Short-term electricity grid maximum demand forecasting with the ARIMAX-SVR Machine Learning Hybrid Model

Hon Fung Chow
CLP Power Hong Kong Limited, Hong Kong, People's Republic of China

ABSTRACT

This paper proposes and discusses the viability of a short-term grid maximum demand forecasting model combining autoregressive integrated moving average with regressors (ARIMAX) and support vector regression (SVR). Grid demand forecasting is essential to generation unit scheduling, maintenance planning and system security. Traditionally, grid demand is forecasted using multivariate linear regression models with parameters adjusted to past data. A disadvantage of the linear regression model is that the parameters require regular adjustment, otherwise the prediction accuracy will deteriorate over time. With recent advances in the field of machine learning and lower computational costs, the usage of machine learning in the power industry becomes increasingly practicable. The proposed model is a machine learning model that combines ARIMAX and SVR to exploit their respective effectiveness in predicting linear and non-linear data. In contrast to linear regression models, the machine learning model automatically updates itself when new data is included. The hybrid model is benchmarked against other forecasting models and demonstrated a marked improvement in accuracy, achieving RMSE of 67.7MW and MAPE of 1.32% in a seven-day forecast.

KEYWORDS
Electrical grid demand forecasting; machine learning; ARIMA; SVR; hybrid models; time series forecasting

1. Introduction

The electricity suppliers of Hong Kong are responsible for managing their respective electricity supply grids and generation units. Due to the impracticality of storing electricity with contemporary technology, the power generated must match the demand of the grid sufficiently to preserve grid security and manage generation costs. Generation units such as coal-fired plants and combined cycle gas turbines require adequate time to start up and load up. Frequent cycling (unit start-up and shut-down) and ramping (output power adjustment) increase unit operation and maintenance expenses, which invariably drive up the cost of electricity (Kumar et al., 2012). Power plants operating at sub-optimal loading also adversely affect fuel efficiency and increase emissions (Turconi et al., 2014). Accurately forecasting future demand is thus an important aspect in power grid planning.

2. Motivation

The traditional approach to forecasting future grid demand utilises a multivariate linear regression model which incorporates variables such as ambient temperature, humidity, economic activities etc. (Hong et al., 2010). The final prediction is calculated by a summation of all variables, adjusted by a parameter, or a linear coefficient, which is determined using past data. A multivariate linear regression model is easy to interpret and robust. However, to preserve the prediction's accuracy, these parameters require regular tuning. Over the years, energy consumption behaviour has shifted towards energy-saving, and users are becoming more environmentally conscious (Guo et al., 2018). With the change in consumption behaviours and increasing prominence in renewable energy, the demand for energy has become more volatile and unpredictable. Hence, the parameters would require more frequent updates to suit the public’s need.

The goal of this study is to demonstrate the capability and viability of an ARIMAX-SVR hybrid forecasting model for the prediction of daily grid maximum demand in short term. A supervised machine learning algorithm such as an SVM can be used to build a regression model, also known as an SVR (Drucker et al., 1996). The greatest advantages that a machine learning model has over a conventional regression model is its automation and self-improvement. When new data is added to the dataset, the machine learning algorithm will spontaneously adapt to the new data without human intervention. Unless the data contains extreme outliers, in general terms, as more data is introduced to the dataset, the better the model performs. This contrasts the traditional model in which performance tends to degrade over time.

3. Data structure

The proposed algorithm hybridises ARIMAX and SVR to enhance the model's ability to forecast the linear component of the data. In the ARIMAX part of the model, two regressors (temperature and workdays) have been included to restrain the forecast variability and reflect the change in grid demand. The temperature data is the daily
maximum ambient temperature acquired from the Hong Kong Observatory.

In this paper, the 900-day daily grid maximum demand record was split into training and testing sets. The model will utilise the training set to learn the pattern in the data and adjust itself through an iterative process, whilst the testing set is used to evaluate the applicability of the model to new data. The training set contains 886 days of data, and the remaining 14 days acted as the testing set. The size of the testing set was chosen to be 14 days since the model’s forecast horizon is restricted by the length of the Hong Kong Observatory’s weather forecast. To use the model to predict upcoming grid maximum demand, the user must first inform the model of the ambient temperatures and identify which day within the forecast horizon is a workday and which is a holiday.

4. Organisation of this study

This study is divided into two segments. The first is a comparison of the hybrid model’s performance with non-hybrid models. The benchmark of the models’ performance is a combination of Akaike Information Criterion and error metrics. Five models were constructed in this study: ARIMA(2,1,3), ARIMAX(3,1,2), ARIMAX(2,1,4), SVR and an ARIMAX-SVR hybrid model. The three non-hybrid ARIMA models and the hybrid model were trained under the same training set, whereas the SVR model was trained with a 28-day training dataset. This is because SVR is prone to be influenced by outliers; the probability of encountering outliers increases with the size of the training set, which then would adversely impact the SVR model’s prediction. The result shows that the hybrid model achieved a marked improvement in terms of error metrics compared to the non-hybrid models, scoring an RMSE of 95.9561, MAPE of 0.0184 and RAE of 0.5067.

The second segment of this paper is dedicated to investigating the influence that the SVR algorithm’s training period length and the length of the forecast horizon have over prediction accuracy in the hybrid model. The hybrid model’s SVR component does not share the same training period as the ARIMAX component, for reasons previously discussed. To gain insight into the optimal training period, the SVR component was trained under different lengths: 14, 28, 56 and 112 days. The best results were obtained when the SVR was trained under 56 days. The forecast horizon also has an impact on forecast accuracy as well. The hybrid model was tasked to predict grid demand in three different forecast horizons: 14 days, seven days, and one day. The model yields the best result when forecasting over seven days. Therefore, based on the result of this study, the SVR component of the model should be trained for 56 days and the forecast horizon should be set to seven days to acquire the most accurate prediction.

5. Literature review

The prevalent approach to forecasting short term electricity demand using machine learning is to apply one algorithm at a time. For example, Al-Musaylh et al. has conducted short-term electricity demand forecasting with MARS (Multivariate Adaptive Regression Spline), SVR and ARIMA separately (2018). In the study, the ARIMA model did not include any regressors. The SVR and MARS models have reported improvements in accuracy in comparison to ARIMA and naïve models. The paper selected three forecast horizons to deploy the model. The forecast results showed that the variance in prediction tends to increase with the length of the forecast horizon. In particular, the ARIMA model performed the worst amongst the models discussed. It is suspected that the poor performance by the ARIMA model is due to the lack of regressors to constrain the forecast, which partly motivated the use of ARIMAX instead of ARIMA model in the ARIMAX-SVR hybrid.

An ARIMA-SVR hybrid model was used to predict the grid demand of Hebei Province in China and demonstrated the hybrid approach’s superiority over a non-hybrid approach (He et al., 2006). The paper deployed the hybrid model to forecast the grid demand over 24 hours and obtained encouraging results. In this study, the ARIMA component of the hybrid model does not contain regressors to regulate the prediction, and the study did not detail the selection of hyper-parameters and training length for the SVR component. Similarly, a study has been done to forecast grid demand in Heilongjiang (Nie et al., 2012) over 49 days. This study did not discuss the reasons for the parameter and training period selection either.

A hybrid approach has been proposed by Yan et al. to combine LSSVM (Least Square SVM) and ARMAX to forecast electricity market clearing prices in Canada, and acquired better predictions in comparison to the non-hybrid approach (2013). The hybridisation of algorithms is similar to the one proposed in this paper. The ARMAX component of the hybrid model did not, however, include a differencing component. Pokora has proposed a hybrid model combining ARIMA and SVR and yielded moderate success on forecasting short term electricity prices (2017). However, the ARIMA algorithm was not moderated by regressors.

To summarise, studies done to apply an ARIMA-SVR hybrid model to forecast various time series data had been focused on demonstrating the improved forecasting accuracy of the hybrid approach over the non-hybrid approach. However, the discussion on the selection of hyper-parameters, forecast horizons and the length of time the SVR model is trained on is lacking. That has motivated this study to discuss their influence on forecast accuracy.
6. Daily grid maximum demand

The daily grid maximum demand dataset features a 900-day record of the highest grid demand on every day from 19-7-2017 to 4-1-2020 in the CLP-operated power grid in Hong Kong. In Figure 1, the data is visualised alongside the trend, seasonal and remainder components of the data using STL decomposition. The y-axis of the trends shows the load demand as a percentage of operating capacity. The data shows multiple orders of seasonality in the Seasonal component in Figure 1. This reflects the cyclic nature of power consumption in Hong Kong: the grid tends to demand more power during hot and humid days to run the HVAC systems in household and commercial buildings, which causes the demand to peak at hot days. Economic activities during weekends and public holidays are significantly lower due to the reduced usage of offices. This causes the demand to slump at holidays and weekends. From the Trend component of Figure 1, it is clear to see that the grid demand varies according to seasons. The Remainder component is the white noise or randomness that cannot be accounted for in the previous two components.

In Figure 2, the autocorrelation factor (ACF) plot shows a scallop-shaped pattern and the lags peaking at lag one, seven, 14, 21 and 28. This pattern suggests the grid demand follows a seven-day cycle, which confirms what is stated above. The correlation factors fall into the range of 0.7 to 0.9, which indicates that the autocorrelation of future data is strongly correlated to the previous data. The decreasing trend in the ACF plot suggests that as the prediction move further forward into the future, the autocorrelation weakens, which will adversely affect the accuracy of ARIMA models. The autocorrelation remains viably high at lag 30, suggesting that the large dataset is beneficial to the ARIMA prediction accuracy. In the partial autocorrelation factor (PACF) plot, the spike in lag one suggests the need of differencing the data to properly construct the ARIMA model.

Figure 1. The 900-Day Grid Maximum Demand undergone STL Decomposition.

Figure 2. Daily Maximum Grid Demand ACF and PACF plots.

7. Forecasting Models

7.1 Autoregressive integrated moving average (ARIMA)

Autoregressive integrated moving average, or ARIMA, is a combination of the autoregressive (AR) model and the moving average (MA) model.

An AR(p) model utilises a linear combination of past values to forecast future values, which is essentially a regression against itself. An AR(p) model can be written as

\[ y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t, \]

where \( c \) is constant, \( \varphi_i \) are parameters, \( y_{t-i} \) are past values and \( \varepsilon_t \) is white noise.

An MA(q) model uses the past forecast errors as predictors in a regression model to predict future values. An MA(q) model can be written as

\[ y_t = c + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t, \]

where \( c \) is constant, \( \theta_i \) are parameters, \( \varepsilon_{t-i} \) are past errors of forecasts and \( \varepsilon_t \) is white noise.

An ARIMA(p,d,q) model combines differencing with AR and MA models, which can be written as
\[ y'_t = \epsilon + \sum_{i=1}^{w} \psi y'_{t-i} + \sum_{i=1}^{d} \theta \epsilon_{t-i} + \epsilon_t, \]  

where \( y'_t \) is the differenced time series and \( d \) is the order of differencing.

One version of ARIMA is ARIMA with regressors, or ARIMAX, which combines a regression model and ARIMA error to take advantage of the information provided by the independent variables in the regression model to enhance the accuracy. The regressors used in this paper are daily maximum ambient temperature provided by the Hong Kong Observatory and an identification of workdays and off days in Hong Kong. The daily maximum ambient temperature is included to account for the increased usage of HVAC systems during hot days and identification of workdays or off days is used to simulate the economic activities of the city.

### 7.2 Supported vector machines regression (SVR)

Supported vector machines (SVM) were originally developed to solve binary classification problems utilising the concept of supported vectors (Cortes and Vapnik, 1995). It was later adapted to solve regression problems. This was achieved by making use of \( \epsilon \)-insensitive loss function, which ignores errors that fall within the \( \epsilon \) band:

\[ L(y, f(x,w)) = \begin{cases} 0, & \text{if } |y - f(x,w)| \leq \epsilon \\ |y - f(x,w)| - \epsilon, & \text{otherwise} \end{cases}, \]

where \( f(x,w) \) represents the original input after being mapped onto feature space.

For a non-linear regression problem, such as the one discussed in this paper, the input is mapped to a higher dimension space, or otherwise known as feature space. With the addition the kernel function, the approximation function can be written as

\[ f(x) = \sum_{i=1}^{w} (a'_i - a_i) k(x, x) , \]

\[ k(x, x) = \exp (-\gamma \|x - y\|^2) , \]

\[ \gamma = \frac{1}{2\sigma^2} , \]

where \( a'_i \) and \( a_i \) are Lagrange multipliers, \( k(x, x) \) is the kernel function and \( \sigma \) is a free parameter. The kernel function selected for the proposed model is a radial basis function (RBF), which is represented by the expression \( \exp(-\gamma \|x - y\|^{2}) \).

The parameters \( \epsilon \) and cost \( C \) are hyper-parameters. They are used to control the behaviour of supported vectors by regularising soft margins and penalising misclassifications. The hyper-parameters’ influence on SVR models are not intuitive. Therefore, an iterative approach called grid search is utilised to test the SVR model under a range of combinations of \( \epsilon \), \( \epsilon \) and \( C \) to look for the combination that offers the least error.

### 7.3 ARIMA-SVR hybrid model

The proposed model seeks to exploit the strengths of ARIMAX and SVR algorithms in modelling linear and non-linear data to obtain better prediction accuracy than using conventional forecasting model. The pipeline of model building and selection adopted was proposed by Hyndman and Athanasopoulos (2018), which entails the use of Kwiatkowski–Phillips–Schmidt–Shin (KPSS) unit-root tests and minimising the Akaike Information Criterion Corrected (AICC). The ARIMAX model is used to fit the training data first, with a segment of the ARIMAX model’s residuals to be fed into the SVR algorithm for training. The final prediction is obtained by combining the ARIMA model’s prediction of the data and the SVR algorithm’s prediction of the ARIMAX residuals.

### 8. Parameter tuning for models

#### 8.1 Parameter selection for ARIMA

The selection of ARIMA components (p, d and q) was conducted iteratively. The parameters p, d and q were assigned with integers ranging from one to five (inclusive) and ARIMA models were constructed with all the combinations of p, d and q. The ARIMA models were evaluated and their respective AICC scores calculated. The p, d and q combinations which passed the KPSS test and scored the lowest AICC scores were selected to be the candidates for the follow-up analysis.
8.2 Parameter selection for SVR

The selection of SVR hyper-parameters involved an iterative technique known as grid search. The technique assigned each hyper-parameter a designated value, constructed an SVR model and subsequently performed cross-validation with the model to calculate the model error. The grid search will assign new values to the hyper-parameters and repeat the aforementioned process until a global minimum of model errors was obtained. The set of hyper-parameters that could achieve minimum model errors was selected to construct the candidate SVR model for follow-up analysis.

Since grid search is a computationally expensive process, the author recommends to first increase the values for γ, ε and C linearly by a factor of two and increase the value of cost C exponentially after each early iteration. As the iteration continues, one will observe the prediction errors from the models first decrease and then increase after reaching an apparent minima. This allows one to obtain the approximate location of the true minima. After cementing the range of γ, ε and C, the hyper-parameters are fine-tuned by altering the assigned values gradually at each following iteration. This approach can reduce computational time and resources and allow the hyper-parameters to converge to minima faster.

9. Model evaluation

The evaluation of the models is based on four parameters: AIC, root-mean-squared error (RMSE), mean absolute percentage error (MAPE) and relative absolute error (RAE). AIC is a measurement based on information theory that evaluates the relative amount of information loss; the lower the AIC, the less likely for a model to overfit or underfit.

\[
AIC = -T \log \left( \frac{SSE}{T} \right) + 2(k + 2),
\]

where T is the number of observations, SSE is the goodness of fit and k is the number of predictors in the model. The prediction errors are measured by RMSE, MAPE and RAE which are given by expression as follows:

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}},
\]

\[
MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|,
\]

\[
RAE = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}},
\]

where \( \hat{y}_i \) is the predicted value, \( y_i \) is the actual value and \( N \) is the number of predictions.

10. Results

The KPSS test was conducted on the raw grid demand time series data and yield a KPSS level of 0.52 and p-value of 0.037. The KPSS test statistics were larger than 0.1 critical value and p-value was lower than 0.05, indicating that the time series was not stationary and require differencing. After performing differencing of order one, the KPSS level decreased to 0.043 and yielded a p-value of 0.10. This indicated the data require at least differencing of order one and the data after differencing is suitably stationary.

The five models constructed in this paper are ARIMA(2,1,3), ARIMAX(3,1,2), ARIMAX(2,1,4), SVR and an ARIMAX-SVR hybrid model. Their respective prediction results are summarised in Table 1. The ARIMA(2,1,3) model is the simplest model amongst the five and serves as a benchmark for other models that utilises a more sophisticated algorithm. The AIC 12695.95 is the highest amongst all four models, indicating the model suffered the most information loss. The error metrics of the ARIMA(2,1,3) model are also the highest. RAE being larger than ones raises a need to improve the model since RAE is a ratio comparing the forecast error to the naïve forecast error. This suggests the ARIMA(2,1,3) model is worse than the naïve approach.

The two ARIMAX models, ARIMAX(3,1,2) and ARIMAX(2,1,4) featured different sets of regressors. The first one featured first-order regressors (daily maximum ambient temperature) + (workdays), while the second one featured a second order regressor, in which an additional term (daily maximum ambient temperature)^2 is added. The addition of a second order regressor lowers the AIC and hence reduces the information loss in the model. The second order model also performs better in terms of error metrics. Both ARIMAX models see improvement in RAEs which are less than one. It was decided it would be imprudent to raise the regressors to higher order due to risk of overfitting. Increasing the order of regressor will lower AIC and thereby reduce the information loss. However, as the complexity of the model increases, so does the risk of extracting noise in the data. Ultimately, it leads to the model fitting the training set very well but performing poorly when predicting future values.

The SVR model is constructed using ε regression. The hyper-parameters, cost, γ and ε, are determined by grid search, which is an iterative process to try out every combination of cost, γ and ε until the minima in error is obtained. The result of the grid search is visualised in Figure 4. The optimal values of cost, γ and ε are four, two and 0.5 respectively. The SVR model is trained over a 28-day training set and yields a forecast result that trumps all
ARIMA and ARIMAX models.

In the ARIMAX-A VR hybrid model, the residuals from the second order ARIMAX model were extracted. The latest 28 days of ARIMAX residuals were injected into the SVR model. The SVR component was treated similarly to the previous SVR model. The hyper-parameters were optimised using grid search. The prediction of the ARIMAX residuals over the 14-day forecast horizon was added to the ARIMAX-A VR hybrid model to compensate for it. The ARIMAX-A VR hybrid yielded a significantly lower RMSE and MAPE. In general terms, as shown in Figure 5, the predictions by model tracks the true values’ pattern closely, and the predictions tend to overestimate the true demand.

The 14-day forecast horizon includes three public holidays, one Saturday and two Sundays. On workdays, e.g. 29/12/2019 and onward, the error is very low. Therefore, the main source of error that contributed to RMSE comes from the off days. In particular, the Christmas holidays (25 to 26/12/2019) see the biggest errors in prediction. Since the Christmas holidays are close to weekends, there appears to be a residual effect from Christmas, and it deteriorates the prediction accuracy on the following weekend (27 to 28/12/2019).

To investigate the effect of forecast horizons and SVR training period, the hybrid model was modified to forecast on the 14, seven, and one-day horizon, under 14, 28, 56, 112 and 224 days of SVR training. Meanwhile, the ARIMAX component is still trained in the same 886-day timeframe. For 14-day SVR training on a 14-day forecast horizon, the SVR component first extracts the ARIMAX residuals from the most recent 14 days, i.e. the 873th day to the 886th day, and undergoes hyper-parameters grid search to optimise before forecasting the ARIMAX residual for the next 14 days. The same testing methodology is applied to other forecast horizon and training periods. The results are summarised in Table 2. Note that RAEs for the tests on the one-day forecast horizon are undefined.

The results in Table 2 show that the errors for the 14-day forecast horizon have a local minimum at 28-day SVR training, before another improvement in RMSE at 224-day SVR training. For seven-day and one-day forecast horizons, both of their respective minimum error is yielded at 56-day SVR training. The error of the seven-day forecast horizon model does not increase again after 56-day SVR training, unlike the one-day forecast horizon where the RMSE is improved again at 224-day SVR training. The best forecast in terms of RMSE is produced by ARIMAX-SVR hybrid with a seven-day forecast horizon and 56-day SVR training, achieving RMSE of 68.0 MW. The best forecast in terms of MAPE and RAE is produced by the hybrid model with a seven-day forecast horizon and 28-day SVR training, achieving MAPE of 0.0121 and RAE of 0.3435.

### Table 1. Summary statistics for model comparison.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>ARIMA</th>
<th>SVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regressors</td>
<td>(temp) + (workdays)</td>
<td>(temp) + (temp)^2 + (workdays)</td>
</tr>
<tr>
<td>Error Metrics</td>
<td>RMSE: 312.9760, MAPE: 4.4258, RAE: 1.3983</td>
<td>RMSE: 229.3549, MAPE: 3.0596, RAE: 0.5854</td>
</tr>
</tbody>
</table>

### Table 2. Summary of influence of SVR hyper-parameters and forecast horizon on the forecast performance.

<table>
<thead>
<tr>
<th>Forecast Horizon (Day)</th>
<th>14</th>
<th>7</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>8</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>ε</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>γ</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>η</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error Metrics</th>
<th>RMSE (MW)</th>
<th>MAPE</th>
<th>RAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>95.9561</td>
<td>87.5442</td>
<td>72.7811</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.0184</td>
<td>0.01789</td>
<td>0.0221</td>
</tr>
<tr>
<td>RAE</td>
<td>0.5067</td>
<td>0.4902</td>
<td>0.5067</td>
</tr>
</tbody>
</table>

The results in Table 2 show that the errors for the 14-day forecast horizon have a local minimum at 28-day SVR training, before another improvement in RMSE at 224-day SVR training. For seven-day and one-day forecast horizons, both of their respective minimum error is yielded at 56-day SVR training. The error of the seven-day forecast horizon model does not increase again after 56-day SVR training, unlike the one-day forecast horizon where the RMSE is improved again at 224-day SVR training. The best forecast in terms of RMSE is produced by ARIMAX-SVR hybrid with a seven-day forecast horizon and 56-day SVR training, achieving RMSE of 68.0 MW. The best forecast in terms of MAPE and RAE is produced by the hybrid model with a seven-day forecast horizon and 28-day SVR training, achieving MAPE of 0.0121 and RAE of 0.3435.
Figure 4. Grid search results for optimal hyper-parameters.

![Grid search results for optimal hyper-parameters.](image)

Figure 5. 14-day forecast by ARIMAX-SVR hybrid (p,d,q): (2,1,4); (Cost, γ, ε): (4, 0.5, 0.5).

![14-day forecast by ARIMAX-SVR hybrid](image)

11. Discussion

In terms of forecast accuracy, the ARIMAX-SVR hybrid outperforms the other ARIMA and SVR models by a considerable margin. The RMSE, MAPE and RAE all see significant improvement over the non-hybrid approach. The non-hybrid ARIMAX models have an inherent disadvantage, which is the assumption of constant variance in prediction residuals. This assumption is violated when there is an unexpected event that disrupts the power demand pattern, such as a typhoon or a pandemic, and the subsequent ARIMA prediction will be adversely impacted. The SVR model, with appropriate selection of kernel function and hyper-parameters, can be trained by ARIMAX residuals and predict the residuals for the following days. Upon summation, the SVR prediction of residuals compensates for the deviation in ARIMAX prediction, allowing for a more accurate result.

To deploy an accurate hybrid forecasting model requires appropriate selection of parameters for ARIMAX and SVR models. For the selection of orders in ARIMAX, experienced users can determine the possible range of (p,d,q) by examining ACF and PACF plots of the data, and conduct iterative testing to arrive at an optimal combination of (p,d,q). The regressors of an ARIMAX model rely on the users’ domain knowledge of the data. For instance, to select regressors for forecasting power grid demand, the users must understand the underlying factors that influence the demand patterns in order to make sensible choices. The user should then proceed to use different combinations of regressors to achieve accurate predictions. The SVR models’ success depends heavily on the choice of kernel function and hyper-parameters (Burgess, 1998), which fundamentally changes how the SVR model behaves. In this study, the RBF kernel function is selected due to its versatility and the choice of hyper-parameters is optimised by grid search. The length of the forecast horizon and SVR training period also have a significant impact on prediction accuracy. From the results, the seven-day forecast horizon offers the most consistent and accurate result. By inspecting the results in Table 2, the prediction errors show signs of fluctuation with the SVR training period. It is not apparent whether a global minimum for prediction errors has been reached and extending the SVR training period further will increase the risk of overfitting. The SVR training periods of 28 days and 56 days have performed consistently better than the other training periods. It is suggested that the following:

$$SVR\ Training\ Period = Forecast\ Horizon \cdot 2^n, \quad (12)$$

be used to iteratively search for the training periods which yield the lowest prediction error. This iterative process can be done automatically using loops in various programming languages. It should be noted that achieving the optimal SVR training period with the training dataset does not guarantee the future forecasts of unknown true values will also yield the lowest error. Therefore, a prudent measure is to constrain the SVR training period to a value comparable to the forecast horizon, since the results in Table 2 appear to be more consistent when the forecast horizon and the SVR training periods are in proximity.

The usage of the hybrid model is not restricted to forecasting daily maximum demand. With a suitable set of regressors and data, it can be adapted to forecast other time series data such as generation unit fuel consumption, hourly grid demand and cost of electricity production. The pipeline of adapting the model to other usage is described in Figure 6. The process involves the use of KPSS test and AIC to select appropriate model error metrics as a benchmark to refine the parameters until an acceptably accurate prediction can be produced consistently.
The main limitations of the hybrid model are its computation time, difficulties in interpreting results and selection of kernel function parameters. The computation time for executing the SVR algorithm is considerably slower than ARIMA. The training process of SVR can take up to several minutes if the training data set is large. The processing time of the model can be reduced if users can limit the range of hyper-parameters during grid search. Nonetheless, this limitation inhibits the usage of hybrid models in real-time prediction problems. The second limitation is that the results of hybrid models are difficult to interpret. This is because SVR is considered a black-box algorithm (Barbella et al., 2009), and users cannot inspect and change the SVR’s action amidst its execution. Unlike the traditional approach of utilising linear regression models, the results are not intuitive. This is partly related to the third limitation, which is the difficulty in deciding upon suitable kernel functions and corresponding hyper-parameters. Since SVR is a black-box algorithm, it is impossible to solve for an optimal set of hyper-parameters mathematically. The user must conduct multiple trials with the established model. The user must be able to interpret the results and tune the kernel function accordingly as well. This poses a challenge to the adoption of the hybrid model in Hong Kong’s power industry, since specialty in data analytics is still rare among engineers.

12. Conclusion

The goals of this study are to propose an ARIMAX-SVR hybrid model to forecast short term grid maximum demand and demonstrate its capability. The hybridisation of ARIMAX and SVR models has shown better forecasting performance than the non-hybrid approach, with the best model yielding a RMSE of 68.0MW. Empirical results show that the selection of forecast horizon and SVR model training period has a significant impact on the final prediction accuracy. In general terms, the performance is the best when the forecast horizon is limited to seven days, and that the selection of SVR training period should depend on the forecast horizon.

The study identified the advantage that the hybrid model attains over the non-hybrid and traditional models is the improvement in forecasting accuracy. For example, the RMSE saw a 53% improvement, decreased to 95.96 MW from 206.1 MW of the best-performing non-hybrid model. However, the hybrid model also has two disadvantages that could discourage its use. They are the increased computation cost and the increased difficulty in model tuning. The added computation cost mainly originates from the grid search for SVR hyper-parameters, and the difficulty in model tuning lies in the amount of trial and error required to obtain the optimal set of model parameters. These disadvantages make the hybrid model non-trivial to build and the deployment of this model requires adequate knowledge of the principles behind the model and domain knowledge of the subject.

Recent advances in the field of data science and machine learning have introduced many promising innovations over the tried-and-true algorithms, such as boosting and ensemble learning. While these new approaches tend to demand more computational resources, computer scientists have been seeking ways to leverage...
GPUs to boost computational speed and optimise their computational cost. Meanwhile, on the front of deep learning, attempts have been made to use convolutional neural networks and long-short-term memory networks to forecast time series data, since deep learning models demand a large amount of labelled data. Machine learning scientists have also been looking into reducing the reliance on labelled data by exploring unsupervised and semi-supervised learning. However, there are still in need of researchers with practical engineering data to test the capabilities of these models and allow engineers to leverage these technologies in real world conditions.

Acknowledgement

The author would like to thank Mr. Lau, Kevin Chi Keung of CLP Power Hong Kong Limited for endorsing the use of data for this study.

Notes on Contributor

Mr Chow Hon Fung is an engineer in CLP working at Black Point Power Station. His specialties include CCGT operation and HRSG maintenance. He is enthusiastic about data science and machine learning. He is currently pursuing their application to enhance business performance. Mr Chow received his M.Sc. degree in Mechanical Engineering from HKU and his B.Eng. degree in Mechanical Engineering from HKUST.

References


