameters:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>0.0008</td>
<td>Weight Decay</td>
</tr>
<tr>
<td>$\rho$</td>
<td>6</td>
<td>Sparsity Penalty</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.05</td>
<td>Sparsity Parameter</td>
</tr>
</tbody>
</table>
Q. In the evaluation phase, the author’s should provide more details about the kind of features (from 81 to 59 as described at the row 322) involved in the reduction process.

R: We thank the reviewer for the time spent in reviewing our work.

Here, the AE reconstructed a new and reduced feature representation pattern that reflects the original data with minimum error. Unlike features selection techniques where the set of features made by feature selection is a subset of the original set of features that can be identified precisely, AE generated new features pattern with reduced dimensions. These explanations are added in the revised manuscript.

Q. I suggest also citing the following papers for underlying the relevance of the topic:


3) Machine learning based network intrusion detection. In Computational Intelligence and Applications (ICCIA), 2017 2nd IEEE International Conference on (pp. 79-83). IEEE.

R: We thank the reviewer for this useful suggestion. The suggested references were included and cited in the revised manuscript as reference numbers [1] and [3].

Q. Finally, a linguistic revision is necessary.

R: The manuscript has been checked by a Native English speaker as well as the academic resource center which is a professional English editing service. Minor corrections were made. Thanks.