Forecasting of Energy Consumption Based on Gaussian Mixture Model and Classification Techniques

Pronóstico de Consumo Energético Basado en Modelo De Mezcla Gaussiana y Técnicas de Clasificación

D. Imbajoa[1], A. Arciniegas[2], J. Revelo[3], D. H. Peluffo-Ordóñez[4], P. D. Rosero-Montalvo[4]

[1]GIIEE Research Group, Ingeniería Electrónica, Universidad de Nariño, Colombia. Email: deivy311@udenar.edu.co
[2]GIIEE Research Group, Ingeniería Electrónica, Universidad de Nariño, Colombia. Email: andresarciniegas@udenar.edu.co
3GIIEE Research Group, Ingeniería Electrónica, Universidad de Nariño, Colombia. Email: javierrevelof@gmail.com
[4]Ingeniería Electrónica en Redes de Comunicaciones - FICA, Universidad Técnica del Norte, Ecuador.
Email: dhpeluffo@utn.edu.ec, pdrosero@utn.edu.ec

RECIBIDO: abril 21, 2017. ACEPTADO: junio 02, 2017. VERSIÓN FINAL: noviembre 01, 2017

ABSTRACT

The estimation of energy demand is not always straightforward or reliable, as one or several classes may fail in the prediction. In this study, a novel methodology of load forecasting is proposed. Three different configurations of Artificial Neural Networks perform a supervised classification of energy consumption data, each one providing an output vector of unreliable predicted data. Under the clustering method k-means, multiple patterns are identified, and then processed by the Gaussian Mixture Model in order to provide higher relevance to the more accurate predicted samples of data. The accuracy of the prediction is evaluated with the several error rate measurements. Finally, a mixture of the generated forecasts by the methods is performed, showing a lower error rate compared to the inputs predictions, therefore, a more reliable forecast.

PALABRAS CLAVE: Demand Forecasting, Artificial Neural Networks, Gaussian Mixture Model, k-means, Support Vector Machine

RESUMEN

La estimación de demanda energética no siempre es sencilla o confiable, pues una o varias clases pueden fallar en la predicción. En este estudio, una nueva metodología de pronóstico de carga es propuesto. Tres diferentes configuraciones de Redes Neuronales Artificiales realizan una clasificación supervisada de datos de consumo de energía, cada una aportando un vector de salida de información de predicción poco confiable. Bajo el método de clustering *K-Means,* se identifican diferentes patrones, que son luego procesados por un Modelo de Mezcla Gaussiana, para proporcionar mayor relevancia a los datos predichos y acertados. La precisión de la predicción es evaluada con diferentes medidas de error. Finalmente, se realiza una mezcla de los pronósticos generados, mostrando una tasa de error más baja que las predicciones de entrada, y, por consiguiente, un pronóstico más confiable.

KEYWORDS: Pronóstico de demanda, Redes Neuronales Artificiales, Modelo de Mezcla Gaussiana, k-means, Máquina de Soporte Vectorial

1. INTRODUCTION

In recent years, facts such as electrical power supply in non-interconnected areas, policies of environmental care, and population growth have led to the continuous increase of the electricity demand. Particularly, in Colombia, 71% of its 16.000 MW of installed capacity is generated by hydroelectric power plants. 2015 had a peak load of 10.000 MW [1]. Nevertheless, meteorological phenomena as *El Niño* have considerably reduced the contribution of the aforementioned plants by approximately 25%. This lack of energy has been mainly covered by thermal power generation (coal, gas, and liquids). In addition, the participation of non-conventional renewable

IX Simposio Internacional sobre Calidad de la Energía Eléctrica SICEL 2017

Noviembre 1, 2 y 3
Bucaramanga Colombia

energies is close to 3%, which is very small, given the listed high potential for wind and solar power in Colombia [2].

Contemplating global issues as the increasing pollution and climate change, renewable energies have been penetrating in energy portfolios as a relevant power solution. However, they are highly volatile as they depend on the weather, posing problems associated with the dispatchable generation and the grid reliability in general. Accordingly, high accuracy forecast systems are required for planning the expansion or reduction of the production capacity [3], and it is evidently correlated to the energy price in the market. For this reason, electricity forecasts have become a fundamental input to energy companies' decision-making mechanisms [4].
A vast collection of methods has been tested for electricity forecasting, both regarding demand and offer capacity. Some of these techniques include computational intelligence models, such as Artificial Neural Networks, Fuzzy Neural Networks and Support vector machines, which have taken place to solve problems as traditional methods (namely., statistical) may fail or work improperly. They combine learning and evolution elements to match the often-called nonlinear models and dynamic systems representing the time series to be predicted. The presented approach takes some of these techniques and performs a mixture through the Gaussian Mixture Model, to achieve better forecasting results.

The rest of this paper is organized as follows: Section II highlights relevant works in forecasting through a mixture of models. Section III outlines the input data, the individual methods included in the next performed mixture process, and describes the methodology of the proposed approach, as well as the optimization process for weights tuning. Some experimental results are shown in Section IV. Finally, Section V gathers the final remarks, and Section VI some recommendations for future works.

2. RELATED WORKS

The idea of combining forecasts goes back to the late 1960's, with a relevant success. Nonetheless, despite its popularity, the combination of forecasts has not been discussed extensively in the context of electricity markets to date. In contrast, there has been much observed in different areas of information, especially Economy and meteorological sciences. Some interesting remarks are gathered in [4], and a mixture of different opinions in Multi-labeler scenarios is developed in [11].

3. PROPOSED LOAD FORECASTING APPROACH

Hourly electricity consumption information is stored in an N-sized vector $\mathbf{y}$, where N is the number of measurements in the time series. Three different forecasting methods are used to generate three consumption prediction vectors $\hat{\boldsymbol{y}}_i$. These can be gathered into an $N \times 3$ data matrix $\hat{\mathbf{Y}}$ such that $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \hat{\mathbf{y}}_3]$. Since the objective of this work is to display how the mixture produces a more accurate forecasting approach than its individual components, these predictions have been tuned solely by a heuristic process with 80%, 70% and 60% of training, and not through a dedicated optimization technique.

Table 1. Input Parameters of Classifiers.

	X	Parameters description
xr	$\mathbf{x}_1$	Hour of the day. 24-hour format
	$\mathbf{x}_2$	Working day identifier
	$\mathbf{x}_3$	Day of the week
xc	$\mathbf{x}_4$	Load at same time the day before
	$\mathbf{x}_5$	Load at the same hour in previous week
	$\mathbf{x}_6$	Local holidays identifier
	$\mathbf{x}_7$	Cumulative load since previous 24 hours
	$\mathbf{x}_8$	Cumulative load since previous week (168 hours)

Where x_i represents the i^{th} vector of matrix $\boldsymbol{X}$.

3.1 Input data

The described method is a data driven model. Besides any given configuration, the identification of the size and frequency of the selected data is a crucial step for the proper operation of a prediction method [5]. A matrix $\boldsymbol{X}$ is composed by generated input vectors described in Table I. In this case, as each x_i array of $\mathbf{x}_4$ and $\mathbf{x}_8$ depends on the previous 168^{th} sample and we´ll define as xc, the size of $\boldsymbol{X}$ is $(N-168) \times d$, where d is the number of input vectors, with $d = 8$ in this respect.

3.2 Unsupervised classification

This type of classifications is commonly used in scenarios where it is required an estimation of density, in order to identify certain features from the data. It is known as *clustering* and it is used to find natural groupings in data. One method of these type is called k-means, which seeks local rather than the global minimum solutions. It is considered a solution a set of certain conditions where the mathematical criterion is minimized.

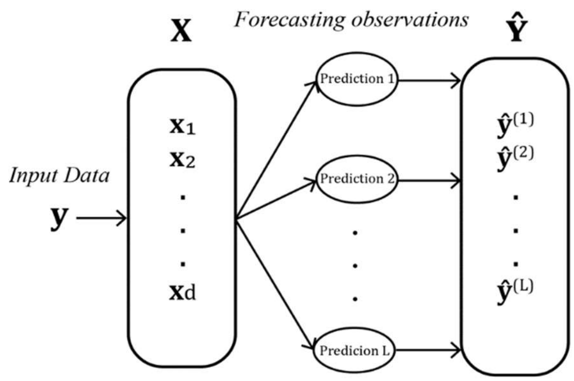

Figure 1. Formation on Forecasting Observations from input data. Adaptada de [12]

Its simplicity and pragmatism have made it a prominent method in machine learning. In this study, this method provides support to prediction methods.

A number of clusters $K \geq 2,$ is determined by the Calinski Harabasz criterion [13], which is a technique to identify the presence of clusters, consisting of points, in a multidimensional Euclidean space, and its applications in the considered taxonomy. For shorthand notation, we use:

$$\boldsymbol{C} = Kmeans(\boldsymbol{X}, K), \tag{1}$$

where C represents the clusters to be used to train the supervised classification model, aiming at the identification and positioning of centroids q_j, which fulfils the condition:

$$2 \leq \mathrm{j} \leq \mathrm{K}. \tag{2}$$

3.3 Supervised classification

In machine learning, Supervised Classification, is the most common scenario associated with regression and ranking problems. The classifier receives a set of labeled examples as training data and makes predictions for all the unseen and desired points. Different information contexts, used in different scenarios, are unsupervised learning, semi-supervised learning, transductive inference, and others [6].

Support Vector Machine (SVM): This technique has shown to be a suitable alternative to approach this problem, mainly due to their versatility regarding supervised classification. This process trains an SVM model to be defined as SVM_t, with $\mathbf{X}$ and $\boldsymbol{C}$ as input parameters.

$$\boldsymbol{C}^l = (C_1^l \;\; C_j^l \;\; \cdots \;\; C_K^l) = SVM_t(\widehat{\boldsymbol{X}}^{\boldsymbol{l}}) \tag{3}$$

Where $\widehat{\boldsymbol{X}}^{\boldsymbol{l}}$: $[xr, \mathrm{xc}^l]$

$\hat{x}_i^{(l)}$: Where x_i represents the i^{th} vector of the matrix $\widehat{\boldsymbol{X}}^{\boldsymbol{l}}$.

$xc^{\boldsymbol{l}}$: Depends on each observation, see *Table 1*

$\mathbf{C}^{\boldsymbol{l}}$: Clusters for each observation.

C_1^l : Cluster for each observation centered in q_j^l

Artificial Neural Network (ANN): It is defined as information processing systems which have common specific characteristics associated with biological networks, in order to achieve more robust performance [7]. They have been successfully applied in multiple different fields, and particularly in nonlinear regression models and forecasting [8].

A Multilayer Perceptron, or MLP model, is configured with a Layer of d input vectors, a layer M of output neurons, and one or more hidden layers, which is described in Figure 1. Under this structure, the connections between neurons feed forwards invariably. Generally, a sigmoid function is used in the neurons of the hidden layer, to provide the capability of learning potential nonlinear functions to the network. The MLP training is considered a supervised technique, and can be developed using the Levenberg-Marquardt back propagation; the classical gradient descent algorithm, or nonlinear optimization algorithm, in order to accelerate the convergence speed of weights [9].

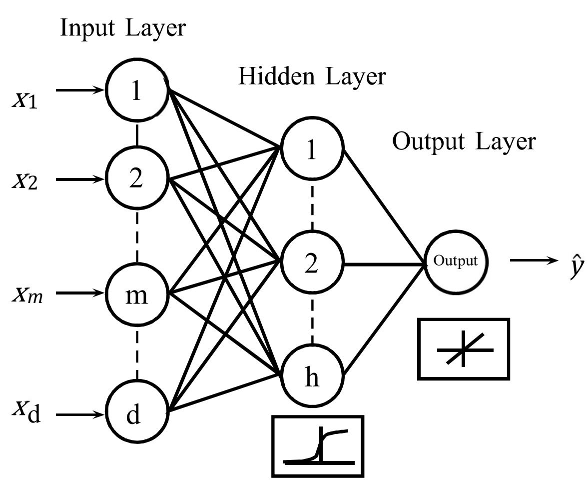

Figure 2. Multilayer Perceptron in Artificial Neural Network

The used model in this work is an MLP model with one hidden layer, implementing sigmoidal activation functions in the hidden layer, which was configured with twenty neurons, with a single output neuron. The training method used in this approach was the Levenberg-Marquardt backpropagation.

3.4 Proposed Load Forecasting Approach

1) Prediction parameters setting:

Matrix $\mathbf{Q}$ is generated by a set of Gaussian distributions centered in $q_j^{(l)}$, formed by clusters belonging to $\boldsymbol{C}^l$, as described in Section 3.3.

$$\mathbf{Q} = \begin{pmatrix} q_1^{(1)} & q_1^{(2)} & \cdots & q_1^{(L)} \\ q_2^{(1)} & q_2^{(2)} & \cdots & q_2^{(L)} \\ \vdots & \vdots & \ddots & \vdots \\ q_K^{(1)} & q_K^{(2)} & \cdots & q_K^{(L)} \end{pmatrix} \quad (4)$$

The proposed model is limited by the following restriction, which guarantees the existence of at least one cluster found in equation 1:

$$p(\boldsymbol{Q}) = \sum_{k=1}^{\boldsymbol{K}} \sum_{l=1}^{L} q_{\mathrm{k}}^{(l)} = 1. \quad (5)$$

2) Forecasting model: The Gaussian Expectation Maximization Clustering (GEMC) is part of the Density-Based Clustering (DBC) methods, and has as an objective function the linear combination of Gaussian distributions centered on the centroids of each group. The respective membership functions of each element are:

$$m_{GEMM}\left(q_j^{(l)}/x_i^{(l)}\right) = \frac{p\left(\hat{x}_i^{(l)}/q_j^{(l)}\right)p\left(q_j^{(l)}\right)}{p\left(\hat{x}_i^{(l)}\right)} \quad (6)$$

Notice that the membership function is a probability value, thus Bayes' rule can be applied to its estimation. Considering $p\left(\hat{x}_i^{(l)}\right)$ as the probability of occurrence of an event $\hat{x}_i^{(l)}$, q, is mathematically described as follows:

$$\eta^{(l)}\left(\hat{x}_i^{(l)}\right) = p\left(\hat{x}_i^{(l)}\right) = \sum_{j=1}^{k} p\left(\hat{x}_i^{(l)}/q_j^{(l)}\right)p\left(q_j^{(l)}\right) \quad (7)$$

In this approach, $\eta^{(l)}\left(\hat{x}_i^{(l)}\right)$ is used to represent the weight for each event $x_i^{(l)}$, corresponding to each predictor $\hat{y}^{(l)}\left(\hat{x}_i^{(l)}\right)$. The term $p\left(\hat{x}_i^{(l)}/q_j^{(l)}\right)$ is the probability of occurrence of an event $\hat{x}_i^{(l)}$ given a gaussian distribution, centered in centroid of $q_j^{(l)}$j.

The probability a priori of a group $p\left(q_j^{(l)}\right)$, whose centroid is $q_j^{(l)}$, is depicted as:

$$p\left(\hat{x}_i^{(l)}/q_j^{(l)}\right) = \frac{1}{\det\left(\Sigma_j^{(l)}\right)^{\frac{1}{2}}}(2\pi)^{-d/2}e^{-\frac{1}{2}\left(\hat{x}_i^{(l)}-\mu\right)\left(\Sigma_j^{(l)^{-1}}\right)\left(\hat{x}_i^{(l)}-\mu\right)^{\mathrm{T}}} \quad (8)$$

where μ is the arithmetic mean of cluster centered in $q_j^{(l)}$; the term denotes the argument matrix determinant. is the dimension; Σ represents the covariance matrix, and $det(\cdot)$ denotes the argument matrix determinant.

$$\hat{y}_{\mathrm{F}}(\widehat{\boldsymbol{X}}) = \sum_{l=1}^{L} \eta^{(l)}\left(\boldsymbol{X}^{l}\right) \cdot \hat{y}^{(l)}\left(\boldsymbol{X}^{l}\right) \quad (9)$$

Finally, each performed prediction is affected directly by the weights provided by the proposed method, as depicted in Equation 9, obtaining $\hat{y}_{\mathrm{F}}(\boldsymbol{X})$ as final result.

4. RESULTS AND DISCUSSION

4.1 Test Database

The effectiveness of the introduced method is applied to two different data sets, represented by vectors $\mathbf{y}_1$ and $\mathbf{y}_2$. The first series belongs to the hourly energy consumption of a medium size hospital, with $N = 2256$. The second and larger data set belongs to a big local supermarket with $N = 6166$. These were furnished from the local energy company database, and are express in Kilowatts-hour (kWh), accumulated hourly.

4.2 Performance measures

To evaluate the efficiency of the method, several measures of accuracy applied to univariate time series data are analyzed. Let y_t denote the observation at time t, and $\hat{y}_t$ denote the forecast of y_t. Then the forecast error is defined as $e_t = y_t - \hat{y}_t$. The most widely used measure of fit in the field of time series forecasting is the Mean Absolute Percentage Error (MAPE), shown in equation 8. To support the results of the presented technique, several others measures of fit were evaluated, such as the scale-dependent Mean Square Error (MSE) and Root Mean Square Error (RMSE) in equations 6 and 7, described broadly in [10]. Finally, standard deviation, symbolized by σ, is evaluated as well as a measurement of data concentration.

$$MSE = \frac{\sum_{t=1}^{N}(e_t^2)}{N} \quad (10)$$

$$RMSE = \sqrt{MSE} \quad (11)$$

$$MAPE = \frac{\sum_{t=1}^{N}\left|\frac{e_t}{y_t}\right| * 100}{N} \quad (12)$$

4.3 Experimental results

The forecasting accuracy in terms of the aforementioned measurements is shown in **Table 2** and **Table 3**. There is an overall reduction of the error rates among the observations. The proposed method, highlighted in yellow, shows a slight, but important, improvement in the classification. A result comparison is depicted in **Figure 3**, where the MSE values of Dataset No. 1 are a plot in a set of histograms where the concentration of the data is clearly displayed. Associated with Table 2, the low value of σ, confirms the shape of the charts, where an MSE value of 1.8429 states a remarkable result.

Table 2. Comparative table. Prediction methods results

y_1	ANN 1	ANN 2	ANN 3	G.M. Mixture
MSE	2.3501	1.9262	2.3950	**1.8429**
RMSE	2.5247	2.9741	3.0178	**2.5644**
MAPE	26.3337	31.2309	31.6450	**27.0576**
σ	4.9050	5.9298	6.0030	**4.1547**

Table 3. Comparative table. Prediction methods results

y_2	ANN 1	ANN 2	ANN 3	G.M. Mixture
MSE	5.2845	4.2145	4.0124	**2.9184**
RMSE	7.8465	5.1251	6.112	**3.8512**
MAPE	32.5541	35.4488	36.0154	**30.1624**
σ	4.1243	5.4429	7.4581	**6.1972**

Analyzing the results of the applied method in Dataset No. 2, although the value of σ is not the lower, showing a greater dispersion of data, the average error of the resulting forecasted vector is lower than the respective inputs.

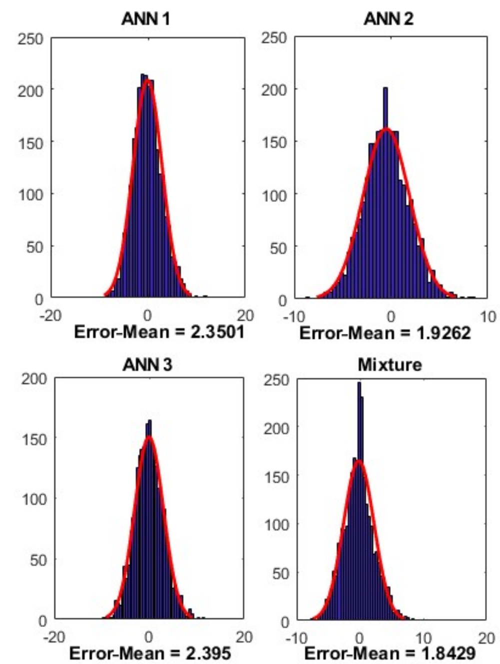

Figure 3. Comparison of Error Values (MSE) between the ANN generated inputs and Mixture in y_1.

5. CONCLUSIONS

This paper assesses energy consumption forecasts in a multiple observation of different predicting models scenario. It takes the example of three forecasting approaches derived from an Artificial Neural Network supervised classification and proposes a methodology to perform a combination of them, based on the Gaussian Mixture Model. This study shows the advantage of this procedure as it recovers the most accurate values of each input method, discarding the wrong approximations. Different methods for defining the ideal number of centroids for clustering may improve the results.

It is pointed out that the production of more accurate power consumption forecasts could be a key factor for the electricity market's industrial strategies, in order to implement effective controls as the Demand Response, or Smart Grid data managing.

6. RECOMMENDATIONS AND FUTURE WORK

The process of amalgamation may have a different number of observations as its input, coming from different and more statistical methods, besides ANN. In this paper, the observations were datasets generated from eight characteristic vectors with data from the original dataset only. With a proper selection of input variables, the proposed method could be extrapolated to different areas as energy price and power generation, including features as weather information and measures from the market economy.

Furthermore, it was observed that the time base and the forecasted period of time also affect the result, as there are long-term and short-term forecasting. Consequently, the input data must be evaluated along with the derived input vectors to establish the best forecasting strategy.

REFERENCES

[1] W. Nustes and S. Rivera, "Colombia: Territory for investment in nonconventional renewable energy to electric generation," *Revista Ingeniería, Investigación y Desarrollo*, vol. 17, pp. 37–48, 2017.

[2] E. Gaona, C. Trujillo, and J. Guacaneme, "Rural microgrids and its potential application in Colombia," *Renewable and Sustainable Energy Reviews*, vol. 51, pp. 125–137, 2015.

[3] R. H. Inman, H. T. Pedro, and C. F. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 51, pp. 125–137, 2015.

[4] R. Weron, "Electricity price forecasting: A review of the state-of-the art with a look into the future," *International Journal in Forecasting*, vol. 30, pp. 1030–1081, 2014.

[5] D. Keles, J. Scelle, F. Paraschiv, and W. Fichtner, "Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks" *Applied Energy*, vol. 162, pp. 218–230, 2016.

[6] M. Mohri, A. Rostamizadeh, and A. Talwalkar, "Foundations of machine learning*". MIT press*, 2012.

[7] R. Jammazi and C. Aloui, "Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling" *Energy Economics*, vol. 34, pp. 828–841, 2012.

[8] S. Samarasinghe, *Neural Networks for Applied Sciences and Engineering. From Fundamentals to Complex Pattern Recognition*, ser. 13. 6000 Broken Sound Parkway NW, Suite 300: Auerbach Publications, 2007.

[9] J. J. Montaño Moreno, "Artificial neural networks applied to forecasting time series," *Psicothema*, vol. 23, no. 2, 2011.

[10] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, pp. 679–688, 2006.

[11] Imbajoa-Ruiz, D. E., Gustin, I. D., Bolaños-Ledezma, M., Arciniegas-Mejía, A. F., Guasmayan-Guasmayan, F. A., Bravo-Montenegro, M. J., & Peluffo-Ordóñez, D. H. (2016, November). Multi-labeler classification using kernel representations and mixture of classifiers. In Iberoamerican Congress on Pattern Recognition (pp. 343-351). Springer, Cham.

[12] J. J. Montaño Moreno, A.P. Pol, P. M. G. *Artificial neural networks applied to forecasting time series.* Pscicothema. Vol. 23, No. 2, pp 322-329. 2011

[13] T. Caliński & J. Harabasz. *A dendrite method for cluster analysis.* Communications in Statistics Vol. 3, Iss. 1, 1974