Human Eye Gaze Patterns: Multi-model Learnable Discernment of Manipulated and Unmanipulated Digital Images Investigation

Tiger Chen

The Australian National University, Research School of Computer Science
Acton, ACT, Australia
u6380238@anu.edu.au

Abstract. Digital images are ubiquitous throughout the digital age and so is the software available that manipulate them. Implications arise from the ability of people to determine the manipulation of images. An investigation has found participants unable to discern manipulated images accurately (poor to moderate) even after being trained in image editing techniques. However, from the eye gaze data from the same study we have found that we are able to train multiple network architectures (neural network, casper, casper evolutionary algorithm variant) to perform marginally better than the human participants. The performance of the machine learning model implies the latent potential of peoples gazes in discerning image changes. Nevertheless, there is a lot of room for expansion on the evolutionary algorithm techniques applied to the casper network. The hybrid method should be evaluated on a multitude of problem domain data sets to get a better grasp of its performance.

Keywords: Eye gaze tracking · Manipulated images · Casper neural network · Evolutionary algorithm.

1 Introduction

In the modern age, digital images are prevalent on all platforms ranging from social networks, news, education and entertainment. Furthermore, image editing software/techniques are available to the common user in comparison to the past where only specialised individuals/organisations had the equipment and expertise. The ease of manipulation of digital resources coupled with the rapid spread of information on the internet landscape has created a volatile environment for misinformation. These can range from relatively harmless image modifications such as for the sake of comedy or aesthetics in areas of the modelling industry. On the nefarious end of the spectrum is the spread of fake political/factual information which has the power to change nations and cause great distress [8]. Therefore, it is crucial to investigate and understand how we perceive images.

Research has been conducted in the area of human discernibility of manipulated and unmanipulated digital images [1]. In the previous work the authors found the participants had a low efficacy to determine modifications even after image manipulation techniques had been explained. Besides noting their guesses, additional data was collected about the subjects eye gaze.

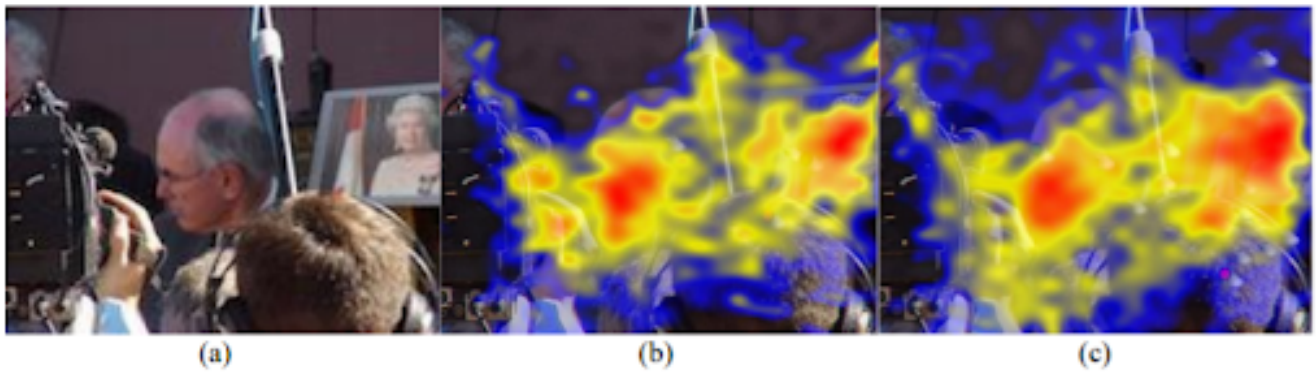

Fig. 1. Example heat map comparison: base image (a), unmanipulated image (b) and manipulated image (c).

Though, the participants could not detect modifications accurately, the goal of this experiment is to determine if the unconscious and conscious tracking of the eye encodes more information about the state of the image. In order to achieve this a neural network model is trained on the eye tracking data due to its generalisation ability.

2 Methodology

Three types of models will be applied to the eye gaze data set. The first is a conventional feed forward neural network used as a baseline before more exotic methods are applied. Next is a Cascade network algorithm employing progressive RPROP (Casper). This type of network architecture is chosen because it is a generative model that automatically 'grows' while learning the task which circumvents traditional model size selection issues. In addition, information can be gleaned based on how large the network expands in terms of differing experimental setups. Its connections can also be analysed more readily due to its simplistic architecture to investigate key features influencing outcomes. Lastly, to address shortcomings of the Casper network evolutionary techqniques will by applied to it to create a hybrid. To evaluate the networks the confusion matrix of the classifications will be computed on the train and test sets. In addition the loss and test accuracy of the models will be tracked providing insight on hyper parameter influence and generalisation/over fitting.

2.1 Feed Forward Neural Network

Out of the neural network models being used the baseline feed forward network is the most well known/basic and will not be covered in depth, but what is important is its specific configuration. In this application the feed forward network has two linear layers with sigmoid activation functions. It also applies cross-entropy loss because of the classification task and an Adam optimizer. Now onto the next architecture being trained.

2.2 Casper

The casper algorithm is a type of cascade network and bears similarities to Cascade Correlation (Cascor). It is a constructive learning algorithm where neurons are successively added. The depth of the network extended and the layers remain composed of a single node. All of the connections from the previous layer, including its inputs are propagated to the next. Thus at the output node its incoming connections are from the model inputs and outputs of each hidden neuron.

New neurons are added when the RMS error has decreased 1% of the previous error. Nevertheless, this must occur within a certain timeframe dictated by the heuristic '15+P*N'. The 'P' parameter is user defined while 'N' is the number of hidden neurons giving bigger networks a longer time window to meet the criteria.

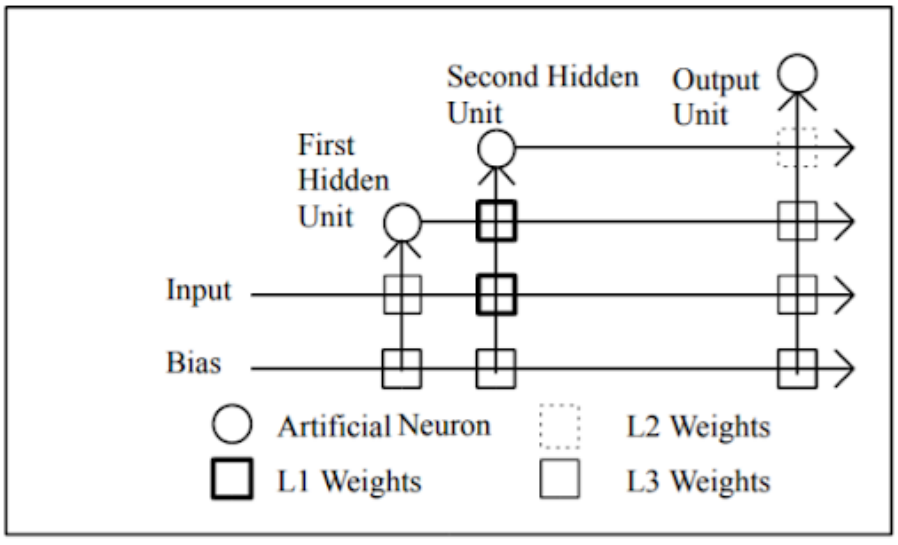

Fig. 2. Casper architecture: A second hidden unit has just been added.

What differs Casper from Cascor is how the network learns. In Cascor, when a new neuron is added, the previous weights are frozen. However, this has shown to have a few disadvantages. Firstly the early frozen neurons may be left in an unrefined state and be poor feature detectors. In turn causing the networks to grow excessively large as later neurons have to make up for the inefficiencies [6]. While with Casper, none of the weights are frozen when a

new neuron is added. Nevertheless, this brings up the issue of too much interference coming from the other weights, this phenomenon is aptly coined the 'herd effect' [7]. The 'herd effect' does not allow the new neuron to make up for the majority of the residual error in the network as all of them have the same learning rate.

An modified adaptive learning rate algorithm RPROP [4, 5] is used to get around the 'herd effect' by setting different learning rates. Unmodified RPROP starts the weights with an initial learning rate that adapts depending on the sign of the gradient and not its magnitude at each epoch. The modification is to reinitialise neurons with specific learning rates to spur varying growths when expanding. As seen in the figure above, the weights are categorised into three learning rates where L1 >> L2 > L3 meaning the second hidden neuron will learn much faster with respect to the older hidden neurons.

Casper also utilises weight decay with simulated annealing as seen in the SARPROP algorithm [3] to improve convergence and generalization [2]. The simulated annealing factor is controlled by the number of epochs since the last neuron was added (Hepoch). The overall presence of the weight decay is controlled by the coefficient 'k'.

$$\delta E/\delta w_{ij} = \delta E/\delta w_{ij} - k * \text{sign}(w_{ij}) * w_{ij}^2 * 2^{-0.01 * HEpoch}$$

2.3 Casper Evolutionary Algorithm Hybrid

The Casper algorihtm provides a novel strategy for a growing cascade network and avoids issues of weight freezing and static architecture build paradigms. Nevertheless, it also has its drawbacks which were made apparent when applying it to the eye gaze data set. Specifically, the problem lies with how the model determines adequate growing conditions. As explained earlier, if the RMS error has decreased 1% of the previous error within '15+P*N' epochs the network expands by a neuron. In practice this has lead to hyper parameter custom tuning per the problem in order to get desirable sizes that can learn the solution properly. Though the added parameters to change is cumbersome it is not the crux of the complication. From empirical testing of Casper on the eye gaze data set a configuration could not be found that had moderate growth. What was observed was either little or explosive growth that stunted the training process. One reason for the irregular growth was the noisy RMS error spiking and not exhibiting a consistent enough decreases.

A few concepts from evolutionary algorithms (EA) were applied to form a hybrid method that attempts to address these issues. The main EA concepts leveraged are population, hall of fame and generations with a twist. With population, we are no longer training a single network and the hall of fame will keep track of the best performers that have existed. Finally, when it comes to generations the twist is instead of applying crossover and mutations all the members in the populations will grow. This presents a caveat to crossover operations because the shape discrepancies from generation to generation make it difficult to produce meaningful re-combinations. The essence of the changes is growth is detached from network performance metrics and is controlled to occur consistently. In summary, the strategy starts with an population of Casper networks set at its initial size which are trained for a set number of epochs. Once the number of training iterations have been reached, the next generation is birthed by growing the old population. However, before modifying the existing population they are compared to the hall of fame and added based on their test accuracy. The new generation is trained for the same number of epochs with the process repeating. The benefit is different network sizes are iteratively explored and saved depending on their performance.

Technically, what is occurring is just a basic hyper parameter iteration of the cascade length and utilising populations is unnecessary because crossover is not happening. However, it leaves the groundwork for full evolutionary system to be implemented in future work. The plan is to perform crossover between the models at the end of each generation by using innovative proven techniques [10]. Furthermore, mutations can be introduced by randomising weights by chance and is equivalent to drop out, thus improving generalisation. On the other hand this may be unnecessary as casper already has weight decay built in and served the same purpose.

2.4 Implementation

Only the casper implementation is covered in depth as the feed forward network and EA hybrid are trivial. Hyper parameters for all three are listed.

Pytorch is an open source library for machine learning and was used to implement Casper. Its developmental structure is based on interlinking high level modules which serve to form the layers of the network. At a glance the

cascade and dynamic structure of Casper makes it seem unsuited to be implemented elegantly and pragmatically implemented with the framework. However, this aspect is tackled by utilising the ModuleList container to store the expanding network. The next issue is the layer by layer definitions. Normally a single learning rate is set per layer, but new hidden neurons have different learning rates for the weights and bias. Fortunately, the optimizer has parameter options to specify a different rate for bias as it is treated as a separate parameter group.

Nevertheless, the output unit has a different learning rate for the connection to the new neuron (L2). Parameters for the weights group cannot be subdivided for individual weights. Thus, another layer is added between the new neuron and output called the L2 buffer layer. Its only purpose is to represent the weights and L2 learning rate of the output while the connections coming into the output are set to unity and frozen. This avoids unnecessary manual modifications to the gradient to achieve the same purpose as this could affect function calculations down stream. RMSPROP is used instead of RPROP because of its improved performance in mini batch applications [9], but its core level behaviour remains the same. Hyperparameters chosen in the Casper implementation are 'num_epoch = 300', a 'P = 5' for the growth timer, 'k = 0.005' weight decay factor and 'L1 = 0.2, L2 = 0.005, L3 = 0.001' learning rates as set in the Casper paper [2]. For the feed forward network, it used 'num_epochs = 300', 'hidden_size = 5' for the two layers and 'learning_rate = 0.01'. For the Casper EA hybrid the 'generations = 10' with 'num_epochs_per_gen = 25' and a 'num_population = 5'.

3 Data

The data being trained on is the eye gaze tracking collected when the participants observed the images [1]. It contains seven columns which are the participant identification number, total number of fixations by the participant when looking at the image, total fixation duration, total number of fixations looking at the modified area, total fixation duration at the modified area, manipulation ground truth and participant vote. The goal is to predict if the image is manipulated based on the eye gaze tracking information, therefore the identification number and participant vote is removed during preprocessing. Additional preprocessing includes normalising the fixation data and leaving the manipulated classifier untouched. Finally the dataset is divided into two sets split '80/20' for training and testing.

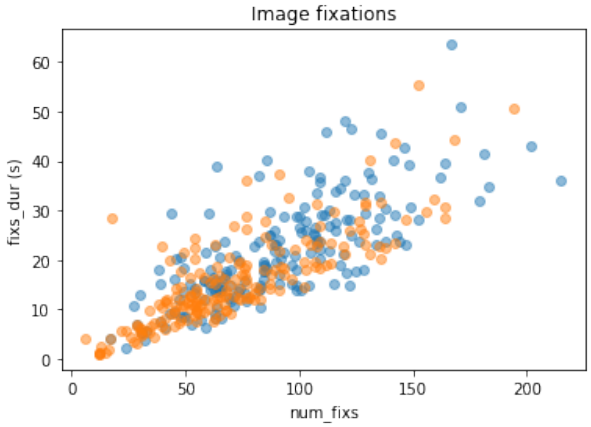

Fig. 3. Fixations in general.

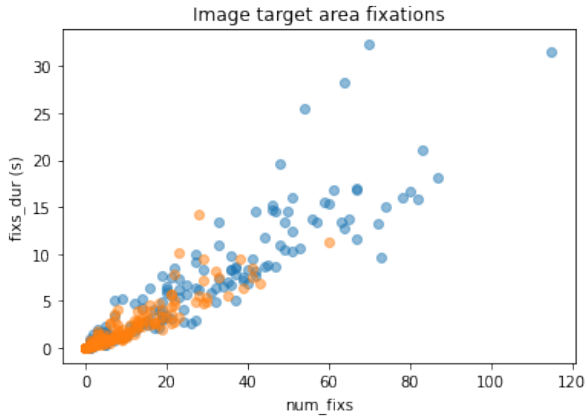

Fig. 4. Fixations in the target manipulated area.

4 Results and Discussion

Training the networks with the data has a training accuracy of ∼75% and test accuracy of ∼70%. This is only marginally better than the participant results with an average success rate of 56% [1] from the eye gaze investigation. The confusion matrices further describes the accuracy distribution with approximately ∼20% misclassifications for unmanipulated images and ∼50% misclassifications for manipulated images. The approximations are generated from running the model a handful of times and observing the general behaviour and are in no manner a thorough and proper investigation of the models performance on the dataset. Properly done, the experiment would have run many trials (a hundred) with different starting random weight initialisations. The dataset will also have been distributed more fairly to ensure even classes of images in the training/test sets to avoid model skews.

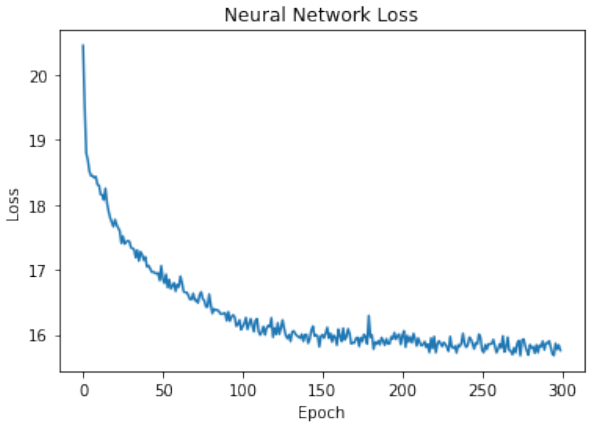

Fig. 5. Feed forward network loss.

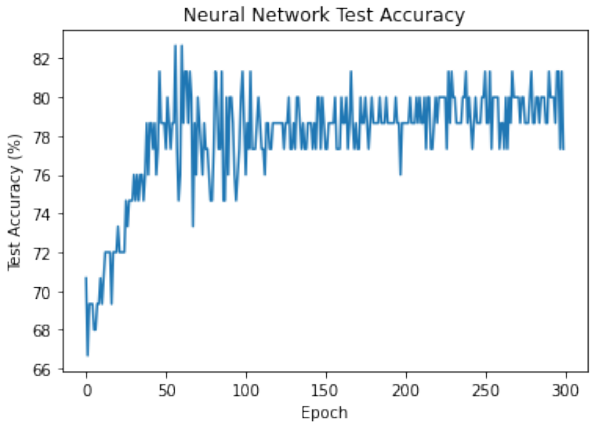

Fig. 6. Feed forward network test accuracy.

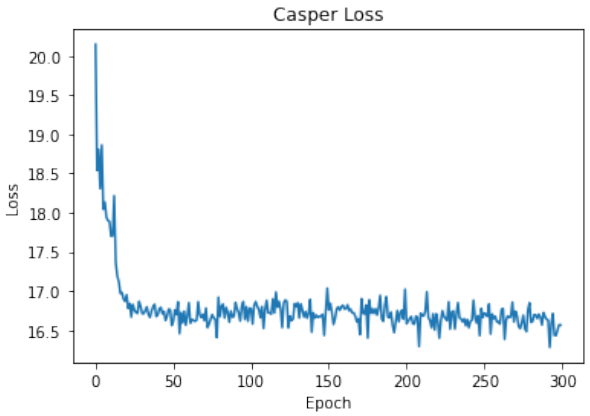

Fig. 7. Standard Casper network loss.

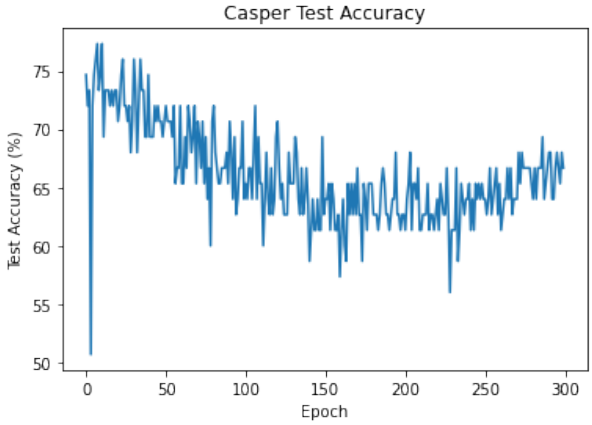

Fig. 8. Standard Casper test accuracy.

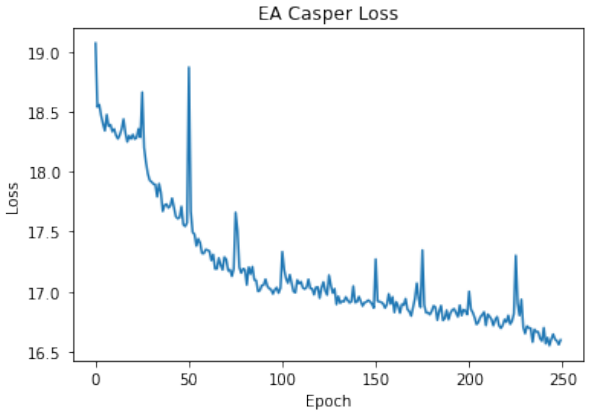

Fig. 9. Evolutionary algorithm Casper hybrid loss.

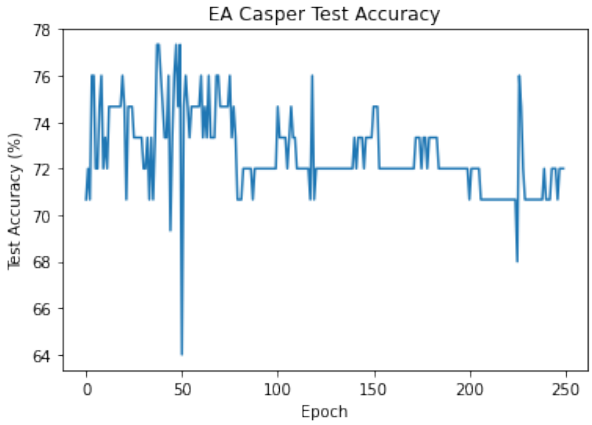

Fig. 10. Evolutionary algorithm Casper hybrid test accuracy.

While being trained the test accuracy was recorded to determine over fitting. What was found instead is that all models exhibited fluctuating test accuracy results after very few epochs passing. This may be because the manipulated and un-manipulated image data points were observed to be very poorly separable in the data set figures. Therefore, shifting the classification in a region will have a chance at negatively spiking the test accuracy based on its distribution and overlap. Another distinction is the loss seen in the EA Casper hybrid. Since the

network is grown each generation every 25 epochs the loss can been seen spiking. These spikes are also reflected in its test accuracy graph respectively. Below, metrics were gathered observing the optimal Casper network size from the hall of fame of specimens. The highest accuracy of ∼80% achieved was with a network grown eight times, but this also saw the lowest accuracy reported.

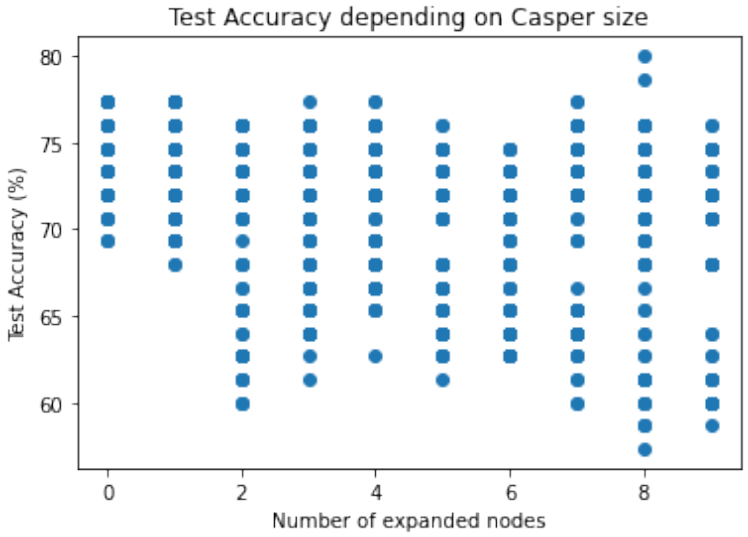

Fig. 11. EA Casper hybrid test accuracy vs network size comparison.

Assuming the limited metrics gathered are accurate then it can be inferred the movement of our eye gazes behaviour encodes more information on the modification of digital images than our conscious self realises. Though there are still many unanswered questions regarding the depth the participants were trained in detecting image modifications and how much further can we improve our accuracy. The initial results indicate a latent potential to improve merely from consciously noticing our eye gaze patterns. These results come with a caveat since two of the predictors are based on the participants' fixation at the target area. From the person's perspective, knowledge of the target area is unknown therefore it is inaccurate to infer a person's learning based on data they do not possess. A second trial should be run on the model with the target fixation columns removed and the performance reevaluated. Further results to gather include the network size and weights. The first allows us to see how limitations in input predictors affect the models growth. The other gives clues on important internal features and inputs which can be extrapolated to techniques applied by people.

5 Conclusion and Future Work

The preliminary results gathered implies eye gaze information encodes clues in determining image modification attributes. All the neural networks boasts a ∼15% improvement in identification from participants trained in image editing techniques who were only successful 56% of the time. The dataset provided as well is lacking in information as it contains summaries of the eye gaze fixations and not the time series information. If this data is available the network may be able to better detect image modification. Nevertheless, there is a lot of room for expansion on the evolutionary algorithm techniques applied to the casper network such as adding crossover and mutation functionality. The hybrid method should also be evaluated on a multitude of problem domain data sets to get a better grasp of its performance.

References

1. Caldwell, S., Gedeon, T., Jones, R., Copeland, L.: 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) (2015).
2. Treadgold, N.K., and Gedeon, T.D.: A Cascade Network Algorithm Employing Progressive RPROP (2006).

3. Treadgold, N.K., and Gedeon, T.D.: A Simulated Annealing Enhancement to Resilient Backpropagation. In: Proc. Int. Panel Conf. Soft and Intelligent Computing, Budapest pp. 293-298 (1996).
4. Riedmiller, M. and Braun, H.: A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm. In: Ruspini, H., (Ed.) Proc. of the ICNN, San Francisco, pp. 586-591 (1993).
5. Riedmiller, M.: Rprop - Description and Implementation Details, Technical Report, University of Karlsruhe (1994).
6. Kwok, T., and Yeung, D.: Experimental Analysis of Input Weight Freezing in Constructive Neural Networks. In: Proc. IEEE Int. Conf. Neural Networks. pp. 511-516. (1993).
7. Fahlman, S.E., and Lebiere, C.: The cascade-correlation learning architecture.. In: Advances in Neural Information Processing II, Touretzky, Ed. San Mateo, CA: Morgan Kauffman, pp. 524-532. (1990).
8. Shen, C., Kasra, M., Pan, W., Bassett, G., Malloch, Y., O'Brien, J.: Fake images: The effects of source, intermediary, and digital media literacy on contextual assessment of image credibility online. In: SAGE New Media & Society 2019 (Vol. 21) (2018).
9. Ruder, S. (2017) An overview of gradient descent optimization algorithms, https://arxiv.org/pdf/1609.04747.pdf. Last accessed 24 April 2021
10. García-Pedrajas N., Ortiz-Boyer D., Hervás-Martínez C., (2005) An alternative approach for neural network evolution with a genetic algorithm: Crossover by combinatorial optimization, https://www.sciencedirect.com/science/article/pii/S0893608005002297. Last accessed 2 June 2021