Department of Computer Science
COSC 4P80 - Artificial Neural Networks

Examination of the Self Organizing Map

Prepared by: Kevin Olenic
Student #: 6814974
Instructor: Dave Bockus
Date: April 12, 2023

Abstract

A Self Organizing Map (SOM) is an artificial neural network that uses unsupervised learning to create a two-dimensional map capable of clustering data. Its purpose, conduct exploratory data analysis of high-dimensional datasets and create visualizations. This report will examine the network's analytical ability, using the engine data from the previous assignment while employing three different sizes of topologies. Thus helping determine the SOM's capabilities and effectiveness at creating clusters that can classify the data.

Contents

1 Introduction

In this section, the goal of this paper and relevant topics such as Self Organizing Maps and neighbourhood functions are explained to improve the reader's understanding of the concepts covered in this report.

1.1 Goal of this Report

The goal of this report: examine the Self Organizing Maps (SOM) artificial intelligence. This is accomplished by developing, training and testing the algorithm with information on electric motors. Using this data, we examine the SOMs ability to classify linearly separable data (i.e. data that falls into one of two categories) into clusters.

1.2 Self Organizing Map

The Self Organizing Map (SOM), introduced in 1980 by Teuvo Kohonen, is an artificial neural network using unsupervised learning to produce a two-dimensional, discretized representation of an input space from provided training data comprised of multiple feature vectors, known as a map, also referred to as a kohonen map.

These differ from other neural networks as they utilize competitive learning and neighbourhood functions, preserving the topology of the input space, unlike feed-forward neural networks, which use error-correction techniques such as back-propagation with gradient descent.

The SOM operates by having each data set in the topology compete for the right to represent an input. Thus training can be expedited, by assigning the weights of each data point, in the map the values of one of the input data values.

After the initialization, the algorithm passes each member of the training data through the SOM for the number of epochs defined one at a time by having a single input data selected randomly for training. Thus each member of the training data passes through the SOM once during an epoch. Upon selecting input for training, it is passed through the entire network to identify the node possessing the least distance between them (i.e. the node that best represents the input), the node that wins is known as the best matching unit (BMU) and is used in the next stage of training.

The next stage in training the network involves altering the BMUs' weights so they become a closer match to the data selected from the training set. This entails taking a portion of the difference from the two vectors corresponding weights and adding it to the BMUs' weight values, shifting them toward the weights of the inputs.

This change is not only applied to the BMU, as some nodes near the BMU have their weights awarded, shifting them towards the inputs values, but less than the BMUs' shift.

Allows the map topology to grow and shift, causing it to form different sizes to represent clusterings of unique data.

The change in weights is represented by the equation below, where $w_{i,(t+1)}$ is the value of the weight at the next stage, $w_{i,(t)}$ is the current weight value, α_t is the learning rate at time t, d(V,X,r) is the influence the winning node (V) has on the node (X), being adjusted in the neighbourhood and $x_i - w_{i,(t)}$ is the difference between the inputs weight and the nodes current weight:

$$w_{i,(t+1)} = w_{i,(t)} + \alpha_t * d(V, X, r) * (x_i - w_{i,(t)}) \tag{1}$$

Once the weights of the BMU and its neighbours are adjusted, the algorithm will adjust the learning rate and lattice (i.e. the size of a node's neighbourhood) for the subsequent epoch.

The change in these variables are outlined in the below equations, where α_t is the learning rate at time t, α_0 is the initial learning rate, t is the current epoch, T is the max number of epochs, σ_t is the size of the neighbourhood at time t and σ_0 is the initial size of the neighbourhood at the beginning of the algorithm:

$$\alpha_t = \alpha_0 * e^{-t/T} \tag{2}$$

$$\sigma_t = \sigma_0 * e^{-t/T} \tag{3}$$

An additional modification to the structure of the SOM to improve clustering is the use of wrapping, wherein the nodes at the edges of the map are connected to the nodes at the opposite edge, creating a unit distance of one, ensuring all nodes in the network possess a minimum of four neighbours with a unit distance of one in the map.

The below image depicts how wrapping would affect the neighbourhood of an edge node in the Kohonen map.

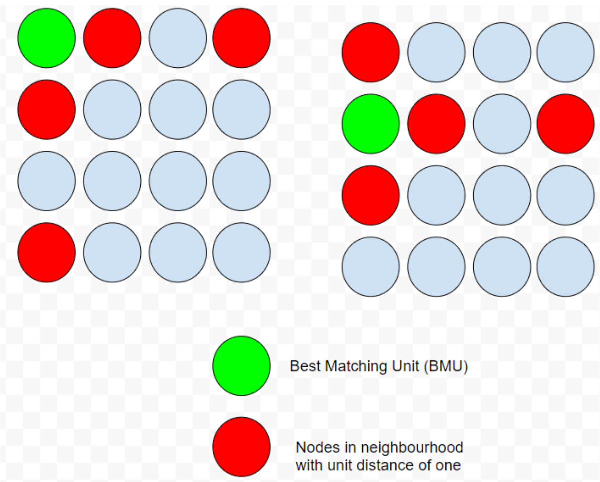

Figure 1.1: SOM Wrapping

1.3 Data

The data used in the examination of the SOM in this report will consists of the amount of current an electric motor draws from its power source. Under normal circumstances, the current drawn by the motor will have a feedback component which causes the motor to have a smooth acceleration when starting. In the case of a bad motor, this feedback will be dirty (noisy). Separating out clean and noisy acceleration data can be used to determine if the motor is good or bad. Good motors are classified with a value of one zero (0) while bad motors are classified with a one (1).

2 Experimental Research

This section of the report covers the experiments performed on the SOM, examining how the number of nodes in the topology affects the SOM's ability to classify. Networks of size twenty-five, thirty-six and forty-nine are used during testing, with the networks' clustering ability measured via a heat map. During testing, two datasets are used, with one set containing data with sixteen input vectors, and the other containing thirty-two input vectors, allowing inspection of the effect input containing additional information has on the SOM network. For this report, heat maps will measure the level of activation an input has using the Gaussian function to determine if it is within the curves activation range of a node based on its relation to its position in the clustering.

2.1 Topology: 25 Nodes

This section of the report reviews the SOMs ability to cluster data using a map containing twenty-five nodes, which results in a five by five Kohenen map.

4.01E-02	1.41E-06	5.18E-05	-6.52E-03	3.61E-03
6.02E-04	2.22E-04	4.12E-04	1.54E-06	6.99E-03
1.17E-09	5.26E-04	2.83E-03	3.17E-07	2.10E-07
1.44E-02	1.12E-02	-1.50E-03	-4.45E-02	-6.08E-06
6.17E-04	5.65E-02	-2.34E-02	-4.23E-02	-8.63E-02

(a) Data set containing 16 bins

0.23182	5.75E-02	-0.13559	0.046433	2.13E-02
3.93E-01	3.14E-01	9.17E-02	0.058514	5.00E-02
0.731454	0.440792	5.57E-01	2.31E-01	0.665734
0.550913	0.520813	9.89E-01	3.47E-01	5.05E-01
6.07E-01	6.77E-01	0.61591	0.132542	0.173997

(b) Data set containing 32 bins

Figure 2.1: Topology with 25 nodes

The above picture shows the heat maps for the twenty-five-node network using the data set containing the sixteen and thirty-two bins of engine data. Upon examination, one can see that the heat map for the data set containing sixteen bins shows that the network has very few nodes capable of classifying either the good or bad engine data, as most of the nodes in the map classify as neither. However, looking at the heat map that used the dataset containing thirty-two bins, the network has a greater spread of nodes associated with good and bad engines and a few that classify as neither. Revealing the amount of information used in the training of the network impacts the algorithm's ability to associate a node with a class of data.

2.2 Topology: 36 Nodes

The second experiment involves expanding the amount of nodes in the map to thirty-six nodes making a six by six Kohenen map to be used in the classification of the data.

-0.014025	1.67E-01	0.002201	9.19E-06	0.001086	1.08E-06
-0.001327	0.001782	2.86E-05	2.26E-04	2.85E-07	3.28E-07
2.02E-03	7.21E-02	3.60E-04	1.09E-04	5.46E-07	6.59E-05
-2.23E-04	-0.00122	-0.00635	1.96E-04	1.44E-04	0.025428
-1.06E-02	-1.08E-01	-2.06E-02	-5.91E-04	-5.59E-06	-9.36E-06
0.0052442	-0.11056	-0.03121	8.84E-04	1.31E-08	3.68E-09

(a) Data set containing 16 bins

-0.037794	0.071969	0.185755	0.081349	0.262261	0.447859
0.7551414	0.569213	0.572158	0.09486	0.191016	0.526694
0.5336274	0.656923	0.69665	0.375932	0.167104	0.204087
0.8450526	0.935744	0.902938	0.587674	0.623218	0.995869
-0.375007	-0.13783	0.192098	0.145095	0.229555	0.007316
-1.269728	-0.67488	-0.04717	0.001756	0.026922	-0.05697

(b) Data set containing 32 bins

Figure 2.2: Topology with 36 nodes

From the heat maps, we can see the effect the two data sets have on the expanded map, surmising that for the data containing sixteen bins, the expansion caused a small improvement in the network's ability to associate a node to a particular class of data, as there now exists a few extra nodes in the heat-map that associate to either good or bad engine data.

The same can be said for the data set containing thirty-two bins, as the heat map shows for this data set a similar effect, as the more data we provided, the closer a node is associated with that class of data, and with additional nodes, we see an increase in the number of nodes classified to either a good or bad engine category.

This improvement occurred, most likely, from the increase in nodes in the network, as this caused more inputs to be specifically tailored to a node in the network, creating a stronger activation value for that classification in the heat map.

2.3 Topology: 49 Nodes

For the third and final experiment of this report, an examination on the effects of an additional expansion on the nodes will be conducted, where the increased network will now contain forty-nine nodes, creating a seven by seven Kohenen map structure.

1.92E-06	0.015906	-0.00727	-1.54E-04	1.57E-08	7.34E-04	4.70E-02
1.37E-05	0.00146	0.007799	0.008162	1.13E-04	2.11E-06	4.77E-07
5.33E-04	8.30E-06	0.052384	0.013939	0.255028	3.30E-04	2.18E-09
6.12E-06	7.89E-08	9.73E-05	0.002058	0.001164	3.60E-02	6.43E-05
2.31E-06	1.68E-07	0.001884	4.78E-02	-0.00499	-1.30E-03	-3.14E-05
0.00186	-0.01825	8.70E-04	-5.52E-02	-0.04126	-3.19E-02	-5.55E-03
0.081782	0.04594	0.033945	4.82E-02	4.69E-02	-3.04E-03	-9.45E-03

(a) Data set containing 16 bins

0.357171	0.556436	0.216355	0.116582	0.449154	0.56318	0.636644
1.029132	0.744761	0.879232	0.251908	0.762545	0.905791	0.510178
0.841472	0.794682	0.940683	0.718615	0.211322	0.844955	0.547834
0.781307	0.569761	0.467487	1.003797	0.633929	-0.18464	0.402415
0.760582	0.880869	0.367174	0.286136	0.607182	0.340918	0.189421
-0.77655	-0.43874	0.219909	0.047226	0.203967	0.087678	-0.10056
-1.40748	-1.00797	-0.26896	0.132717	0.080282	0.038997	0.038898

(b) Data set containing 32 bins

Figure 2.3: Topology with 49 Nodes

The heat maps above show that adding the additional nodes to the network produced the same results as the previous experiment. For example, the heat map depicting the data containing sixteen bins of data has the same level of improvement as a few of the nodes in the network possess a better activation towards the good engines, with others having a better activation towards the bad. Similarly, with the data-set containing thirty-two bins, the increase in nodes caused even more nodes to be tailored to one of the input data, causing those nodes to have a better activation value for a specific class of engine data.

3 Conclusion

In conclusion, through the experiments performed in this report on the Self Organizing Map neural network, we have seen the capabilities of the map under different implementations. During this report, experimentation involved testing data with different sizes of input for use in training and using different map sizes as we used maps with twenty-five, thirty-six and forty-nine nodes.

Through experiments performed above on the SOM model, the following information on the SOM model was made clear. First, when training a SOM network, the more data used, the better its classification ability becomes, observable by comparing the heat maps in the above experiments. Heat maps using the data containing thirty-two bins produced heat maps containing a mostly even distribution of nodes capable of classifying data as either good, bad or neither, while maps representing the SOM using the input with fewer data had the majority of the nodes classified as neither good nor bad.

Secondly, the more nodes included in the Kohenen map work towards improving the classification ability of data since increasing the map size causes the individual nodes to become tailored to specific training data. Leading more towards memorization than generalization, which may impact its ability to classify data of the same type, but not included in the training data.

Based on the above results, the SOM is a valuable tool for classifying linearly separable data, capable of creating clusters with enough generalization to cluster data not included in the training set. However, a lack of training data or insufficient distinction between the classes will create difficulties for the SOM to generate maps with clusters capable of classifying the data.

Thus the SOM does possess a great deal of usefulness as a resource for classification problems, but also some significant limitations for certain types of data.

List of Figures