An Organized Committee of Artificial Neural Networks in the Classification of Human Chromosomes

Sadina Gagula-Palalic
Computer Sci. and Engineering
Faculty of Engineering and Natural Sciences
International University of Sarajevo,
Bosnia and Herzegovina

Mehmet Can
Professor
Faculty of Engineering and Natural Sciences
International University of Sarajevo,
Bosnia and Herzegovina

ABSTRACT
Neural networks are organized in committees to improve the correctness of the decisions created by artificial neural networks (ANN's). In the classification of human chromosomes, it is accustomed to use multilayer perceptrons with multiple (22-24) outputs. Because of the huge number of synaptic weights to be tuned, these classifiers cannot go beyond a level of 92% overall correctness. In this study we represent a special organized committee of 462 simple perceptrons to improve the rate of correct classification of 22 types of human chromosomes. Each of these simple perceptrons is trained to distinguish between two types of chromosomes. When a new data is entered, the votes of these 462 simple perceptrons and additional 22 dummy perceptrons create a decision matrix of the size 22×22. By a special assembling of these votes we get a higher rate of correct classification of 22 types of human chromosomes.

General Terms
Pattern Recognition, Classification

Keywords
Classification of human chromosomes, perceptrons, committee machines, image profile, metaphase

1. INTRODUCTION
Chromosomes are tightly coiled microscopic rod-like structures of DNA and protein that are found in the nuclei of Eukaryotic cells. Each chromosome consists of two sets of genetic material connected at the centromere. Human cells contain 46 chromosomes, arranged into 23 pairs, 22 for body chromosomes and one pair for sexual chromosomes in the nuclei. Technicians in the cytogenetics labs analyze human chromosomes in cells to determine possible genetic diseases. They count the number of chromosomes, studying the banding patterns in each one. This is a cumbersome process completed manually by a microscope, and the naked eye.

There are a huge number of researches to replace technicians in the cytogenetic labs with software and computers. Some of them use image processing techniques for segmentation of human cells photographs in metaphase, and artificial neural networks in chromosome classification and pairing.

Artificial intelligence and machine learning methods have been widely used techniques to improve the performance of the computer-assisted chromosome detection and classification systems. Because of their capability to recognize objects based on incomplete or partial information, Artificial Neural Network (ANN) is the most popular tool. Its architecture is simple and training process is simple [1], [2]. A large number of different ANNs have been tested in classification of human chromosomes, which include supervised neural network architecture. Multi-layer neural networks are studied in [3-9] and Hopfield network [10]; fuzzy neural techniques [11-15]; and unsupervised architecture of nonlinear maps [16], self-organizing feature maps [17] and mutual information maximization based training method [18].

In chromosome classification and pairing, back propagation training method is used to train ANNs. In multi-layer feed-forward ANNs, the number of output neurons is equal to the number of human chromosome types. The number of input neurons is equal to the dimension of the input data, which is the number of features used for classification. Often, principal components replace real features to reduce the dimension of the input data, and hence the computation cost. The number of hidden layers, number of hidden neurons, steepness of the activation function, learning rate, momentum factor, number of learning iterations and upper bound of training error are chosen by the user experimentally.

While the proper choice of these parameters is important for the performance and robustness of an ANN used in chromosome classification [20], studies indicated that ANN performance was slightly lower than that obtained using simpler statistical methods [21-24]. Unnecessary complexity of the ANN architecture, and overtraining of ANNs dramatically reduces the robustness of the ANN in chromosome classification. One study [1] obtained 0% error rate in the training data set, using multilayer perception based ANN, but 24.2% error rate in the testing data set. To increase ANN performance, another study showed that by reducing the complexity of an ANN, its testing accuracy can be increased from 75.8% to 88.3% [6].

One of the other more sophisticated neural networks proposed and tested in this area is a fuzzy Hopfield neural network. It holds fuzzy clustering capability and learning mechanism of acquiring knowledge about the human chromosomes from noisy inputs. In a test involving 100 human chromosomes Ruan [10] succeeded to achieve a very low non-identification rate of 3.33%.

In this paper, using a special ensemble of simple multilayer perceptrons, we reached the same score as Ruan [10].

2. PROBLEM FORMULATION

In the cytogenetics, lab technicians examine the human cell. They take several cell samples from the patient and prepare the slide for a very high magnifying power microscope. They identify chromosomes as pairs and study the banding patterns. This process might be repeated for several cells in the sample. This is a time consuming cumbersome process and needs the replacement with software, and computers.

We made the phase of image processing, the extraction of features used in the classification phase as a subject of another paper.

3. DATA SET

In our study, we use the Copenhagen Chromosome Data Set [19], which contains data for 4400 chromosomes, 200 from each of 22 types. For chromosome classification, we use the p-arm length, total length and gray level profiles of length 26 for the shortest, and 100 for the longest chromosome. We completed shorter chromosomes to the length 100 by padding zeros. We also realized that principal components compresses the data enormously and economizes the computation cost. Experimenting with several numbers of principal components, we found out that 10 principal components are optimal. At the end the length of the feature vector for each chromosome is 12, containing the p-arm length, total length of the chromosome and 10 principal components.

4. METHODOLOGY

4.1 Architecture of ANN

We represent the network consisting of 12 inputs x[i], i=1,…,12, 12 neurons in the hidden layer and one neuron in the output layer as shown in the Fig 1. A special organized committee of 462 simple perceptrons is used to improve the rate of correct classification of 22 types of human chromosomes. Each of these simple perceptrons is trained to distinguish between two types of chromosomes. These multilayer perceptrons use Back-Propagation algorithm.

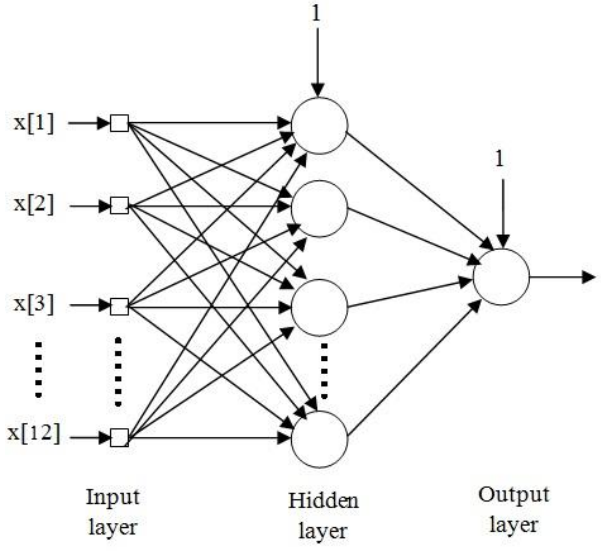

Fig 1: Neural network architecture for a simple multilayer perceptron

4.2 Assembling votes

When new data enters the network, the perceptron trained to distinguish type *i* and type *j*, creates the output 1, if this new data is of type *i*, and creates the output -1, if this new data is of type *j*, where i = j = 1…22. If the new data is neither of type i, or type j, it also creates an output either 1, or -1. Hence the votes of these 462 simple perceptrons and additional 22 dummy perceptrons create a decision matrix of the size 22×22. By a special assembling of these votes we get a higher rate of correct classification of 22 types of human chromosomes.

Let us name a simple perceptron which is trained to cluster a mixture of chromosomes of type i, and j by P(i, j). After the training stage, assume a new data of unknown type entered into all of the perceptrons of the committee. If the new data belongs to a chromosome of type k, then almost all perceptrons $P(j,k), j = 1,2,\dots,k-1,k+1,\dots,22$ will give an output of 1, while almost all perceptrons $P(k,i), i = 1,2,\dots,k-1,k+1,\dots,22$ will give an output of -1. Other perceptrons will also give an output of 1 or -1. We will add dummy perceptrons $P(j,j), j = 1,2,\dots,22$ which always give output 0 to the committee.

When the votes of 22×22 perceptrons arranged as a 22×22 dimensional matrix, the pair of k-th column vector, and k-th row vector is the closest to the expected pair of vectors

$$\{[1,1,\dots,0,1,\dots,1],[-1,-1,\dots,0,-1,\dots,-1]\}$$

in which zero appears in the k-th position in the two vectors.

For example, when a new data that belongs to a chromosome of type 10 is entered into the committee of 22×22 simple multi layer perceptrons, then in our example 19 of all 21 perceptrons $P(i,10), i = 1,2,\dots,k-1,k+1,\dots,22$ gave an output of 1, while all perceptrons $P(10,j), j = 1,2,\dots,k-1,k+1,\dots,22$ gave an output of -1. In this, no other {jth row, jth column} pairs can compete with the {10th row, 10th column} pair in the similarity of the expected ideal pair of

$$\{[1,1,\dots,0,1,\dots,1],[-1,-1,\dots,0,-1,\dots,-1]\}$$

in which zero appears in the 10th positions of the two vectors (see Fig 2). The Euclidean distances of crosses to ideal cross are called the ranks of crosses. If we represent ranks with gray levels, competing crosses are as shown in Figure 3.

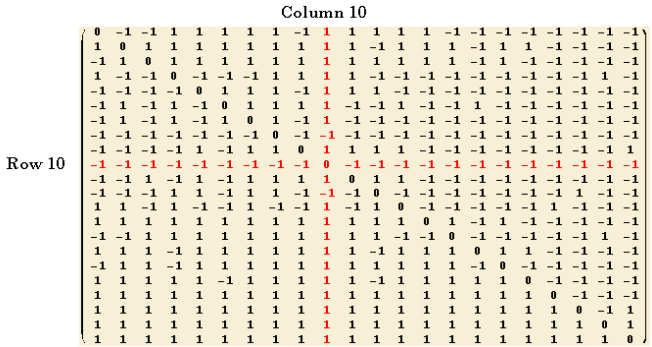

Fig. 2: Example of the decision matrix

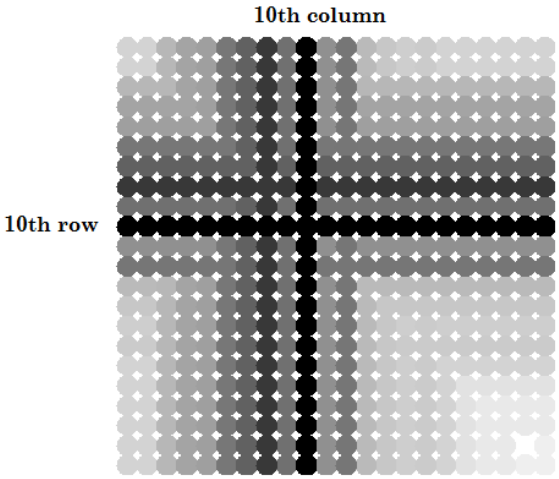

Fig3: Competing crosses. The darkest cross is the one which consist of 10th row and 10th column that wins the competition. The nearest competitor to type 10 is type 8.

5. RESULTS

During the training of 462 simple multilayer perceptrons, it is possible to complete training with zero error. But this leads to overtraining that causes lower rates in testing. From each chromosome type 50 random samples are chosen for training. The same numbers of random samples are also chosen for validation and testing. We have seen that it is possible to go over 97% correct classification rates with this special committee of perceptrons.

Table 1. Correct classification rates during training and testing. Using a validation data set, the overtraining is prevented

Chromosome Type	Correct Classification Rate (%)	
	Training	Testing
1	100	99
2	97	100
3	96	99
4	95	91
5	93	89
6	96	99
7	93	94
8	94	89
9	93	90
10	95	92
11	98	96
12	91	97
13	97	99
14	89	98
15	99	95
16	99	100
17	99	99
18	94	99
19	89	97
20	90	90
21	97	99
22	96	98
Average	95	95.86

6. CONCLUSION

In this study we presented a special organized committee of 462 simple perceptrons used to improve the rate of correct classification of 22 types of human chromosomes. Each of these simple perceptrons is trained to distinguish between two types of chromosomes. When a new data is entered, the votes of these 462 simple perceptrons and additional 22 dummy perceptrons create a decision matrix of the size 22×22. By a special assembling of these votes we get a higher rate of correct classification of 22 types of human chromosomes, with an average of 95.86% correct classification when tested on Copenhagen Chromosome Dataset.

7. REFERENCES

[1] Mitchell T. M., "Machine Learning", McGraw-Hill, pp. 81-116, 1997

[2] Haykin S., "Neural Networks: A Comprehensive Foundation", Second Edition, Prentice-Hall, Inc., Simon & Schuster, 1999

[3] Lu, Y. and Y. Ya, "An expert system for banded chromosomes recognition", Proceedings of the Annual International Conference of the IEEE Engineering in Engineering Medicine and Biology Society (IEMBS), Seattle, WA, pp. 1789-1790, 9-12 November, 1989

[4] Wu, Q., P. Suetens and A. Oosterlinck, "Chromosome classification using a multi-layer perception neural network", Proceedings of the 12th Annual International Conference of IEEE Engineering in Medicine and Biology Society, pp. 1453-1454, Vol. 12, No. 3, 1990

[5] Errington P.A., Graham J., "Classification of Chromosomes using a Combination of Neural Networks", IEEE International Conference on Neural Networks (ICNN), San Francisco, CA, pp. 1236-1241, Vol. 3, 28th March – 1st April, 1993

[6] Delshadpour S., "Reduced size multilayer perceptron neural network for human chromosome classification", Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS), pp. 2249-2252, Vol3, 17-21 September, 2003

[7] El Emary I. M. M., "On the Application of Artificial Neural Networks in Analyzing and Classifying the Human Chromosomes", Journal of Computer Science 2 (1): 72-75, 2006

[8] Wang X., Zheng B., Li S., Mulvihill J. J., Wood M. C., and Liu H., "Automated Classification of Metaphase Chromosomes: Optimization of an Adaptive Computerized Scheme", Journal of Biomedical Informatics, 42 (1): 22–31, 2009

[9] Can, M. and Gagula-Palalic S., "Application of Ensemble Machines of Neural Networks to Chromosome Classification", South East Journal of Soft Computing (SEJSC), Vol. 1, No. 2, pp. 31-35, 2012

[10] Ruan X., "A classifier with the fuzzy Hopfield network for human chromosomes", Proceedings of the 3rd World Congress on Intelligent Control and Automation, Hefei, Vol. 2, pp. 1159-1164, 28 Jun – 2 Jul, 2000

[11] Ruspini E., "A Fast Method for Probabilistic and Fuzzy Cluster Analysis using Association Measures", Proceedings of the 6th International Conference on System Sciences, Hawaii, pp. 56–58, 1973

[12] Ramstein G., Bernadet M., Kangoud A. and Barba D., "A rule-based image analysis system for chromosome classification", Proceedings of the 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS), Paris, Vol. 3, pp. 926-927, 29 Oct – 1 Nov, 1992

[13] Keller J. M., Gader P., Sjahputera O. and Caldwell C. W., "A fuzzy logic rule based system for chromosome recognition", Proceedings of the 8th IEEE Symposium on Computer-Based Medical Systems (CMBS), Lubbock, TX, pp. 125-132, 9 – 10 Jun, 1995

[14] Sjahputera O. and Keller J. M., "Evolution of a fuzzy rule-based system for automatic chromosome recognition", IEEE International Fuzzy Systems Conference Proceedings, Seoul, South Korea, pp. 129-134,22 – 25 Aug, 1999

[15] Sarosa M., Ahmad A. S., Riyanto B. and Noer A.S., "Optimization of Neuro-Fuzzy System Using Genetic Algorithm for Chromosome Classification", ITB Journal of Information and Communication Technology, Vol. 1, No. 1, pp. 56-69, 2007

[16] Lerner B., Guterman H., Aladjem M. and Dinstein I., "A comparative study of neural network based feature extraction paradigms", Pattern Regognition Letters 20, pp. 7-14, 1999

[17] Kyan M. J., Guan L., Amison M.R. and Cogswell C.J., "Feature extraction of chromosomes from 3D confocal microscope images", IEEE Transactions on Biomedical Engineering, Vol. 48, No. 11, November, 2001

[18] Mousavi P., Fels S. S., Ward R.K. and Lansdorp P.M., "Classification of homologous human chromosomes using mutual information maximization", International Conference on Image Processing, Thessaloniki, pp. 845-848, Vol. 2, 7 – 10 Oct, 2001

[19] Lundsteen, C., Phillip J. and Granum E., "Quantitative analysis of 6985 digitized trypsin G-banded human metaphase chromosomes", Clinical Genetics, 18, pp. 355-370, 1980

[20] Cho J., "Chromosome classification using back propagation neural networks", IEEE Engineering in Medicine and Biology, Vol. 19, pp. 28-33, 2000

[21] Granum E. and Thomason M.G., "Automatically Inferred Markov Network Models for Classification of Chromosomal Band Pattern Structures", Cytometry 11, pp. 26-39, 1990

[22] Sweeney W.P. and M.T. Musavi, "Application of neural networks for chromosome classification", Proceedings of the 15th Annual International Conference of IEEE Engineering in Medicine and Biology Society, pp: 239-240, 1993

[23] Conroy J.M., Tamara G.K., O'Leary D.P. and O'Leary, T.J., "Chromosome Identification Using Hidden Markov Models: Comparison with Neural Networks, Singular Value Decomposition, Principal Components Analysis, and Fisher Discriminant Analysis", Center for Computing Sciences, Institute for Defense Analyses, Bowie, Maryland, USA, 2000

[24] Schwartzkopf, W.C., "Maximum Likelihood Techniques for Joint Segmentation-Classification of Multi-spectral Chromosome Images", Doctoral Dissertation, The Faculty of the Graduate School of The University of Texas at Austin, 2002