Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Time Series Forecasting for Electricity Consumption using Kernel Principal Component Analysis (kPCA) and Support Vector Machine (SVM)

To cite this article: Verlly Puspita and Ermatita 2019 *J. Phys.: Conf. Ser.* **1196** 012073

IOP Conf. Series: Journal of Physics: Conf. Series **1196** (2019) 012073 doi:10.1088/1742-6596/1196/1/012073

Time Series Forecasting for Electricity Consumption using Kernel Principal Component Analysis (kPCA) and Support Vector Machine (SVM)

Verlly Puspita[1], ***Ermatita**[2]
[1,2]Faculty of Computer Science, Universitas Sriwijaya, Indonesia

Email: verllyp@gmail.com, ermatitaz@yahoo.com

Abstract. Time series data are collected based on certain periods which have constants value (e.g. daily, weekly or monthly), it can be used to forecast or predict future circumstance. Prediction is one of the objectives of the time series analysis by identifying the model from previous data and assuming the current information will also occur in the future. In Big Data trend, huge amount of time series data come from different heterogeneous sources and multiple application domains. This present a new challenge for time series forecasting. Time series modelling has been widely applied and proposed in various fields to improve its accuracy and efficiency of forecasting. This paper discusses time series forecasting for electricity consumption using kernel principal component analysis (KPCA) and Support Vector Machine (SVM). This research will be measured with Mean Absolute Error and Mean Squared Error.

1. Introduction

The result of time series data analysis is usually in the form of prediction, classification, clustering, segmentation and detection. The most challenges in analysing time series data are high dimensional data, noise on data, non-stationary data, and lack of privacy protection [1]. That is clear that the success of time series forecasting depends on using the right model. Time series data in predicting electricity consumption have been applied using the K-means [2], K-means and FCM-based [3] and Polynomial-Fourier series model (P-FS) methods [4]. In the study explained that the use of K-means method can determine the best pattern, but when grouping a lot of data there will be duplicate cluster. Likewise with the P-FS model that cannot accommodate a lot of information that has different data structures. Gonzalez and Zamarreno [5] predict short-term electricity loads using artificial neural network feedback methods. Prediction errors with respect to the output are used to train the network. This model produces a maximum Mean Absolute Percentage Error (MAPE) of 2.88. Another method that has the potential to forecast or predict is Support Vector Machine (SVM).

SVM has been proven to be able to overcome the problem of overfitting, so that it can achieve high generalization performance in solving various time series forecasting problems [6]. In addition, the advantages of using SVM [7] compared to other forecasting methods such as Artificial Neural Network (ANN) are able to handle the size of the data sample and avoid local minimum. In developing forecasting using SVM, the first most important step is feature selection. According to Mark Claesen, all available indicators can be used as inputs in SVM, but features that are irrelevant or correlated can have a negative impact on generalization performance. Therefore, the Kernel Principal Component Analysis (KPCA) method for feature extraction is added.

IOP Conf. Series: Journal of Physics: Conf. Series **1196** (2019) 012073 doi:10.1088/1742-6596/1196/1/012073

This paper will describe the models or methods used in time series forecasting; kernel principal component analysis and support vector machine. The analysis of result using SVM method will also be described in this paper. The organizations of this paper are as follows. In section 2 (Literature Review), present the theory of Time Series, kernel principal component analysis and SVM. In section 3 (Methodology), explain the how this research will be working. In section 4 (Experiment and Result), present the case study and show the obtained result. In section 5 (Conclusion), is the last part of this paper.

2. Literature Review

2.1. Time Series

Time series Models and forecasting methods have been studied by various people and detailed analysis can be found in. Time Series Models can be divided into two kinds [8]. Univariate Models where the observations are those of single variable recorded sequentially over equal spaced time intervals. The other kind is the Multivariate, where the observations are of multiple variables. A common assumption in many time series techniques is that the data are stationary. A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. Stationarity can be defined in precise mathematical terms, but for our purpose we mean a flat looking series, without trend, constant variance over time, a constant autocorrelation structure over time and no periodic fluctuations. There are a number of approaches to modelling time series.

2.2. Support Vector Machine

SVM is a supervised machine learning algorithm that mostly used in classification purposes but can be used to the task of regression and time series forecasting. SVM uses a linear model to find the best hyperplane as a separator of two classes on a vector input. The best hyperplane can be determined by calculating the hyperplane margin value. Margin is the distance between hyperplane and the nearest pattern of each class. Pattern that closest to the maximum margin of this hyperplane is called support vector [9]. If both classes -1 and +1 and a hyperplane is a d-dimensional can defined as:

$$\vec{w} \cdot \vec{x} + b = 0 \tag{1}$$

Pattern $\vec{x_i}$ in negative (-1) and positive (+1) samples can be formulated:

$$\vec{w} \cdot \vec{x} + b \leq -1 \tag{2}$$
$$\vec{w} \cdot \vec{x} + b \geq +1 \tag{3}$$

Quadratic programming is used to find the largest margin value, $\frac{1}{||\vec{w}||}$ by finding the minimal point.

$$min_{\vec{w}} \tau(w) = \frac{1}{2}||\vec{w}||^2 \tag{4}$$

Using the Lagrange multiplier then the primal form of quadratic programming can be transformed into a dual form with the following equation:

$$L(\vec{w}, b, \alpha) = \frac{1}{2}||\vec{w}||^2 - \sum_{i=1}^{l} \alpha_i (y_i((\vec{x_i} \cdot \vec{w} + b) - 1)) \tag{5}$$

2.3. Kernel Principal Component Analysis

Traditional PCA only allows linear dimensionality reduction. However, if the data has more complicated structures, which cannot be simplified in a linear sub-space, traditional PCA will become invalid. Fortunately, Kernel PCA allows us to generalize traditional PCA to nonlinear dimensionality reduction. A kernel principal component analysis (PCA) was previously proposed as a nonlinear extension of a

PCA. The basic idea is [10] to first map the input space into a feature space through nonlinear mapping and then compute the principal components in that feature space.

Generally, this transformation is unknown, and is very difficult to understand, so the dot product calculation according to Marcer's theory can be replaced with $K(x_i,$ kernel function which defines implicitly transformation Φ. This is called the kernel trick which is formulated as follows:

$$K(x_i,x_j) = \Phi(x_i)^T \Phi(\quad (6)$$

Gaussian kernel is one of the kernel tricks. The Gaussian Kernel function that is formed is as follows:

$$K(x_i,x_j) = \exp\left(-\frac{1}{2}\left(\frac{\|x_i-x_j\|}{\sigma^2}\right)\right) \quad (7)$$

2.4. Related Works

Zhao and Magoules [11] reviewed the prediction of electricity consumption in buildings using technical expertise methods, time series statistical methods, and artificial methods. Sandel and his colleagues [12] used regression methods to predict electricity consumption one day in the future in office buildings. Fumo and Biswas applied single, multiple and quadratic regression analysis to predict energy consumption per hour and per day by using historical data on energy consumption and weather parameters as input. Various time series models such as ARIMA and SARIMA have been developed over the years. The ARIMA model designed by Box and Jenkins [13] is the most commonly used model and an effective statistical tool for time series forecasting [14]. In particular, ARIMA combines three processes, namely autoregressive (AR) as a linear function, moving average (MA) as a linear function based on a combination of linear and integrated errors (I) as the average variation of constant data.

Non-linear applications in machine learning techniques (machine learning) to predict energy consumption and optimize the model of Artificial Neural Network (ANN) have been proposed by Lee and his colleagues [15] to predict hourly electricity consumption in a building. Linear and non-linear hybrid methods have also been used to forecast time series data by Lee and Tong. For example, Khashei and Bijair [16] proposed an ANN-ARIMA model for time series forecasting. In his research, the ARIMA model was used to identify and enlarge linear structures in the data, while the ANN model was used to capture non-linear components of the data. The results of his research explained that this hybrid model requires quite a lot of data in order to provide an effective modeling process. In addition, the ANN model has several problems in controlling various parameters and overfitting. Another machine learning methods to improve prediction accuracy and at the same time to avoid overfitting data are Support Vector Machine (SVM) methods. SVM is widely applied in other fields such as function estimation, image classification and time series data prediction [17,18,19]. SVM characteristics are not only intended to make the classification process better, but also to produce a good generalization of training data [20].

3. Methodology

3.1. Description of Electricity Dataset

The dataset from the UCI Machine Learning Repository. The data source of this study is donated by George Hebrail with the name of Individual Household Electric Power Consumption data collected from December 2006 to November 2010 (47 Months). This data displays electricity consumption per minute in every house in Paris. This dataset has 1.25% missing values. As an index of accuracy the forecasting method uses Mean Squared Error (MSE) dan Mean Absolute Error (MAE).

3.2. Modelling

With the fundamental theoretical background given in the previous chapter the model of this research will be explained. The experiments were conducted using kernel principal component analysis (KPCA)

IOP Conf. Series: Journal of Physics: Conf. Series **1196** (2019) 012073 doi:10.1088/1742-6596/1196/1/012073

for selecting features and reduce dimension of data and Support Vector Machine (SVM) for prediction. the conceptual framework for the use of methods in forecasting electricity consumption as follows:

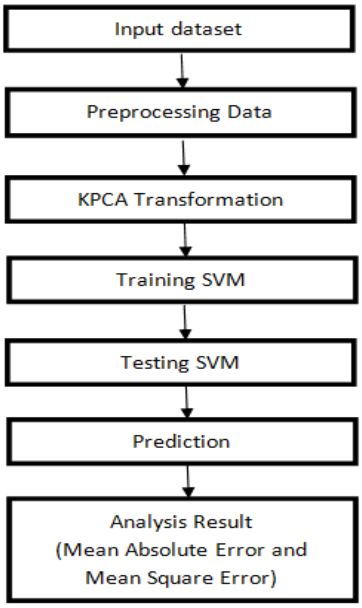

Figure 1. Research Methodology

4. Experiment and Result

4.1. Time Series Model Identification

The first step is to make time series plot for load data of Individual Household Electric Power Consumption from 2006 to 2010 using the software R. This plot is useful to see the stability of the data.

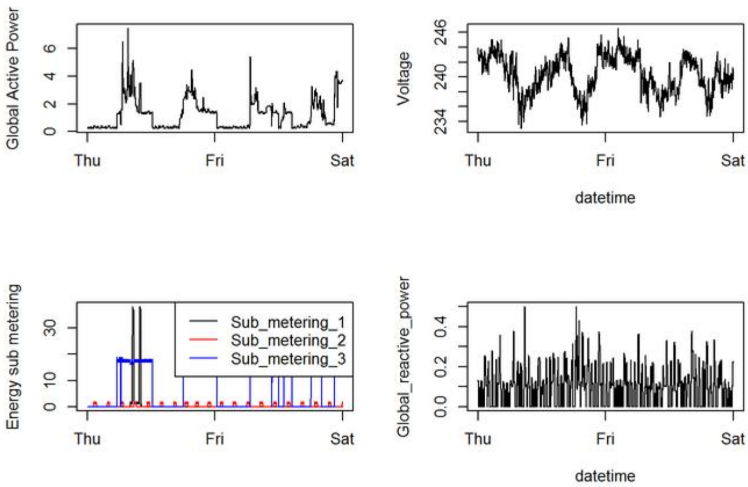

Figure 2. Time Series Plot

Figure 2 above shows that the characteristics of time series data sets are multivariate because they have many variables or features, besides that the data set has seasonal or seasonal movements. Therefore, data sets for Individual Household Electric Power Consumption can be said to be non-linear. The kernel PCA is useful for extracting non-linear data structures.

4.2. Normalization Data

The electricity consumption data used has a different range of values. For example, global active power has values ranging from one to ten, global reactive power is under ten and the global intensity value is above ten. Therefore, the initial data obtained will be transformed so that there is no longer a data variable that dominates other data variables. This transformation process is used in the training and testing stages. This data transformation is done so that the data is at 0-1 intervals, so that convergence is faster achieved. In this study, normalization of the data used is min-max. Min-max technique will be used to normalize feature data so that they are in the range of 0 to 1. Examples of feature results after normalization can be seen in Table 1. The normalization results below are data at 17:53:00 - 18:02:00.

Table 1. Normalization Data.

No	GAP	GRA	Volt	Global	SM1	SM2	SM3
1	0.4499	0.5226	0.2825	0.4559	0.000	0.013	0.85
2	0.5779	0.5449	0.2310	0.5751	0.000	0.013	0.8
3	0.5795	0.6224	0.2165	0.5751	0.000	0.027	0.85
4	0.5811	0.6275	0.2357	0.3886	0.000	0.013	0.85
5	0.3884	0.66	0.3182	0.3678	0.000	0.013	0.85
6	0.3721	0.6525	0.2902	0.3886	0.000	0.027	0.85
7	0.3924	0.65	0.2931	0.3886	0.000	0.013	0.85
8	0.3922	0.6375	0.2987	0.3886	0.000	0.013	0.85
9	0.3886	0.6375	0.2463	0.3886	0.000	0.013	0.85
10	0.3880	0.6224	0.2408	0.4870	0.000	0.027	0.8

As can be seen in Table 1, GAP (Global Active Power), GRA (Global Reactive Power), Voltage, Global Intensity in ampere, SM1 (Sub Metering 1) in watt/hour that corresponds to kitchen, SM2 (Sub Metering 2) corresponds to laundry room and SM3 (Sub Metering 3) corresponds to electric water-heater and an air-conditioner, after normalization the values live between 0 and 1.

4.3. KPCA Transformation

In this research used Gaussian Kernel PCA to extract 9 features from the training data. For Gaussian kernel PCA, the most important parameter is the σ in the kernel function [21] defined by Eq. 8. The Gaussian kernel is a function of the distance $||x-y||$ between two vectors x and y. Since there exists no principled method for finding the kernel width σ in unsupervised learning, we chose the default value $\sigma = 1$ which lead to visually satisfying results. Kernel PCA succeeded to reduction the data from 20628 $\times$ 210 to 1680 $\times$ 2.

4.4. Prediction using Support Vector Machine

Predictions of electricity consumption in household's process was using java programming language. Prediction time is for 30 minutes ahead. SVM model parameters are determined by experience of trial and error. Furthermore, the epsilon, gamma and probability parameters are set to find the best MSE and RMSE values. Predicting results are as shown in Fig 3. The following is a prediction in table 2.

In Fig. 3 the horizontal axis is consumption electricity data and the vertical axis is sum of consumption electricity each data. The red graphic is actual and the blue one is predicted result. Actual data and prediction have quite different comparison. The data around 1 to 300 are almost close to the actual data value, while the data around 500 do not approach the actual data. So, the MAE and RMSE values are 11.48 and 13.86.

Table 2. MAE and RMSE of Prediction

ε	Gamma	P	MAE	RMSE
0.001	0.5	1	11.71	14.04
0.0001	0.5	1	14.32	18.51
0.00001	0.5	1	11.84	14.10
0.000001	0.5	1	11.69	14.75
0.0000001	0.5	1	11.53	13.89
0.00000001	0.5	1	11.58	14.63
0.00000001	0.4	1	11.56	13.92
0.00000001	0.3	1	11.72	14.86
0.00000001	0.2	1	11.48	13.86
0.00000001	0.1	1	12.53	15.64
0.00000001	0.2	0.9	12.43	15.53
0.00000001	0.2	0.8	11.71	13.97
0.00000001	0.2	0.7	14.38	18.61
0.00000001	0.2	0.6	11.22	14.20
0.00000001	0.2	0.5	11.60	13.98

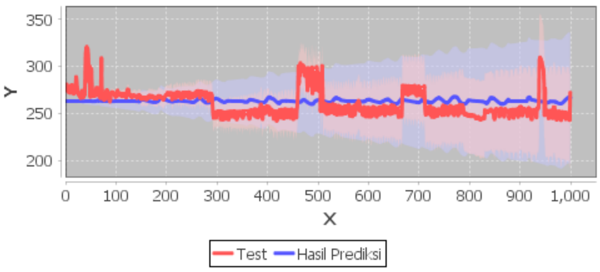

Figure 3. Comparison between actual and predicted consumption of electricity

5. Conclusion

In this paper, we discussed the theories of Time Series, kernel PCA and Support Vector Machine (SVM), then focused on prediction consumption electricity of household. Time Series use for to see the stability of the data. And the result the data is non-linier. With this information, the kernel PCA is useful for extracting non-linear data structures. we have tested kernel PCA dimensionality reduction on non-linier data of Electric Power Consumption, and found that Gaussian kernel PCA succeeded to reduction the data from 20628×210 to 1680×2. The main conclusion of this case study is that it is possible to obtain good energy consumption forecast using methodologies such as SVM in complex scenarios as discussed in this paper. Support Vector Machine reach good enough generalisation properties using the structural risk minimisation principle. The strength of SVM is controlling the parameters. The best achieved for studied are with epsilon 0.00000001, gamma 0.2 and probability 1, and the result are MAE 11.48 and RMSE 13.86.

As future work we propose to explore and experiment with various techniques in the pre-processing phase, particularly in defining the best training data set for processing of SVMs in order to improve consistency and accuracy of results.

IOP Conf. Series: Journal of Physics: Conf. Series **1196** (2019) 012073 doi:10.1088/1742-6596/1196/1/012073

6. References

[1] Molaei SM, and Keyvanpour MR 2015 An analytical review for event prediction system on time series International Conference on Pattern Recognition and Image Analysis

[2] Rasanen T, Kolehmainen M 2009 Feature-based clustering for electricity use time series data (Springer) pp 401-412

[3] Benitez I, Diez JL, Quijano A, Delgado I 2016 Dynamic clustering of residential electricity consumption time series data based on Hausdorff distance (Elsevier, Electric power system, vol 140) pp 517-526

[4] Tsai CL, Chen WT, and Chang CS 2016 Polynomial-fourier series model for analyzing and predicting electricity consumption in buildings (Elsevier, Energy and Building, vol 127) pp 301-312

[5] Pedro A, Gozalez and Jesus M, Zammarreno 2005 Prediction of hourly energy consumption in buildings based on a feedback artificial neural network (Elsevier, Energy and Building vol 37) pp 595-601

[6] Lijuan C 2003 Support Vector Machine expert for time series forecasting (Elsevier Neurocomputing vol 51) pp 321-339

[7] Yan X, and Chowdhury NA 2013 A comparison between SVM and LSSVM in mid-term electricity market clearing price forecasting (IEEE Canadian Conference of Electrical and Computer Engineering)

[8] NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

[9] Sembiring K unpublished Penerapan Teknik Support Vector Machine untuk Pendeteksian Intrusi pada Jaringan

[10] Ahmadinejad MM and Sherly E 2016 A Comparative Study on PCA and KPCA Methods for Face Recognition (International Journal Of Science and Research (IJSR) vol 5) Issue 6

[11] Wang Q 2014 Kernel Principal Component Analysis and its Application in Face Recognition and Active Shave Models (Ressealear Polytechnic Institute)

[12] Zhao H, and Magoules 2012 A review on the prediction of building energy consumption (Elsivier, Renewable and Sustainable Energy Reviews vol 16) pp 3586-3592

[13] Box, and Jenkins 1986 Forecasting and Time Series Analysis Using the Statistical System (UTS Module)

[14] Zhang GP 2011 Time series forecasting using a hybrid ARIMA and neural network model (Elsevier, Neurocomputing col 50) pp 159-175

[15] Sandels C, Widen J, Nordstrom L, and Andersson E 2015 Day-Ahead prediction of electricity consumption in a Swedish office building from weather, occupancy and temporal data (Elsivier, Energy Building)

[16] Lee Y, and Tong L 2012 Forecasting nonlinier time series of energy consumption using hubrid dynamic model (Elsevier, Applied Energy vol. 94) pp 251-256

[17] Cao LJ, and Tay F 2003 Support Vector Machine with adaptive parameters in financial time series forecasting (IEEE, vol 14) pp 1506-1518

[18] Xie L, Li G, Peng L, and Chen Q 2017 Hyperspectral image classification using discrete space model and support vector machine (IEEE)

[19] Tan G, Yan J, Gao C, and Yang S 2011 Prediction of water quality time series data based on least squares support vector machine (Elsevier, Procedia Engineering vol 31) pp 1194-1199

[20] Adhikari R, and Agrawal RK An introductory study on time series modelling and forecasting (unpublished)

[21] Khashei M, and Bijari M 2010 A novel hybridization of artificial network and ARIMA models for time series forescating (Elsivier, Applied Soft Computing vol 11) pp 2664-2675