Connectionists and Statisticians, Friends or Foes?

Arthur Flexer

The Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A−1010 Vienna, Austria
arthur@ai.univie.ac.at

Abstract:
This investigation on relationships between the field of artificial neural networks (connectionism) and statistics starts with a look on relevant work based on a classification of possible points of contact. Then follows a distinction between connectionism seen as a tool for data analysis (engineering connectionism) and seen as a model for human thinking or, as one might say, a tool for cognitive or biological modeling (explanatory connectionism). It will be argued that statistics will have a major impact on the former but a rather minor on the latter. As a consequence, the gap between applied neural network research and research concerning cognitive modeling with artificial neural networks will become even bigger than it already is. Statistics will be adopted as the theory of engineering connectionism and therefore entail its development fom a purely empirical to a fullgrown theoretical science. Explanatory connectionism has its own problems and will have to undergo its own changes. Consequently, it will finally be seen as a science of its own independent from mere data analysis purposes.

1.) Introduction

Already in the early work that started the second rise of connectionism in the mid−1980's, analogies between neural networks and statistical methods were drawn (e.g. Stone 1986). However, it took quite a few years to bring such commonalities to the center of (connectionist) awareness. Statisticians realized, that "neural networks [...] are now beginning to be used in a wide range of subject areas traditionally thought by statisticians to be their domain" (Ripley 1993). Out of this situation of competition emerged an ever growing discussion and argument between statisticians and connectionists about the relationship of the two fields of science. The claim that connectionism is in fact nothing but statistics, is brought forward more and more often both in scientific publications (e.g. "In one sense neural networks are little more than non−linear regression and allied optimization methods" Ripley 1993) and in more informal media (e.g. the comp.ai.neural−nets newsgrpoup on the internet). Since at the same time the 'media hype' about neural networks vanishes and in consequence lots of interest and funding is going back to statistics, the question on the connectionist side seems to arise as to what a statistician is to a connectionist: a friend or a foe?
Since such criticism on connectionism is often expressed generally without any further discrimination, it is highly overdue to try shed some light on the proper treatment of the relationship between statistics and connectionism. Such a systematic account is even more urgent, because the already very sharp argument seems to lead to a condemnation of the idea of connectionism in its entirety.

2.) Points of contact between connectionism and statistics

The following overview is divided into paragraphs based on a classification of possible points of contact. The first section is concerned with the theoretical comparison of artificial neural networks and statistics based on a formalization in a joint famework. The second section deals with the empirical comparison of connectionist and statistical methods in the routine of data analysis and is therefore about competition. The third section is concerned with the role that statistics can play in neural network experimentation, especially in the evaluation of such studies.

2.1) Theoretical comparison

To allow for a comparison of neural network models and statistical models on a theoretical basis, both kinds of models must be described in one joint theoretical framework. Up to now, most of the work in this line of research has been done by statisticians (see e.g. Ripley 1992, Sarle 1994). Naturally, the theory used in such endeavours is that of statistical mathematics.

Ripley (1993) concentrates on ”the engineer's viewpoint” in connectionism that tries to find ”new computing and pattern recognition paradigms” in contrast to the biologist's view. He gives an overview of the statistical aspects of a variety of neural network algorithms and compares their performance on one data set to statistical procedures (see also 2.2).
Concerning the theoretical comparison, his emphasis is on feed forward neural networks. He comments on problems like the rather unexplored convergence behaviour of the backpropagation algorithm. He elaborates on the problem of getting stuck in local minima instead of reaching the global minimum whilst minimizing the total squared error (i.e. the sum of the squared differences between network outputs and target outputs).
According to Ripley, this minimization of the squared error could be achieved more effectively by a nonlinear least squares algorithm or by general minimization algorithms such as quasi–Newton and conjugate gradient methods and simulated annealing than by relying on the backpropagation algorithm.

Sarle (1994) provides a translation from connectionist to statistic terminology and points out the common underlying principles. Because connectionist researchers ”routinely reinvent methods that have been known in the statistical or mathematical literature for decades or centuries”, Sarle is able to provide equivalents in the form of statistical models for most of the more common neural network models.
The following overview is a compilation of Sarle's comparisons:

Neural Network Model	Equivalent Statistical Model
Simple Linear Perceptron	Multivariate Multiple Linear Regression
Simple Nonlinear Perceptron	Logistic Regression
Adaline	Linear Discriminant Function
Multilayer Perceptron	Simple Nonlinear Regression, Multivariate Multiple Nonlinear Regression
Unsupervised Hebbian Learning	Principal Components
Adaptive Vector Quantization	Least–squares Cluster Analysis
Learning Vector Quantization	Variation of Nearest–neighbour Discriminant Ananlysis
Radial Basis Functions	Kernel Regression Methods

He also discusses the statistical view of a few other network models, for which he does not find precise equivalents such as counterpropagation, self–organizing maps or ART networks.
However, Sarle continues stating that the training algorithms for the network models are very often inefficient. One reason being that ”they are designed to be implemented on massiveley parallel computers but are, in fact, usually implemented on common serial computers”. Another more severe argument is that

connectionist researchers ”often fail to understand how these methods work” because of their lack of understanding of the underlying statistical principles.
For instance, multilayer perceptrons can be trained more efficiently with general purpose nonlinear modeling or optimization programs. Sarle even names the statistical procedures of a common statistical software package to be used for training of the various connectionist models.

The reformulation of neural network models in terms of a statistical framework is often followed by modifications of those models themselves:
Röscheisen et al. (1991) are able to achieve better performance of their neural controller by recasting the problem in Bayes' decision theory leading to a specific network model and training regime. Tresp et al. (1994) deduce an improved algorithm for the training of feedforward neural networks with incomplete data based on Maximum Likelihood considerations.

So the conclusion is that neural network models can be described properly within a statistical framework because of the inherent similarities between statistics and neural networks. What often follows is that statistical procedures are better suited for network training than the originally proposed connectionist learning algorithms. Moreover, the neural network models themselves often have to be altered to achieve optimal performance.

2.2) Empirical comparison

The most common use of neural networks for engineering purposes is that of data analysis. In both classification via supervised learning and clustering via unsupervised learning the main goal is optimal performance of the chosen method for a given problem. Therefore, working on a data analysis problem, the applied neural networks have to compete with appropriate other, e.g. statistical, methods. Only by not restricting oneself to the methods of one single technology, one can hope for optimal solutions.
Which statistical methods are to be chosen for such a comparison should be a result of previous theoretical studies (see 2.1). Because statistical methods are far better understood and there is a general lack of theory within the field of connectionism (see e.g. Dorffner et al. 1991), one should stick to statistical methods as long as they perform at least as good as their connectionist competitors.

Generally, one must say that the quality of research in the area of connectionism definitely needs improvement. In his study, Prechelt (1994a) examines 113 articles from two connectionist ”top journals” (*Neural Computation* and *Neural Networks*) for the amount of experimental evaluation they contain. 34% of all articles feature no comparison with other algorithms at all and only 19% compare to more than two known algorithms.

An extensive empirical comparison between 23 statistical, machine learning and neural network methods on 21 different data sets is given in Michie et al. (1994). The different methods employed include algorithms like linear and quadratic discriminants, nearest neighbour and Bayes approaches on the statistical side, various decision tree− and rule based methods on the machine learning, and multi−layer perceptrons, Kohonen and radial basis networks among others on the neural networks side.
Unsurprisingly, no general best method can be identified for the diverse data sets in this study. With the help of statistical and information−based measures of the datasets, it is tried to find some connection between the features of the data sets and the performance of the different methods. Regarding neural networks, the authors conclude that they perform quite well on most of the problems. A radial basis network is the winning algorithm on one data set and for 15 of the 21 data sets at least one neural network algorithm ranks in the top five. However, the authors also point out certain difficulties like computational inefficiency, laborious adjusting of parameters and hard to understand behaviour of the completed networks because of their high degree of nonlinearity.

Ripley (1993) compares various statistical and neural network methods and a decision tree classifier on one data set. Looking at the classification accuracy, some of the networks performed as well as the statistical methods. However, the author also reports about the ”degree of frustration” involved with parameter tuning and getting the back−propagation algorithm to converge.

Most of the other empirical studies found in neural network literature do not include statistical methods. Some of them are restricted to comparison of connectionist to machine learning methods (e.g. Schaffer (1993), Shavlik et al. (1991)), most of them remain within the field of connectionism (e.g. Prechelt 1994b who provides performance figures for a number of neural networks as a starting point for further exploration using his benchmark collection of datasets).
A comprehensive review of such previous empirical comparisons is provided in Michie et al. (1994).

So the bottom line of this section is that in general comparison of connectionist and statistical methods does not happen as often as it should. If such a performance comparison is undertaken, connectionist methods often perform quite well, although there are a lot of difficulties in applying the algorithms.

2.3) Statistics as a tool for connectionist experimentation

Because most of the neural network learning algorithms remain too complex for rigorous formal treatment, connectionism is mostly performed as an experimental science (see Kibler & Langley (1988) for a discussion within the framework of the related field of machine learning). Such experiments should involve systematic variation of independent variables (e.g. learning rate, number of hidden units) and examination of their effects on dependent variables (e.g. accuracy, learning behaviour). Statistics can be of great help in the proper design and evaluation of those experiments.

When measuring the accuracy of any classification rule, it is widely known that the error rate estimated from the same set of data as that used for construction of the classificator is usually over−optimistic (see e.g. Michie et al. (1994)). In statistical terms it is said that such error rates tend to be *biased*. Therefore it is at least necessary to use different sets of data for training and testing.
An overview of other statistical procedures like the bootstrap algorithm or cross−validation to minimize such a bias is given in Michie et al. (1994).
Paass (1993) discusses the use of bootstrap especially in the context of feedforward neural networks. By applying bootstrap, confidence intervals for the prediction distribution of inputs unknown to the network during training can be derived.

With neural network algorithms, it is necessary to tune some parameters (e.g. learning rate, number of layers, numbers of units, etc.) to get the best performance. As Michie et al. (1994) point out, a division of the available data into three different sets is recommended. One should hold back approximately 20% of the data and divide the remaining data in a set for training and a set for testing, and then tune the parameter using those two sets. The final network should use both training and test data for learning with the now optimized parameters and should then be finally tested with the remaining, never before used, 20% of the data. If the use of such a third independent data set is omiited, the obtained error rates will again be biased and over−optimistic.

Connectionist experimentation involves the comparison of many different test runs. Only by means of such multiple test runs, the effects of the varying of parameters, of different algorithms employed, different kinds of data used and so forth can be observed.
Often such comparisons are made by looking only at absolute numbers like the accuracy achieved in a certain classification task. But only by statistical testing it can be ensured that the observed performance differences are caused by the varied independent variables (e.g. network parameters, kind of network method, etc.) and not by mere chance (i.e. whether the observed phenomena are *significant* in a statistical sense or not).

The following papers use some sort of statistics to compare performance results.
Finnoff et al. (1992) and Hergert et al. (1992b) use "a (robust) modification of a t−test statistic" for comparison of different algorithms. Hergert et al. (1992a) compare some network pruning methods with a four field table $\chi 2$−test. Shavlik et al. (1991) investigate different algorithms on a variety of data sets and employ a two−way analysis of variance (ANOVA) to isolate the source of variation in the results. Additionally, a t−test for paired differences is used.
Egmont−Petersen et al. (1994) give an excellent overview on statistical evaluation of neural network classifiers especially for the medical domain, including e.g. standard errors and confidence intervals for sensitivity and specificity.

All the problems connected with empirical research and experiment design are well known to statisticians. A closer look at connectionist literature reveals that there is only little awareness of such issues within the neural network community.

3) What kind of impacts on what kind of connectionist research?

In the previous section, the relationship between statistics and connectionism has been elaborated based on a survey of possible points of contact. From that, an overview of possible impacts of statistics on connectionism that can be expected or that can already be observed is straightforward.

The possible "benefit" from statistical methods is threefold:
i) They can make some neural networks obsolete because these are less well suited to solve certain problems.
Statistics can show that some networks are not the optimal choice for certain problems by means of theoretical analysis (see 2.1) or through direct empirical competition (see 2.2).
ii) They can lead to better training algorithms and better models for neural networks.
This can be achieved by reformulation of neural networks in a statistical framework (see 2.1).
iii) They can help to improve the quality of connectionist research.
The quality of neural network research can profit from statistical knowledge about design and evaluation of empirical research (see 2.3).

Before further elaborating on these three points, it is necessary to consider what kind of connectionist research will "benefit" in these ways in order to avoid mistakes due to unsystematic and undifferentiated treatment of connectionism.
The field of connectionism is usually divided into two subfields (see e.g. Winston 1992 for a discussion in the more general framework of artificial intelligence):
The first one persecutes an engineering goal trying to solve real−world problems using connectionist methods. The emphasis here is on the construction of hopefully good solutions and not so much on how these are achieved. From now on, we will call this subfield *engineering* connectionism.
The second subfield is concerned with the understanding and explanation of human intelligence, either with an emphasis on biological or cognitive explanations. This is attempted by building computational models and examination of their behaviour, where the emphasis is clearly on how certain performances are achieved. In what follows, we will call this subfield *explanatory* connectionism.

Concerning engineering connectionism, there can be impacts from all three "benefits".
ad i): If there exists a (statistical) method that solves a given problem better than neural networks, then there is no reason to stick to connectionist methods.
ad ii): As the main objective of engineering connectionism are optimal engineering solutions, every help from statistics that leads to the improvement of algorithms and models is welcome.
ad iii): Again, any help from statistics towards proper empirical connectionist research should be highly appreciated.

Concerning explanatory connectionism, the possible impacts of statistics are rather minor.
Explanatory connectionism is trying to build computational models of cognitive or biological phenomena. Therefore, each part of such a model, like the architecture (units, layers, weights, etc.) or the algorithms (updates of weights and activations), represents something of what is being modelled. For instance units may stand for biological neurons, weights for synapses, activations for membrane potentials, update rules for certain biochemical processes. Even if the intention is to model something on a higer cognitive level of abstraction, there still is the connection of each part of the model to the phenomenon which it represents. This implicates strong constraints on what impacts from statistics can be awaited on that type of research.

ad i): Connectionist models cannot be replaced by statistical models just because of inferior performance. To warrant a replacement, a statistical model must be shown to be a model of the phenomena one is interested in.
ad ii): As has been outlined above (see 2.1), the reformulation of connectionist models in statistical terms can give great insight into mechanics and performance of neural networks. This, of course, is of great value also to explanatory connectionism. But again, changes within the connectionist models cannot be based solely on such performance considerations. Because there exists a mapping from all the parts of the connectionist model to the phenomena that are being modelled, changes in the connectionist model must always be plausible from the point of view of those modelled phenomena.
ad iii): This is the only point in which explanatory connectionism can profit from statistics in the same way as engineering connectionism or any other empirical investigation.

Even if such a division into the two subfields is employed in comparisons of statistics and connectionism, statements are nevertheless often applied generally to the whole field of connectionism (e.g. "This paper is concerned with artificial neural networks for data analysis", "If artificial neural networks are intelligent, then many statistical methods must also be considered intelligent", both Sarle 1994).

4) On the road to ... where?

In the previous sections a survey of likely impacts of statistics on connectionism has been given. In what follows, possible future directions and changes of connectionism will be sketched, again based on the division into engineering and explanatory connectionism proposed above.
The fact that already an increasing amount of research in the engineering subfield includes statistics in its endeavours, is an indication of an impact that is already taking place. Therefore it can be expected that in the near future such references and links to statistics will become obligatory and, hopefully, a must for acknowledgement of scientific soundness.
This will become a difficulty for all those connectionist researchers who want to pursue engineering goals with neural networks and do not have the rigorous statistical background that could become a must−have very soon. Those people involved in connectionism that stem from very diverse scientific fields "and are engineers, physicists, neurophysiologists, psychologists, or computer scientists who know little about statistics" (Sarle 1994) will then become what some think that they already are: laymen or amateur statisticians.

Another impact that can already be observed is that many neural network models will simply not be used for data analysis any more, or at least not in their original form. Models that are simply not suited for engineering tasks will either vanish completely or be severly altered with the help of statistical knowledge (see 2.1).

Hopefully, because of the improvement of the quality of connectionist experimentation (see 2.3) and the increasing competition (see 2.2), the steady growth of the number of different models and algorithms in connectionism will be stopped. This will happen because the majority of these methods will be recognized as what they really are: Highly specialized solutions to mostly single (and often artificial) problems.

Therefore, only those connectionist models and algorithms will remain in the engineering subfield, that survive the increased competition or that contain new and interesting ideas even from the point of view of a statistician.
For instance, Multi Layer Perceptrons have been recognized as being ”especially valueable because [...] the complexity of the model” can be varied ”from a simple parametric model to a highly flexible, nonparametric model” (Sarle 1994). Recurrent networks are already being examined for their commonalities with NARIMA (nonlinear autoregressive–moving average) models. The underlying statistical principles of such networks are still unexplored, especially in the case of more complicated models (e.g. multi–recurrent networks, see Ulbricht 1994).
An often especially by statisticians noticed drawback of neural networks is the conceptual opaqueness of the representations in the hidden units. This has been termed the ”explanation problem” (Partridge 1987) and addresses the fact that it is often very difficult for the designer of a network to understand the network's behaviour. For instance it is usually very hard to describe what kinds of inputs have what kinds of influence on the network's weights and outputs. This, of course, is a direct consequence of the distributed nature of the representation in the hidden units, which is a notion that lies at the very heart of connectionism. Prem (1995b) proposes the use of ”symbol grounding” to enable the networks to explain their behaviour by themselves. Therefore, such networks can be said to employ some kind of infering statistics without giving up their distributedness.

To sum things up for engineering connectionism, there is an increasing probability that statistics will become the theoretical basis of this subfield of connectionism. This will entail the development of engineering connectionism from a purely empirical to a fullgrown theoretical science.

But what about what we have termed explanatory connectionism? This subfield of connectionism has its own problems and is consequently already undergoing major changes of its paradigms (see Dorffner 1995 for an overview).
Prem (1995a) argues that in the study of what has been termed ”new artificial intelligence” (including fields like artificial life, behaviour based robotics and also neural networks) a move towards the original aims of artificial intelligence is being made. Such a ”new artificial intelligence” stresses the physical embodiment of the designed systems (i.e. the necessity to build robots), the aspect of cognitive modeling and a greater proximity to biology. As a consequence, connectionist paradigms within this ”new wave” of artificial intelligence employ new architectures and techniques that have less and less in common with their former engineering twins and therefore with statistical methods as well.
As has been outlined above, statistics can give insight into mechanics and performance of neural networks and help to improve the quality of experimentation even for explanatory connectionism. But this will not stop the widening of the gap between engineering and explanatory connectionism.

5) Conclusion

The starting point of the debate about the relation between statistics and connectionism was a lot of justified criticism on connectionism brought forward by statisticians. This criticism has often been stated very bluntly: ”The marketing hype claims that neural networks can be used with no experience and automatically learn whatever is required; this, of course, is nonsense.” (Sarle 1994). It is the belief of the author, that the whole field of connectionism can only profit from such sharp criticism.

In accordance with the three possible ”benefits” from statistics outlined in the previous section, every connectionist working within the engineering subfield should keep the following three questions in mind during all of his research work and consequently relate to them in all his publications:
Is there a statistical method that can solve my problem more optimal?
How is my neural network model related to statistical methods?
What is the proper design and evaluation of my neural network experiment?

The engineering part of connectionism will either mature to a fullgrown independent theoretical science or simply be seen as a part of statistics.

As a consequence, the modeling of cognitive or biological phenomena with neural networks will finally be realized as being an endeavour on its own only remotely linked to engineering purposes. A lot of research in the short history of connectionism has been handicapped by the following double strategy: People originally interested in biological or cognitive modeling had to show some additional engineering payoff of their neural networks (e.g. usability for data analysis) in order to keep the funding agencies and institutions interested. As a consequence they ended up with neural network models that are neither plausible in terms of biology or cognition nor are they really suited for engineering purposes. The current criticism from statisticians will help to clarify those confusions and stop further misuse of the original ideas of connectionism.

Therefore, statisticians should rather be seen as friends that give connectionists some good advice and help them to untangle what has been muddled years ago, and not as competing foes.

Acknowledgements

The author wishes to thank Erich Prem for ongoing discussions about the idiosyncrasies of connectionist research. The following persons made valuable comments on previous versions of this work: Johannes Fürnkranz, Paolo Petta, Lutz Prechelt and Erich Prem. The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Science and Research.

Literature:

Dorffner, G., Prem, E., Ulbricht, C., Wiklicky, H. (1991), Theory and Practice of Neural Networks, in Brauer, W. & Hernandez, D. (eds.), Verteilte künstliche Intelligenz und kooperatives Arbeiten, Springer Berlin, Informatik−Fachberichte 291.

Dorffner, G. (ed.) (1995), Neural Networks and a New AI, Chapman & Hall, London.

Egmont−Petersen, M., Talmon, J.L., Brender, J., McNair, P. (1994), On the quality of neural net classifiers, Artificial Intelligence in Medicine 6, 359−381.

Finnoff, W., Hergert, F., Zimmermann, H.G. (1992), Improving Generalization Performance by Nonconvergent Model Selection Methods, in Aleksander, I. & Taylor, J. (eds.), Artificial Neural Networks, 2, North−Holland, Amsterdam, pp.233−236.

Hergert, F., Finnoff, W., Zimmermann, H.G. (1992a), A Comparison of Weight Elimination Methods for Reducing Complexity in Neural Networks, in IJCNN International Joint Conference on Neural Networks, Baltimore, IEEE, pp.980−987.

Hergert, F., Zimmermann, H.G., Kramer, U., Finnoff, W. (1992b), Domain Independent Testing and Performance Comparisons for Neural Networks, in Aleksander, I. & Taylor, J. (eds.), Artificial Neural Networks, 2, Nort−Holland, Amsterdam, pp.1071−1076.

Kibler, D. & Langley, P. (1988), Machine Learning as an Experimental Science, Machine Learning, 3(1), 5−8.

Michie, D., Spiegelhalter, D.J., Taylor, C.C. (eds.) (1994), Machine Learning, Neural and Statistical Classification, Ellis Horwood, England.

Paass, G. (1993), Assessing and Improving Neural Network Predictions by the Bootsrap Algorithm, in Hanson, S.J., et al. (eds.), Neural Information Processing Systems 5, Morgan Kaufmann, San Mateo, CA.

Partridge, D. (1987), What's Wrong With Neural Architectures, Computer Science Dept., Univ. of Exeter, Research Report 142, 1987.

Prechelt, L. (1994a), A Study of Experimental Evaluations of Neural Network Learning Algorithms: Current Research Practice, Fakultät für Informatik, Universität Karlsruhe, Technical Report 19/94.

Prechelt, L. (1994b), PROBEN1 – A Set of Neural Network Benchmark Problems and Benchmarking Rules, Fakultät für Informatik, Universität Karlsruhe, Technical Report 21/94.

Prem, E., (1995a), New AI: Naturalness Revealed in the Study of Artificial Intelligence, in Dorffner (1995).

Prem, E., (1995b), Understanding Complex Systems: What can the speaking lion tell us?, in Steels, L. (ed.), The Biology and Technology of Autonomos Agents, Springer, NATO ASI Series.

Ripley, B.D. (1993), Statistical Aspects of Neural Networks, in Barndorff–Nielsen, O.E, Jensen, J.L., Kendall, W.S., (eds.), Networks and Chaos: Statistical and Probabilistic Aspects, London: Chapman & Hall.

Röscheisen, M., Hofmann, R., Tresp. V. (1992), Neural Control for Rolling Mills: Incorporating Domain Theories to Overcome Data Deficiency, in Moody, J.E., et al. (eds.), Neural Information Processing Systems 4, Morgan Kaufmann, San Mateo, CA.

Schaffer, C. (1993), Technical Note: Selecting a Classification Method by Cross–Validation, Machine Learning, 13(1), 135–143.

Shavlik, J.W., Mooney, R.J., Towell, G.G. (1991), Symbolic and Neural Learning Algorithms: An Experimental Comparison, Machine Learning, 6(2).

Sarle, W.S. (1994), Neural Networks and Statistical Models, Proceedings of the Nineteenth Annual SAS Users Group International Conference, Cary, NC: SAS Institute, pp 1538–1550.

Stone, G.O. (1986), An Analysis of the Delta Rule and the Learning of Statistical Associations, in Rumelhart D.E. & McClelland J.L., Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Vol 1: Foundations, MIT Press, Cambridge, MA.

Tresp, V., Ahmad, S., Neuneier, R. (1994), Traininig Neural Networks with Deficient Data, in Cowan, J.D., Tesauro, G., Alspector, J. (eds.), Neural Information Processing Systems 6, Morgan Kaufmann.

Ulbricht, C. (1994), Multi–recurrent Networks for Traffic Forecasting, Proceedings of the AAAI'94 Conference, Seattle, Washington, Volume II, pp. 883–888.

Winston, P.H. (1992), Artificial Intelligence, Addison–Wesley, Reading, MA.