Neural network system, JED, offers solution for predicting single-copy sales

TOM HESKES and BERT KAPPEN
Foundation for Neural Networks (SNN)
Nijmegen, The Netherlands

With declining newspaper sales in recent years, single-copy sales are getting more and more attention. In this report we will describe JED, a software system for efficient distribution of single copies. JED uses historical sales figures to predict how many newspapers can be sold in the future.

Ideally, each outlet would sell all its newspapers except one. This extra return copy informs the distributor that the outlet has not been sold out. The ideal situation is, of course, not a realistic target, since newspaper sales is a highly irregular process. Therefore, distributors will always send a few more copies than they expect to sell.

This last sentence says it all. How do you know what to expect? And how many more copies do you then deliver? These questions need a statistical solution.

JED stands for "just enough delivery" and has been developed by SNN, the Foundation for Neural Networks, for "De Telegraaf", a major Dutch newspaper company. Development of the system took about one year. It is currently in operation to handle part of the distribution and is scheduled to replace the current software system – which is based on more traditional techniques for time-series analysis (linear regression) – in the near future.

How JED works

The main idea behind JED is to train a model on the historical sales figures of each individual outlet and predict

as closely as possible the number of copies to be delivered to each outlet without selling out.

Firstly, the historical data is input into the model. It includes variables, such as the most recent sales figures, the sales figures during the same period last year, trends in the sales figures, weather predictions (which might affect sales) and so on. To optimise the input / output relationship, a set of parameters called *weights* are adjusted to the historial data. *Secondly*, the historical data runs through the model and gives a prediction of upcoming sales. It also gives the percentage of uncertainty of this prediction (for example, a plus-five or minus-five percent of this prediction) for the distributor. The uncertainty is estimated from the difference between the model's prediction of sales and the observed sales of the past. *Finally*, the distributor takes the prediction and uncertainty of this prediction to determine how many copies to deliver to each outlet, once again with the idea of not selling out.

For example *(see Figures 1 and 2)*, if the sales strategy is aimed toward a significant reduction of sellouts, the number of extra copies has to be relatively large. If the model's prediction is 50 copies with a 5 percent plus or minus uncertainty, the distributor would likely send 55 copies.

Figure 3 shows the flow of information for a single outlet. Note that the correct number of returned copies, or any information about the number of sold copies for each individual outlet is essential to the system. Without this information, it is impossible to build a prediction model.
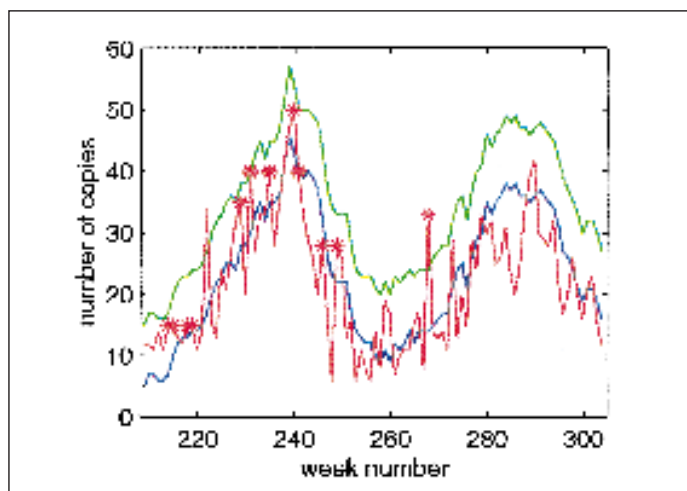

Figure 1: Outlet showing seasonal fluctuations. The observed sales (red line) can be compared with JED's predictions of the sales (blue line). For each prediction, JED has been trained on two years of the most recently available data. The delivery suggested by JED (green line) has been tuned to have a chance of less than 10 percent of selling out. These are historical sales figures over 1994 and 1995 when JED was not yet in operation. The stars indicate sellouts.
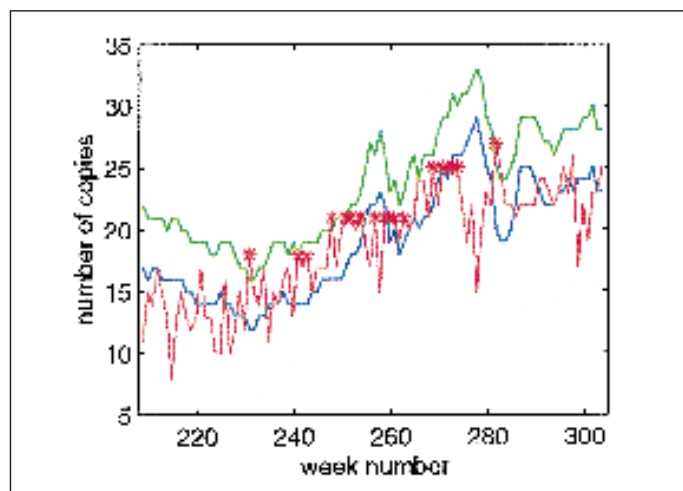

Figure 2: Outlet showing increasing sales. JED's predictions (blue line) manage to follow the increasing sales (red line), although there is inevitably some delay due to the time needed to collect the return copies (see the figures around week number 250). The large number of sellouts (stars) between week numbers 250 and 280 could have reduced significantly if JED's suggestions (green line) had been followed. See figure 1 for more explanation.

Inside the model – neural networks

The models in JED are **neural networks**, or more specifically, **multi-layered perceptrons**. The multi-layered perceptrons consist of three layers of units or neurons *(see Figure 4)*.

– The first layer represents the input data (variables). These are multiplied by the aforementioned set of parameters called *weights*.
– The result is taken through a nonlinear function and yields the activity of the second layer of so-called *hidden units*. These hidden units help give a more complex relation to the output of the model, thus making it more accurate.
– The same principle transforms the activities of the hidden units into the activities of the output unit which represents *expected sales*.

While training the model, the weights are adjusted to minimise the difference between the expected and observed sales.

Training and predicting with neural networks is fairly similar to fitting and predicting with more traditional techniques for time-series analysis such as Box-Jenkins or ARMA modeling. The advantages of neural networks are 1) the possibility to represent nonlinear relationships in the data, and 2) the robustness against noise, which basically means it allows the ability to give a more accurate prediction.

JED is by no means a standard neural network application. Without going into too many technical details, there are two aspects worth mentioning. Firstly, there is a huge amount of data and neural networks to train. Each outlet has its own neural network and requires about two years of weekly data to make a reasonably accurate model. As a consequence, training and prediction procedures must be very robust and fully automatic. Secondly, special attention must be paid regarding historical sellouts. No returned copies would indicate that more newspapers could have been sold than the amount delivered. By changing the standard definition of this training error, a downward trend of the models in which sales is structurally underestimated can be prevented.

Tests prove positive

A test was conducted to compare JED and Telegraaf's system in predicting the sales of a glossy weekly magazine at 1000 outlets for a period of several months. JED's neural-network approach realised approximately 1.7 percent more sales than Telegraaf's current system.

JED needs more testing to substantiate these results, and improvements are certainly possible. For example, tests have shown that the accuracy of the predictions improves significantly if the time needed to collect all returned
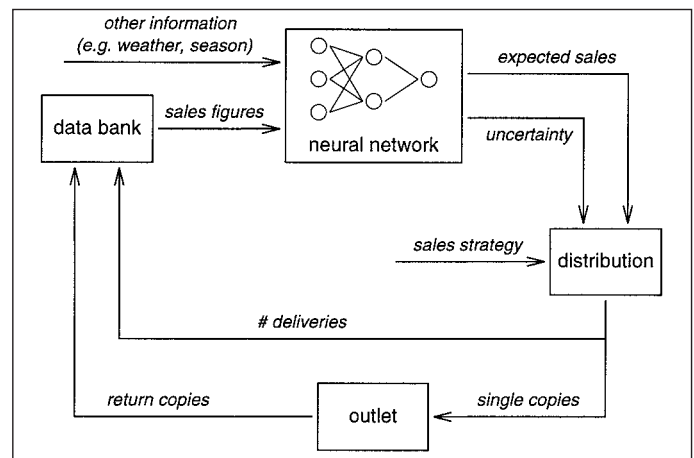

Figure 3: Shows the flow of information for one outlet. A neural network has been trained on historical sales figures, stored in a data bank, supplemented by other information such as weather or seasonal information. Using the most recent figures, the network gives a prediction of upcoming single-copy sales with an estimate of the accuracy of this prediction. These figures are combined with the sales strategy (with the focus on reducing sellouts or diminishing the number of return copies) to yield the number of single copies to be delivered. Information about the realised sales, based on the delivery and the return, is again stored in the data bank for future training and prediction.

copies goes down from four to two weeks. Some of its shortcomings hope to be resolved in upcoming versions.

Further information about JED can be obtained from Tom Heskes and Bert Kappen of SNN. SNN was founded in 1989 by the Dutch Ministry of Economic Affairs with the aim to facilitate the transfer of knowledge between Dutch universities and industry with regard to neural networks. SNN has a leading position in Europe, both in research on neural networks, as well as in the realisation of neural network applications in business. Address: Geert Grooteplein 21, NL 6525 EZ, Nijmegen, The Netherlands, phone: ++31-24-3614245, fax: ++31-24-3541435, e-mail: snn@mbfys.kun.nl. Or call IFRA's Bénédicte Lamy at +49-6151-733764.
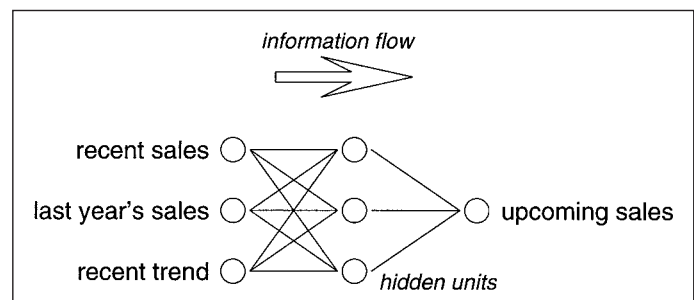

Figure 4: Shows the multi-layered units of the neural network.