Detecting Deception using a Cascade Neural Network Architecture and Evolutionary Algorithm for Feature Selection

Kelly Hedemann

Research School of Computer Science, Australian National University, Canberra Australia
u5202179@anu.edu.au

Abstract. In an increasingly online world, it is essential that we can trust the information being presented through televised media sources. However, research shows that humans have difficulty consciously detecting dishonestly. This paper explores the types of neural network architectures which are best suited to analyzing physiological signals for detecting deception. The results show success with both accepted back propagation networks and cascading networks built one neuron at a time. The models are then further optimized by applying a genetic algorithm for feature selection, which result in improvements in accuracy of at least 9 percentage points across all tested models. The cascading neural network trained with cross entropy and feature selection produced the best results with an accuracy of 73% on the validation data, using only 59 of the 119 features available for training.

Keywords: Cascade Correlation, Neural Networks, Subjective Belief, Feature Selection, Genetic Algorithms

1 Introduction

Communication is a key facet of every aspect of human life. As a result of widespread isolation and social distancing measures introduced during the Coronavirus pandemic, people are relying increasingly on receiving this communication online or remotely. In receiving this communication, we need to be able to assess whether we can trust the information being conveyed. This has been particularly important regarding scientific information during a time of rapidly evolving research into the use and safety of Coronavirus vaccines. For example, Palomo (2021) argues that increasing disinformation in the media increased the number of deaths during the beginning of the Coronavirus pandemic.

Research to date has shown that people may be better at detecting dishonesty subconsciously (Van't Veer, 2016). The idea being that people can subconsciously be aware of subtle cues of dishonesty but may not fully be able to consciously form the opinion that someone is being deceptive. Zhu et al (2018) took this concept and investigated the merit of using physiological signals as an unconscious veracity detector. Their research investigated whether viewers could detect the 'doubt effect' – where a presenter's subjective belief in some information has been manipulated.

To do this, the researchers set up an experiment where a presenter is manipulated to doubt the veracity of the information they are about to present. In this case it was a piece of scientific evidence. Right before they presented the information, they were told it was 'a bit bogus' but asked them to present it anyway. This was the manipulated subjective belief condition.

The researchers then recorded physiological data for participants viewing the videos to see if they could unconsciously pick up on the deception. The original paper focused on pupillary responses – the size of the pupil as the participants viewed videos both where the presenter's belief had been manipulated and where the presenter had no reason to doubt the material. A neural network was then used to try and predict whether a presenter believed the information they were delivering based on the physiological signals of viewers.

Zhu et al (2018) concluded by saying that human data is complex. This paper expands upon that work done, using an updated and newly collected data source. The paper will experiment with applying different deep learning architectures to the data and the use of genetic algorithms for feature selection. This includes standard feed forward neural networks trained using back propagation as well as the cascade correlation algorithm proposed by Falhman and Lebiere in 1991. If human data is complex than potentially a network topology that can make its only determinations about depth and the inclusion of feature selection to focus on the most influential fields will be advantageous in this topic of research.

2 Method

2.1 Neural Network Architecture

The paper experiments with three network architectures and examines feature selection using genetic algorithms. The first model is a baseline neural network using back propagation. The final two models experiment with cascade network topologies.

Model 1: Baseline one-layer neural network – 20 hidden neurons
Model 2: Cascade Correlation (CasCor) network where the hidden unit was trained using covariance
Model 3: Cascade network where the hidden unit was trained using cross entropy

All neural networks had an output layer with two output neurons and were trained with the Adam optimizer using backpropagation and the Cross Entropy loss function. A learning rate of 0.01 was applied to the models and 500 epochs completed. Each model was trained using stratified k-fold cross validation with 4 folds to ensure stability of results.

2.2 Cascading Architecture

The Cascade Correlation network architecture (CasCor) was proposed by Fahlman and Lebiere (1990) to overcome limitations they saw with the standard backprop algorithm. The CasCor algorithm begins with a simple network of input and output units only and then creates a cascading network by adding hidden units one at a time, each as a separate layer (Dandurand et al, 2007). Each hidden unit is trained based on maximizing the covariance with the residual error of the network to create the most effective feature to reduce the overall error and then added to the network.

CasCor is designed to overcome the moving target problem which is caused by the ability of all hidden unit weights to be updated repeatedly during training. As networks become deeper, neurons in the inner layers are impacted by different weights in the network shifting different ways and continuously changing their purpose. CasCor proposes freezing the hidden neurons once they are added to the network to overcome this issue. Figure 1 shows the architecture of a CasCor network following the addition of two hidden units (Fahlman & Lebiere, 1990).

Model 2 in this paper experiments with the add-one element of the cascade algorithm but maintains the more standard back propagation instead of implementing the quick prop method proposed by Fahlman (1988). Hidden units are added one at a time and trained by maximising the covariance with the residuals from the previous iteration of the overall network. The hidden neuron weights are then frozen as per the original CasCor. The broader network is then trained using back propagation and a cross entropy loss function.

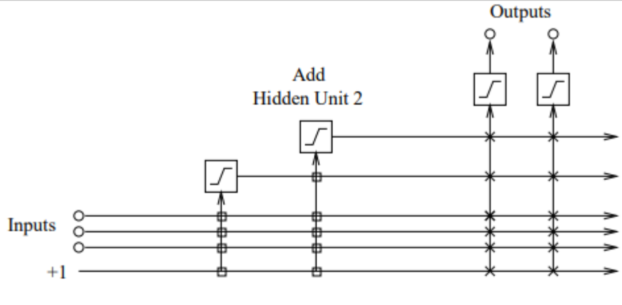


Figure 1: The Cascade architecture, initial state and after adding two hidden units. The vertical lines sum all incoming activation. Boxed connections are frozen, X connections are trained repeatedly.

Model 3 further experiments with the cascade methodology by focusing on the cascading ‘add one’ approach of CasCor without using the covariance method. In this final model a hidden neuron is added and then trained using cross entropy loss, following that initial training of the neuron its weights are frozen.

2.2 Genetic Algorithm for Feature Selection

Feature Selection can be beneficial where there is high dimensional data consisting of irrelevant or highly correlated features (Nasreen, 2014). The aim is to select a small subset of relevant features to improve the accuracy and computation time of machine learning models by reducing the resources required to search for patterns in what may be unnecessarily high dimensional space. This is relevant to the analysis of physiological signal data which collects many thousands of observations per experiment and participant. For a researcher it is challenging to know which measurements or which statistical representation of measurements will be useful in classifying the outcome.

A **Genetic Algorithm** (GA) is used in this paper to experiment with feature selection. The idea of the GA for feature selection is to initially select a random population of solutions which each contain a subset of features to use in training the relevant model. Each of these subsets is referred to as an individual or chromosome. The individuals are then assessed based on their ‘fitness’ in predicting the solution and the best of the individuals are selected to move on to the next generation (García-Dominguez et al, 2020). For each new generation selection, mutation and crossover are applied to the individuals to create new subsets of features to test, each time improving the overall performance of the fitness function. These techniques are designed to reflect the process of natural selection where the fittest individuals are selected for reproduction for the next generation and are then genetically altered to introduce diversity and allow continued improvement in reproduction (Mallawaarachchi, 2017)

The fitness function evaluated to find the most suitable set of features was the accuracy of the model on the test set. The models were then assessed by evaluating the accuracy on an unseen validation set, which was not used in training the model, nor in the selection of features. All results reported in this paper are based on accuracy on the validation set and are compared to baseline models which include all 119 available features.

The GA was implemented using the DEAP library (Distributed Evolutionary Algorithms in python) and was based on adapted code from Colianni (2017) which is available on github. Crossover was set to one point crossover between individuals with a mutation probability of 0.05. Selection was determined by a tournament of three and the GA was evaluated using a population (solution space) of 25 individuals over ten generations which was a fixed stopping point for the algorithm.

3 Results and Discussion

3.1 Data Inspection

The dataset contains 368 observations, relating to eight individuals who watched eight videos. Each row represents the set of 119 features extracted from participants' physiological signals when they watched one specific video. Four types of physiological data were collected:

- 34 features capturing the participants Blood Volume Pulse (BVP)
- 23 features capturing the participants Galvanic Skin Response (GSR)
- 23 features capturing the participants Skin Temperature (ST)
- 39 features capturing the participants Pupillary Dilation (PD)

All features were numeric with a mix of integer and continuous variables across all four physiological categories. There was no further information provided from the experimenters regarding the meaning or collection of features, thus limited transformations were applied to the data prior to modelling.

Zhu et al (2018) point out that different individuals will have naturally varying pupil sizes, thus the fields need to be normalised across all videos viewed by the individual to capture the variation for the individual, not natural variation in individual pupil size. It is reasonable to extend this to all the supplied physiological features, thus normalisation has been applied to each of the physiological features.

The target is a binary field capturing whether the presenter in the video doubted the content they were presenting. 1 represented no manipulation and a 0 represented manipulation of the presenter’s belief when presenting the information. Zhu et al (2018) implemented a leave-one-participant-out method to split the training and test set. However, during this experiment, while all participants watched 16 videos, they did not necessarily watch a balanced number of belief manipulated videos. For example, participants 7, 11 and 23 did not watch any videos where the presents belief had been manipulated, while participants 14 and 21 only watched videos where the presenters’ beliefs had been manipulated. Considering this it was deemed more appropriate to apply stratified k-fold cross validation to split the data into train, test and validation. Using stratified k-fold ensured there was a balanced proportion of the target class in the validation data.

3.2 Findings and Discussion

Not all models with all features performed better than chance – the CasCor model achieved accuracy of less than 50% indicating it was not a useful model. The baseline NN model and the Cascade models (model 1 and 3) performed better with an accuracy of 55.41% and 54.05% respectively on the validation set. Table 1 shows the results of each model with and without feature selection implemented.

Table 1: Accuracy on validation set by model architecture, with and without feature selection

Model Architecture	Accuracy using all features (baseline)	Accuracy with feature selection	Number of features selected	Percentage point improvement with feature selection
Model 1: Baseline NN	55.41%	64.86%	53	9.45
Model 2: CasCor	47.40%	56.76%	59	9.36
Model 3: Cascade	54.05%	73.00%	59	18.94

Feature selection was successful across all model types with at least a nine-percentage point increase in the accuracy achieved by each of the models. The Cascade model which implemented the frozen add-one hidden neuron approach but trained the new neuron using back propagation had the most significant improvement in performance with the additional inclusion of feature selection. This model achieved an almost 19 percentage point increase in accuracy on the validation set by only including 59 of the 119 features (~50%).

With a 73% accuracy rate the Cascade model was the best performer for this experiment. This suggests there is value in implementing the cascade architecture but that it may be dependent upon careful feature selection. However, it should be noted that this result was only based on training the relevant models on one split of the validation data due to resource constraints with the time taken to apply the GA on a basic CPU. There is some risk that result was due to an advantageous cut of the test and validation data, this is potentially supported by Figure 2 which shows the performance at 73% as a somewhat outlier result. In comparison, Figure 3 shows the GA test and validation accuracy results for the Baseline NN (Model 1), presenting a more consistent result across test sets. Further experimentation is required to confirm the value in the Cascade model.

Figure 2: Validation and Test set accuracy – Model 3

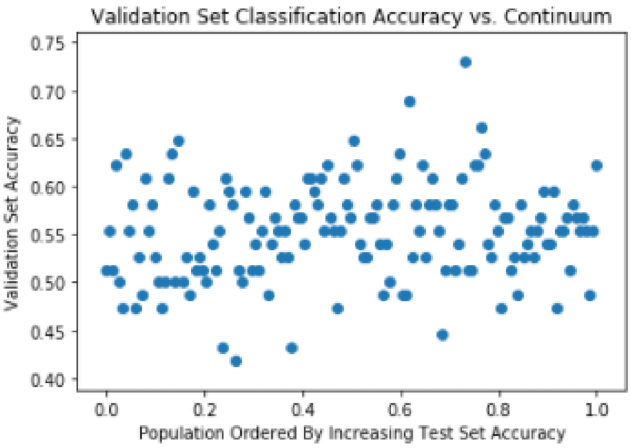
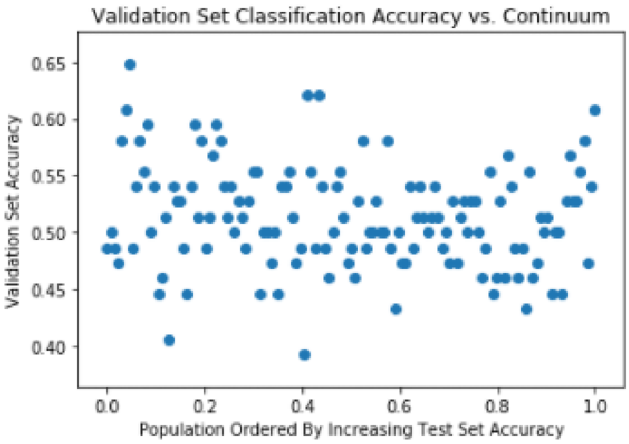
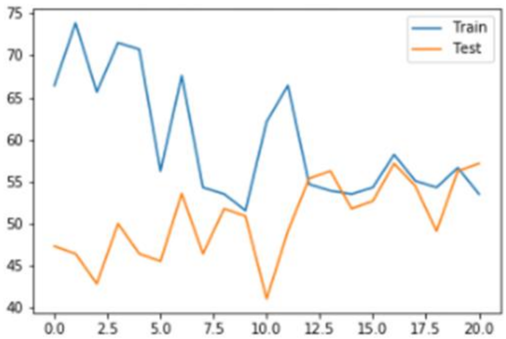


Figure 3: Validation and Test set accuracy – Model 1



The CasCor model was by far the most problematic, as the training accuracy and loss swung up and down as hidden neurons were added. The advantage of this architecture is that the model determines its own topology, adding hidden neurons until the network achieves a minimum loss threshold. This was not achieved in this research, the loss never consistently decreased, and the training continued indefinitely with no lasting improvement in training or test accuracy (Figure 4).

Figure 4: CasCor model training and test accuracy



The implementation of cascade correlation conducted for this paper did not train a pool of candidates and instead added only one at a time and trained that individual neuron based on maximising covariance. This has resulted in a poorer implementation as the use of the pool of candidates is designed to reduce the chance that a single unit will be poorly randomly initialised and get stuck in a local minima during training. If this happens then a suboptimal unit can be permanently added to the network with frozen weights (Fahlman & Lebiere, 1990).

A criticism of the hidden-unit cascading network is often associated with poor generalisation (Prechelt, 1997). On a small dataset such as this one this issue may be particularly relevant as weights need the ability to recalculate once other signals are picked up. Finally, the major limitation of the CasCor network was that it is not widely used, so there are very few examples or pre-built resources available. This makes this type of network more difficult to implement and modify outside of computer science academia.

4 Conclusion and Future Work

The results show success with both accepted back propagation networks and cascading networks built one neuron at a time. However due to complexity and training instability the standard back propagation network is preferred. As noted previously the CasCor model was limited by the basic implementation of the hidden neuron. Future work on implementing a pool of candidates may improve the performance of this model.

The application of feature selection was successful across all model types suggesting future work using genetic algorithms to reduce the feature space for physiological data would be of interest. This may involve running experiments with different or bigger data sets, testing different statistical aggregations of the raw data, and exploring different parameter settings for the genetic algorithm. Additionally, investigating which optimal features are selected and comparing the consistency of these features across different models or GA solutions may provide insight into which physiological features are generally most useful in detecting deception. This may assist researchers in setting up future experiments as they better understand which features are most important and prioritise collecting them most accurately.

Zhu et al (2018) found that the ability to consciously detect dishonesty varied by presenter, hypothesising that this may be substantially influenced by the presenter's ability to disguise their doubt. This makes sense as an experienced presenter may not be phased by a suggestion of 'bogus'-ness and would continue with the job without giving any tells. Further work using data and networks that can take presenter effects into account would be valuable.

5 References

1. Palomo, M.: How disinformation kills: philosophical challenges in the post-Covid society. HPLS 43, 51 (2021), https//doi-org.virtual.anu.edu.au/10.1007/s40656-021-00408-4. (2021)
2. van't Veer, A.: Effortless morality: cognitive and affective processes in deception and its detection. Dissertation, Tilburg. University (2016)
3. X. Zhu, T. Gedeon, S. Caldwell, R. Jones and X. Gu, Deceit Detection: Identification of Presenter's Subjective Doubt Using Affective Observation Neural Network Analysis, 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 2020, pp. 3174-3181, (2020), doi: 10.1109/SMC42975.2020.9283210.
4. Fahlman SE, Lebiere C.: The cascade-correlation learning architecture. In: Touretzky DS (ed) Advances in neural information processing systems, vol 2. Morgan Kaufmann, Los Altos, pp 524–532 (1990)
5. Fahlman, SE.: An empirical study of learning speed in back-propagation networks, Carnegie Mellon University, Computer Science Department (1988)
6. Nasreen, S.: A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning, SAI. (2014)
7. García-Dominguez, A., Galván-Tejada, CE., Zanella-Calzada, LA., Gamboa-Rosales, H., Galván-Tejada, J., Celaya-Padilla, JM., Luna-García, H., Magallanes-Quintanar, R.: Feature Selection Using Genetic Algorithms for the Generation of a Recognition and Classification of Children Activities Model Using Environmental Sound, Mobile Information Systems, vol. 2020, Article ID 8617430, 12 pages. https://doi.org/10.1155/2020/8617430. (2020)
8. Mallawaarachchi, V.: Introduction to Genetic Algorithms – Including Example Code, https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3. (2017)
9. Fortin, FA., De Rainville, FM., Gardner, MA., Parizeau, M. and Gagné, C.: DEAP: Evolutionary Algorithms Made Easy, Journal of Machine Learning Research, pp. 2171-2175, no 13. 9. (2012)
10. Coliann, S.: https://github.com/scoliann/GeneticAlgorithmFeatureSelection. (2017)
11. Zhu X., Qin Z., Gedeon T., Jones R., Hossain M.Z., Caldwell S. Detecting the Doubt Effect and Subjective Beliefs Using Neural Networks and Observers' Pupillary Responses. In: Cheng L., Leung A., Ozawa S. (eds) Neural Information Processing. ICONIP 2018. Lecture Notes in Computer Science, vol 11304. Springer, Cham. (2018). https://doi.org/10.1007/978-3-030-04212-7_54
12. Dandurand, F., Berthiaume V. & Shultz, T. R.: A systematic comparison of flat and standard cascade-correlation using a student–teacher network approximation task. Connection Science, 19:3, 223-244, (2007). DOI: 10.1080/09540090701528951
13. Prechelt, L:. Investigation of the CasCor family of learning algorithms. Neural Net., 10: 885–896. (1997)