IJHSR

ijhighschoolresearch.org

RESEARCH ARTICLE

AI-Based Power Demand Forecasting of California Counties

Maxwell Y. Chen, Shourya Bose, Yu Zhang

Department of Electrical and Computer Engineering, University of California, 1156 High Street, Santa Cruz, California 95064, USA; maxwellyhchen@gmail.com

ABSTRACT: Electricity is an indispensable form of energy in almost every aspect of our life. Balancing power supply and demand is critical in maximizing energy efficiency and preventing power outages. Towards this end, the ability to make reliable power demand predictions represents a key step, and artificial intelligence and machine learning are emerging tools. In this study, the power demands of selected counties in California are analyzed for the past 30 years by various models, including linear regression, polynomial regression, and autoregressive integrated moving averages (ARIMA). The simulation results show that ARIMA is an effective tool in predicting future power demand, with performance noticeably enhanced compared with those by linear and polynomial regressions.

KEYWORDS: Systems Software; Algorithms; Power Demand; ARIMA; Prediction.

Introduction

Electrical energy is a major engine that drives the economy and plays a critical role in almost every aspect of our life. Currently, the world consumes about 2.7 terawatts (TW) of electricity a year, which accounts for ca. 15% of the total energy consumption, including natural gas, oil, coal, and alternative energy sources like solar, wind, hydropower, etc.;[1] in the United States alone, approximately 0.4 TW of electricity is used in a year.[2] With such a huge demand, it is essential to develop effective strategies to minimize the waste of electricity by balancing supply and demand since it is costly and challenging to develop viable technologies for large-scale electricity storage. Towards this end, making sound predictions of power demand represents a crucial step. Yet conventional technologies for power demand forecasts are primarily empirical in nature, which compromises the accuracy and reliability of the results. The development of smart grid technology is a key building block in a sustainable economy.

Within this context, artificial intelligence (AI) has emerged as an attractive tool that can be used to make predictions of power demand in the future based on past usage patterns. AI is a computer science technology that renders it possible for the computer to perform tasks that traditionally require human intelligence. For instance, in a recent study,[3] Awalin and coworkers developed a machine learning platform based on Microsoft Azure cloud to predict energy consumption. The models were constructed with an algorithm based on three methods, i.e., Support Vector Machine (SVM), Artificial Neural Network (ANN), and k-Nearest Neighbor (k-NN). Data from two tenants in a commercial building were used for model training and testing, and the results show that energy consumption characteristics varied between the tenants. In another study, four random forest algorithms are used in a big data environment to make accurate and feasible predictions of household energy demand, where the consumer's socioeconomic status is found to play a key role in energy use. Wang *et al.* proposed a stacking model by integrating various base-prediction algorithms into "meta-features" such that the resultant model can analyze the data from different structural and spatial perspectives.[5] The results show an improved performance of the stacking method compared to conventional ones, such as Random Forest, Gradient Boosted Decision Tree, Extreme Gradient Boosting, SVM, and k-NN, in terms of accuracy, generalization, and robustness.

Autoregressive integrated moving average (ARIMA) represents another effective method in data analysis and trend prediction of power demand.[6] ARIMA is a time series technique with an implicit assumption that the future will follow the past pattern. Thus it can predict future values based on past values by smoothing time-series data using lagged moving averages. In the present study, the power demand in six select counties in California is analyzed from 1990 to 2019, three in Northern California (Alameda, Santa Clara, and San Mateo) and the other three in Southern California (Los Angeles, Riverside, and San Diego). Compared to linear and polynomial regressions, ARIMA exhibits markedly enhanced fitting to the actual data and can predict future power demand in the counties.

Methods

The power demand data were obtained from the California Energy Commission for the period of 1990 to 2019,[7] and the corresponding population data were retrieved from the Census Bureau.[8-10]

Three tools were used in this study, including (a) Pycharm is, an integrated development environment (IDE) used to write and execute code in Python, (b) NumPy, a plugin for Python that adds math functions and other useful tools, and (c) Matplotlib, which is, a plugin for Python and a library for the numerical mathematics extension NumPy that allows the ability to plot graphs.

Results and Discussion

Table 1: Average power consumption (Watts, W) per capita per year within 1990-2019 in six select counties in California.

Santa Clara	Alameda	San Mateo	Los Angeles	Riverside	San Diego
610 ± 26	812 ± 53	604 ± 63	3383 ± 388	1972 ± 420	4930 ± 360

Based on the power demand and population data, we first analyzed the average power consumption (Watts, W) per capita per year from 1990 to 2019. From Table 1, it can be seen that the per-capita power consumption varies rather markedly from one county to another and that the Southern California counties consume far more electricity per person than the Northern California counties, probably because of the drastically more extensive use of air-conditioning in the former. For instance, in the three northern counties, each person consumes 600- 800 W per year compared to 2000 - 5000 W in the southern counties. However, the variation within a specific county is mostly under 10% over the past 30 years. This suggests that per-capita electricity consumption is unlikely to be a significant factor in determining the power demand for these six counties.

Therefore, the study focused on a specific county's total power demand. Figure 1 shows the total power demand (gigawatts, GW) in these six counties over the past 30 years (red symbols). The data were first fitted with linear regression, y = mx + c, with m being the slope and c being the intercept. It can be clearly seen that while linear regression roughly captures the general trends of Alameda, Santa Clara, Riverside, and San Diego, drastic deviations can be seen for several data points, in particular, in the most recent years. For San Mateo and Los Angeles, linear regression is clearly not a good fit for the data.

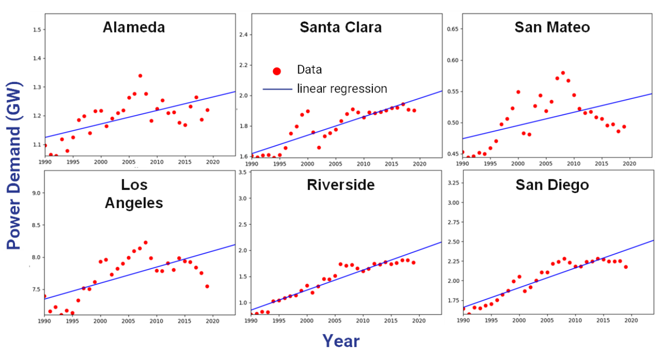

Figure 1: Power demands in six counties in California from 1990 to 2019. Symbols are actual data, and lines are linear regressions.

Further analysis was then carried out with a more complicated fitting model based on polynomial fitting. The datasets were divided into two parts to test the model's validity. Data points from the first 20 years were used as the fitting data for polynomial regression, and data points in the last ten years were used as the comparison data to assess the accuracy of the polynomial fits. In Figure 2, the red symbols are the training data, green symbols are the comparison data, and solid lines are the corresponding polynomial fits. One can see that of the six counties, the polynomial model over-predicted the power demand in comparison to the actual data for Alameda and Santa Clara counties. In contrast, it clearly failed to reflect the actual data trends for the other four counties. This suggests that polynomial modeling is not a reliable tool for predicting power demand, either.

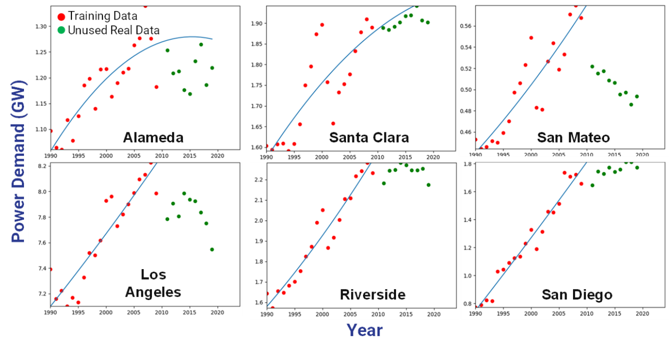

Figure 2: Power demands in six counties in California from 1990 to 2019. Red symbols are the actual data in the first 20 years for training, green symbols are the actual data in the last ten years for comparison, and blue lines are polynomial regressions.

In sharp contrast, a significant improvement in data prediction was achieved by using ARIMA models. Santa Mateo, Alameda, and Los Angeles were chosen as the illustrating examples as their power demand patterns were the most complicated in the series. As depicted in Figure 3, their actual datasets were divided into two parts: the first 25 years' data were used as the training data (red symbols), and the remaining five years' data were used as the comparison data. The ARIMA prediction data are shown in green, which can be seen to exhibit good agreement with the actual data for these three counties, suggesting that ARIMA indeed can be used as a reliable tool in making predictions about power demands.

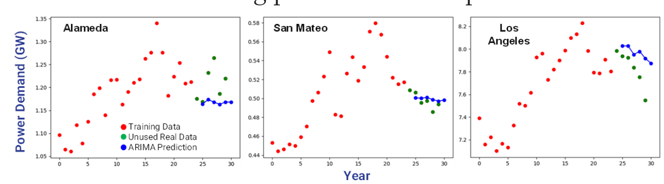

Figure 3: Power demands in three counties in California from 1990 to 2019. Red symbols are the actual data in the first 25 years for training, green symbols are the actual data in the last five years for comparison, and blue symbols/lines are ARIMA predictions.

Therefore, a further analysis was performed where the entire datasets were used as the training data to make predictions about the future. Figure 4 depicts the ARIMA power demand predictions for the next five years. For Alameda County, the power demand is expected to remain relatively steady, whereas it increases slightly for San Mateo and Los Angeles counties. In addition, all counties show moderate fluctuations in power demand in the next five years.

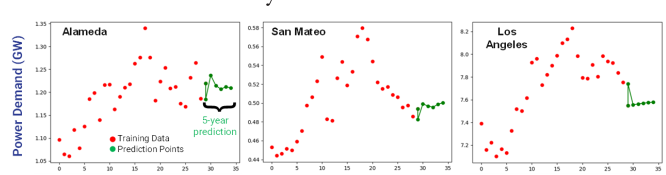

Figure 4: Power demands in three counties in California from 1990 to 2019. Red symbols are the actual data for the entire time period, and green symbols/lines are ARIMA predictions for the next five years.

DOI: 10.36838/v5i1.16

Conclusions

Power demand varies from one county to another, and ARIMA was demonstrated to be an effective tool in modeling and predicting power demand for various counties in California. The ARIMA performance is markedly better than linear and polynomial regressions. Based on the ARIMA predictions, the three select counties of Alameda, San Mateo, and Los Angeles are anticipated to see a moderate increase in power demand, with slight fluctuations, in the next five years. This study's results highlight ARIMA's unique significance in power demand forecasting, where additional variables may be included to provide a more comprehensive socioeconomic context. Further research is underway, and results will be reported in due course.

Acknowledgments

M.Y.C. would like to thank UCSC for the opportunity to participate in the research project.

References

1. https://www.statista.com/statistics/280704/world-power-consumption/
2. https://www.statista.com/statistics/201794/us-electricity-consumption-since-1975/#:~:text=Electricity%20consumption%20in%20the%20United,of%20electricity%20by%20power%20plants.
3. Shapi, M. K. M.; Ramli, N. A.; Awalin, L. J. Energy consumption prediction by using machine learning for smart building: Case study in Malaysia, *Dev. Built Environ.*, 2021, 5, 100037.
4. Caceres, L.; Merino, J. I.; Diaz-Diaz, N. A Computational Intelligence Approach to Predict Energy Demand Using Random Forest in a Cloudera Cluster, *Appl. Sci.*, 2021, 11, 8635.
5. Wang, R.; Lu, S.; Feng, W. A novel improved model for building energy consumption prediction based on model integration, *Appl. Energy*, 2020, 262, 114561.
6. Jahanshahi, A.; Jahanianfard, D.; Mostafaie, A.; Kamali, M. An Auto Regressive Integrated Moving Average (ARIMA) Model for prediction of energy consumption by household sector in Euro area, *AIMS Energy*, 2019, 7, 151-164.
7. http://www.ecdms.energy.ca.gov/elecbycounty.aspx
8. https://www.census.gov/data/tables/time-series/demo/popest/1990s-county.html
9. https://www.census.gov/data/tables/time-series/demo/popest/intrcensal-2000-2010-counties.html
10. https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-total.html

Author

Maxwell Y. Chen is a junior at the Georgiana Bruce Kirby Preparatory School. He wants to pursue a computer science major in college and is very interested in modern technological advancements. In his spare time, he likes to experiment with software and hardware.

Shourya Bose is a graduate student pursuing a Ph.D. in the Department of Electrical and Computer Engineering (ECE) at UCSC under the supervision of Prof. Yu Zhang.

Yu Zhang received his Ph.D. degree in ECE from the University of Minnesota and worked as a postdoctoral fellow at the University of California Berkeley and Lawrence Berkeley National Laboratory before joining UCSC as an assistant professor of ECE. His research interests span the spectrum from theory to real-world applications.

DOI: 10.36838/v5i1.16