Enhancing the Forecasting and Capacity Planning Capabilities in a Telecommunications Company

Ada M. Álvarez Socarrás, Arturo Berrones Santos, M. Angélica Salazar Aguilar, Carlos A. Álvarez Herrera, Mauricio Cabrera Ríos

División de Posgrado en Ingeniería de Sistemas, Facultad de Ingeniería Mecánica y Eléctrica, Universidad Autónoma de Nuevo León, Av. Pedro de Alba s/n, Cd. Universitaria 66451, San Nicolás de los Garza, Nuevo León, México

Autor Responsable

Abstract

Companies throughout the world are required to plan, under budgetary restrictions, how to meet their demand while remaining at a competitive cost level. In this work, two planning phases –demand forecasting and capacity planning- are approached through operations research tools for the yearly operation of a telecommunications company’s network in Mexico. The end result takes the form of a decision support system that integrates the tools proposed for these tasks. The use of the system is demonstrated here with an example, although, its real application has already reported significant savings for the company.

Keywords: Forecasting; Artificial neural networks; Capacity planning; Inventory control; Decision support systems; Telecommunications networks

1. Introduction

Capacity planning is a problem that appears in many different fields and industries. In general, the process of capacity planning starts with a demand forecast followed by a mathematical model and a subsequent optimization task. In this work, an approach that integrates all of these phases is presented in the context of the rapidly growing field of telecommunications.

This study is motivated by a real capacity expansion planning problem in an optical fiber network. Given a certain amount of historic information on the utilization of the network as well as the network’s general characteristics, the objective is to plan
capacity expansions in time for a short planning horizon of up to twelve months as defined by the company. First, a series of demand forecasts are generated using Artificial Neural Networks (ANNs). Then the forecasts are fed into a mathematical model that considers the stated planning horizon. The modular structure used in this work facilitated the development of a Decision Support System, which is also demonstrated.

This manuscript is organized as follows: in Section 2, the general capacity planning problem is described; Sections 3 and 4 describe the methods used for the forecasting and planning phases respectively; Section 5 discusses the integration of the methods previously described in a decision support system. Finally, Section 6 contains the conclusions and describes potential extensions to this research.

2. Problem Description

The general capacity expansion problem aims to find the lowest cost combination of equipment necessary to meet some forecasted demand. In telecommunications networks with nodal electronic equipment that does not require fiber optic, the planner must decide where and when to add equipment to meet the demand, as well as the characteristics of the equipment to be purchased. In the other hand, when optical fiber is required for transmission, the problem is reduced to planning the installation of the necessary electronic circuits in the equipment. The equipment cost dominates the optical fiber cost, since the fiber lines are usually installed in underground ducts already in place and the distances to the distribution points and the connection centers are relatively short.

This work focuses on the case where it is necessary to expand the capacity in a node connected to the network through optical fiber. This implies that the company must purchase and install equipment to deliver the bandwidth units required for future periods to meet the forecasted demand at the lowest possible cost.
Decisions regarding capacity expansion are made before the actual services are requested, therefore the bases for these decisions are demand estimates or forecasts. Knowing that all estimates are prone to error, telecommunications companies try to deal with this uncertainty by purchasing additional equipment to buffer out potential demand peaks. It is due to this coupling between planning and forecasting that a good decision support system must deal with and include both in an adequate manner. This work, then, includes developing such decision support system for short term planning, where the planning horizon considered goes as far as 12 months.

The company was able to gather historic information on the utilization of the network in terms of bandwidth. This information, organized in time series, was used to build artificial neural network models to generate forecasts for the next 12 months. The forecasted demand helped to provide the company with the lowest-cost equipment combination and its purchasing schedule based on the costs and capacities of the candidate equipment pieces that were commercially available. The prescribed solution was aimed to provide a user-specified level of protection against sudden demand peaks.

In the following sections, the details of the forecasting methods are discussed first and then those of the capacity planning phase.

### 3. Demand Forecasting

The objective in this part was to develop a model to approximate the behavior of the demand that allowed to reliably forecast 12 months. The output of such model was to be fed into the capacity planning model described in the following section.

The monthly traffic log in particular nodes of the network was available through the company for the past six years. The behavior of one of the services is shown as a time series in Figure 1.

Figure 1
Artificial Neural Networks (ANNs) have been shown to be effective handling nonlinear dependencies due to their functional nature. This capability along with the fact that it was necessary to have a general forecasting method that could accommodate a wide variety of time series behaviors, made the ANNs very attractive for this particular project.

The use of ANNs for time series forecasting, however, does require making several decisions including the algorithm used to find the ANN fitting parameters, the architecture of the ANN -especially concerning the number of hidden neurons-, as well as defining how many historic data points or lags have to be taken into account to generate the forecasts. The lack of a standard methodology to make these decisions simultaneously motivated the development of a method based on statistics and optimization techniques to decide upon all of these ANN parameters at once. The method to deal with these decisions was presented in (Salazar et al., 2006). Essentially, it is a method based on design of experiments and nonlinear optimization concepts to effectively decide upon the parameters of the ANN. A brief description of the method is provided later in this paper. The repeatability of the method as well as its effectiveness in finding competitive forecasting ANNs was key to this phase of the project. The type of ANNs used in this phase are described next.

3.1 Description of the ANN

A feedforward ANN with three layers of neurons, namely input, hidden and output layer respectively, was chosen for this project. Being required to forecast 12 demand periods sets the number of neurons in the output layer to twelve. However, it was necessary to determine the number of neurons in the input layer as well as that in the hidden layer. The number of neurons in the input layer would follow from establishing how many historic points had to be considered to forecast the ensuing 12 periods, while the number of neurons in the hidden layer would follow from establishing an adequate number of neurons for the ANN to be able to learn the training patterns while keeping its ability to generalize or predict. These two decisions however were nontrivial and had to be dealt with through the method described later in this paper.
In order to train such ANN, the Levenberg-Marquard algorithm \((lm)\) was used due to its convergence speed and quality of solutions determined experimentally (Salazar et al., 2005). The available data to build the ANN were divided into two sets: (1) one training set for the ANN to learn the inherent functional relationships, and (2) one validation set to test the prediction capability of the ANN. The training set used up about 70% of the available data while the rest was used for validation purposes. The partition was done randomly following a normal distribution to avoid biasing the results. The 70-30 split is recommended elsewhere in the literature (Zhang et al., 1998; Granger, 1993).

The forecasting performance of the ANN was measured through the mean square error (MSE). Values for the training set (MSE\(_T\)) and for the validation set (MSE\(_V\)) were kept separately to measure pattern learning and prediction capability respectively.

Although it is a common practice to begin the search of the ANN fitting parameters (weights) in a random manner, in this work several predetermined initial points were used. The rationale is that this search is indeed an optimization problem where the fitting error is to be minimized. However, the function of the error is nonlinear and nonconvex, thus making the problem sensitive to the algorithm’s starting point. Starting the optimization procedure from different points increases the chance of finding an attractive solution.

### 3.2 Determination of the ANN parameters

The parameters that needed to be set for the ANN were: (1) the number of neurons in the hidden layer, \(\text{neurons}\); (2) the number of historic data points that would be used for the forecast, \(\text{lags}\). The latter would actually determine the number of neurons in the input layer as explained previously. Parameter \(\text{lags}\) is an important one because it determines the extent to which future behavior is correlated with past behavior.

As described in (Salazar et al., 2006), a full factorial experimental design was setup to sample the experimental region and generate response surfaces of the MSE\(_T\)
and MSE_V as functions of neurons and lags. Under each particular combination of neurons and lags several ANNs were built according to the multiple starting points policy defined previously. The ANN with the lowest MSE_V was registered for the particular combination, since it was prediction capability what really mattered in this application.

The results of the experiments were used to build regression models that related the MSE values to the variation of neurons and lags. Defining models that fit the experimental data well proved difficult, and was approached through a series of redefinitions of the experimental area. The complete procedure can be consulted in (Salazar et al., 2006).

The regression models found were then used as objective functions in optimization problems, one per regression model. The aim was to find the specific values of the parameters neurons and lags that minimized the objective function. The procedure leads to combinations of these parameters that guarantee an adequate ANN prediction performance. One key advantage of the method used to set the ANN parameters, is that of reproducibility.

The ANNs built following the method described above allowed obtaining better forecasts than those obtained through traditional linear statistics methods for the series provided by the company (Villarreal, 2006).

The forecasts obtained with these ANNs were then used as inputs for a capacity planning phase described in the following section.

4. Capacity Expansion

Telecommunication companies must guarantee a high service level to their customers while making sure that the network is not extremely underutilized. It is, therefore, the task of the network manager to determine the necessary additional
bandwidth, the time at which it must be provided, as well as the type and quantity of equipment to provide it.

The relevant information to handle such planning problem is related to the equipment costs, configurations and capacities, the suppliers’ lead times, as well as the forecasted demand for the planning horizon. Because the characteristics of this problem are similar to those encountered in inventory planning, it was decided to apply inventory control techniques to approach it.

Inventories, in general, are made up by articles destined for consumption. The main purpose of inventory control is to adapt the supply to the different customers’ levels of demand. Inventory for a particular product, then, is given by the mathematical difference between the amount of products on hand for consumption (supply) and the amount of products actually being consumed (demand). Inventory control deals with establishing when and how many products should be replenished (Solow and Mathur, 1996). The objective then is to satisfy the demand at a certain service level while keeping the inventory cost as low as possible, just like it is for the telecommunications industry. The inventory, for the telecommunications case, must be understood as the mathematical difference between the installed capacity (supply) and the actual utilized capacity (demand). Idle capacity, then, becomes the inventory for the telecommunications industry.

With such remarkable similarities, there was no question that it would be beneficial to capitalize on the inventory control techniques ability to provide effective solutions at low computational cost to solve the capacity planning problem for the telecommunications company. This approach was the central idea in (Álvarez and Cabrera, 2006).

Figure 2 shows the time evolution of demand (D) and capacity (C) for one of the services of the company. The following can be readily identified in this figure: there is an initial installed capacity (C₀) that, in this case, is enough to accommodate the current level of demand. As time progresses, the demand grows, and the time in which additional equipment must be ordered from the supplier is met. This is the reorder point (r). The amount of capacity that must be purchased (Q) is to be determined in terms of
bandwidth units. The reorder point is calculated taking into account the suppliers’ lead time (L) and a particular stockout protection level.

Figure 2.

Several inventory control models were explored and analyzed in (Álvarez, 2006), resulting in a recommendation to use the Service Level Model for the company’s requirements. This model receives a user-determined service level that correlates to the protection level that the company wants to provide their users given a stochastic demand.

In the Service Level Model, the inventory level is monitored continuously, and the objective is to determine the reorder point such that there is a certain guarantee of not running out of stock, or in this case, of capacity. This guarantee is associated with the service level predefined for the customer. One important condition for this model to work properly is that the demand must be normally distributed. Although this condition was not met by the demand in this application, a series of tests showed that no major problems arose from these deviations.

The application of this inventory control model provided the company with the capability to plan its capacity expansions in the 12-month period requested in time and quantity. The analysis in this phase was carried out with capacity units. The output of this module in capacity units was then fed into an equipment selection problem, which is described next.

4.1 Equipment Acquisition

Once the amount of bandwidth units to be added in a node have been determined as well as the times at which these should be purchased, the next step is to select the equipment that must provide this additional capacity. This decision entails an optimization problem. The particular problem for this situation for a particular demand point is as follows:
Find \( x_i, y_i, w_i \quad i = 1, 2, \ldots, rt \) to

Minimize \( \sum_{i=1}^{rt} (a_i x_i + b_i y_i + c_i w_i) \)

Subject to

\[
\begin{align*}
\sum_i p_i y_i & \geq D \\
x_i, x_i & \geq y_i \quad \forall i = 1, 2, \ldots, rt \\
S_i w_i & \geq y_i \quad \forall i = 1, 2, \ldots, rt \\
x_i, y_i, w_i & \in Z^+ \quad i = 1, 2, \ldots, rt
\end{align*}
\] (1)

where \( rt \) is the number of router types available in the market, and for a router type \( i \), the following definitions hold: \( x_i \) is the number of routers to be purchased, \( y_i \) is the number of cards to be purchased, \( w_i \) is the number of card separators required to hold the required number of cards, and \( D \) is the demand in terms of bandwidth units; \( a_i, b_i, \) and \( c_i \) represent costs per router, router card, and card separator respectively; \( t_i \) is the maximum capacity of a card in bandwidth units and \( s_i \) is the maximum capacity of cards per card separator.

Problem (1) is to be solved for each particular order of bandwidth units to be purchased as the demand point \( D \), thereby, effectively providing a detailed schedule for capacity expansion with the particular type of equipment.

All of the previously described modules were coded for convenience of the company in a Decision Support System that is described in the following section.

5. Decision Support System

A comprehensive computational tool that is capable to address the previously discussed forecasting and planning aspects in a common framework was developed in the .NET platform. The software is able to use historical data to train and execute the ANN scheme discussed in Section 3. The results can then be taken as demand inputs for the capacity planning module of Section 4. In order to specify the instances of the involved optimization models, the software has an integrated database that stores the necessary information about the equipments and the installed capacity and that interacts
in run-time with the planning module. The main software components are shown in Figure 3.

Figure 3.

The communication between the different software components is controlled by a master program that takes advantage of the .NET Microsoft libraries in order to allow an adequate information flow. The implementation of the numerical algorithms is such that while being consistent with the platforms at the disposal of the company, it is possible to call suitable mathematical libraries for the heaviest calculations. For instance, the neural network in the forecasting module is mainly programmed in C++, making use of the FANN library, freely distributed under the LGPL license (http://leenissen.dk/fann/). The planning module, on the other hand, is implemented in Microsoft Visual Basic for Applications code, and it makes use of the routines included in the Microsoft Excel add-in Solver.

5.1 Example

We illustrate the information flow of our decision support system using the simulated data presented in Figure 4.

Figure 4.

The simulation has been carried out integrating an hourly model for information packet flow aiming to generate a monthly time series. At the time scale of weeks, the model is described by

\[ x_t = a \left| \sin \frac{2\pi}{168} t \right| + \varepsilon_t \]

where \( t \) stands for time. The time unit is an hour, so Equation (2) represents a weekly seasonal component of amplitude \( a \) with an additive perturbation given by \( \varepsilon_t \). Instead of the usual approach in which the perturbation is modeled by a stochastic process, we have considered the following chaotic deterministic map,
\[
\varepsilon_t = b(s_t - \bar{s}) ,
\]
\[
s_t = \beta + s_{t-1} + cs_{t-1}^m , \quad 0 < s_{t-1} \leq d \\
= \frac{s_{t-1} - d}{1 - d} , \quad d < s_{t-1} \leq 1 ,
\]
\[
c = \frac{1 - \beta - d}{d^m}
\]

where \(\bar{s}\) is the temporal average of the map. In this way, \(\varepsilon_t\) is a perturbation with zero mean and an amplitude controlled by \(b\). The parameters \(\beta, c, m\) and \(d\) are real constants. The map \(s_t\), known as the “intermittency map”, has been proposed as a model for packet traffic in telecommunication networks (Erramilli et al., 1994). It has been argued that the intermittency map is capable to display the most important statistical properties observed in real traffic, which are not explainable by a linear stochastic process (Erramilli et al.1994, 1996). In our proposed integration up to the time scale of months the model also displays a behavior that is not consistent with a linear stochastic process, which we have actually observed in real monthly time series. Details on this are reported in (Olivares, 2006) and are currently under research by the authors. The fact that there is evidence of nonlinear behavior in packet traffic gives an additional motivation for our forecasting algorithm based on ANNs.

The parameter values used in the example are: \(a = 1.0\), \(b = 0.8\), \(\beta = 0.0005\), \(m = 2\), and \(d = 0.7\). Each point of the time series is given by the sum of 720 time steps. At this monthly time scale also acts an additional positive linear drift of 1.5. The resulting series consists of 84 points, shown in Figure 4. The first 72 points in the data have been used to train and define the architecture of the ANN. The algorithm finds an optimal value of three hidden neurons for the forecasting model. The final 12 points are compared with the forecasted points in Figure 5. The dashed line connects forecasted values, while the continuous line is for the actual values. The standard deviation of the prediction error is 18.6, while the standard deviation of the differences between successive steps in the observed data is 38.5. This means that our forecast beats the
random walk by a 48%, which is a strong evidence for the capabilities of the obtained ANN to discover the correct dependencies in the data.

Figure 5.

For illustrative purposes, Figure 6 shows some of the user interface screens for this particular example. The top one is the interface for the forecasting module. The results of the forecasting module pass to the planning module, which is accessed by the user through a Microsoft Excel Application as shown in the bottom screen shot.

More specifically, the time series is entered in the blank column of the first screen and then, the user is able to define over how many different neural network architectures the system must search in order to find an optimal forecasting model. The best 12 – period forecast passes automatically to the planning module shown in the screen of the bottom. This module is coded for Microsoft Excel in Visual Basic for Applications. Given that the initial capacity is of 610 units and that the lead time is two months in this example, the planning module reports that it is necessary to place an order of 108 units in the first month in order to keep an 80% service level. This policy is perfectly adequate for the demand that has actually occurred. The resulting policy and the necessary information of the equipments stored in the corresponding databases passes automatically to the sub-module of equipment acquisition. In this module an optimal equipment selection is calculated according to the model given in Subsection 4.1.

It is worth to mention that updates in the database’s records can also be performed in run-time, so the user is able to analyze different equipment portfolios for a given demand forecast in the same run.

Figure 6.

6. Conclusions
The application of operations research techniques to tailor to the planning requirements of a telecommunications company has been presented here. Two planning phases, demand forecasting and capacity planning, were approached through artificial neural networks, inventory control and mathematical programming. An assessment of the tool described here by the company has resulted in an estimated yearly savings of close to ten million dollars in telecommunications equipment, largely due to the experienced reliability of the forecasts that allows a better accuracy of the equipment purchasing plan.

As the company wishes to extend this work in order to consider larger planning horizons, we are currently designing and developing the appropriate tools to tackle this request.

References


http://leenissen.dk/fann/