Neural networks and pattern recognition: two steps closer

Robert P.W. Duin, Pattern Recognition Group,
Delft University of Technology, Delft, The Netherlands

The automatic recognition of patterns in sensory data has been studied from the introduction of computers. From the middle of the sixties it was an independent discipline having its own workshops and conferences. Around 1973 a first wave of about 10 books was published, including the classic ones by Fukunaga [1] and Duda and Hart [2]. At that moment the predecessor of the present neural network, the perceptron, had already been heavily studied and abandoned as not useful for practical problems. After 1976 virtually no important new monographs appeared in the field, with as a single, outstanding exception the one by Devijver and Kittler [3].

Just before 1990 the pattern recognition research community was shocked by the enormous success of the revived perceptron: the neural network. A single glance made clear that such machines contradicted the classic design principles for trainable pattern recognizers. Moreover, many of the papers reporting enthusiastically on the neural net successes appeared to be poor studies from a methodological point of view: no real independent test results, no comparison with simple traditional methods, no references to the classical pattern recognition literature. Moreover, it appeared that many of the successes could be obtained as well by a simple technique like the nearest neighbour rule. As a result, neural networks were avoided for many years, preserving the gap from both sides. The heavily extended reprint of Fukunaga in 1990 [5] did not contain any references to neural networks, nor did McLachan's study on discriminant analysis and pattern recognition in 1992 [6].

This regrettable state of affairs has now come to an end. It has become clear to most researchers in the pattern recognition field that artificial neural networks have interesting properties. The neural network community seems to have realized that there was a world before the back-propagation rule. After the earlier works by Pao [4] and Schalkoff [7], more books are now being published bridging the gap. Two recent ones are "*Neural networks for pattern recognition*" by C.M. Bishop [8] and "*Pattern recognition and neural networks*" by B.D. Ripley [9]. From a pattern recognition point of view the books are similar with respect to the topics that are treated. On the basis of a list of 27 possible subjects constructed by this reviewer Ripley scored 13 and Bishop 12 with just a few differences. For missing some of the some recent ones they cannot be blamed: bagging and the support vector classifier are very new developments, of interest to the researcher, but too recent for crystallization in a book. For other issues both authors seem to have no interest because these issues are too fundamental: the no-free-lunch theorems and the compactness hypothesis. This is really a pity as a discussion on these topics might reveal the essential assumptions underlying the applicability of pattern recognizers and neural networks. On the other side of the spectrum both authors disregard the opportunity to write a truly practical book, directly of use to the engineer and applied physicist that have to solve the real world pattern recognition problem: there is no interface with applications in speech recognition, image analysis or computer vision. As a result, classifiers based on the Hausdorff distance or on the edit distance are not discussed. To be fair, the books by Bishop and Ripley share this lack with almost all other pattern recognition books of the past. One may hope, however, that this gap will be bridged someday.

If we consider the differences in the list of topics treated in the two books then we see for example that Ripley treats some nonneural techniques such as belief networks and decision trees while Bishop more extensively discusses neural networks, their optimization and related generalization issues. For instance, he uses much more space for the overtraining phenomenon. It is to be regretted, however, that both books fail to remove clearly the common but wrong belief that it is worthwhile to find the global minimum in the classical neural network training procedure. This is good only for very large training set sizes compared to the number of weights. On the whole it can be stated that Ripley has written a more general book on pattern recognition, treating neural networks as one of the topics. Bishop's book can be better characterized as a neural network introduction placed in a pattern recognition context.

Considering the qualities of the books, it is immediately apparent that the one by Ripley looks better and is more accurately printed. It is a pleasure to browse through it, get caught by a sideline comment or attracted by a discussion. A special feature is the large number of original proofs. This will be mainly of interest to the scientist searching for the fundamentals. An extensive reference list and an author index serve the same purpose. A glossary and examples may make it accessible to students in engineering. However, Bishop serves this purpose better. He has written a text book, introducing techniques, relating them to the theory and explaining their pitfalls. Moreover, a large set of exercises makes it attractive for the teacher to use the book.

In conclusion, the two books should be warmly welcomed by the neural network and pattern recognition communities. Treating almost the same field they have their own style, goal and emphasis. They may also have a somewhat different readership. For instance, Bishop can be recommended to students and engineers in computer science, Ripley to researchers with an interest in statistical pattern recognition. If possible, however, take both!

[1] K. Fukunaga, Introduction to statistical pattern recognition, Academic Press, New York, 1972.

[2] R.O. Duda and P.E. Hart, Pattern classification and scene analysis, John Wiley and Sons, New York, 1973.

[3] P.A. Devijver and J. Kittler, Pattern recognition: A statistical approach, Prentice/Hall, London, 1982.

[4] Yoh-Han Pao, Adaptive pattern recognition and neural networks, Addison-Wesley, Reading, Massachusetts, 1989.

[5] K. Fukunaga, Introduction to statistical pattern recognition, second edition, Academic Press, New York, 1990.

[6] G.J. McLachlan, Discriminant Analysis and Statistical Pattern Recognition, John Wiley and Sons, New York, 1992.

[7] R. Schalkoff, Pattern recognition: Statistical, structural and neural approaches, John Wiley and Sons, New York, 1992.

[8] C.M. Bishop, Neural networks for pattern recognition, Clarendon Press, Oxford, 1995.

[9] B. Ripley, Pattern Recognition and neural networks, Cambridge University Press, Cambridge, 1996.