The Importance of Hyperparameter Optimisation
for Facial Recognition Applications

Hannah M. Claus

University of Bedfordshire, School of Computer Science and Technology
University Square,
Luton, LU1 3JU
United Kingdom
hannahmclaus@gmail.com

Abstract

This paper explores the importance of using optimisation techniques when tuning a machine learning model. The hyperparameters that need to be determined for the Artificial Neural Network (ANN) to work most efficiently are supposed to find a value that achieves the highest recognition accuracy in a face recognition application. First, the model was trained with manual optimisation of the parameters. The highest recognition accuracy that could be achieved was 96.6% with a specific set of parameters used in the ANN. However, the error rate was at 30%, which was not optimal. After utilising Grid Search as the first automated tuning method for hyperparameters, the recognition accuracy rose to 96.9% and the error rate could be minimised to be less than 1%. Applying Random Search, a recognition accuracy of 98.1% could be achieved with the same error rate. Adding further optimisation to the results from Random Search resulted in receiving an accuracy of 98.2%. Hence, the accuracy of the facial recognition application could be increased by 1.6% by applying automated optimisation methods. Furthermore, this paper will also deal with common issues in face recognition and focus on potential solutions.

Introduction

Recognising human faces appears to be a standard skill for human beings. However, an algorithm requires a considerable amount of time and resources spent on training and testing. After recognising a human face, it can be linked to a unique identity. Creating such a system that combines biometric authentication and facial recognition can help in preventing criminal attacks or identifying criminals. However, it is important to create a safe application that has the best recognition accuracy and the lowest error ratio possible. For this reason, the ANN needs to be implemented in a way that guarantees the best outcomes. This can be achieved by finding the best values for the hyperparameters that control the training process of the ANN. However, there are too many possibilities and combinations of values that could be used and only the best ones are needed. In order to find an optimal set of parameters, the model can be tuned supported by an optimisation technique like Grid Search or Random Search. These optimisation procedures are of great impor-

tance because they can improve the ANN's recognition accuracy just by the slightest change of values and therefore, this can influence whether a criminal can be found or not. Thus, effective optimisation of the hyperparameters and using a diverse set of training input is necessary in order to safely implement a facial recognition application. However, facial recognition has certain issues that will be discussed further.

Related Work

There is bias in the data, there is bias in society, hence, the algorithm can only mirror the data that it receives (Buolamwini et al. 2018). A recent paper (Bergstra and Bengio 2012) demonstrated the differences between Grid Search and Random Search. There it is stated that Random Search explores the space of hyperparameters more widely because it chooses the values of the parameters randomly. In this case, a significant difference can be found between both search algorithms. Grid Search might promise to find the optimal set eventually, but only in the defined grid and limited through time. Random Search on the other hand does not promise to find the one perfect set of parameters but a wide range of different options that are chosen randomly. Thus, a broader spectrum of sets can be tested and an approximation can be made.

Methodology

The testing and training datasets for the ANN included 1500 images of 30 different people in total, 50 images each in different lighting conditions (Georghiades, Belhumeur, and Kriegman 2001). The ANN used a multi-layer perceptron classifier which was supported by principal component analysis to reduce the multidimensionality but retain the information transferred (Karamizadeh et al. 2013).

The experiments are divided into four categories: 1. Manual Optimisation, 2. Grid Search, 3. Random Search, 4. Final Approach. The experiments will deepen their complexity with each approach, starting with the easiest one.

The Optimisation Process

Manual Optimisation After the first iterations, a pattern could be detected, which indicates that the numbers of principle components (PCs) between 200 and 300 show the best

performances. Hence, the ANN can have an average value of 250 PCs. The best performance was achieved by using 250 PCs and 5000 hidden neurons (HNs), with a recognition accuracy of 96.6% and an error of 30%.

Grid Search For Grid Search, a specific grid space had to be defined. It was recommended to use a small number of dimensions, to prevent the grid from becoming too complex. It tries all possible combinations and evaluates the results using cross validation. Hence, the grid contained ranges for the following parameters: PCs, HNs, activation function, batch size, and weight optimisation (solver). The best recognition accuracy achieved 96.9% with an error of less than 1%.

Random Search Random Search chooses its values randomly and can therefore explore a wider range of possible combinations with a lower chance of using the same values twice. Using 250 PCs to simplify the search, three new parameters were added to the existing ranges: momentum, learning rate (type), initial learning rate. The results show that Random Search was able to find specific values for the parameters that improved the accuracies by 1-2% compared to the previous experiments. The best result was at 98.1% with, again, less than 1% error.

Further Optimisation By choosing the two best sets of parameters from Random Search that share the most values, new experiments were conducted manually. All accuracies that were achieved during this optimisation had an average recognition accuracy between 98.1% and 98.2% and an error of less than 1%.

Results

Optimisation Method	Accuracy	Error
Manual Optimisation	96.6%	30.0%
Grid Search	96.9%	<1.0%
Random Search	98.1%	<1.0%
Further Optimisation	98.2%	<1.0%

Table 1: The results of all implemented methods

All methods that were used provided resulting sets of parameters that achieved at least 96% accuracy, with the best one achieving 98.2% as can be seen in table 1. Important to mention is the fact that the aspect of time limited the search for the optimal set of parameters for this face recognition task.

Discussion

Analysing the search for an optimal set of parameters showed that it matters which optimisation method is used. A lot of time ran into the capture of the outcomes of each iteration in the optimisation process. In total, 27 full hours were spent on manual optimisation, 30 hours on Grid Search, and 31 hours on Random Search. Provided that more time is available to conduct deeper experiments, an even better recognition accuracy could be achieved.

The results show the range of values for the hyperparameters that could be identified through these experiments. Grid Search might be able to provide the best values for the parameters eventually, but the cost is too high to still be called an efficient optimisation. However, Random Search does not guarantee to find the one best set of values, but there is a wider range of better results in less iterations. And combining this with manually tuning the already optimised search, a good range can be defined as shown.

Common Issues in Face Recognition Recent studies have found that algorithms used for machine learning show biases towards classes like race and gender (Buolamwini et al. 2018). A biased dataset can have severe impacts and create an even stronger imbalance within society because the ANN might discriminate in favour or against certain groups, which is already happening in the present (i.e. against women, people of colour, or specific ethnic groups). After looking closely at the provided dataset (Georghiades, Belhumeur, and Kriegman 2001), it proved the missing diversity of people. The dataset has a total of 30 different people, of which only approximately 30% represent women and people of colour. Since 1997, when the dataset was created, the demographics of society have changed, therefore, the dataset needs to be changed, as well. Although a perfect training dataset does not exist, there are ways to achieve better performances by using different databases, for instance the "Gender Shades" database (Buolamwini et al. 2018).

Conclusion

Several experiments were conducted to display a variety of solutions. The risk of overfitting the model could be ignored because of the convenient script that stopped the training process before it could fit to the peculiarities of the images. By increasing the complexity of the ANN, hence, by increasing the number of hidden neurons, the testing accuracies were improved. Two different methods of tuning the values of the hyperparameters were experimented with. By manually optimising the values from Random Search, an accuracy of 98.2% could be achieved. In other words, the combination of both automated optimisation followed by manual optimisation for an approximation of a better accuracy turned out to be the best solution.

References

Bergstra, J.; and Bengio, Y. 2012. Random search for hyperparameter optimization. *Journal of Machine Learning Research*, 13: 281–305.

Buolamwini, J.; Gebru, T.; Friedler, S.; and Wilson, C. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81: 1–15.

Georghiades, A.; Belhumeur, P.; and Kriegman, D. 2001. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6): 643–660.

Karamizadeh, S.; Abdullah, S.; Manaf, A.; Zamani, M.; and Hooman, A. 2013. An Overview of Principal Component Analysis. *Journal of Signal and Information Processing*, 4: 173–175.