Classification of manipulated and unmanipulated images using an expanded human eye gazing data with a recurrent neural network, using pruning techniques

Angus Wickham
The Australian National University, Research School of Computer Science
Acton, ACT, Australia
u6375392@anu.edu.au

Abstract To investigate the extent at which using a recurrent neural network (RNN) with an expanded eye gazing dataset [3] can outperform model demonstrated in the last paper [4], which was a shallow neural network on a much smaller dataset. An RNN will be used to classify whether or not 4 different images have been manipulated or not using the complete eye gazing data of 80 participants. The results conclude that an RNN with an expanded dataset performs worse than that of the shallow neural network, and some minor gains in accuracy are achieved when pruning by distinctiveness [2] is applied.

Introduction

With the increasing use of photo editing software such as photoshop, edited and manipulated images are becoming more and more prolific and harder to classify as fake. With the increased use of social media in the last decade, these manipulated hoots are now also finding channels to flow through and populate where people may interact with them. Therefore, it is ever more pressing to try and identify these manipulated images.

Some research has been conducted on how well humans are at classifying those images [3], showing that humans have poor performance are accurately identifying whether or not a photo is edited. It also presented that humans have even more trouble in trying to locate what part of the photo is manipulated [3]. This paper found that there was some association with that fact that higher fixation times on certain image were associated with greater accuracy on correctly classifying that image. A preceding paper has demonstrated that a simple feedforward network can outperform that of a human, achieving an accuracy of 60% without the use of pruning. With the use of pruning this same model can increase its performance further and reach an accuracy of 65% [4].

This paper proceeds to take it one step further by using an expanded dataset, containing of 30,000 data points. A recurrent neural network will be used on this data and the results will be compared to that of shallow feedforward network mentioned above. Pruning by distinctiveness will be applied in a similar manner as above and the results will be examined to investigate whether or not pruning will increase the model's performance.

Method

Devising the Classification problem

The expanded dataset contains of 30,000 data entries, each representing a participant's single fixation at a particular image. Each data point contains time series data, stating when every single fixation started and ended. Because of the characteristics of the data, the dataset can be thought of as hundreds of sequences, with each sequence comprising of individual gains, and each sequence representing the entire time some participant viewed some image. Therefore, a RNN is suited for this type of classification where the inputs are not of fixed size and are sequential.

The dataset comprises of 9 columns of data, and as mentioned above the aim of the paper is investigate the extent to how humans are able to perceive manipulated images. Unlike the previous dataset used in [4], there is no column for whether or not the image viewed at each entry is manipulated or not and therefore must be added to the database. This was achieved by using record linkage between the two databases, based on the participant and image ID's. Using this method, it was possible to map to each and every data entry in the new dataset, whether or not that particular fixation was directed at a manipulated image.

Therefore, the resulting classification problem consists of using input data in the form of sequences fed through an RNN to solve a binary classification

Data Preprocessing

An extensive amount of preprocessing was conducted on the data to ensure it was in the proper form to be fed through the RNN.

Firstly, columns were discarded that weren't useful for the classification problem. In this instance this only meant discarding one column: 'Fixations_ID' which was purely an ID for each and every fixation in the dataset and therefore gives nothing of value towards the RNN. The remaining columns were all normalised between 0 and 1 as the majority of columns were all in a different scale, which would hurt the performance of the RNN.

Secondly, the data must be transformed to ensure the proper use of sequences for the RNN. This was done by sorting the database by the columns 'Start Time and 'Stop Time', to ensure all sequences would be fed into the model in the correct order. The dataset was then grouped by the columns 'Particpant_ID' and 'Image_ID' to create the sequences themselves.

The data was then split into training and testing sets, representing a 80:20 split, respectively.

in models that are underfitted and perform poorly in classification.

Neural Network Model

The model used in this paper consists of a single layer recurrent neural network.

It uses Backpropagation and is trained over 50 epochs. The neural network has six inputs, as described above and will be using a binary output which is whether or not the image is manipulated. The activation function being used on the recurrent layer is the tanh function, while the final activation function being used is the sigmoid function since this is a binary classification problem. Fifty neurons are used in the fully connected layer. A large number of neurons are used in the hidden layer to ensure the use of pruning by distinctiveness can be properly applied to this scenario.

Pruning by Distinctiveness

The extra technique that is being applied upon this data set is the pruning by distinctiveness technique. Distinctiveness is defined as the similarity between two neurons weight vectors. This is determined by finding the angle of similarity between every single pair of neurons in a single hidden layer. If the angle is less than 15° or greater 165, then the pair of neurons are deemed too similar or too complimentary and therefore one must be pruned. These two thresholds were chosen from the paper when the technique ins used and am therefore using them for consistency and to easily conclude the benefit or lack of benefit that this technique may provide. This pruning technique is also applied repeatedly throughout the training process. The amount of pruning was varied and record to try and illustrate the effects of increasing pruning on a network.

Measures to evaluate performance

To evaluate and compare how the neural network initially compares against the human baseline as well as how. A pruned network fares compared to an unpruned network; the accuracy of each model is used to evaluate their performance. The aim of this paper is to try and evaluate whether a neural network using the data of a human eye gazing, is more accurate at deciding if an image is manipulated or not compared to a human. Therefore, the accuracy measure s extremely important as it is evaluating which model can correctly classify manipulated images and is extremely useful for comparisons.

Results

Results were firstly done to compare the human baseline present in the paper to the implemented neural network without pruning. The results are in the able as follows:

Table 1. Neural network (without pruning) against the human baseline.

Technique	Accuracy (%)
Feedforward neural network	60.00
RNN	60.53

Secondly, neural networks with different stages of pruning were implemented and tested. Networks were pruned, once, twice, thrice, and four times during the training period and were then evaluated against each other. Those results are as follows

Table 2. Neural networks, with and without pruning, with different stages of pruning.

Technique	No. of pruning stages	Accuracy (%)
RNN (w/o pruning)	N/A	60.53
RNN (w/ pruning)	1	62.01
RNN (w/ pruning)	2	62.03
RNN (w/ pruning)	3	62.03
RNN (w/ pruning)	4	62.03

The threshold for angles when pruning by distinctiveness were then restricted further, becoming 35 and 150. The results are as follows:

Table 3. Neural networks, with and without pruning, with different stages of pruning, and the use of more restrictive angle thresholds.

Technique	No. of pruning stages	Accuracy (%)
RNN (w/o pruning)	N/A	60.53
RNN (w/ pruning)	1	58.43
RNN (w/ pruning	2	58.43
RNN (w/ pruning)	3	58.43
RNN (w/ pruning)	4	58.43

Discussion

By using the eye gazing data of participants, the RNN has marginally outperformed that of the feedforward neural network on unseen images. This may be the result of

the increased amount of data accessible to the RNN compared to that of the neural network, however, what is surprising is the marginal increase in performance by the RNN. With the plethora of data available to the model is unexpected that the increase in performance compared to the neural network is only 0.53%, while the data increase by a magnitude of 100. However, it must be noted that the data used by the RNN is vastly different in format compared to that used by the neural network and this may suggest that using sequences in this problem domain are only slightly more effective than that of using smaller datasets with simpler neural networks.

Once pruning is added initially, surprisingly very little changes. At all the different stages of pruning, there was little effect to the accuracy of the model where it increased to 62.03% and then plateaued at that level of accuracy. This seems to suggest that recurrences of pruning have little effect, insinuating that once some neurons are pruned, they tend to stay distinctive and not become too similar or too complimentary. These results also suggest that the thresholds for distinctiveness may need to be adjusted to determine if they are restrictive enough. In Table 3, the thresholds were increased to be more restrictive and it seems more pruning occurred. This led to some slight decreases in performance, decreasing the model's accuracy to 58.43%. There were also no changes when additional stages of pruning were created, suggesting again that neurons, once pruned tend to stay within the distinctive thresholds. It also suggests that by increasing the thresholds for neurons not being distinctive meant that neurons that were actually contributing to the performance of the model were being pruned, leading to a deterioration in accuracy as seen in Table 3.

Compared to other work done in his domain with thaw same data given from [2], other paper has produced more impressive results. [1] used a shallow feedforward network. They also used the same pruning technique to evaluate it performance. They were able to produce results of an accuracy of 67% which is quite higher than the results found in this paper. This is most likely down to the way the pruning method is used in comparisons to how it is used here. They pruned over much smaller hidden layers, which may have allowed greater performance due to not overfitting, which may be rife in the model outlined in this paper.

Conclusion

As a result, it appears that an RNN with an expanded dataset can achieve a greater accuracy than a neural network with a smaller dataset. What is surprising, however, is how marginal the improvement is. This may come down to imperfections of the model of the ineffectiveness of the expanded dataset in this problem domain. The model in this paper is definitely not perfect and therefore may suggest that the problem resides in the model where it may be overfitted and not learning when being trained. This, however, does not discredit the thought that the expanded dataset may also be the problem and is causing the RNN to have degraded performance.

Pruning by distinctiveness also proves to have little effect on the RNN, and even degrades performance when the threshold for what isn't distinctive is increased. This suggests that network is largely comprised of distinct neurons and when a neuron is pruned it remains distinctive afterwards. As a result, pruning by distinctiveness merits little gain and seems to be inefficient for this problem domain.

Future work

The ability to correctly predict whether an image is manipulated or not has been shown possible in this paper and that the use of deep learning techniques can further the performance of correct classification. More work needs to be done, however, delving into why this RNN model has only given marginal increases in performance when the amount of data it has access is substantially larger than previous models. Future would could be conducted on testing other RNN's to see if they can be further optimized for this problem domain. Other deep learning techniques that can work on sequences, such as convolutional neural networks and long short term memory networks can yield betters results.

These future techniques may also be tested with other pruning techniques that might give more performance benefit than pruning by distinctiveness that was used in this paper.

As demonstrated, if the data of one's eye gazing can correctly detect if an image is manipulated if not, then it has huge implications for future classification of fake documents, forged photos and the like, resulting in huge implication for how future designers will design human computer interaction.

References

1. Tan, Z. and Plested, J., 2019. Classification of Humans' Perception of Manipulated and Unmanipulated Digital Images Using Feedforward Neural Network with Network Reduction Technique.
2. Gedeon, T. D. (1995, November). Indicators of hidden neuron functionality: the weight matrix versus neuron behavior. In *Artificial Neural Networks and Expert Systems, 1995. Proceedings, Second New Zealand International Two-Stream Conference on* (pp. 26-29). IEEE.
3. Caldwell, S., Gedeon, T., Jones, R. and Copeland, L., 2015, July. Imperfect understandings: a grounded theory and eye gaze investigation of human perceptions of manipulated and unmanipulated digital images. In *Proceedings of the World Congress on Electrical Engineering and Computer Systems and Science* (Vol. 308).
4. Wickham, A, 2020 Classification of manipulated and unmanipulated images using human eye gazing data with a feedforward network, using pruning techniques.