Comment

S. Amari

First of all, I would like to thank the Editor for giving me an opportunity to present my personal view on this interesting paper connecting the interdisciplinary field of neural networks and statistics. I also congratulate the authors for their excellent job of reviewing this difficult field in a very compact and comprehensive way.

The brain is an enormously complex system in which distributed information is processed in parallel by mutual dynamical interactions of neurons. It is still difficult, and challenging, to understand the mechanisms of the brain. Recently, the importance and effectiveness of brain-style computation has been widely recognized by the name of neural networks. Roughly speaking, there are three different research areas concerning neural networks. One is the experimental area based on physiology and molecular-biology, which is progressing rapidly and steadily. The second area is engineering applications of neural networks inspired by the brain-style computation where information is distributed as analog pattern signals, parallel computations are dominant and learning guarantees flexibility and robustness of computation. This area has opened new practical methods of pattern recognition, control systems, time-series analysis, optimization, memories, etc. The third area is concerned with theoretical (or mathematical) foundations of neurocomputing, which search for the fundamental principles of parallel distributed information systems with learning capabilities. From this standpoint, the actual brain is a biological realization of these principles through a long history of evolution.

Statistics has a close relation with the second applications area of neural networks, as the present authors have so clearly shown (also see Ripley, 1993a). Statistical methodology is indeed a very important tool for analyzing neural networks. On the other hand, neural networks provides statistics with tractable multivariate nonlinear models to be studied further. It also inspires statistical sciences with the notions of learning, self-organization, dynamics, field theory, etc. which statistics has so far paid little attention to. On the other hand, statistical sciences provides one of the crucial methods for constructing theoretical foundations of neurocomputing (e.g., Amari, 1990, 1993a). Without these foundations, it is difficult for neural network technology to take off from the present rather "easy and shallow" technology to a more fundamental one.

Artificial neural networks research has experienced ups and downs; up in the early sixties where the perceptron and the adaline were proposed and again a big up in the middle of the eighties until now. It is said that the dark period was around the seventies where little attention had been paid to ANN and that the Minsky-Papert critique gave rise to this down. However, I believe this prevailing story is merely a myth. We can point out the lack of supporting technology as the background of this fall. Computer technology had developed greatly through the sixties and seventies. Researchers on pattern recognition and artificial intelligence thought that it was easier and more powerful to use symbol processing in modern computers rather than to use neural networks technology. This was true, and information processing technology including artificial intelligence had been constructed successfully upon modern computers. However, hardware technology had further developed in the eighties such that it could support neural parallel computation. It was not a dream to construct neurochips or even neurocomputers. There are, of course, many other intellectual reasons to support the resuscitation in the eighties.

In the seventies, most researchers did not think that engineering applications of neural networks were realizable. The background technology was not yet matured at the time. However, it was not a dark period in theoretical study because many of the ideas were proposed in the "dark period" that were rediscovered or developed further to be the fundamental methods supporting the neural network methods today.

For example, the generalized delta rule for a multilayer perceptron was proposed in 1967 (Amari, 1967) where analog neurons were used and the stochastic descent algorithm was applied. The idea was also introduced in a Russian book (Tsypkin, 1973). I believe that there were not a few researchers who knew the idea at that time. It was

S. Amari is Professor, Department of Mathematical Engineering, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan.