Towards An Enhanced Backpropagation Network for Short-Term Load Demand Forecasting

Andrei D. Coronel[1], Cielito C. Olegario[2*], Bobby D. Gerardo[3], Ruji P. Medina[4]
[1]Ateneo de Manila University, Philippines
[2]Technological Institute of the Philippines, Quezon City, Philippines
[3]West Visayas State University, Philippines
[4]Technological Institute of the Philippines, Quezon City, Philippines
[*]corresponding authors: cielito.olegario@wmsu.edu.ph

Abstract – *Artificial neural networks (ANNs) are ideal for the prediction and classification of non-linear relationships however they are also known for computational intensity and long training times especially when large data sets are used. A two-tiered approach combining data mining algorithms is proposed in order to enhance an artificial neural network's performance when applied to a phenomenon exhibits predictable changes every calendar year such as that of electrical load demand. This approach is simulated using the French zonal load data for 2016 and 2017. The first tier performs clustering into seasons and classification into day-types. The second tier uses artificial neural networks to forecast 24-hour loads. The first tier results are the focus of this. The K-means algorithm is first applied to the morning slope feature of the data set and a comparison is then made between the Naïve Bayes algorithm and the k-Nearest Neighbors algorithm to determine the better classifier for this particular data set. The first tier results show that calendar-based clustering does not accurately reflect electrical load behavior. The results also show that k-Nearest Neighbors is the better classifier for this particular data set. It is expected that by optimizing the data set and reducing training time, the learning performance of ANN-based short-term load demand forecasting.*

Keywords: *Artificial Neural Network, Data Mining, Forecasting*

Article History
Received 14 January 2019
Received in revised form 21 January 2019
Accepted 21 January 2019

I. Introduction

Load forecasting is a difficult task because of the time-series data being used and the consequent non-linear relationships between the variables. However, it can be noted that this data also exhibits various seasonality levels marked by yearly recurring predictable changes[1]. Artificial neural networks have received considerable interest because of their application to a diverse range of applications including load forecasting. Short-term load forecasting (STLF) is defined as the prediction of load demands ranging from an hour to a week and is needed for distribution schedules, maintenance planning and scheduling, seasonal peaks and contingency analysis, unit commitment, and overall demand side management[2].

The multi-layer perceptron (MLP) is said to be the most widely used ANN-based predictive model for time series data[3]. This artificial neural network can be trained to predict and classify by applying learning algorithms such as the Backpropagation method which focuses on determining a local minimum of the error function. This model determines the local minimum of the error function by computing the error between the neural network output and pre-determined output. Weight adjustments are done between two connecting layers and the model propagates backwards from output to input layer[26,40].

Artificial Neural Networks are known to be computationally intensive to train. This is evident when large datasets such as that of daily load data is used. How then can the computation intensity of an artificial neural network be reduced such that there can be a reduction in training time? The researchers seek to address this question by adopting an approach that ensures the training, testing, and validation sets are at optimal. This is done by first applying a clustering algorithm to determine periodicity (seasonality) and then using an appropriate classifier on the resulting clusters to create day-types. This is on the premise that calendar information alone does not reflect load behavior.

The objective for this stage of the project is two-fold: (a) to determine the optimum number of clusters to reflect the load behavior of this data set and (b) to determine which between the Naïve Bayes and k-Nearest Neighbor algorithm classifiers can yield a higher degree of accuracy in categorizing consequent day-types.

II. Related Literature

Machine learning models with no previous domain knowledge and resembles systems of interconnected biological neurons which be trained to communicate with one another and do prediction and classification are known as artificial neural networks. This is done by mapping input patterns into model output vectors and doing weight adjustments such that each input vector is associated correctly to its corresponding exemplar output vector [13]. A neural network is said to be able to successfully interpret data and learn more quickly if the input sets undergo pre-processing prior to network training.

However it can be noted that there is a dearth in studies that have considered the impact of improving the quality of data on ANN modeling performance [4]. The proposed approach can therefore be applied to any forecasting problem where the phenomenon to be predicted exhibits yearly calendar-based periodic changes (seasonality)[1], [5].

The uni-directional neural activation flow is done layer by layer and the hidden layers accept the data which are modified through weight value adjustments where correct decisions are reinforced and incorrect ones are weakened. The output layer then accepts the adjusted value but they still undergo some weight modification on the output and hidden layer connection. The network learns through iteratively decreasing the error measure associated to the outputs of a given training sample[6]. The selected learning algorithm determines the appropriate number of hidden units whereas a stopping criterion or an acceptable range determines when the process stops[7].

The focus is ensuring that the ANN's outputs achieve accuracy such that they are as near as possible to its exemplars with the smallest amount of error produced by optimum weight combination[6]. Improvements in learning performance, specifically, the ability of the neural network to produce the nearest approximated solutions , have been achieved using various techniques such as early stopping criteria[7], activation function change[8], principal components analysis[9], and the application of clustering or classification algorithms.

The researchers have considered the most commonly used clustering and classification algorithms in their most basic forms in order to determine if the combined approach in itself can be an effective tool in achieving better approximations and reducing computational intensity by optimizing the data set.

A. K-Means Algorithm

Clustering is the process of using a similar criterion to group together data into sets in order to accomplish data reduction. The K-means algorithm can be considered as one of the easiest to use and the most popular among the clustering techniques[10],[11]. This unsupervised algorithm splits an anonymous data set into a fixed number of K clusters with each cluster having a centroid that represents an object within the said cluster[10].

Each data set object is assigned based on the nearest distance measurement of a particular cluster. The process of classification and centroid adjustment is continued until the centroid values are stabilized [9], [17]. The minimization of the squared error or total intra-cluster variance is the objective of the algorithm (1)

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \tag{1}$$

where: J = squared error function, n=number of cases, $\left\| x_i^{j} - c_j \right\|$ is the distance function, k=number of clusters, and c is the centroid for cluster j. The squared error function is the Euclidian distance function sum between g cluster centroid and its corresponding vector.

Studies such as those of Kim, Koo, and Park [12]. Cooley et al[13]. and Nuchprayoon [14] have used K-means clustering to determine the load profile of a system or to determine load patterns in Korea, Australia, and Thailand respectively.

To determine the optimal number of K clusters, researchers such as Dent et.al[2] and Viola[15] have used the elbow method where the within-cluster sum of squares is calculated and plotted graphically to determine the elbow that marks the optimal number of clusters.

Tanwar et al.[16] argue that better results can be yielded by clustering techniques if extracted features from the data are used. They used features like the minimum-max load, daily variance, peak load, morning slope, evening slope, and daily mean were used in their study.

In their study of six selected countries, they found out that using the morning slope (load difference between 10am and 6pm) gave the highest degree of correct clustering. This view is also supported by Hernandez[11] and will be used for this paper as well.

B. Classifiers

The classifiers for consideration in this study are considered the most basic and easiest to use, namely: the Naïve Bayes algorithm and k-Nearest Neighbors algorithm. A training set is used to define the various class label attributes that would classify a sample in the first stage.

The resulting model can be represented as mathematical formulae, classification rules or decision trees. An estimation of the model's accuracy can be made by determining the percentage of test set samples that are have been correctly classified [6].

K-Nearest Neighbors (k-NN) is a supervised learning algorithm that approximates the mapping function in order to produce accurate predictions that reflect precise classifications when new inputs are presented. This technique has also been used together with the Backpropagation algorithm for classification problems[17].

A suitable distance measure is to determine the k-nearest neighbor to be able to classify an input vector X. This vector is assigned to the class by finding the majority of k-Nearest Neighbors(k-NN). While there are various distance measurements, this paper uses the Euclidean distance. The training phase and the testing phase comprise the k-NN algorithm.

In the former, data values with associated labels representing the class are given to the classifier to train it. In the latter, the algorithm is given a set of unlabelled data points. The classifier then generates the k-nearest data values to determine the class of unlabelled data points.

The Naïve Bayes algorithm, like the k-NN method, is also a supervised learning technique. It follows an assumption of independence[18], [19].

This is on the premise that a particular class feature presence is not related to the presence of any other feature. This theorem enables the posterior probability P(c|x) calculation from the prior probabilities of predictor P(x) and class P(c) as well as the prior probability of predictor given a particular class P(x|c)

$$P(c|x)=[P(x|c)\ P(c)]/P(x) \tag{2}$$

Naïve-Bayes theorem's strengths lie in its ease for use, speed and that it is deal for categorical values[20], [21]. However, it is also said that zero frequency can occur if a variable's category is not seen in the training data set[22]. It has been used in a wide range of applications ranging from multi-media analysis[18], [19], [23] to medical diagnosis[24], [25].

Bayes, Kanojia and Motwani[20] in their studies compared both classifiers. They found that in doing transcript analysis, t the former performed better when the sample number and learning parameters are small. However, the Naïve-Base algorithm yielded better results as the sample set increased.

III. Methodology

The overall proposed training algorithm is based on the premise that electrical loads exhibit periodicity (seasonality) and can be classified into day-types which can be used to form predictive models such as those for short-term load forecasting.

The researchers propose the combination of clustering and classifying data mining techniques as an added step to the backpropagation method.

This method then becomes a two-phase procedure with the first phase being the selection phase (clustering and classification) and the second phase is the training part (Fig. 1). This paper is a work in progress and focuses on the results of the first tier which is the use of clustering (K-means) then classification.

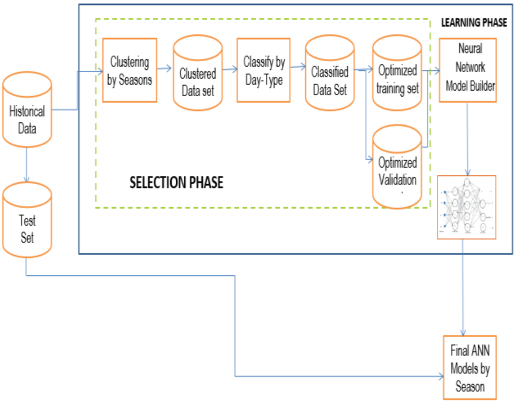

Fig. 1. Proposed approach to enhanced neural network training

The optimization of the input sets to be used the neural network model building (training phase) is expected to contribute to the improvement of the learning performance.

The final outputs of the said project are dedicated multilayer perceptrons that have been trained to forecast a 24-hour electrical load curve as draw from season and day-type.

The selection phase consists of the clustering process and the classification process. The primary data set is the ENSO-based hourly electrical load data of France for the period between 2016 and 2017.

This data set has been split into the training (70%) and validation (30%) sets. The difference between the 10a.m. and 6 a.m. readings otherwise known as the daily morning slopes have been used as the data set for clustering.

A. Clustering

1. With the objective of grouping the data into appropriate clusters, the morning slope and the average monthly minimum and maximum temperature values were used. Data was sorted into 2 to 5 clusters using XLStat in order to calculate the ideal number of clusters for K-means. Load curves were classified based on the Euclidean distance between their cluster centers. Each time this process was completed, a new cluster center point became updated according to the load profiles in the cluster. The process was repeated until a stable condition was met.
2. The elbow method has been used in order to determine what the optimum number of clusters is. A linear graph then represents the sum of squared errors plotted against the number of clusters. The elbow of the curve signifies the optimum number of clusters to be used for the analysis.
3. The K-means algorithm was then re-run. As a comparison, the first data set used consisted solely of daily morning slope data. The second set consisted of the daily morning slope readings with the inclusion of the minimum and maximum temperature recorded.

B. Classification Phase

For this phase, the entire 24-hour daily load data was used as the input sets for k-Nearest Neighbors and Naïve-Bayes respectively. This paper presents an initial simulation of one cluster for the classification process.

k-Nearest Neighbors Model Construction

1. Using the chosen K-means sample cluster, the data set was derived from the chosen K-means sample cluster and was divided into 70% training and 30% validation sets respectively. The Euclidean distance measure between a sample *A* and all the training set samples was then calculated.
2. *A* was then assigned by majority vote among its k nearest neighbors. The values of 1, 3, 5 and 7 k-neighbors were considered for comparison. These odd numbers were used in to address the possibility of a tie among the nearest neighbors. The mean absolute percentage error (MAPE) will then be determined to see which k number of neighbors would give the best classification.
3. Each data set feature considered was then run using k-NN.

Naïve Bayes Classifier Application

1. For the same considered cluster, there are three (3) possible outcomes: 0 (Monday-Fridays), 1 (Saturday), and 2(Sunday). The posterior probability P(c|x) of each outcome is computed for each load data where:

$$P(c|x)=[P(x|c)\ P(c)]/P(x)$$

2. A comparison is then made by computing the MAPE.

IV. Results

In order to determine the optimum K number of clusters, two sets were considered. One set consisted solely of the daily morning slope values and the other contained both the daily morning slope data as well as the minimum and maximum temperatures recorded for the particular day.

It can be seen in the graph for the elbow method that both run Using the elbow method, both runs yielded 3 k clusters as the optimum number as reflected in Fig. 2.

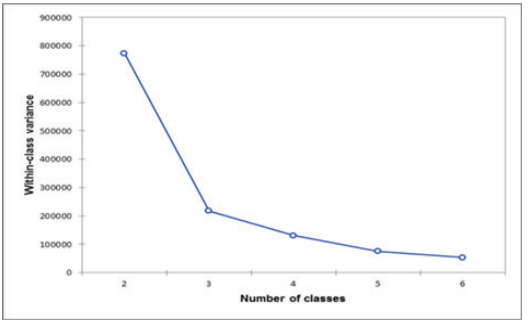

Fig. 2. Elbow method to determine optimum clusters

Both runs yielded almost identical curves which pointed to 3 as the most optimal number of clusters for the data set being considered and is represented in Fig. 2.

This confirms the premise that a better electrical load data profiling can be done, and this cannot solely be dependent on pre-determined or accepted calendar seasons or calendar days.

Summer is typically between June to August in France while the months from September to November as considered as part of Autumn. December to February comprise the Winter months and the remaining months are considered part of Spring[26]. The results of both runs cluster the months in the same groups (Fig 3).

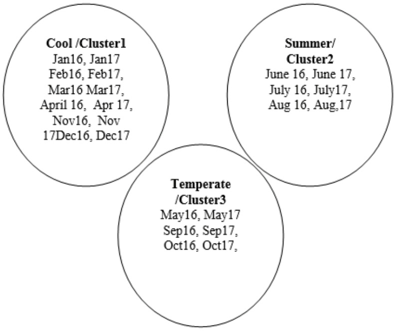

Fig. 3. Seasons by clusters

Cluster 1 represents the cluster for cooler temperatures and having similar power consumption (Fig. 3). The second "season" can be considered as consisting of months where the weather is temperate. It can be noted that the only cluster reflective directly of calendar-based clustering is the summer cluster consisting of June to August. One thing that can be observed from the results also is that while it is expected that there is a yearly power demand increase for a particular month, by using the daily morning slope, the K-means results showed periodicity with the same months being clustered together despite being belonging to different calendar years. This confirms other studies that advocated the daily morning slope as an ideal clustering feature in place of other related hourly load features or the hourly load itself [16], [27].

The next step was to determine which among K-Nearest Neighbors and Naïve Bayes would be better in classifying the clustered load curves into day-types. To test this approach, the cluster containing the load data from the temperate months (May, September, and October) was used. This totalled to 4416 data points consisting of 184 records with 24 entries (hourly load) per record as the data set. From this data set, 39 records were used as the prediction set. The Naïve Bayes Classifier was able to successfully classify the weekday types into the correct category but was weaker in classifying the weekend days (Saturdays and Sundays). The K-Nearest Neighbors algorithm performed better with a k-Nearest Neighbors of 5 yielding the lowest MAPE at 4.17 (Table 1). In comparing the work of [28] which was used as a benchmark for the k-NN approach, where they used the Korean load to profile day-types using their own classification, it can be seen that this approach also achieved lower than 5% MAPE with K=5.

TABLE 1
COMPARISON OF CLASSIFICATION ACCURACY

Day Type	Actual	Naïve Bayes	K1	K3	K5
0	39	39	39	39	39
1	8	7	7	7	8
2	8	6	7	7	7
	MAPE	12.5	8.33	8.33	4.17

The results of this study show that the proposed combined approach of using of K-means clustering and a classifier like the k-nearest neighbors approach in the context of seasonality is a viable alternative to simple calendar-based grouping into "regular" seasons.

V. Conclusion

The researchers propose the use of a neural network training algorithm that is a hybrid of three data mining techniques namely clustering, classification, and backpropagation in order to forecast a 24-hour load curve. The former consists of applying the K-means algorithm to daily morning load feature of a 24-hour electrical load data set and using the clusters/seasons to draw data-types. A comparison was made between the k-Nearest Neighbors approach and Naïve Bayes theorem to determine which would yield a higher accuracy in classifying data types using a sample cluster. The results show the k-Nearest Neighbors is the better classifier together with K-means. The next phase is the artificial neural network builder or training stage (learning phase). This approach is proposed to improve the learning performance of neural networks in the context of seasonality by optimizing the data sets. This paper focuses on the results of the first stage and is a work in progress. Specifically, it is envisioned that by using the results of the first phase, the network builder will be able to perform faster and produce better results. It is said that good clustering applications and classifying techniques can be judged by their predictive power. For this case, the results show that the use of the daily load morning slope is effective with or without the use of temperature values for K-means clustering. The results also show that when clustering load data, a better approach can be the use of a feature such as the daily morning slope in order to accurately reflect load behavior. By doing this, it is envisioned that a more accurate forecast can be made.

The researchers recognize the fact as is typical of a hybrid approach there is the possibility that some other critical load factors may not have been taken into consideration. It remains a task for future work to incorporate said factors. The immediate steps following this work is learning phase implementation which marks the formation of predictive artificial neural network models that forecast a 24-hour load curve.

Acknowledgements

This work is supported in part by the Commission on Higher Education (CHED) K-12 Faculty Development Program and the Western Mindanao State University.

References

[1] H. Cho, Y. Goude, X. Brossat, and Q. Yao, "Modeling and forecasting daily electricity load curves: A hybrid approach," *J. Am. Stat. Assoc.*, vol. 108, no. 501, pp. 7–21, 2013.

[2] I. Dent, U. Aickelin, and T. Rodden, "The Application of a Data Mining Framework to Energy Usage Profiling in Domestic Residences Using UK Data," no. June, pp. 1–10, 2011.

[3] L. Hernández, C. Baladrón, J. M. Aguiar, B. Carro, A. Sánchez-Esguevillas, and J. Lloret, "Artificial neural networks for short-term load forecasting in microgrids environment," *Energy*, vol. 75, pp. 252–264, 2014.

[4] M. Y. El-Sharkh and M. A. Rahman, "Forecasting Electricity Demand Using Dynamic Artificial Neural Network Model," in *rd International Conference on Industrial Engineering and Operations Management*, 2012, pp. 1691–1694.

[5] F. Gorunescu, "Data mining: Concepts, models and techniques," *Intell. Syst. Ref. Libr.*, vol. 12, pp. 467–534, 2011.

[6] C. A. R. De Sousa, "An overview on weight initialization methods for feedforward neural networks," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2016–Octob, no. July, pp. 52–59, 2016.

[7] J. T. Lalis, B. D. Gerardo, and Y. Byun, "An Adaptive Stopping Criterion for Backpropagation Learning in Feedforward Neural Network," *Int. J. Multimed. Ubiquitous Eng.*, vol. 9, no. 8, pp. 149–156, 2014.

[8] K. Halawa, "A method to improve the performance of multilayer perceptron by utilizing various activation functions in the last hidden layer and the least squares method," *Neural Process. Lett.*, 2011.

[9] G. B. Dela Cruz, B. D. Gerardo, and B. T. Tanguilig III, "An Improved Data Mining Mechanism Based on PCA-GA for Agricultural Crops Characterization," *Int. J. Comput. Commun. Eng.*, vol. 3, no. 3, pp. 221–225, 2014.

[10] R. Green, I. Staffell, and N. Vasilakos, "Divide and Conquer? k-means clustering of demand data allows rapid and accurate simulations of the British electricity system," *IEEE Trans. Eng. Manag.*, vol. 61, no. 2, pp. 251–260, 2014.

[11] L. Hernández *et al.*, "Artificial Neural Network for Short-Term Load Forecasting in Distribution Systems," *Energies*, vol. 7, no. 3, pp. 1576–1598, 2014.

[12] B. Koo, S.-W. Lee, W. Kim, and J. H. Park, "Comparative Study of Short-Term Electric Load Forecasting," in *2014 5th International Conference on Intelligent Systems, Modelling and Simulation*, 2014, pp. 463–467.

[13] D. Colley, N. Mahmoudi, D. Eghbal, and T. K. Saha, "Queensland Load Profiling by Using Clustering Techniques," no. October, pp. 1–6, 2014.

[14] S. Nuchprayoon, "Electricity Load Classification Using K-Means Clustering Algorithm," in *5th Brunei International Conference on Engineering and Technology (BICET 2014)*, 2014, pp. 1–5.

[15] L. G. Viola, "Segmentation of household load-profiles with K-means clustering algorithm Segmentation of household load-profiles with K-means clustering algorithm," no. July, pp. 0–9, 2016.

[16] A. Tanwar, E. Crisostomi, P. Ferraro, M. Raugi, M. Tucci, and G. Giunta, "Clustering Analysis of the Electrical Load in European Countries," pp. 1–8, 2015.

[17] C.-R. Chen and U. Kartini, "k-Nearest Neighbor Neural Network Models for Very Short-Term Global Solar Irradiance Forecasting Based on Meteorological Data," *Energies*, vol. 10, no. 2, p. 186, 2017.

[18] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier," *Int. J. Inf. Eng. Electron. Bus.*, vol. 8, no. 4, pp. 54–62, 2016.

[19] I. H. Sarker, M. A. Kabir, A. Colman, and J. Han, "An improved Naive Bayes classifier-based noise detection technique for classifying user phone call behavior," *Commun. Comput. Inf. Sci.*, vol. 845, no. AusDM 2017, pp. 72–85, 2018.

[20] D. Kanojia, "Comparison of Naive Basian and K-NN Classifier," vol. 65, no. 23, pp. 40–45, 2013.

[21] K. Chai, H. T. Hn, and H. L. Cheiu, "Naive-Bayes Classification Algorithm," *Proc. 25th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 97–104, 2002.

[22] I. Witten and E. F. Frank, *Data Mining: Pratical Machine Learning Tools and Techiniques - 2nd edition*. 2005.

[23] M. Ali Fauzi, A. Z. Arifin, S. C. Gosaria, and I. S. Prabowo, "Indonesian news classification using naïve bayes and two-phase feature selection model," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 8, no. 3, pp. 610–615, 2017.

[24] F. Ortuño and I. Rojas, "Bioinformatics and biomedical engineering: Third International Conference, IWBBIO 2015 Granada, Spain, April 15-17, 2015 proceedings, part I," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9043, no. April, 2015.

[25] R. Amirkhah, A. Farazmand, S. K. Gupta, H. Ahmadi, O. Wolkenhauer, and U. Schmitz, "Naïve Bayes classifier predicts functional microRNA target interactions in colorectal cancer," *Mol. Biosyst.*, vol. 11, no. 8, pp. 2126–2134, 2015.

[26] F. Winter and A. Outlook, "French Winter Adequacy Outlook 2016-2017," no. November 2016, pp. 1–14, 2016.

[27] P. Ferraro, E. Crisostomi, M. Tucci, and M. Raugi, "Comparison and clustering analysis of the daily electrical load in eight European countries," *Electr. Power Syst. Res.*, vol. 141, no. October 2017, pp. 114–123, 2016.

[28] B. Koo, M. Kim, H. Lee, J. Park, and C. Kim, "Using Data Mining Technique," *IEEE Conf. Publ.*, pp. 153–157, 2013.