ELSEVIER

Journal of Microbiological Methods 43 (2000) 1–2

Journal of Microbiological Methods

www.elsevier.com/locate/jmicmeth

Editorial

Neural computing in microbiology

Jonas S. Almeida[a,b,*], Peter A. Noble (Guest Editors)[b]

[a] ITQB and FCT/Univ. Nova de Lisboa, R Quinta Grande 6, 2780-156 Oeiras, Portugal
[b] Belle Baruch Institute for Marine Biology and Coastal Research, University of South Carolina, Columbia, SC 29208, USA

Artificial neural networks (NNs) are artificial intelligence techniques that are able to identify predictive associations directly from experimental data. The architecture of NNs and the artificial learning algorithms that generate them mimic the information processing and knowledge acquisition of natural nervous systems. NNs are adaptable and can be trained to recognize associations among complex data without requiring any prior assumptions of the mechanistic associations among the data. This property is particularly useful for analyzing biological data because NNs are able to find patterns in fuzzy, noisy, and nonlinear data which elude conventional linear-based statistical methods.

Recently, there has been a tremendous increase in the number of articles in microbiology journals in which NNs have been used to analyze data (Fig. 1A). Many of these articles have used NNs to recognize patterns in nucleotide and protein sequences (Fig. 1B; 24%), while a substantial number of articles have used NNs to classify microbes (or microbial communities) based on specific structural features and/or biochemical properties (Fig. 1B; HPLC, GC, PLFA, Ident./morphol., PMS, 49%). The remaining articles deal with biomass predictions and ecological modelling (Fig. 1B; 11% and 16%, respectively).

The increased application of NNs in microbiological articles may be attributed to the increased availability of ‘‘user-friendly’’ software in the latest versions of general purpose statistical programs such as Statistica and SPSS. Increased utilization of NNs may also be due to the increased availability of personal computers and recent increases in computer processing unit (CPU) speed. Several years ago, NN software required a significant amount of CPU time and mainframe computer resources, making NN analysis prohibitively expensive. Now, personal computers are capable of performing the same analysis.

In 1998, a workshop held at the General Meeting of the American Society of Microbiology (‘‘Neural computing in microbiology’’) dealt with the fundamentals of neural computing and the utility of NNs for examining complex and/or fuzzy microbiological data. Discussions among the participants and speakers revealed the need for articles dealing with fundamentals of neural computing so that microbiologists can easily transition from theoretical aspects of neural computing to practical applications of NNs for their research. Discussion at the workshop also revealed that NNs can be employed to analyze many different types of microbiology data.

This issue of the Journal of Microbiology Methods was organized to address some of the questions raised at this workshop. The Basheer and Hajmeer article provides a foundation in the fundamentals, design, and application of NNs. The focus of the Basheer and Hajmeer article was to familiarize the reader with NNs and to serve as a useful practical guide for those who are interested in learning about

---

*Corresponding author.
E-mail address: jalmeida@itqb.unl.pt (J.S. Almeida).

0167-7012/00/$ – see front matter
PII: S0167-7012(00)00200-1

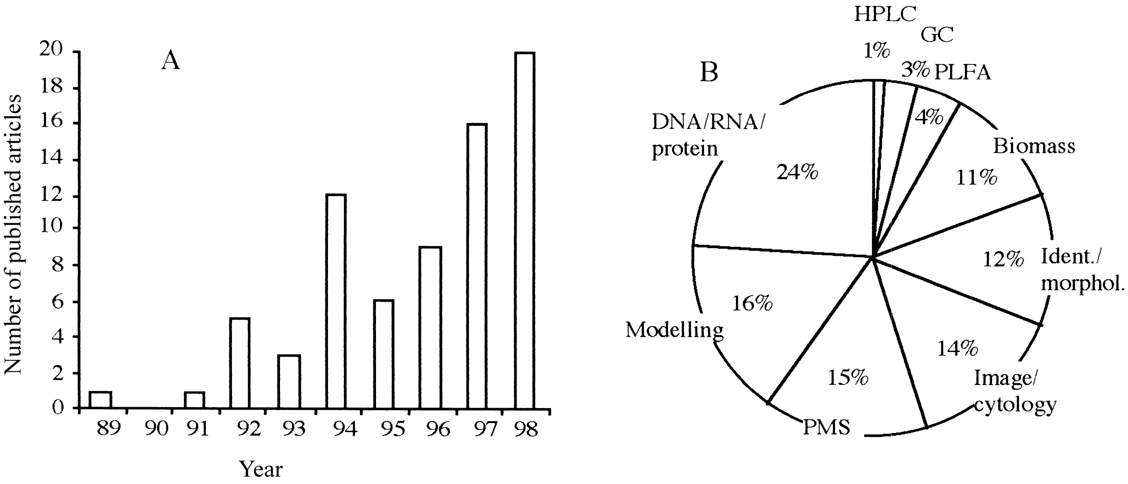

Fig. 1. A, Number of peer-reviewed scientific articles dealing with microbiology and NNs by year. B, Percent of articles by major topic. Categories: DNA/RNA/protein, articles dealing exclusively with predicting properties and recognizing patterns of DNA, RNA or protein molecules; modelling articles dealing with predicting or estimating the effects of microbial products; PMS, articles dealing with interpreting pyrolysis mass spectrometry data; image/cytol. articles dealing with recognizing patterns in microscopic images or in flow cytometry data; ident/morphol., articles dealing with the identification of microbes and/or particular morphologies: biomass, articles dealing with predicting biomass; PLFA, articles dealing with interpreting phospholipid fatty acids data; GC, articles dealing with interpreting gas chromatography data; HPLC, articles dealing with interpreting high performance liquid chromatography data.

NN. Also presented, is an overview of various types of NNs and information on developing NNs for analyzing microbiological data.

The Al-Haddad et al. article examines the effects of training set size and imbalanced training sets on the accuracy of identification of microalgae. When training a NN, it is important to determine effects of training set size and the number of target patterns in the data set because they both influence the ability of the NN to accurately discriminate among the targeted patterns. This article shows that the greater number of target patterns, the lower the accuracy of the NN and that the greater the imbalance in the number of training patterns per target the greater the effect. Strategies to compensate for imbalanced training sets are discussed.

The Giacomini et al. article demonstrates the use of NNs to identify environmental bacteria based on gas-chromatographic and electrophoretic data. The authors used two different types of NNs (supervised and unsupervised) to show that both approaches can be useful for classifying bacteria.

The Wilkins et al. article deals with the limitations of NNs to correctly classify targets by determining the effects of sampling errors and by examining the ability of NNs to classify new patterns not used to train the NN. The authors provide an algorithm to estimate the confidence in which the abundance of a target microbial population is predicted relative to the total populations.

Acknowledgements

PAN and JSA are supported by NSF grant MCB-9802342.