

Poster B-60

Classification of snake venom-neutralizing effects of medicinal plants via artificial neural networks



Authors:

Gabriela Félix dos Santos (*Department of Physics and Mathematics of Ribeirão Preto - University of São Paulo*)

Renato Tinós (*Department of Physics and Mathematics of Ribeirão Preto - University of São Paulo*)

Silvana Giuliatti (*Faculty of Medicine of Ribeirão Preto - University of São Paulo*)

Short Abstract: Artificial Neural Networks (ANNs) are employed to map relationship between amino acid sequences of snake venoms and anti-venom medicinal plants. The inputs of the ANN are obtained by encoding the amino acid sequences by two different methods. The validation results indicate that 75% of the patterns were correctly classified.

Long Abstract:

Envenomation by snake bites is considered a serious problem of public health, not only in Brazil, but in all Latin America [SOARES et al., 2004]. In Brazil, many accidents are caused by serpents of the genus *Bothrops* and *Crotalus*, that include the snakes popularly known as jararaca and rattlesnake. Frequently, envenomation by snake bites is treated by the administration of antiophidic serum. Despite the success of this treatment, it is important to search for new venom inhibitors to complement the action of antivenoms, e. g., for neutralizing local tissue damage.

In the Brazilian popular medicine, several plants are employed for the treatment of ophidian envenomation, in spite of only a few having their anti-ophidian activity investigated by scientific experiments [SOARES et al., 2004]. The objective of this project is to help the investigation of the relations of inhibition between medicinal plants and animals' venoms. In this work, Artificial Neural Networks (ANNs) are employed to map the relations between amino acid sequences of snake venoms and medicinal plants with action anti-venom. ANNs are a practical mechanism largely used to map unknown functions using inputs and desired outputs of a training set [HAYKIN, 2001]. The investigation of the mapping performed by the ANN between sequences of venoms and plants can help the scientists to recognize the sites in the protein sequences where the interaction with substances of the medicinal plants occurs, and to study the anti-venom activity of new plants.

Artificial Neural Networks (ANNs) have been successfully applied to different problems in Bioinformatic, e. g. analysis of sequences of DNA, RNA, and proteins, prediction of molecular structures, and alignment of sequences [BALDI & BRUNNAK, 1998]. The ANN used here is an MLP (multi-layer perceptron). An ANN is composed of basic units of processing, the artificial neurons. In an MLP, the neurons are distributed in one or more hidden layers, and one output layer. The direct use of the protein sequences as inputs of the MLP is not practical as the sequences have different length. In order to solve this problem, two techniques have been used: the 2-gram or 2-tuples encoding method and the 6-letter

exchange group encoding method [WANG et al, 2001]. The 2-gram or 2-tuples encoding consists in extracting every pattern of two consecutive amino acids residues in a protein sequence and to use as inputs of the MLP the sum of the occurrences of each one of them. Thus, each sequence will have 400 attributes (20²), since every combination of two amino acids is employed.

The 6-letter exchange group encoding method is based on the 2-gram encoding method, however it takes in consideration exchanges that occur during the evolution. The amino acids are divided in 6 groups of exchanges based on the substitution matrix PAM and, then, 36 attributes (6²) are used as inputs of the MLP.

In this work, the software WEKA with the 10-fold cross validation method was used to train and validate the MLP. In the experiments, 93 patterns of two classes were employed. The classes considered in the experiments are related to medicinal plants which anti-venom actions have already been scientifically investigated [SOARES et al., 2004]. The classes considered in this paper correspond to the plants *Casearia sylvestris* and *Eclipta prostrata*. The attributes used as inputs of the MLP were obtained by applying the 2-gram encoding or the 6-letter exchange group encoding is the amino acid sequences of the serpents: *B. alternatus*, *B. atrox*, *C.d. cascavella*, *B. insularis*, *B.jararaca*, *B. jararacussu*, *B. moojeni*, *L. muta*, *B. neuwiedi*, *B. pirajai*, *C. rhodostoma*, and *C.d. terrificus*. The protein sequences were obtained from the public database NCBI.

In the first set of experiments, the 2-gram encoding was employed. In the MLP, 50 neurons in only one hidden layer were used, and the learning and momentum rate were respectively set to 0.01 and 0.9 during the training phase. The 10-fold cross validation results obtained after 100 iterations indicated that 63% of the patterns were correctly classified.

In the second set of experiments, the 6-letter exchange group encoding was used. In the MLP, 20 neurons in only one hidden layer were used, and the learning and momentum rate were respectively set to 0.05 and 0.3 during the training phase. The 10-fold cross validation results obtained after 100 iterations indicated that 75% of the patterns were correctly classified.

The results obtained in these experiments indicated that it is possible to relate the amino acid sequences of the venoms to medicinal plants. In future works, new experiments using amino acid sequences of other species will be done, witch can result in better results for the classification task. The system should be applied for sequences of other species, e.g. species of bees, and for plants which anti-venom action were not investigated yet. The relations between venoms and plants obtained in the experiments should help the scientists in the research of inhibition mechanisms of the plant substances.

Acknowledgements

The authors would like to thank CNPq and Fapesp for the financial support to this project.

References

BALDI, P. & BRUNNAK, S. (1998). "Bioinformatics: The Machine Learning Approach".

Cambridge: MIT Press.

HAYKIN, S. (1994). "Neural Networks: A Comprehensive Foundation". New York: Macmillan.

SOARES, A. M. ; JANUÁRIO, A. H ; LOURENÇO, M. V. ; PEREIRA, A. M. S. & PEREIRA, P. S. (2004). "Snake venom-neutralizing effects of Brazilian plants". *Drugs of the Future*, 29 (11).

WANG, J. T. L. ; MA, Q. ; SHASHA, D. & WU, C. H. (2001). "New techniques from extracting features from protein sequences". *IBM Systems Journal*, 40 (2).