Recurrent Neural Networks for mobile phone cell planning using topological information

Laurent Bougrain, Frédéric Alexandre

Recurrent Neural Networks for mobile phone cell planning using topological information*

Laurent Bougrain, Frédéric Alexandre

LORIA-INRIA Lorraine
Campus Scientifique - BP 239 54506 Vandoeuvre-lès-Nancy cedex
e-mail:bougrain@loria.fr, falex@loria.fr

Abstract : Real-world problems are often characterized by the large size of their input space, the noise which is added to these data and by the complexity of the underlying physical laws. These laws are often continuous and more information can be brought to the understanding of the problem if contextual hints are added. This contextual information yields close behaviors. This topology is itself often linked to the topology of the input space. This paper explores how recurrent models can improve prediction in a radio communication problem with such contextual information.

keywords: recurrent neural networks, contextual data, radio communication.

I An industrial problem

Cell net planning is a strategic stage for telecommunication operators. Choosing location and size of the glazed zone of transmitting stations is a key point to optimize the development of a radio mobile net. Cell planning depends on the attenuation of radio electrical waves. The attenuation is predicted as a function of the landscape and the characteristics of the antenna. Such a prediction can be carried out by classical artificial neural networks like multilayer perceptrons (MLP) (see figure 1).

II Previous studies

Previous studies [1-2] have shown that the forecasting of the attenuation is better if various prediction models are specialized on each class obtained by clustering the data with unsupervised algorithms (e.g. kohonen, growing neural gas, k-means) or using a contextual model such as an orthogonal weight estimator [7] which uses a part of its attributes as context.

*This research was supported by the Centre National d'Etudes des Télécommunication through contract n°97 1B008.

III Using a contextual information

The values of radio electric fields are continuous values. Close situations have close attenuation values. So in this forecasting application, using contextual information by adding proximity information to the current data allows to define contextual models. Useful information to describe the context can be given by topological information or by the predictive values of the nearest neighbors. The latter solution has been chosen for several reasons. First, less parameters are involved in this solution; this yields a smaller size for the model. Second, this solution causes the predictive model to be independent from the representation of the topological data.

IV Preparing the data

From a technical point of view, neighborhood information is considered valid up to one hundred meters. 44,000 standardized patterns with 32 attributes were extracted from transmitting, receiving, and a national geographic database, describing terrain in France. 160 series, longer than 10 elements, gathering about 39000 patterns, are used as the learning set. 17 series, gathering about 5000 patterns, are used as the test set.

V Recurrent neural networks

Recurrent neural networks (RNNs) use information generated by the previous situations, seen as the context, to act upon the current computation. The context is made up of a set of neurons with a feedback connection to the same layer or to an earlier layer. In the Jordan's model, the context is a combination of the output units of the previous situation and the previous context [6]. In the Elman's model, the context is given by the hidden units of the previous situation [3].

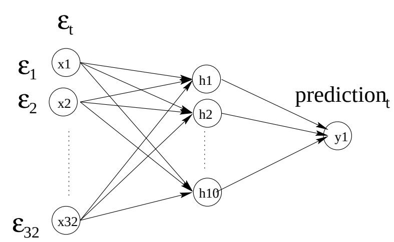

Figure 1: Multilayer perceptron: architecture

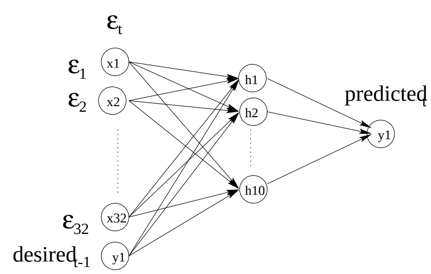

Figure 2: PDV: architecture

V.1 Interest of the method

To check our hypothesis that a contextual information generated by the neighborhood situation improves the results, the desired value of the neighbor was presented as context (figure 2). Let us call this model the previous desired value model (PDV).
The attenuation of radio electrical wave is predicted with a precision of 98.3% on the test set compared to 96.5% with a MLP which does not take context into account (table 1). Of course, this result does not represent an improvement in itself, considering that the desired value of the neighbor will not be known during the application stage, but this result validates our hypothesis and it represents the predictive limit of a Jordan's model.

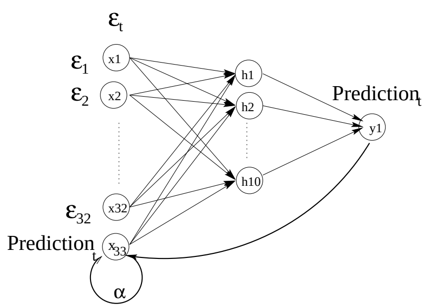

Figure 3: Jordan's network: architecture

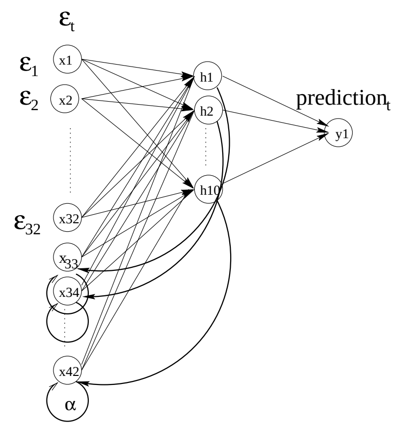

Figure 4: Elman's network: architecture

V.2 Jordan's model

In the Jordan's model (figure 3), the context at time t, c(t), depends on the output unit at time t-1, o(t-1), and on the previous context c(t-1) with the following rule : c(t) = o(t-1) + alpha*c(t-1) [5]. The size of the context window depends on the alpha parameter. For this application, there is just one value to forecast. The network architecture contains 32+1 units on the input layer, 10 units on the hidden layer and 1 unit on the output unit. The analysis of the results (table 1) shows that the Jordan's model takes the context into account with confidence because the results are better than the ones obtained by a MLP with a similar architecture (32x10x1). The generated context is pertinent. Changing the definition of the context by varying the alpha parameter does not influence the results significantly.

V.3 Elman's model

With the Jordan's model, the context represents a thirty third of the input data. That can be too small to accurately affect the prediction. With the Elman's model (figure 4), the context at time t is the hidden layer at time t-1. To compare with the MLP performances, the architecture of the chosen Elman's model is (32+10)x10x1. So the context represents the quarter of the input data. The results are similar than for the Jordan's model (table 1). The added context seems more relevant for larger learning corpus.

Results	Learning		Test		epoch
	μ	σ	μ	σ	
MLP	4.248938	5.444293	4.589846	5.944992	1612
Jordan	4.173493	5.353248	4.297221	5.549206	944
Elman	4.014523	5.154875	4.310753	5.550944	3503
PDV	2.540030	3.416124	2.247765	3.055724	1174

Table 1: results

VI Time analysis: beyond the global mean square error

Recurrent neural networks use values calculated at previous time. If these previous values are not pertinent, the computing will not give pertinent values. So the prediction will be less and less good. Figures 5 and 6 show the performance with time. The performance is the mean, plotted to a scale of percentage, of all square errors at a specific date i.e. for a particular position in series. Series contain from 10 to 1500 situations. Long series are rare. For on old date, the average is based upon few values and goes less significant. So only the first 350 positions appear in the figures. Confidence interval I is ± 0.20 db for the learning corpus and ± 0.55 db for the test corpus according to the formula :

$$I(\alpha, N) = \frac{X + \frac{Z_\alpha^2}{2N} \pm Z_\alpha \sqrt{\frac{X(1-X)}{N} + \frac{Z_\alpha^2}{4N^2}}}{1 + \frac{Z_\alpha^2}{N}}$$

where N is the corpus size, X the performance, $Z_\alpha = 1.96$ with $\alpha = 95\%$ [4].
A smoothing on a time window, 20 positions wide, was applied to reduce error fuzziness. The performance is constant in time. There is no degradation.. No one among both RNNs has an optimal period for performance.

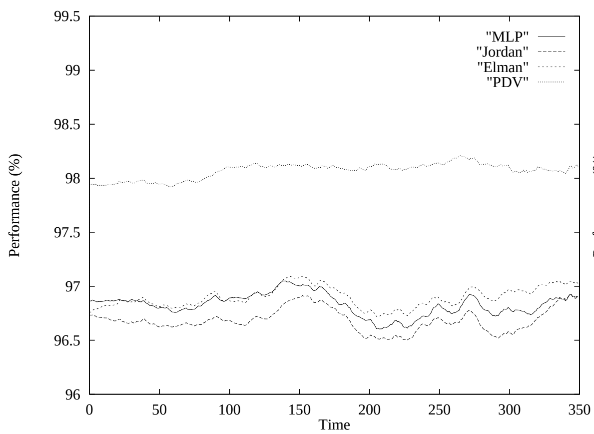

Figure 5: Performance evolution: learning corpus

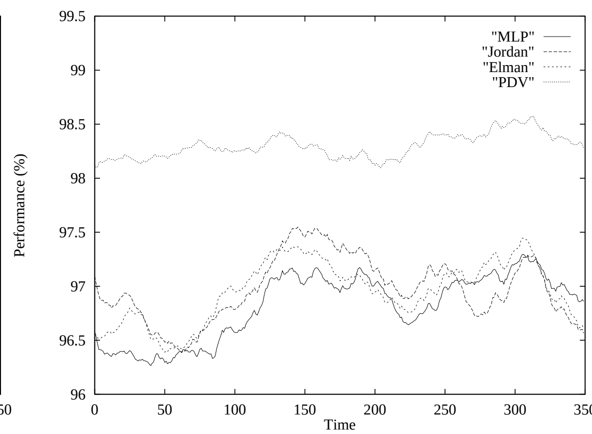

Figure 6: Performance evolution: test corpus

VII Consistency of the forecasting values

Initially, these recurrent neural networks have been preferred to a simple MLP, because the use of an additional information could increase the performance. But by choosing a contextual information linked to the neighborhood forecasting it is possible to obtain a better regularity in forecasting values, according to the reality, stating that close geographical positions have close attenuation values of the radio electrical waves.
To observe the influence of the neighbor attenuation on the continuity of the forecasting values, the mean of forecasting gap between two adjacent positions, for all series, have been measured (table 2). This mean was also calculated on the forecasting measured in reality.

Model	Mean gap	
	learning	test
Practice	2.735251	2.290719
MLP	2.118404	1.805893
Jordan	1.491032	1.32030
Elman	1.594415	1.283445
PDV	2.232227	1.982311

Table 2: Forecasting gap between two adjacent situations (dB)

Firstly, by comparing the two columns, and in particular the practical values, we observe that the forecasting values of the test corpus are more consistent than the ones of the learning corpus. We can imagine that the adjacent positions of the test corpus are closer. So the forecasting values could be better than reality for the contextual neural networks when they are applied on the test corpus. Secondly, PDV network seems able to track the discontinuity of the attenuation better than MLP but lesser than in practice. Forecasting values from RNNs are more consistent than MLP. So RNNs are a good choice if we want to have a smoothing of the outputs of close inputs in this radio communication problem.

VIII Conclusion

We have shown the interest of adding a context to the input data and we have evaluated the limit that can be reached. The Elman's model is sometimes more efficient than the Jordan's model because its context representation is more influent than the Jordan's one. The performance -of the system was improved. There is no degradation of the performance in time. Beyond performance and convergence aspects, the predictions obtained by a recurrent model are consistent the ones with the others in relation to their proximity. Taking into account older contextual information, by adding context at t-n, will smooth more the prediction but the error will not be reduced because the predictions will not be able to follow the variability of the attenuation of radio electrical waves. The influence of the information given by the neighborhood of the current position to estimate makes a smoothing of the predictions. Using a spatiotemporal smoothness constraint, high value errors can be avoided.

References

[1] L. Bougrain, N. Pican and F. Alexandre. Rôle du contexte dans le modèle OWE: un réseau de neurones artificiels utilisant des connexions axo-synaptiques. Proc. of NeuroSciences pour Ingenieur 1998, (Munster, 1998)

[2] L. Bougrain and F. Alexandre. Unsupervised Connectionist Clustering Algorithms for a better Supervised Prediction: Application to a radio communication problem. Proc. of International Join Conference on Neural Networks (Washington 1999)

[3] J.L. Elman. Finding structure in time. Cognitive Science 14 (1990) 179-211

[4] M. Gardner and D. Altman, Statistics with Confidence: Confidence Intervals and Statistical Guidelines. British Medical Journal Publication (London, 1989)

[5] J. Hertz and A. Krogh and R. G. Palmer, Introduction to the theory of neural computation. Addison-Wesley (1992) 179-182

[6] M.I. Jordan. Serial order: A parallel, distributed processing approach. In Advances in connectionist theory: Speech.(J.L. Elman & D.E. Rumelhart. Erlbaum, Hillsdale, 1989).

[7] N. Pican. An Orthogonal Delta Weight Estimator for MLP Architectures. ICNN'96 (Washington, 1996)