



## Enhanced Self-Organizing Map Neural Network for DNA Sequence Classification

PDF (Size: 631KB) PP. 25-33 DOI : 10.4236/iim.2013.51004

### Author(s)

Marghny Mohamed, Abeer A. Al-Mehdhar, Mohamed Bamatraf, Moheb R. Girgis

### ABSTRACT

The artificial neural networks (ANNs), among different soft computing methodologies are widely used to meet the challenges thrown by the main objectives of data mining classification techniques, due to their robust, powerful, distributed, fault tolerant computing and capability to learn in a data-rich environment. ANNs has been used in several fields, showing high performance as classifiers. The problem of dealing with non numerical data is one major obstacle prevents using them with various data sets and several domains. Another problem is their complex structure and how hands to interprets. Self-Organizing Map (SOM) is type of neural systems that can be easily interpreted, but still can't be used with non numerical data directly. This paper presents an enhanced SOM structure to cope with non numerical data. It used DNA sequences as the training dataset. Results show very good performance compared to other classifiers. For better evaluation both micro-array structure and their sequential representation as proteins were targeted as dataset accuracy is measured accordingly.

### KEYWORDS

Bioinformatics; Artificial Neural Networks; Self-Organizing Map; Classification; Sequence Alignment

### Cite this paper

M. Mohamed, A. Al-Mehdhar, M. Bamatraf and M. Girgis, "Enhanced Self-Organizing Map Neural Network for DNA Sequence Classification," *Intelligent Information Management*, Vol. 5 No. 1, 2013, pp. 25-33. doi: 10.4236/iim.2013.51004.

### References

- [1] P. Khandheria and H. R. Garner, " Developing a Modern Web Interface for Database-Driven Bioinformatics Tools," *Engineering in Medicine and Biology Magazine*, Vol. 26, No. 2, 2007, pp. 96-98.
- [2] P. Baldi and S. Brunak, " Bioinformatics the Machine Learning Approach," 2nd Edition, Massachusetts Institute of Technology, Cambridge, 2001.
- [3] S. K. Shukla, S. Rungta and L. K. Sharma, " Self-Organizing Map Based Clustering Approach for Trajectory Data," *International Journal of Computer Trends and Technology (IJCTT)*, Vol. 3, No. 3, 2012, pp. 321-324.
- [4] L. K. Sharma and S. Rungta, " Comparative Study of Data Cluster Analysis for Microarra," *International Journal of Computer Trends and Technology (IJCTT)*, Vol. 3, No. 3, 2012, pp. 387-390.
- [5] S. K. Bhatia and V. S. Dixit, " A Propound Method for the Improvement of Cluster Quality," *International Journal of Computer Science Issues (IJCSI)*, Vol. 9, No. 2, 2012, pp. 216-222.
- [6] M. S. Babu, N. Geethanjali and B. Satyanarayana, " Clustering Approach to Stock Market Prediction," *Advanced Networking and Applications*, Vol. 3, No. 4, 2012, pp. 1281-1291.
- [7] R. Krakovsky and R. Forgac, " Neural Network Approach to Multidimensional Data Classification via Clustering," *IEEE 9th International Symposium on Intelligent Systems and Informatics (SISY)*, Subotica, 8-10 September 2011, pp. 169-174.
- [8] J. Malone, K. McGarry, S. Wermter and C. Bowerman, " Data Mining Using Rule Extraction from Kohonen Self-Organizing Maps," *Neural Computing & Applications*, Vol. 15, No. 1, 2005, pp. 9-17.

- [Open Special Issues](#)
- [Published Special Issues](#)
- [Special Issues Guideline](#)

[IIM Subscription](#)

[Most popular papers in IIM](#)

[About IIM News](#)

[Frequently Asked Questions](#)

[Recommend to Peers](#)

[Recommend to Library](#)

[Contact Us](#)

Downloads: 144,622

Visits: 361,770

[Sponsors >>](#)

- [9] C. Burge and S. Karlin, " Prediction of Complete Gene Structures in Human Genomic DNA," *Journal of Molecular Biology*, Vol. 268, No. 1, 1997, pp. 78-94. doi:10.1006/jmbi.1997.0951
- [10] C. Math' e, M. F. Sagot, T. Schiex and P. Rouz' e, " Current Methods of Gene Prediction, Their Strengths and Weaknesses," *Nucleic Acids Research*, Vol. 30, No. 19, 2002, pp. 4103-4117. doi:10.1093/nar/gkf543
- [11] S. L. Salzberg, M. Pertea, A. L. Delcher, M. J. Gardner, and H. Tettelin, " Interpolated Markov Models for Eukaryotic Gene Finding," *Genomics*, Vol. 59, No. 1, 1999, pp. 24-31. doi:10.1006/geno.1999.5854
- [12] T. Golub, D. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. Mesirov, H. Coller, M. Loh, J. Downing and M. Caligiuri, " Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring," *Science*, Vol. 286, No. 5439, 1999, pp. 531-537. doi:10.1126/science.286.5439.531
- [13] I. Inza, P. Larranaga, R. Blanco and A. J. Cerrrolaza, " Filter versus Wrapper Gene Selection Approaches in DNA Micro-Array Domains," *Artificial Intelligence in Medicine*, Vol. 31, No. 2, 2004, pp. 91-103.
- [14] T. Hastie, R. Tibshirani, M. B. Eisen, A. Alizadeh, R. Levy, L. Staudt, W. C. Chan, D. Botstein and P. Brown, " Gene Shaving as a Method for Identifying Distinct Sets of Genes with Similar Expression Patterns," *Genome Biology*, Vol. 1, No. 2, 2000, pp. 3.1-3.21. doi:10.1186/gb-2000-1-2-research0003
- [15] Q. Sheng, Y. Moreau and B. De Moor, " Biclustering Micro-array Data by Gibbs Sampling," *Bioinformatics*, Vol. 19, No. S2, 2003, pp. 196-205. doi:10.1093/bioinformatics/btg1078
- [16] F. Ronquist and J. P. Huelsenbeck, " MRBAYES 3: Bayesian Phylogenetic Inference under Mixed Models," *Bioinformatics*, Vol. 19, No. 12, 2003, pp. 1572-1574. doi:10.1093/bioinformatics/btg180
- [17] P. Larranaga, B. Calvo, R. Santana, C. Bielza, J. Galdiano, I. Inza, J. A. Lozano, R. Arma?nanzas, R. Santafe, A. Perez and V. Robles, " Machine Learning in Bioinformatics," *Briefings in Bioinformatics*, Vol. 7, No. 1, 2006, pp. 86-112. doi:10.1093/bib/bbk007
- [18] C. Iliopoulos, K. Perdikuri, E. Theodoridis, A. Tsakalidis and K. Tsihlias, " Algorithms for Extracting Motifs from Biological Weighted Sequences," *Original Research Article Journal of Discrete Algorithms*, Vol. 5, No. 2, 2007, pp. 229-242. doi:10.1016/j.jda.2006.03.018
- [19] M. Tompa, N. Li, T. L. Bailey, G. M. Church, B. De Moor, E. Eskin, A. V. Favorov, M. C. Frith, Y. Fu, W. J. Kent, V. J. Makeev, A. A. Mironov, W. S. Noble, G. Pavesi, G. Pesole, M. Regnier, N. Simonis, S. Sinha, G. Thijs, J. van Helden, M. Vandenbogaert, Z. Weng, C. Workman, C. Ye and Z. Zhu, " Assessing Computational Tools for the Discovery of Transcription Factor Binding Sites," *Nature Biotechnology*, Vol. 23, No. 1, 2005, pp. 137-144. doi:10.1038/nbt1053
- [20] S. B. Needleman and C. D. Wunsch, " A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins," *Journal of Molecular Biology*, Vol. 48, No. 3, 1970, pp. 443-453. doi:10.1016/0022-2836(70)90057-4
- [21] H. J. B?ckenhauer and D. Bongartz, " *Algorithmic Aspects of Bioinformatics*," Springer, Berlin, 2007.
- [22] A. Polanski and M. Kimmel, " *Bioinformatics*," Springer, Berlin, 2007.
- [23] M. Ishikawa, T. Toya, M. Hoshida, K. Nitta, A. Ogiwara and M. Kanehisa, " Multiple Sequence Alignment by Parallel Simulated Annealing," *Computer Applications in the Biosciences*, Vol. 9, No. 3, 1993, pp. 267-273.
- [24] J. Kim, S. Pramanik and M. J. Chung, " Multiple Sequence Alignment Using Simulated Annealing," *Computer Applications in the Biosciences*, Vol. 10, No. 4, 1994, pp. 419-426.
- [25] J. M. Keith, P. Adams, D. Bryant, D. P. Kroese, K. R. Mitchelson, D. A. E. Cochran and G. H. Lala, " A Simulated Annealing Algorithm for Finding Consensus Sequences," *Bioinformatics*, Vol. 18, No. 11, 2002, pp. 1494-1499. doi:10.1093/bioinformatics/18.11.1494