Screen Size Classification from Eye Gaze Search Data: Combining Bimodal Distribution Removal with Back-Propagation Neural Network and Genetic Algorithm

Haolin Zhou

Research School of Computer Science,
Australian National University,
u7085374@anu.edu.au

Abstract. In recent years, many screens with different sizes are developed, regarding which eye gaze behavior would be different. Different kinds of search performance and search behavior was observed and most of them are correlative to the screen size. [1] This paper investigates the performance between the neural network with or without applying the bimodal distribution removal (BDR) technique on the binary classification of screen size by the eye gaze search features. This research provides the basic idea of how bimodal distribution removal works on the small dataset and how it handles the outliers. Additionally, the Genetic Algorithm was used for feature selection. BDR could remove outliers according to the error of the sample to improve the performance of the neural network. The performance of the Back-Propagation Neural Network could be improved after adding the BDR and Genetic Algorithm respectively in this data set.

Keywords: neural networks, bimodal distribution removal, outlier detection, eye gaze search, screen size, genetic algorithm, evolutionary algorithms

1 Introduction

More and more people use mobile devices to search on the website rather than using computers with large screens. Eye gaze detection could provide information on human interaction for research of screen design. When people are searching on different size of the screens, the eye gaze has different reactions. Investigating the classification of the size of screens by eye gaze information could be used in further analysis in human-computer interaction. Furthermore, the performance of the classification task can verify the correlation of the features and the target.

The previous research about the effect of screen size to the searching performance that has done by Jaewon Kim, Paul Thomas, etc. used only statistical methods such as ANOVA to evaluate the correlative of each feature to the screen size and search task [1]. The people who participated in the experiment have been asked to search for navigation or information problems on large and small screens. Their eye gaze information was collected and generated as different features. The chosen data set contains 644 rows and 29 columns which is about the eye gaze of people when they are having different search tasks on different size of the screen. It contains 27 features that can be used to predict the class of screen size. The features of the data set are divided into 2 parts, search performance and search behavior. The search performance consists of search speed and accuracy. Fixation, scan path, scanning direction, regression, skip, search strategy and number of the visited page are concluded search behavior.

The neural network could be used to solve the binary classification problems with good performance [5]. The genetic algorithm is a good choice to select the model with the best performance by adjusting the hyper parameters or selecting the most useful features [7]. The dataset contains 27 features while some of them could be useless. When deciding the features, the genetic algorithm has shown good performance for optimization [8]. By using the genetic algorithm for selection, we could make a relatively optimal decision so that could observe the performance of the models with a better condition [9].

Moreover, the features that has been selected automatically by the genetic algorithm could be compared with the correlative analysis of the features in this data set and verify the idea of that has done by Jaewon Kim, Paul Thomas, etc. [1]. However, the outliers in the data set would affect the performance of the neural network [2]. In such case, many statistical techniques are introduced. Bimodal Distribution Removal (BDR) is an efficient technique that avoid the weakness of Absolute Criterion Method, Least Median Squares (LMS) and Least Trimmed Squares (LTS) [3].

In the following sections, the method would be introduced first. It includes the idea of how to preprocess the data with z-score normalization and feature selection, the structure and the setting of the back-propagation neural network, the principle of bimodal distribution removal technique, and the evaluation method. Secondly, the performance of BDR would be evaluated with discussions based on the experiments. At last, the conclusion would be made and the interesting investigations could be done in the future are suggested.

2 Method

2.1 Data Pre-processing

The data set has been loaded into a data frame and separated randomly into 80% of training data and 20% of testing data. The column ‘subject’ and the header of the original file are removed from the data frame since they have no contribution to the neural network model. By encoding characteristic features to numerical data, the whole data set could be transmitted to the neural network.

The transformations are shown as follow:

1. The values in the feature ‘size’ are transferred to 0.0 and 1.0 from “L” and “S” respectively.
2. The values in the feature ‘task’ are transferred to 0.0 and 1.0 from “info” and “nav” respectively.
3. The values in the feature ‘task’ are transferred to 0.0, 1.0 and 2.0 from “DF”, “BF” and “MX” respectively.

2.1.1 Z-score Normalization

Z-score normalization is a common method of data processing. It can be used to convert data of different magnitudes into a unified Z-score for comparison. [6] The data is subtracted from its mean by column and divided by its variance. The result is that for each column all data are clustered around 0 with a variance of 1. Every single data would be calculated by the formula as follows:

$$z = \frac{x-\mu}{\sigma} \quad \textbf{(1)}$$

Where μ is the mean and σ is the standard deviation of the data in a column. The data procced by z-score would distributed as the normal distribution in Fig.1.

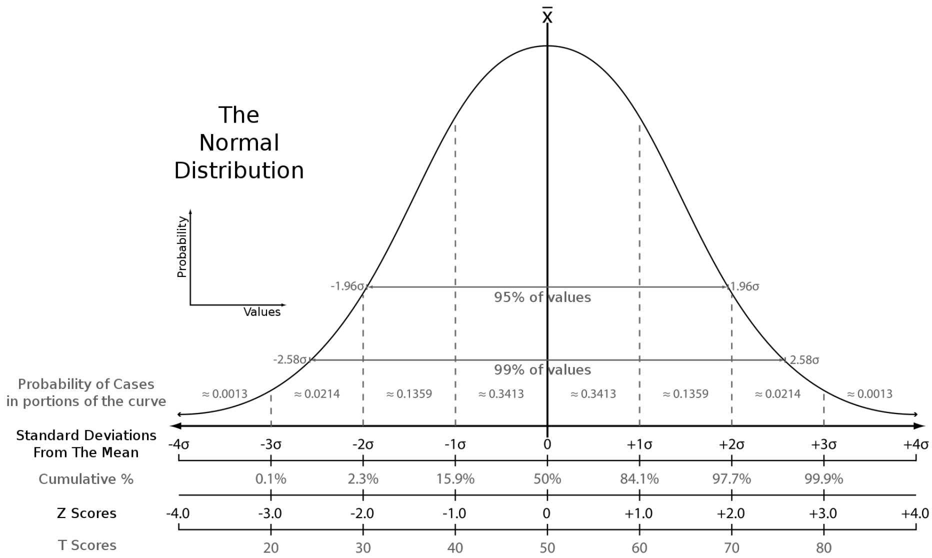

Fig. 1. The normal distribution generated by z-score [4].

2.1.2 Manual Feature Selection

Feature selection is to select the best subset of features for a specific task from the data set. Feature selection can not only reduce overfitting, but also improve the generalization ability of the model. Generally, it will accelerate the training speed of the model and get better performance. Fortunately, [2] used statistical methods to investigate the correlation between each feature and the target, as well as the correlation between features and features. As such, the manual feature selection is based on the semantic analysis from [2]. There are 27 features in this dataset. The features “Accuracy”, “Wrong answer”, “Minimal scanpath value”, “Compressed scanpath value”, “Compressed/Minimal” and “Page visit” are removed since they are not correlative to the size of screens [2]. The genetic algorithm is used for automatic feature selection which would be discussed in section 2.4.

2.2 Back-Propagation Neural Network

Neural Network is an algorithmic mathematical model that imitates the behavioural characteristics of animal neural networks and performs multiple parallel information processing. It relies on the complexity of the system and adjusts the interconnected relationship between many internal neurons to achieve the purpose of information processing [4]. A simple two-layer neural network with one hidden layer was constructed to solve the binary classification problem in this data set. The hidden layer is set to include 50 hidden neurons. The number of input and output neurons are 22 and 2 respectively since there are 22 features in total. The sigmoid function is used as the activation function for hidden layer as common setting for classification tasks. Since the model is built to solve the binary classification problem, the loss function is defined as the cross-entropy loss function. Stochastic Gradient Descent (SGD) is chosen to be the optimizer. The neural network is trained with the back-propagation algorithm. With the learning rate 0.01, after 500 epochs the neural network is trained.

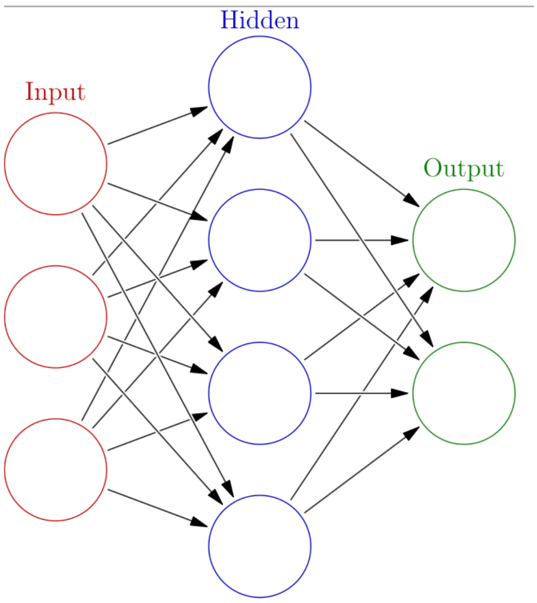

Fig. 2. Visualization of a simple two–layer Neural Network.

2.3 Bimodal Distribution Removal

The Bimodal Distribution Removal technique is used for the outlier detection using the distribution of error which suggested by P. Slade and T.D. Gedeon. The number of epochs is 500 as mentioned above. However, the Bimodal Distribution Removal would be applied every 100 epochs during training. If the Bimodal Distribution Removal is applied too frequent, the performance of the model would be worse since the useful samples are removed as the outliers. Applying Bimodal Distribution Removal too early would cause this problem as well.
To remove the outlier, the evaluation of the samples that needs to remove is defined as follows:

$$errors > \delta_{ss} + \alpha\sigma_{ss} \qquad 0 < \alpha < 1 \tag{2}$$

Where δ_{ss} is the mean, σ_{ss} is the standard deviation, and α is a constant that between 0 and 1. If the pattern has the error that is greater than the value calculated by (2), it would be removed.
The hyper parameters of Bimodal Distribution Removal are set as follows:

- Variance threshold: 0.00001

- Alpha: 0.53

When the variance threshold is achieved, the training would be halted to prevent overfitting. The variance threshold is 0.00001 as the model would have a better performance than just roughly use 0.01. The higher the threshold is, the earlier the training would halt. The training is uncompleted when the variance threshold is set as 0.01 in this case. When the variance threshold is 0.01 the training stopped too early so that the accuracy is low.

2.4 Genetic Algorithm

Genetic Algorithm (GA) is a computational model that simulates the biological selection process of Darwin's biological evolution theory and genetic mechanism. It is a method to search for the optimal solution by simulating the natural evolution process [10].

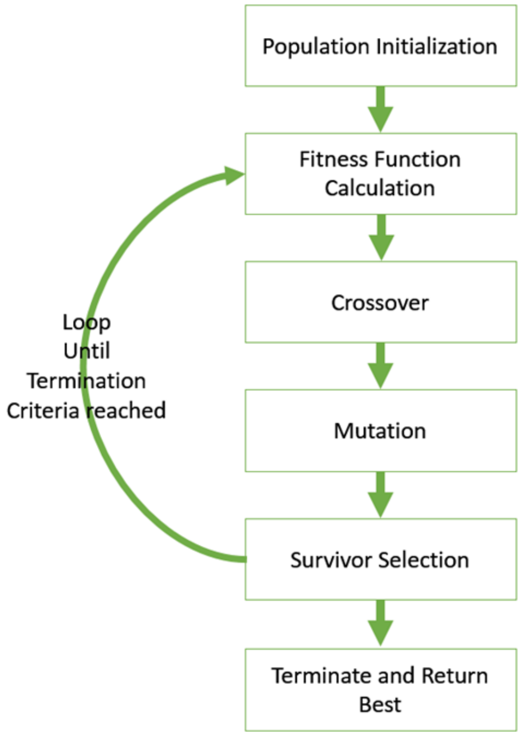

Fig. 3. The procedure of the Genetic Algorithm

From Fig.3 we can see after the first-generation population has been generated, the generations evolve to produce a better approximate solution, according to the principles of fitness survival. In each generation, the choice is based on the individual's suitability for the problem domain. By combining crossover and mutation, a population representing a new solution set is generated. This process allows the descendants of the population to adapt to the environment more naturally than their predecessors. The best individual of the last population is decoded and can be used as an approximately optimal solution to the problem.

The main challenge of this dataset is there are too many features which make the decision of feature selection harder. The Genetic Algorithm shows its value on feature selection. The format of the DNA is a 28-bit array which conations binary numbers. It will show like [0 1 1 0 0 0 0 1 0 1 0…]. The corresponding removal features are represented as “1”. Precisely, when the index of the feature is as 1, the feature would be removed while it would be kept.

Since most of the feature are correlative to the screen size, the number of removal features is set as 6. If remove too much features the performance of the model would be worse. When initializing the population, there would be 6 remove features in the DNA. To ensure the number of removal features not greater than 6, the offspring would be checked after generated. If the remove features are more than 6, the amount of remove features would be reduce to 6 randomly.

To generate the population, the population size is 50 since there could be more chance to have child with better performance in one generation. The number of generation is 10. Ten generation is enough to get relatively optimal solution in this case. The parent with better performance would be selected by the fitness function. The performance of the model is evaluated by the test accuracy. Thus, the fitness function is defined as to return a list highest accuracy after applying the model for each DNA in the population. To be more precise, the accuracy would be calculated for 5 times for each model. The average of the test accuracy would represent the accuracy of the model.

Crossover and mutation would be applied randomly after the best parents are selected. The mutation rate and the cross rate has been set as 0.05 and 0.8 respectively. The offspring would be created iteratively and the optimal offspring would occur after fitness, crossover and mutation in an amount of generations.

2.5 Evaluation

The performance of the model is evaluated by the accuracy as in this binary classification model as the outcome of the binary classification model is simply 0 and 1. As mentioned above, the data set has been divide into training set and testing set randomly by the ratio 0.8: 0.2. The test accuracy is according to the following formula:

$$Test\ Accuracy = \frac{the\ number\ of\ correct\ predictions}{the\ number\ of\ the\ test\ samples} * 100\% \quad \textbf{(3)}$$

To evaluate the performance between simple neural network model and the model with bimodal distribution technique, we focus on the test accuracy.

3 Results and Discussion

3.1 The Performance of Back- Propagation Neural Network with and without Bimodal Distribution Removal

The training and testing accuracy of the neural network with or without bimodal distribution removal are calculated every 10 epochs during the training.

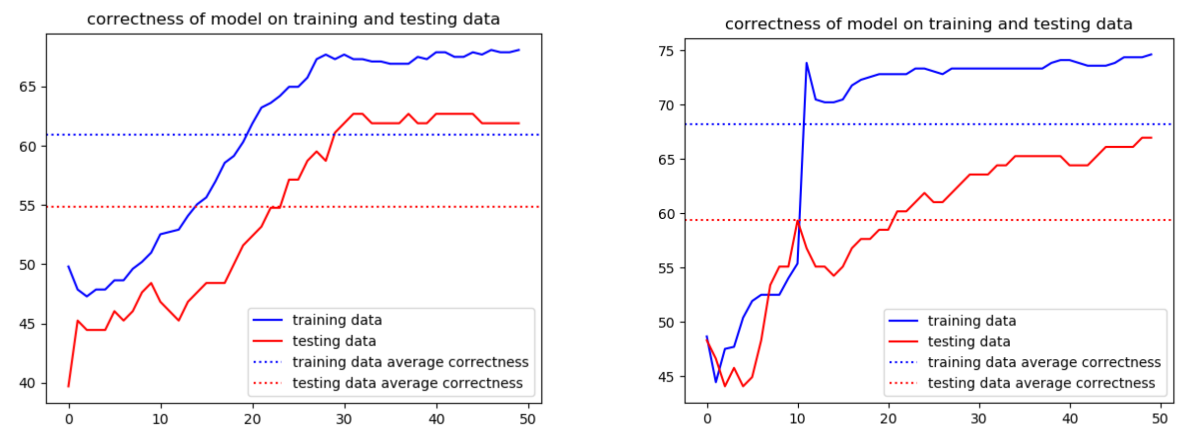

Fig. 4. The performances of simple neural network (left) and the neural network with bimodal distribution removal (right) are shown respectively.

As shown above, the curve of accuracy of the simple neural network is gradually rise and reach a peak at the later of training. The testing accuracy of the simple neural network is usually around 60% from the observation over 50 times of training respectively. The average accuracy of the training and testing accuracy is 55% and 61% approximately. The ability of the simple neural network is relatively weak to deal with the data set with many features.

The testing accuracy of the neural network is higher after applied the bimodal distribution removal which achieved around 65%. The bimodal distribution removal technique has a good performance in this data set. The training and testing accuracy increase rapidly when the epoch is at 100 since most of the outliers are removed at epoch. The training accuracy is higher than the testing accuracy. The reason is the outliers are only removed in the training set. The outliers are still existed in the testing set.

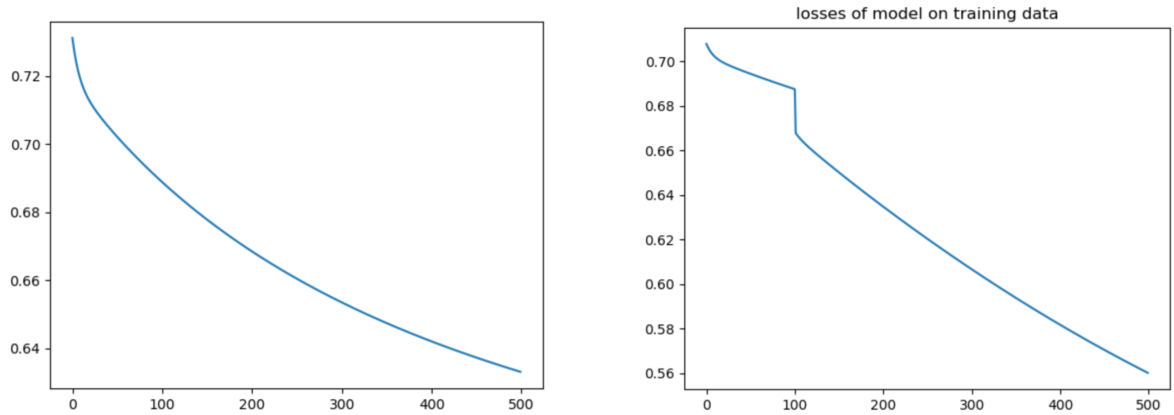

Fig. 5. The value of loss relative to the epoch in simple neural network (left) and the neural network with bimodal distribution removal (right).

Figure 4 could provide the information that how the loss changes during the training. The loss of the simple neural network is decreased smoothly and reach a lowest value at 500 epochs. As such, the training accuracy and the testing accuracy increase at the mean time. The neural network is trained effectively.

The loss of the neural network with bimodal distribution is decrease gradually and reached the lowest point at the 500th epoch. When the training is at the 100th epoch, the loss decreases rapidly since some of the outliers have been removed. It only appears once since the BDR removes most of the outliers at the first time.

3.2 The Performance of Back- Propagation Neural Network with Bimodal Distribution Removal after Automatic Feature Selection by the Genetic Algorithm

The accuracy of each generation is shown in figure 6 and figure 7 which could easily show that the genetic algorithm has good performance.

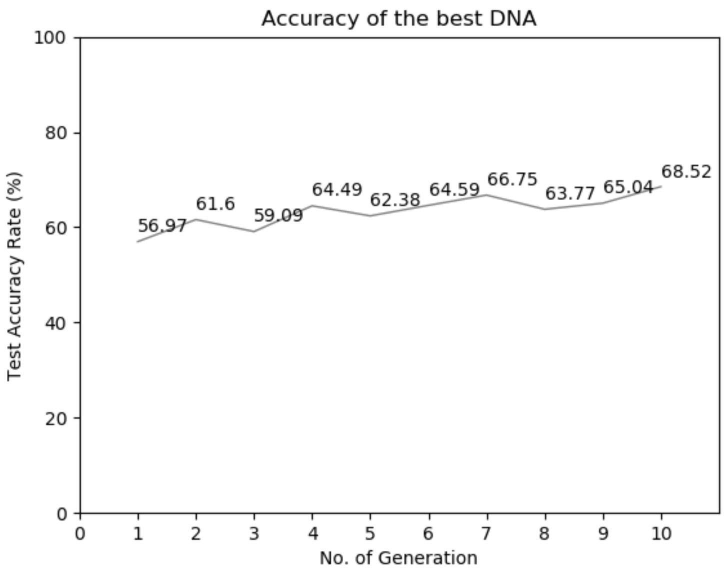

Fig. 6. The best test accuracy of the population in each generation in simple neural network.

According to figure 6, the accuracy is increasing among generations. The model reached the test accuracy 68.52% at the last generation which is greater than the accuracy that could achieved by simple neural network without feature selection.

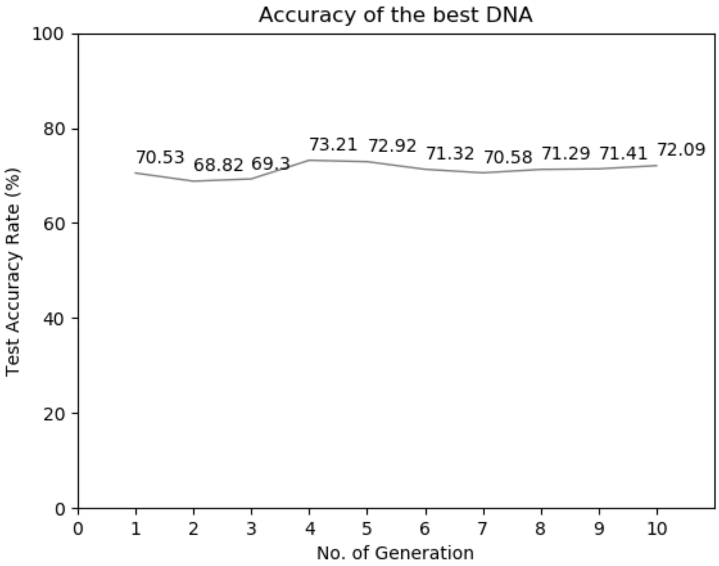

Fig. 7. The best test accuracy of the population in each generation in neural network with bimodal distribution removal.

The performance of the model has great improvement after combining the bimodal distribution technique and the genetic algorithm. We could also find from figure 7 that the genetic algorithm does not always improve the performance of the next generation. However, the genetic algorithm would reach the approximate optimal solution after an amount of generations. The features that have been removed by the genetic algorithm are similar to the features that have been removed in the manual feature selection which could verify the conclusion of [1].

4 Conclusion and Future Work

The simple back-propagation neural network could handle the binary classification problem. Adding the bimodal distribution removal technique on this data set, the performance is better than the simple neural network, leading to the conclusion that the bimodal distribution technique is suitable for this data set. The percentage of the outliers in this data set is relatively high as it contains around 600 samples. The performance of the bimodal distribution removal technique is depending on the outliers. When the outliers are useful training samples, the bimodal distribution removal technique would perform relatively worse. According to the result of the experiment, the correlation of the selected features and the size of the screens has also been justified. The genetic algorithm has good performance on finding the best set of the features [7]. The great improvement it made could make the comparison more valuable since it provided relatively optimal solutions.

In the future, the relationship between the number of outliers and the performance of the bimodal distribution removal technique could be investigated further. As is described in [1], each feature and how they correlative to the size of the screen (small or large) and the type of the tasks (navigation or information). What have been generally verified is the relationship between the features and the size of screen. However, the correlation of the features and the type of the tasks is meaningful to have some experiment on. The time of the feature selection done automatically by the genetic algorithm is longer than expected. The algorithm could be optimized to reduce time. What's more, the current work only shows that all the features that are not correlative to the size of the screen could be removed as feature selection. Besides, there are more investigations to be done on the features. Using the features one by one to as the input of the neural network could verify the degree of correlation of each feature.

References

1. Kim, Jaewon, Paul Thomas, Ramesh Sankaranarayana, Tom Gedeon, and Hwan-Jin Yoon. "Eye-tracking analysis of user behavior and performance in web search on large and small screens." Journal of the Association for Information Science and Technology 66, no. 3 (2015): 526-544.
2. Azme Khamis, Zuhaimy Ismail, Khalid Haron, and Ahmad Tarmizi Mohammed. "The Effects of Outliers Data on Neural Network Performance". Journal of Applied Sciences, 5: 1394-1398 (2005)
3. P. Slade and T.D. Gedeon. 1993. Bimodal Distribution Removal. In: Mira J., Cabestany J., Prieto A. (eds) New Trends in Neural Computation. IWANN 1993. Lecture Notes in Computer Science, vol 686. Springer, Berlin, Heidelberg

4. Beale, Hagan Demuth, Howard B. Demuth, and M. T. Hagan. "Neural network design." Pws, Boston (1996).
5. Bischof, Horst, Werner Schneider, and Axel J. Pinz. "Multispectral classification of Landsat-images using neural networks." IEEE transactions on Geoscience and Remote Sensing 30.3 (1992): 482-490.
6. Cheadle, Chris, et al. "Analysis of microarray data using Z score transformation." The Journal of molecular diagnostics 5.2 (2003): 73-81.
7. Jihoon. Y., Vasant. H.: Feature Subset Selection Using a Genetic Algorithm (1997)
8. Richa. M., Gaganpreet. K.: Neural Networks using Genetic Algorithms (2013)
9. Duch, W., Adamczak, R., Grabczewski, K.: Extraction of logical rules from training data using backpropagation networks. In: The 1st Online Workshop on Soft Computing. pp. 19–30 (1996)
10. Goldberg, David E. Genetic algorithms. Pearson Education India (2006).