

Application of artificial neural networks in the simulation with genetic data

Leila Maria Ferreira¹, Janaína de Andrade Silva², Gustavo César Sant'Ana³,
Geraldo Magela de Almeida Cançado⁴, Aluizio Borém⁵, Juliano Lino Ferreira⁶

¹PhD student in Agricultural Statistics and Experimentation, Federal University of Lavras, Brazil,

²PhD in Computational Modeling, Federal University of Juiz de Fora, Brazil,

³Post-Doctor in Bioinformatics, Agronomic Research for Development, France,

⁴Senior Researcher at the Brazilian Agricultural Research Corporation (Embrapa) - Center for Genetic Engineering and Molecular Biology, Campinas-SP, Brazil,

⁵Teacher of the Department of Plant Science, Federal University of Viçosa, Brazil,

⁶Researcher at the Brazilian Agricultural Research Corporation (Embrapa Livestock South), Bagé-RS, Brazil

Abstract: The objective of this work was the concept of applying artificial neural networks in the study of genetic data, in order to make the identification of the microsatellite markers for a particular species of plant to be analyzed more efficient. In this study, was used as an experimental model the data generated for 26 grapevine genotypes were divided into the following populations: *Vitis vinifera*; North American varieties; and interspecific hybrid of rootstocks. After the network training was carried out, an error rate of 0.0003460 was obtained, concluding that the network was able to learn according to the type of data used, even when these data are small.

Keywords: Network, Genotype, Grapevine

Date of Submission: 27-12-2017

Date of acceptance: 09-01-2018

I. INTRODUCTION

From the molecular point of view, a genetic marker (or locus marker) serves to identify a site or region of a chromosome. An ideal genetic marker should present a series of attributes: high polymorphism, good reproducibility, detection of large numbers of unlinked loci, and simple inheritance. However, the simplicity and low costs of the method are determining factors for the routine application of a molecular marker. Molecular markers are facilitating studies of genetics, taxonomy and plant evolution, providing a significant advance in scientific knowledge. The main implications of this advance in knowledge are reflected in the power, precision and speed in the manipulation of genetic variability. Thus, plant breeding may benefit in various ways with the application of molecular markers.

Recently, a series of statistical analyzes (Principal coordinate analysis - PCoA, Structure, UPGMA, Closest neighbor, Statistical summaries as content of polymorphic information - PIC, allelic diversity, probability of identity - PI and exclusion - PE, allele number, frequency of alleles, among others) has been routinely and widely applied in the evaluation of data obtained with molecular markers. With the advancement of the multidisciplinary, new opportunities are created for the use of neural networks, including the evaluation of data obtained by molecular markers. Artificial neural networks have been gaining more and more visibility in the studies associated with genetic data, showing an efficient technique of analysis. Among the several works in this area are: Khan et al. (2001), Kan et al. (2004), Lancashire et al. (2010), Narayanan et al. (2004), Coppedè et al. (2010), Ornella and Tapia (2010), Genoud et al. (2009), Cajas et al. (2009), Camacho et al. (2012), Pandolfi et al. (2001), Khoshroo et al. (2014), Pan et al. (2013) and Silva et al. (2014).

According to Crestan (2017) there are several definitions for neural networks, also called neurocomputers, connection networks or parallel distributed processors. "A neural network is a massively distributed parallel processor being made up of simple processing units, which have the natural propensity to store experimental knowledge and make it available for use" (Haykin, 1999).

A neural network resembles the human brain in two respects:

- Knowledge is acquired by the network based on its environment through a learning process;
- Connecting forces between neurons, known as synaptic weights, are used to store the acquired knowledge.

With a focus on artificial neural networks, one of the most fascinating research areas today is the simulation of a human being's cognitive abilities. Machines are designed to exhibit intelligent behavior, as if they were human reactions. The intelligence of the human being is the most advanced within the universe of living organisms and the place that welcomes this intelligence within the human body is the brain. The basic

entities of the brain are the neurons, interconnected in networks which allows the exchange of information between them by means of electric pulses, thus creating biological intelligence. One obvious ambition arising from this premise is the attempt to imitate the structure and functioning of the brain in an artificial environment. This means that the research attempts to understand and map the functioning of neuron-resident intelligence and to transfer this knowledge to an artificial structure, for example a combination of hardware and software, thus transforming biological neural networks into artificial neural networks (Rauber, 2017).

An artificial neural network is formed by the combination of several artificial neurons. The inputs can be connected in many neurons, resulting in a series of outputs, where each neuron represents an output. Comparing with the biological system, these connections represent the contact of the dendrites with other neurons, thus forming the synapses. Connections make the output signal from one neuron into an input signal from another, or else direct the output signal to the outside world.

The main benefits of neural networks are: a) generalization, which refers to the fact that the neural network produces adequate outputs for inputs that were not present during the training; b) computational power, through its massively parallel structure; ability to learn.

The objective of this work was the use of artificial neural networks in the learning of genetic data from microsatellite markers.

II. MATERIAL AND METHODS

To simulate the network, the Scilab 5.2.2 program was used, with a total of 364 data obtained from the combination of 7 microsatellite markers (VVMD7, VVMD25, VVMD5, VVMD27, VVMD31, VVS2 and VZag62) commonly used in grapevines (Schuk et al., 2009) in 26 different genotypes, from the germplasm bank of the EPAMIG - Grape and Wine Technology Center, Caldas-MG, belonging to the genus *Vitis* (Table 1), whose genome is diploid, that is, with one an allele of origin maternal and other of paternal origin.

The data used were as follows:

- 14 input data;
- 2 neurons in the middle layer;
- 3 neurons in the output layer referring to the number of classes;
- 0.3 rate of learning;
- 0.4 momentum.

III. RESULTS AND DISCUSSION

The data obtained by the simulation demonstrated the efficiency of the use of the microsatellite markers to train the neural network. Even though the simulation was applied with few data for the training of the network, its learning capacity was considered high, reaching an error value close to zero (0.0003460) (Figure 1). It is concluded that the ideal simulation should have at least 10 thousand data (combination of number of markers multiplied by the number of genotypes) to enable the network to present more accurate results. However, the results obtained with the universe of applied data demonstrate that this work, even in an initial phase, was efficient. A bank larger microsatellite database should be tested in order to improve network accuracy, allowing for more accurate predictions of genetic behavior via markers.

Table 1. Grapevine genotypes used in this study.

Cup				
ID ^a	Name	Origin	Species	Genealogy
1	Moscato	Italy	Hybrids	Couderc 13 x July Muscat
2	Chardonnay	France	<i>V. vinifera</i>	Pinot Noir x Gouais Blanc
3	Syrah	France	<i>V. vinifera</i>	Dureza x Mondeuse
	Cabernet			
4	Sauvignon	France	<i>V. vinifera</i>	Cabernet Franc x Sauvignon Blanc
5	Merlot	France	<i>V. vinifera</i>	Magdeleine Noire des Charentes x Cabernet Franc
6	Cabernet Franc	France	<i>V. vinifera</i>	?
7	Sauvignon Blanc	France	<i>V. vinifera</i>	Traminer x ?
8	Niágara Rosada	Brazil	<i>V. labrusca</i>	Mutation of Niágara Branca
9	Niágara Branca	USA	<i>V. labrusca</i>	Concord (<i>V. labrusca</i>) x Cassady (<i>V. vinifera</i> x <i>Vitis labrusca</i>)
10	Bordô	USA	Hybrids	Isabella x <i>V. labrusca</i>
11	Violeta	Brazil	Hybrids	BRS Rúbea x IAC 1398-21
Rootstock				

ID ^a	Name	Origin	Species	Genealogy
12	Gravesac	France	Hybrids	161-49 x 3309
13	Kober 5BB	Austria	Hybrids	<i>V. berlandieri</i> x <i>V. riparia</i>
14	Rupestris du Lot	France	<i>Vitis</i>	<i>V. rupestris</i>
15	101-14	France	<i>rupestris</i>	<i>V. riparia</i> x <i>V. rupestris</i>
16	R99	France	Hybrids	<i>V. berlandieri</i> x <i>V. rupestris</i> (<i>Berlandieri Las Sorres</i> x <i>Rupestris du Lot</i>)
17	420 A	Italy	Hybrids	<i>V. berlandieri</i> x <i>V. riparia</i>
18	1045 - Paulsen	France	Hybrids	<i>V. berlandieri</i> x (<i>V. rupestris</i> x <i>V. Vinifera</i>)
19	161-49	France	Hybrids	<i>V. riparia</i> x <i>V. berlandieri</i>
20	R110	USA	Hybrids	<i>V. berlandieri</i> x <i>V. rupestris</i>
21	3309	Italy	Hybrids	<i>V. riparia</i> x <i>V. rupestris</i>
22	1103 Paulsen	France	Hybrids	<i>V. rupestris</i> x <i>V. berlandieri</i>
23	SO4	Germany	Hybrids	<i>V. berlandieri</i> x <i>V. riparia</i>
24	Traviú	France	Hybrids	<i>V. riparia</i> x (<i>V. rupestris</i> x <i>V. cordifolia</i>)
25	IAC766	Brazil	Hybrids	106-6 x <i>V. caribea</i>
26	IAC572	Brazil	Hybrids	[(<i>V. tiliaefolia</i> x <i>V. riparia</i> x <i>V. rupestris</i> ('101-14 Mgt'))]

ID^a – identification

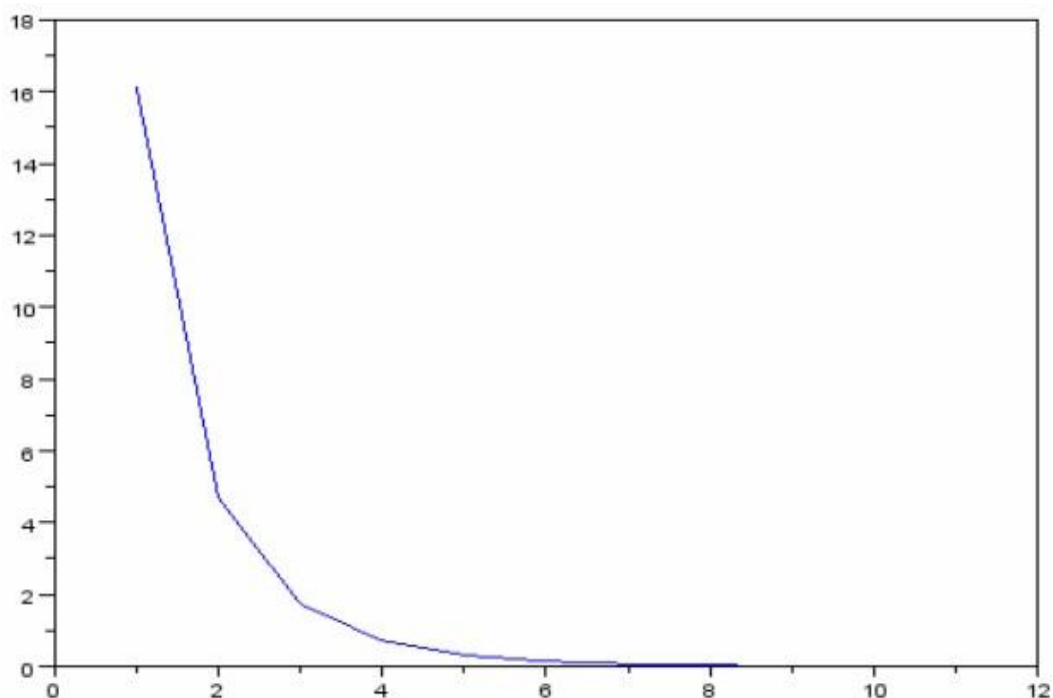


Figure 1. Neural network training error. Error magnitude: 0.0003460.

Table 2. Errors associated after each cycle of operation of the Scilab program 5.2.2.

Column	Error
1	16.121236
2	4.6892856
3	1.7334808
4	0.7154670
5	0.3157074
6	0.1450748
7	0.0677903
8	0.0313731
9	0.0138964
10	0.0055817
11	0.0018215

ACKNOWLEDGEMENTS

The authors would like to thank the financial support and scholarships granted by FAPEMIG, CAPES, CNPq, FINEP and EMBRAPA.

REFERENCES

- [1]. Cajka, T.; Hajslova, J.; Pudil, F.; Riddellova, K. Traceability of honey origin based on volatiles pattern processing by artificial neural networks. *Journal of Chromatography A*, 2009, 1216: 1458-1462.
- [2]. Camacho, J. M. G.; Campos, G.; Pérez, P.; Gianola, D.; Cairns J. E.; Mahuku, G.; Babu, R.; Crossa, J. Genome-enabled prediction of genetic values using radial basis function neural networks. *Theoretical and Applied Genetics*, 2012, 125: 759-771.
- [3]. Coppedè, F.; Grossi, E.; Migheli, F.; Migliore, L. Polymorphisms in folate-metabolizing genes, chromosome damage, and risk of Down syndrome in Italian women: identification of key factors using artificial neural networks. *BMC Medical Genomics*, 2010, 3: 1-10.
- [4]. Crestan, G; Araki, L. E.; Tolosa, T. A. G.; Assis, W. O.; Júnior, W. C. S. L. Reconhecimento de Trajetória com redes neurais, 2017. Disponível: <http://maua.br/files/artigos/reconhecimento-de-trajetorias-com-redes-neurais.pdf>. Acessado: nov, 2017.
- [5]. Genoud, T.; Cruz, M. B. T. S.; Métraux, J. P. Numeric Simulation of Plant Signaling Networks. *Plant Physiology*, 2001, 126: 1430-1437.
- [6]. Haykin, S. *Neural Networks, a Comprehensive Foundation*. Second Edition, Pearson Education, Singapore, India, 1999, 823p.
- [7]. Heslot, N.; Yang, H. P.; Sorrells, M. E.; Jannink, J. L. Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Science*, 2012, 52: 146-160.
- [8]. Kan, T.; Shimada, Y.; Sato, F.; Ito, T.; Kondo, K.; Watanabe, G.; Maeda, M.; Yamasaki, S.; Meltzer, S. J.; Imamura, M. Prediction of Lymph Node Metastasis with Use of Artificial Neural Networks Based on Gene Expression Profiles in Esophageal Squamous Cell Carcinoma. *Annals of Surgical Oncology*, 2004, 11:1070-1078.
- [9]. Khan, J.; Wei, J. S.; Ringnér, M.; Saal, L. H.; Ladayni, M.; Westermann, F.; Berthold, M.; Schwab, M.; Antonescu, C. R.; Peterson, C.; Meltzer, P. S. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, 2001, 7: 673-679.
- [10]. Khoshroo, A.; Arefi, A.; Masoumiasl, A.; Jowkar, G. H. Classification of Wheat Cultivars Using Image Processing and Artificial Neural Networks. *Agricultural Communications*, 2014, 2: 17-22.
- [11]. Lancashire, L. J.; Powe, D. G.; Filho, J. S. R.; Rakha, E.; Lemetre, C.; Weigelt, B.; Fatah, T. M. A.; Green, A. R.; Mukta, R.; Blamey, R.; Paish, E. C.; Rees, R. C.; Ellis, I. O.; Ball, G. R. A validated gene expression profile for detecting clinical outcome in breast cancer using artificial neural networks. *Breast Cancer Res Treat*, 2010, 120: 83-93.
- [12]. Narayanan, A.; Keedwell, E. C.; Gamalielsson, J.; Tatineni, S. Single-layer artificial neural networks for gene expression analysis. *Neurocomputing*, 2004, 61: 217-240.
- [13]. Ormella, L.; Tapia, E. Supervised machine learning and heterotic classification of maize (*Zea mays* L.) using molecular marker data. *Computers and Electronics in Agriculture*, 2010, 74: 250-257.
- [14]. Pan, Y.; Bradley, G.; Pyke, K.; Ball, G.; Lu C.; Fray, R.; Marshall, A.; Jayasuta, S.; Baxter, C.; Wijk, R.; Boyden, L.; Cade, R.; Chapman, N. H.; Fraser, P. D.; Hodgman, C.; Seymour, G. B. Network Inference Analysis Identifies an APRR2-Like Gene Linked to Pigment Accumulation in Tomato and Pepper Fruits. *Plant Physiology*, 2013, 161: 1476-1485.
- [15]. Pandolfi, C.; Mugnai, S.; Azzarello, E.; Bergamasco, S.; Masi, E.; Mancuso, S. Artificial neural networks as a tool for plant identification: a case study on Vietnamese tea accessions. *Euphytica*, 2009, 166: 411-421.
- [16]. Rauber, T. W. *Redes Neurais Artificiais*, 2017. Disponível: <https://inf.ufes.br/~thomas/pubs/eri98.pdf>. Acessado: nov, 2017.
- [17]. Schuck, M. R.; Moreira, F. M.; Guerra, M. P.; Voltolini, J. A.; Grando, M. S.; Silva, A. L. Molecular characterization of grapevine from Santa Catarina, Brazil, using microsatellite markers. *Pesq. agropec. bras.*, 2009, 44: 487-495.
- [18]. Silva, G. N.; Tomaz, R. S.; Sant'Anna, I. C.; Nascimento, M.; Bhering, L. L.; Cruz, C. D. Neural networks for predicting breeding values and genetic gains. *Sci. Agric.*, 2014, 71: 494-498.

Leila Maria Ferreira "Application of artificial neural networks in the simulation with genetic data." *International Journal of Engineering Inventions (IJEI)*, vol. 6, no. 12, 2017, pp. 43-46.