Demand Forecasting in Power Distribution Systems using an Hybrid Method of Case Based Reasoning and Expert Systems

Raúl Vilcahuamán[2 1], Joaquim Meléndez[1], and Josep Lluis de la Rosa[1]

[1] Institut d'Informàtica i Aplicacions, Universitat de Girona,
Campus Montilivi, Edifici P.4. Girona, Spain.
[2] Facultad de Ingeniería de Sistemas.
Universidad Nacional del Centro del Perú,
Calle Real 160, Huancayo, Perú.
r.vilcahuaman@ieee.org, {quimmel,peplluis}@eia.udg.es

Abstract. A novel approach to electric load forecasting using an hybrid method of artificial intelligence is presented. This paper proposes a combining of a numeric extrapolation of data with the methodology of case-based reasoning and expert system. The retrieve of previous cases adapted by an expert system has been overlapped to the load forecasting given by the numeric techniques, presently accepted for this purpose, obtaining a considerable improvement in the forecast. FUTURA software has been developed as a result of this work. It incorporates the different proposed techniques in a modular way while it provides a graphic user interface and the access capabilities to existing data bases.

1 Introduction

The quality of a load forecasting has a significant impact on the efficiency of the operation of any electric utility since many potentially costly operational decisions such as economic scheduling of generating capacity, scheduling of fuel purchases, system security assessment and planning for energy transactions are based on such forecast. The importance of accurate load forecast will increase in the future because of the dramatic changes occurring in the structure of the utility industry due to deregulation and competition. This environment compels the utilities to operate at the highest possible efficiency which, as indicate above, requires accurate load forecast.

The systems used nowadays take into consideration just registered numeric information and the forecast is limited to extrapolation algorithms with rough results in a medium term. The limits that these methods present have motivated the development of FUTURA software which incorporates Artificial Intelligence(AI) techniques in order to benefit from another forecasting knowledge. FUTURA consists of four main modules that use different techniques to solve forecasting partial aspects appealing to the available knowledge to do so.

The first module operates with a curve fitting(CF) algorithm implemented according to the recommendations of the Electric Power Research Institute (EPRI) [11]. The second module uses a case-based reasoning (CBR) to adjust the result obtained by the previous module in the different purchase/sale points of the electric companies for any future day/month/year (date) with a 15-minute resolution. The third module is an expert system whose knowledge base considers indicators such as PBI (gross domestic product GDP) growing, population, urban growing, inflation, electric rates, investments among the main partners. It is used to adapt the former results to the new socioeconomic context. The fourth module has an interface which permits to make reports with respect to analyzed values from both the future or historically-registered ones such as load factor, demand factor, coincidence factor, peak demand, minimum demand, single-line diagram, etc.

FUTURA permits to administer a great amount of data and generate reports to be used in power flows, planning and optimum expansion programs, and calculation of marginal costs in electric power systems. FUTURA is used in Peru nowadays for the study of –Valor Agregado de Distribución 2001 (VAD-2001)-yard stick competition- in order to obtain the electric rates in Peru for the next 4 years.

2 Algorithms for electric load forecasting

Many algorithms have been proposed in the last few decades for performing accurate load forecast. The most commonly used techniques include statistically based techniques, expert systems approaches and artificial neural network algorithms(ANN). The time series and the regression techniques are the two major classes of conventional statistical algorithms, and have been applied successfully in this field for years. The expert systems based algorithm use a symbolic computational approach to automating intelligence. This approach takes advantage of the expert knowledge of the operator. Thus, the load forecasts are obtained based upon logic relationships (typically rules) among the diverse indicators such as GDP (Gross domestic product-PBI) [20],[21], population growth, rates, weather, etc. The application of fuzzy systems also allows to treat theses contexts' own imprecision, and it is often done paying attention to probabilistic criteria as in [2]. Other widely used techniques are the artificial neural networks [2],[4],[5],[9],[12] and the combination of these with fuzzy systems [2]. A major advantage of using ANN over expert systems is its non-dependency on an expert. The principal disadvantage of the systems based on neural networks is that once the system is trained and by any chance the system changes its topology (reconfiguration of the electric network) the system simply does not respond adequately [3],[4],[9],[15],[16]. Generally time series approaches assume that the load can be decomposed into two components. One is the weather dependent and the other is weather independent. Each component is modeled separately and the sum of these two gives the total load forecast. The behaviour of the weather independent load is mostly represent by Fourier series or trend profiles in terms of time func-

tions. The weather sensitive portion of the load is arbitrarily extracted and modeled by a predetermined functional relationship with weather variables.
The proposal of using deterministic annealing (DA) clustering is used to classify the input data in clusters [18]. DA is very similar to simulated annealing (SA) since they both start from the analogy of statistics mechanisms. The main difference between both is that SA uses a probabilistic search while DA uses a deterministic technique. The idea of using methodology belonging to Case-Based Reasoning (CBR), as it is presented in the following sections, is new in this field although it has been proposed for the forecast in other fields (for example [14]).

3 Combination of knowledge-based methods and numeric methods for electric load forecasting

The use of individual methodologies, either based on numeric computing (like curve fitting) or AI techniques, has given partial results [2],[3],[5],[8] for the electric load forecasting. Then it is proposed the development of a system which combines the advantages of each method involved. In real terms the developed work uses firstly the Curve Fitting (CF) technique to obtain a forecasting polynomial curve based on a criterion of quadratic error minimizing. By using such technique a gross or softened forecast is obtained. Next this curve is corrected for each day by applying the methodology of the case-based reasoning (CBR). Since the consumption under normal conditions responds to a monotonous growing function, last year data is recovered to adjust to the result obtained through CF considering an adaptation of such data according to the values that certain indicators (GDP growth, population, urban growth, inflation, electric rates, etc.) take in the forecasting point. For the case adaptation an expert system (ES) is proposed to be used. Following, these forecast stages are detailed.

3.1 Polynomial Model Curve Fitting (CF)

It is based on the extrapolation of historic loads. A strong computing support is required in order to handle a wide data range. In general the Curve Fitting method is applied with the purpose of extrapolating peak loads because of two reasons: First, the peak load is the most important value for planning since the peak load is the one which has a higher impact in the system requirements. Second, the electric companies can easily obtain the annual peak load because they keep records of it.

$$L_n(t) \approx a_n t^3 + b_n t^2 + c_n t + d_n \tag{1}$$

$L_n(t)$ represents the load for the substation n in the instant t. $C_n = \lfloor a_{n,} b_n, c_n, d_n \rfloor$ are the coefficients of the polynomial which better approximate to the load in such substation n. These coefficients are obtained by minimizing the quadratic error mean from the polynomial approximation for the available measurements (vector Ln) according to this expression:

$$C_n = \left[P^T P\right]^{-1}.P^T L_n \tag{2}$$

Being P the sampling matrix (rectangular) defined as:

$$P = \begin{bmatrix} 1^3 & 1^2 & 1^1 & 1^0 \\ 2^3 & 2^2 & 2^1 & 2^0 \\ \vdots & \vdots & \vdots & \vdots \\ N^3 & N^2 & N^1 & N^0 \end{bmatrix}_{N\mathrm{x}(3+1)} \tag{3}$$

With the purpose of reducing the computing cost, FUTURA realizes a data preprocess which consists of considering only the highest and lowest values of the measurements in the peak hour (typically it is taken the 19:00 h according to the EPRI recommendations) and they are separated by days of the week [12]. Three forecasting curves are obtained with them for each day of the week (dL, dM, dX, dJ, dV, dS, dD). These curves are the highest value curve, the lowest value one, and a third one obtained from the mean of the former ones. Let's call $L_{CF}(t)$ to the forecast obtained in the instant t from the polynomial Cn. As it was stated, the interest in the peak hour is based on that the use of the forecast will be the contracting, and if sometime this is inferior to the consumption, the supplying company will punish the customer.

3.2 Case-Based Reasoning Hybrid System

The previous method is obtained from the criteria of quadratic error mean. Then the forecast obtained will be valid as demand tendency or gross estimation. With the purpose of refining this forecast the case-based reasoning has been used as a methodology which proposes to retrieve real data that responds to a similar description of the current problem to be solved.

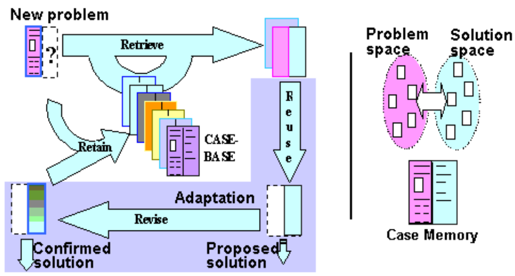

Fig. 1. Cases and CBR cycle: Retrieve, Reuse, Retain and Revise [1]

In our case the problem definition is determined by the forecasting point, the forecasting instant and the contextual conditions (political, economic, etc.) relative to the forecasting point and instant. The forecasting point (substation n) will be the one which will determine the use of one historical data base or another. (Ln(t)=L(t)) is not considered next because we have worked only with substations where previous records existed. In respect to the forecasting time, it will permit differentiate for each year (a) among the months (m) of the year (mEn, mFb, mMr, mAb, mMy, mJn, mJl, mAg, mSp, mOc, mNv, mDc) and the days (d) of the week (dL, dM, dX, dJ, dV ,dS, dD, dE). Where dE corresponds to special days as national holidays and the special days relative to exceptional situations such as elections, catastrophes, sports events (soccer finals, Olympic Games, etc.) and similar events. For each day hourly periods (h) of 15 minutes are differentiated according to the cadence of the available records.

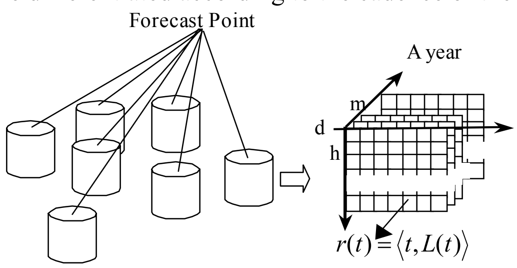

Fig. 2. Historic data in the forecasting points

This way we consider the load records r(t) as instances which associate time instants (t) with load values L(t) in such instants.

$$r(t) = \langle t, L(t) \rangle \quad con \quad t = (a, m, d, h) \tag{4}$$

Once the forecasting curves are obtained through the Curve Fitting method (LCF), these are adapted according to the following relationship:

$$L_P(t) = L_{CF}(t) \cdot \overline{L}_{CBR}(t) \tag{5}$$

Where $\overline{L}_{CBR}(t)$ represents the retrieved and adapted value for each instant to be forecasted.

3.3 Case Retrieval

Up until now the algorithm used for the retrieval is designed to assist medium-term forecasting needs. It can be different if the short term would be considered. This considers the weather and season characteristics in Peru, where the variations in a month are not significant and thus neither are the electricity consumption characteristics. On the contrary the variations which depend on the day of the week and the time are in fact important. From these considerations the daily consumption curves for each month to be forecasted are recovered (15 minutes x 24 hours = 96 points per curve). They are averaged according to the day of the week. That is, the criterion of the k-Nearest Neighbours is considered for the retrieve of a whole month (k = nsem., number of weeks of the recovered month), considering as likeness criterion the nearest year (typically the previous one) and the coincidence of the attributes month (m) and day (d). In order to soften the possible errors, the recovered values for each day of the week are averaged. The figure 3 shows this average of the cases for each day of the week, where m represents the months, d represents the days, and k is an interval of 15 minutes (measurements every 15 minutes).

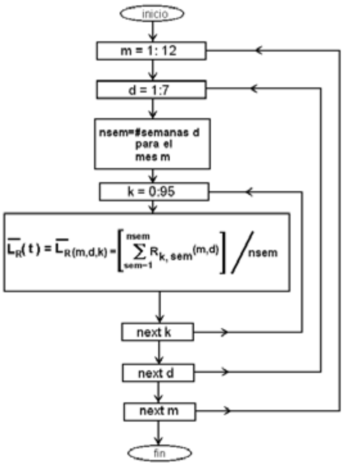

Fig. 3. Case Retrieve for a month

$\overline{L}_R(t)$ is the average case recovered and normalized in respect to the highest value for each day of the week (dL, …, dD). The normalization is necessary to be able to do a modification of the coherent forecast CF and according to the equation (5). The special days (dE) are put aside and do not intervene in the average.

3.4 Expert System for the adaptation of cases

Once the average case , $\overline{L}_R(t)$, is retrieved for each day of the week (dL, …, dD) in the month to be forecasted and with values every 15 minutes, we proceed to its adaptation according to a factor (K(t)) obtained from considering the socioeconomic context of the forecasting point.

$$\overline{L}_{CBR}(t) = \overline{L}_R \cdot K(t) \tag{6}$$

The adaptation factor K(t) is obtained through an expert system (ES) which evaluates the socioeconomic indices of the measurement point in respect to what is considered as a normal situation. This way a fine adjust of the forecasting curve ($L_P(t)$) is obtained for the month under study from the previous values retrieved and adapted. The knowledge base of the expert system (ES) developed for this purpose comes from the knowledge of network operators as long as reports of the electric companies and documents of the regulating entity – Gerencia Adjunta de Regulación Tarifaría (GART- OSINERG) del Perú [13],[17],[20],[21]. The knowledge base considers rules for special loads such as mines, refineries, The Child Effect (fenómeno del niño), etc.

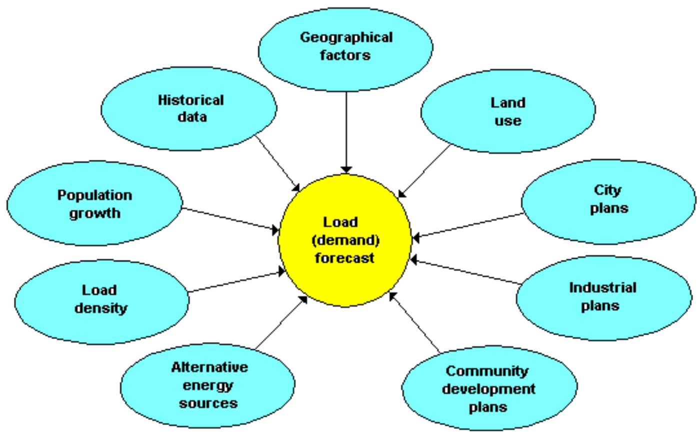

Fig. 4. Factors affecting the load forecast

The adaptation factor K(t) is obtained through rules such us:

Rule 1: IF $(4 < GDP \le 5)$ THEN $K_{gdp} = 1.025$ $cf_{gdp}=90\%$

Rule 2: IF $(3 < POP \le 4)$ THEN $K_{pop} = 1.0015$ $cf_{pop}=87\%$

… ….

Rule n: IF A THEN K_a $cf_a\%$

$$K(t)=K_{gdp}*K_{pop}*K_{ug}*K_{nf}*K_{inf}*K_{er}*K_{in} \tag{7}$$

The K(t) factor under usual situations can introduce variations of up to 20% over the recovered cases. In exceptional cases these variations can reach to 60%. That is the case when considering the effects of the Child Effect -fenómeno del niño (with a rhythm of approximately 10 years).

4 AN EXAMPLE

The next graphic belongs to a load forecast in which the power engineer wants to know the load diagram for October 5 of 2000. In order to do this task we will use the available data for two years (January 1998 –July 2000). It is necessary to mention that the forecast given by FUTURA will be done for the highest demand, average demand, and lowest demand. Under strict rigor the forecast which seems to be interesting is the average value, but for distribution companies the most important forecasting values are the highest and the lowest ones since their energy purchase contracts include a range where there is no overcharge for consumption out of place.

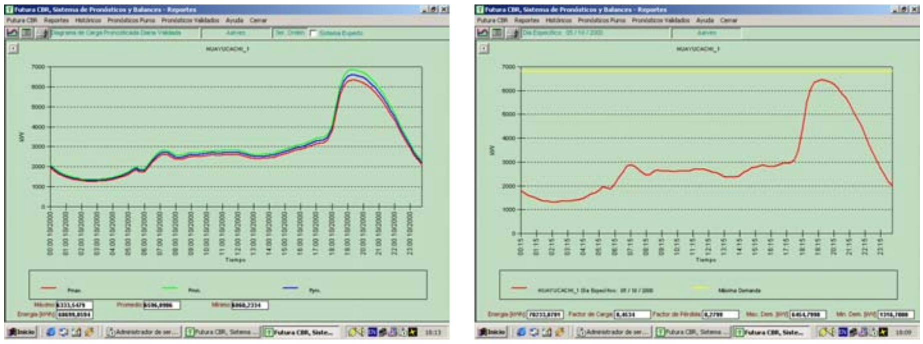

Fig. 5. Load diagram forecasted and real load diagram for October 5

To evaluate the medium term load forecast performance of the proposed methodology it is used to forecast two months a head for one week in October. The load forecast error by days is shown below:

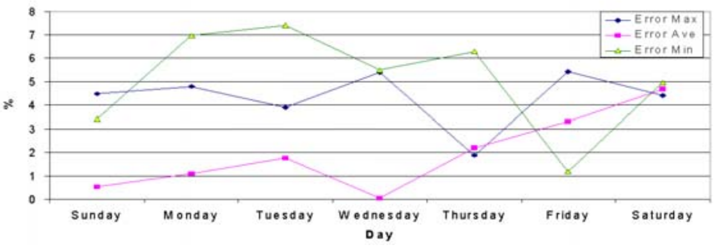

Fig. 6. Error for a week in October.

The figure 7 shown a comparative study of FUTURA in constrast with other models. Met 1: uses CF 2 with real data. Met 2: Uses CF2 with validated data. Met 3: Use CF3 with real data.

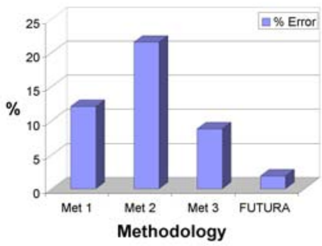

Fig. 7. Comparative study for one week

With FUTURA model we obtained a 1.86% of average error. In our proposed model we use the term "validated forecasting", it means that abnormal weather and systems conditions, such as thunderstorms or distribution outages are treated as abnormal events with bad real-time readings and are not considered.

6 Result exploitation

In Latin America the methodology called Valor Agregado de Distribución –VAD or yardstick competition is mainly used to determine the income gotten from the distribution activity. In it the distribution company competes against an efficient model company. This analysis due to its size and importance is done every four years. Owing to FUTURA's described characteristics, this methodology was used to carry out the VAD 2001 study in Peru for the sectors 3 and 4. The contract was obtained after a rigorous competition, and FUTURA was selected because of its good results in forecasting errors.

Referring to the implementation of the CBR methodology, under this environment, it has been done from the knowledge analysis to be used according to the containers proposed by Richter [19]. The analysis of the sector's own vocabulary and the existing records in the standard data base have led to the definition and case indexation that have been exposed. The mechanism to gather data is done by the temporary neighbor criterion. However, some other alternatives are being studied. Finally, the adaptation of these is done by the ES based on the socioeconomic indices of the measurement point.

7 Conclusions

The proposed methodology improves substantially the electric load forecasting against methods which work individually. This happens because it benefits from the advantages of each method. The advantages contributed by the CBR methodology resides in the flexibility of the knowledge representation (this can be redistributed among the four containers –according to the application needs). We can also mention fewer maintenance efforts, reuse of real measurements, continuous learning, time improvement and adaptation to changes in the environment. The expert system module uses the expert engineers' knowledge in demand growth. The knowledge base can be done as sophisticated as the power systems analyst wants. In order to get the best performance at medium term forecast we use a third grade polynomial. It is important to use data in adequate conditions without abnormal weather and systems conditions, such as thunderstorms or distribution outages. FUTURA is in conditions to administer great demand information amounts. It permits to do consultations and to make reports by magnetic or printed means.

8 Acknowledgements

We thank to the help given by Agencia Española de Cooperación Internacional AECI/ALE 2002 and the resources given by the project "Supervisión experta de la calidad de servicio, SECSE. DPI2001-2198 del MCYT español.

References

1. Aamodt, A., Plaza, E.(1994) Case-Base Reasoning: foundational issues, methodological variations and system approaches. AI Communications. IOS Press, Vol. 7: 1 pp 39-59.
2. Bartkiewicz, W. (2000) Neuro-Fuzzy Approaches to Short-Term Electrical Load Forecasting. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN'00).
3. Bartkiewicz, W. Gontar Z. and Zielinski, J. (2000) Uncertainty of the Short-Term Electrical Load Forecasting in Utilities. Proc. of the IEEE-INNS-ENNS -IJCNN'00.
4. Carpentiero, O., Silva A., Feichas C. (2000) A Hierarchical Neural Model in Short-Term Load Forecasting. Proceedings of the IEEE-INNS-ENNS - IJCNN'00.
5. Charytoniuk, W. y Mo-Shing Chen (2000). Very short-term load forecasting using artificial neural networks. IEEE Trans. on Power Systems, Vol. 15 No. 1, 263 – 268.
6. Charytoniuk, W.; Chen, M.S.; Van Olinda, P. (1997) Nonparametric Regression Based Short-Term Load Forecasting. U. of Texas at Arlington. PE-891-PWRS-0-08-1997
7. Comisión de Tarifas de Energía (2001) Estudio de Costos del Valor Agregado de Distribución. Gerencia Adjunta de Regulación tarifaría(ex CTE) – OSINERG. Lima.
8. Dabbaaghchi I, Cristie R., et al. (1997). Al Aplication Areas in Power Systems, IEEE Expert, January February
9. Dash, P.K. (1997) A real time short term load forecasting system using functional link network. IEEE Transactions on Power Systems, Vol. 12, No.2, 675-680.
10. De la Rosa, J.(1999) Sistemes experts a temps real, Publicación de Universitat de Girona, España.
11. Electric Power Research Institute (1979) Research into Load Forecasting and Distribution Planning, EPRI Report EL-1198, EPRI, Palo Alto, CA.
12. Garcia, A et al.(1994) A Neural System for Short-Term Load Forecasting Based on Day-Type Classification, Proc. ISAP94, pp. 353-360.
13. GART-OSINERG. Anuario Estadístico 1996-1997-1998-1999-2000. Gerencia Adjunta de Regulación Tarifaría.- Organismo supervisor de la inversión en energía GART OSINERG. Lima – Perú.
14. Hansen, B. (2000) Analog forecasting of ceiling and visibility using fuzzy sets. Maritimes Weather Centre, Darmouth, Nova Scotia.
15. Hong-Tzer, Y, et al(1996) Identification of ARMAX Model for short term load forecasting: an evolutionary programming approach. IEEE Trans. on Power Systems, Vol. 11, No. 1, 403-408.
16. Khotanzad, A. et al. (1998) ANNSTLF – Artificial neural network short-term load forecaster – generation three. IEEE Trans. on Power Systems, Vol. 13, No.4, 1413-1422.
17. Lee, H. (1996) Spatial Electric Load Forecasting, Marcel Dekker, Inc.
18. Mori, H., y Yuihara, A. (2001) Deterministic Annealing clustering for ANN-based short-term load forecasting. IEEE Trans. on Power Systems, Vol. 16, No. 3, 545-552.
19. Richter, M. The Knowledge Contained in Similary Measures. University of Kaiserlautern.
20. Vilcahuamán, R. Davila, W. et al. (1999) Sistema experto para el pronostico de la demanda. CONIMERA XXV, Lima, Perú.
21. Vilcahuamán, R. Medina, I. y Trelles, A. (2000) PRONOS: Sistema experto para el pronostico de la demanda. Facultad de Ingeniería de Sistemas. Universidad Nacional del Centro del Perú.
22. Third international conference on case-based reasoning (1999) Engineering Applications of Case-Based Reasoning. Special Issue of The International Journal. Engineering Applications of Artificial Intelligence. Volume 12(6) 1999.