Neural Network Optimization With Salient Feature Reduction and Convolution

Thomas Martin

Research School of Computer Science, Australian National University
27 May 2018

Abstract. A set of sonar readings is used to train a neural network in an effort to automate mine detection. Two improvements to the model are applied. The first is a method of feature reduction that isolates and removes salient attributes. It adds an attribute of random noise to the data which is compared with the saliency of the other attributes to remove redundant input. The second method applies a convolutional neural network in an attempt to extract spatial features from the input data. The method of attribute reduction has the best improvement on performance, with a 3.7% increase in classification accuracy, compared to 0.8% for the convolution neural network (CNN). The poor result of the CNN is equated to a lack of spatial relation in the sonar data.

Keywords: Salience, Feature, Reduction, Convolutional, attribute, Sonar, Classification

1 Introduction

The modern abundance of collected data and vast improvements in processing power, continue to strengthen the performance of multilayer neural networks. The automation of complex classification tasks that previously required trained human operators is slowly becoming reality. One such example is the classification of sonar data intended to differentiate rocks and mines. The features in this dataset represent the energy within a particular frequency band of the sonar signal. It presents 60 different attributes for which it is hard to discriminate the effectiveness of a given feature for classification.

The existence of redundant features minimizes the effectiveness of a neural network and increases its computational cost [1]. In addition, the 'curse of dimensionality' suggests that large feature spaces require more training vectors to properly classify, which this dataset lacks [3]. A simple feed forward multilayer neural network is trained on the sonar data. A method of salient feature selection is applied to remove the input features that have minimal impact on successful classification. The accuracy of the new simplified network will be compared to to the initial network to draw a conclusion on the effectiveness of the feature reduction method.

A convolutional neural network is capable of yielding hierarchies of features for a given input [6]. They are typically applied to image data to detect similar features in different parts of an image. In contrast to the salient attribute reduction method, the use of a deep convolutional neural network (CNN) will instead attempt to extract relevant features from every input d attribute. Instead of discarding useless data, the CNN approach hopes to utilize local correlations to force the extraction of local features [10]. The accuracy of the old simplified network will be compared to the CNN to draw a conclusion on the effectiveness of the feature reduction method.

2 Method

The architecture of the constructed neural network is designed to replicate the neural network created by Gorman [2]. It is a feed forward network with three layers. The input layer has 60 neurons, one for each input signal. There are 24 units in the hidden layer which are activated with a sigmoid activation function. These feed into two final outputs that correspond with the detection of a rock or a mine. The model is trained using the stochastic gradient descent algorithm for back propagation. With a cross entropy loss function used to calculate the difference between the given output and the target output at each training phase. A batch size of three is used as it best mimics the results of Gorman.

Preprocessing of the data includes replacing the ‘R’ and ‘M’ values with 0 and 1 respectively and normalizing the data between 0 and 1. The sonar data contains aspect angles on the object spanning 180 degrees with an average of 5 entries for each aspect angle. An aspect angle independent model was chosen for classification, splitting the data into 13 random sets of testing data. The network is trained 13 times setting aside of the 13 sets each time for testing. The accuracy of the model is then taken as the mean of these accuracies. This technique provides a cross-validated accuracy for the classification model.

The feature reduction method applied to the network implements an attribute saliency metric suggested by Tarr [4]. The saliency metric is an effort to quantify the sensitivity of the networks output to it’s input.

$$\Delta_j = \sum_m \left(w_{jm}^1 \right)^2 \tag{1}$$

The saliency is simply the squared sum of the input weights w from each input j to each hidden nueron m. The value attempts to categorise the affect of each neuron to the trained network by quantifying the amount that each input neuron is able to affect the network. To determine the salient features in the data set, a feature of uniform noise on the interval (0,1), is added to the data set. If the saliency of the resulting noise is the same or greater than the saliency of a feature, then that feature is deemed insignificant and removed. The following algorithm is used for determining significant is features [5].

1. Introduce a noise feature to the original set of feature vectors.
2. Train the network.
3. Compute the saliency of all features.
4. Repeat steps 2 and 3 at least 30 times (with weights being randomly initialized and training and test sets being randomly selected at the beginning of each training cycle).
5. Assume the average saliency of noise is normally distributed and find the upper one-sided 99% confidence interval for the mean value of the saliency of noise using the equation below.

$$Upper\,99\,PercentConfidence = \bar{X} + z\frac{\sigma}{\sqrt{n}}$$

6. Choose only those features whose average saliency value falls outside this confidence interval.
7. Retrain the network with the salient features.

The architecture of the constructed neural network takes inspiration from Kalchbrenner [8] and Kim [9]. It is a Multilayer Convolutional Neural Network with two convolutional layers followed by two fully connected neural layers. The first convolutional considers the 60 input features and creates 60 feature maps with a kernel size of 2 across the input. These are pooled using max pooling which is employed as it preserves the order of the features and can also determine the number of times the feature is highly activated [8]. A second layer of convolution is then applied which adds a deeper layer of feature extraction to the network. Dropout is applied to the second convolutional layer to prevent co-adaption of hidden units during the feed forward back propagation [9]. Finally the two fully connected soft max layers process the feature data after max pooling and output a classification.

3 Results and Discussion

Table 1 demonstrates the testing accuracy of three neural networks with a 13 fold cross-validation. The mean accuracy of the network is 0.1 of a standard deviation from the results of Gorman [2]. As a result it can confidently be stated that the network has successfully replicated the results of Gorman. Figure 1 also demonstrates the spread of accuracies for different test sets. As the data is angle-independent, it is likely that the test sets with low accuracy are a result of the training data missing key instances that teach the correct classification.

Table 1**.** Shows the mean classification for each test set in the angle independent CNN for the sonar data.

	Prediction Accuracy	
Network Type	Mean	Std.
Three Layer Neural Network	83.9	3.07
Neural Network with Feature Reduction	87.6	2.92
Convolutional Neural Network	84.7	2.91

Table 2**.** Shows the mean saliency and standard deviation each of the 60 features for the sonar data, plus 61 the uniformly distributed noise attribute.

Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Δ	5.85144	0.96761	1.87221	4.50766	6.7277	5.60836	7.27879	7.23014	15.59881	6.75337	19.32619	22.16091	2.3481	2.51441	4.90675	8.94896	23.06995	12.17149	3.71208	10.30042
Std	0.96935	0.28403	0.49779	0.57717	1.59098	0.8311	1.18183	1.43028	1.62173	0.92984	1.96871	2.07026	0.81164	0.49867	0.95094	1.40467	3.00889	2.07446	0.91725	2.05112
Feature	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Δ	4.37536	7.87998	10.53546	8.07427	6.451	10.23794	10.52289	5.68623	2.77108	11.87598	24.63837	5.78908	6.76661	4.80574	2.45693	13.9025	17.72604	6.98092	7.16389	11.89789
Std	0.85339	1.43001	1.48469	1.35083	2.2022	1.17572	1.7291	0.81327	0.52017	1.34867	3.28337	1.01636	1.0748	0.95754	0.70559	1.51032	1.99604	2.17934	1.24992	1.95078
Feature	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Δ	14.04562	6.92716	6.57483	4.87274	10.42327	7.82209	2.90205	9.97273	10.56733	27.86912	5.73155	9.37154	14.31029	6.24536	4.14988	4.41523	3.99579	9.54905	4.11672	2.94668
Std	2.35824	1.07079	0.79709	0.89923	1.56771	1.21292	0.59379	1.13742	1.4929	3.08694	0.83392	1.49491	1.87415	1.38769	0.92526	0.8561	1.42363	0.97279	0.88277	0.69962
Feature	61 (Uniform Noise Feature)																			
Δ	6.46676																			
Std	1.50195																			

Figure 2 demonstrates the saliency and standard deviation for the 60 features in the sonar data. The 99% upper confidence interval was found to be 7.2 using equation 2. This interval indicates the attributes that have no more of an affect on the data than random uniformly distributed noise. The following features fall below that interval and are removed from the feature set: {1, 2, 3, 4, 5, 6, 10, 13, 14, 15, 19, 21, 25, 28, 29, 32, 33, 34, 35, 38, 39, 42, 43, 44, 47, 51, 54, 55, 56, 57, 59, 60}. 28 input neurons remain significant and are used to train a second neural network for which the results can be seen in figure 2.

The new classification accuracy for the network with feature reduction has a mean classification accuracy that is 3.7% better than the original network. In relation to the mean of the original network, equation 2 is applied to give a z value of 4. This correlates with over a 99.9% confidence that the two distributions are statistically different, assuming they are normally distributed. Therefore we can confidently say that the feature reduction method successfully increased the accuracy of the network. Additionally, the technique also reduced the overall complexity of the model as it has less inputs and hence the structure of the resulting network was much smaller.

It can be inferred from these results that Tarrs attribute saliency metric is effective for increasing model accuracy [4]. Sentiono used a Neural-Network feature reduction method to greatly reduce the average number of features to 3.87 [1]. They achieved a 93% accuracy with their network, however the feature reduction only created an increase of 1.5% on their accuracy. Comparatively, the salient features method seems more effective at producing a large change in the accuracy. However, Sentiono's method created a more robust network with a better accuracy. Whilst the method was effective, it would be preferable for the network to have it's feature space further reduced as the resulting network was still fairly complex.

Figure 1 shows the classification accuracy of the CNN applied to the data with a 13 fold cross validation. The new classification accuracy for the network with feature reduction has a mean classification accuracy that is 0.8% as in table 1. The CNN is 2.9% worse at correctly classifying. The dataset only had 208 unique instances, with a number of distinct sonar readings that were poorly represented. The salient reduction method simplified the network and allowed this small amount of data to have a greater significance. In contrast the complexity of the CNN was significantly greater than the salient network. The CNN had in the order of 10 times more connections and required 500 more epoch's of training to achieve a similar result. The complexity of the CNN was not sufficiently trained by such a small dataset. The increased complexity of the network should in theory allow for more patterns to be observed and as a result a better classification rate obtained. However, the size of the dataset made this extremely difficult to achieve.

The inability of the CNN to produce significantly improved results likely signifies that the data is not spatially correlated. The success of CNN for feature extraction is reliant on local spatial correlations existing within the data [10]. It was inferred that there would be a spatial pattern within the sonar data as the 60 input attributes were raw data from the sonar readings. Whilst patterns do exist within the data, if no local features are learned from the data the CNN model cannot improve over a basic neural network.

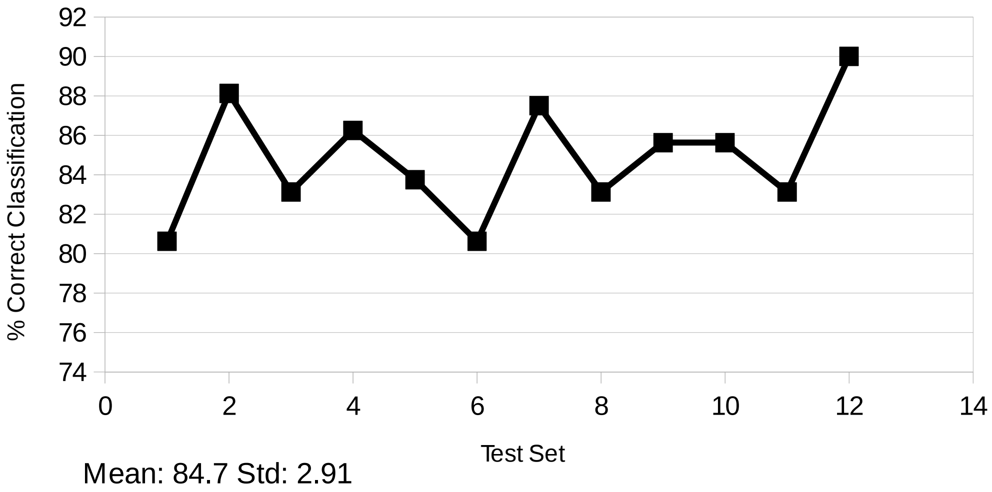

Fig 1. Shows the mean classification for each test set in the angle independent CNN for the sonar data

4 Conclusion

A neural network was constructed that successfully classified given sonar data to the same standard as Gorman [2]. A method of feature reduction was applied to the network to determine statistically insignificant input data. Over half of the input was removed and the resulting network received a 3.7% increase in classification accuracy, whilst reducing the complexity of the neural network. The increase in accuracy on this data set was found to be better than a similar method applied by Setiono on the same data set [1]. However, the salient reduction method left 7 times more features than Setiono's method. The CNN was unable to provide a significant improvement in the with only a 0.8% increase in classification. The low increase in accuracy is attributed to a small dataset and a lack of spatially correlated features in the dataset. The reduction of attributes has been successful in this study and future work should expand on the idea of attribute reduction. In particular, a method of minimizing the size of feature space, whilst still keeping the same or improving the accuracy of the network.

References

1. Rudy Setiono and Huan Liu. Neural-Network Feature Selector. Department of Information Systems and Computer Science National University of Singapore
2. Gorman, R. P., and Sejnowski, T. J. (1988). "Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets" in Neural Networks, Vol. 1, pp. 75-89.
3. P.A. Devijver, J. Kittler (2nd ed.), Pattern Recognition: A Statistical Approach, Prentice-Hall, Englewood Cliffs, NJ (1982)
4. G. Tarr Multi-layered feedforward neural networks for image segmentation (2nd ed.), Ph. D. dissertation, School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB OH (1991)
5. L.M. Belue, K.W. Bauer Determining input features for multilayer perceptrons Neurocomputing, 7 (2) (1995), pp. 111-121
6. J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR, 2015.
7. T. Martin, Neural Network Optimization With Salient Feature Reduction, Research School of Computer Science, Australian National University, 27 May 2018
8. Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. ACL, 2014.
9. Kim, Yoon. 2014. Convolutional neural net-works for sentence classification. arXiv preprint arXiv:1408.5882
10. Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time-series. The Handbook of Brain Theory and Neural Networks, 1995. 3361