# Evaluating prediction models for electricity consumption

Lasse A. Karlsen, Morten Goodwin

Faculty of Engineering and Science, University of Agder

Teknova AS, Grimstad, Norway

**Abstract** A system for visualizing electricity consumption data is presented in this paper, along with the implementation of a pattern recognition approach for peak prediction. Various classification algorithms and machine learning techniques are tested and discussed; in particular, Support Vector Machine (SVM), Gaussian Mixture Model (GMM) and hierarchical classifiers. Most notably, the best algorithms are able to detect 70% of the peaks occurring within the next 24 hours. Also, various ways of correlating energy consumption are considered in the present context. Finally, a few directions for future work are discussed.

## 1 Introduction

Peaks in energy consumption are a major challenge for energy companies, since they appear seemingly randomly and requires the power grid to support any level of load to avoid power outages. In order to control the peaks and prevent potential loss of electricity, they need to be known prior to their occurrence. However, consumption peaks are influenced by numerous factors, e.g. human behavior, and thus hard to predict. We are creating an application that can help visualize energy consumption and potentially predict peaks 24 hours into the future. The application will implement a pattern recognition approach for peak prediction, using supervised machine learning algorithms for classification. The classification output can then be used as input to techniques for controlling the peaks.

## 2 Data

Data is collected through the external DeVID project [1]. The data we are using for this project comes from Hvaler, Norway, collected over a time period of 6 weeks [2]. This is an area with approximately 8000 different installations with various degree of activity. From each installation we have the accumulated energy consumption per hour (kWh). In other words, we have actual real world data to work with.

*This paper was presented at the NIK-2014 conference; see* http://www.nik.no/.

### 2.1 Electricity peaks

Based on mathematical outlier detection, a peak in electricity consumption is defined as follows: For a data array $t$, of which $t_i \epsilon\, t$, $t_i$ is defined as a peak according to the following equation [3]:

$$p(t_i, n) = \begin{cases} \text{peak if } t_i > \mu + n * \sigma \\ \text{not peak otherwise} \end{cases}$$

$\mu$ is the arithmetic average and $\sigma$ is the standard deviation of `t`. n is defined as the extremity of the peak. In other words, an energy value is defined as a peak if the value is more than `n` times the standard deviation above the average consumption.

Peaks in electricity consumption are seemingly random, as they are influenced by complex user behavior and many external factors. This makes peak prediction far from trivial. Peaks are rare, meaning we have little data to learn from.

### 2.2 Third-party data

Third-party data is generated outside the realm of the existing system and is not guaranteed to always be available. It is often produced on other platforms or aggregated through different entities. Many approaches to peak prediction that exists in the literature assumes that data which influences energy consumption are available [4]. However, this can not be taken for granted. Nonetheless, we will try to use weather forecasts in our correlation effort, to see if it may effect the outcome of the prediction.

## 3 Results and discussion

The aim of this project was to create an application that could both visualize energy consumption data and predict potential peaks in the system. Various classification algorithms are implemented and successfully applied on real-life data from a Norwegian smart-grid project. The pattern recognition approach shows promising results for a system that does not rely on third party data. We were also able to correlate the consumption by adding weather forecasts to the training data.

### 3.1 Classification results

The available data set should be large enough for a realistic test of real-life application performance, although many installations have virtually zero energy consumption, or a very distinctive consumption pattern. Yet, a comparison to a selection of installations may give an indication of the overall ability to detect peaks.

Figure 1 shows recall for all algorithms predicting in intervals of 5 hours up to 24 hours into the future with $n$ set to 1. 100 installations with average consumption above 0.5 were tested and the average result is illustrated in the graph. Recall is the fraction of relevant instances that are retrieved, meaning that an algorithm with high recall returned most of the relevant results. Note that the x-axis is not time of day, but how many hours into the future we are trying to predict.

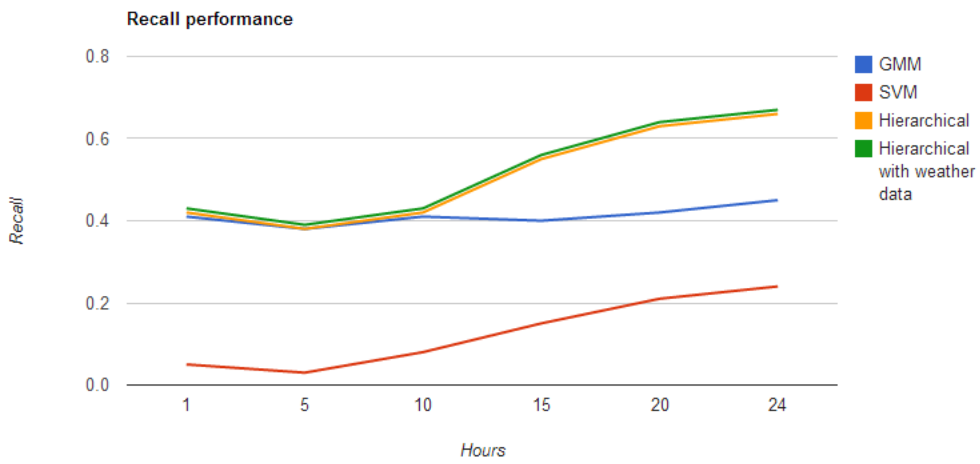


Figure 1: Recall

The GMM classifier is relatively constant, with a small increase when reaching 24 hours. The results from the linear SVM alone are not particularly good. When predicting 1-10 hours into the future, the classifier is rarely able to detect any peaks. However, between 15-24 hours, it is able to predict approximately 25% of the existing peaks.

When combining the two classifiers and predicting 24 hours into the future, the hierarchical classifier is able to detect approximately 70% of the peaks in the data.

Adding weather forecasts to the learning phase have seemingly zero impact. The reason for this might be that the data is from mid 2012, in a month where the temperature was rather constant with little or no fluctuation.

GMM and SVM are standard classification techniques. The hierarchical classifier is an extension which combines GMM and predictions of same day behavior. Since this classifier, based on repetitions, works so well, it is natural to conclude that people live in cycles (e.g. eat dinner at the same time each day). Thus, a classifier utilizing the cyclic behavior in the data fits well the cyclic behavior of the consumers.

## 4 Conclusion

As has been seen, it is possible to implement a peak prediction system based on pattern recognition with reasonable accuracy, even with relatively simple classification algorithms, and without the use of third party data. For testing purposes, a correlation with weather forecasts were implemented. A simple hierarchical classifier, combining linear support vector machines, Gaussian mixture models and weather data, appears to be a promising choice for a system of this form. The classification output is ment to be used in smart grid systems, as an aid to load balancing and smart pricing strategies.

## 5 Future work

As the use of rather simple classification algorithms have shown promising results, it would be natural to consider similar algorithms, either standalone or in a hierarchy. In particular, k-nearest neighbors, naive Bayes and neural networks, all of which focuses on the recognition of patterns and regularities in data. This would also help assess the current system compared to more advanced hierarchical classifiers, as well as improving the overall performance.

Furthermore, larger-scale tests would be interesting - if a system like this is capable of working well with a limited amount of data, image how great it would work with unlimited amounts of data. Studies have shown that electrical installations are able to learn from each other, so new installations can utilize already existing consumption data [3]. However, with larger data sets, more sophisticated searching and processing algorithms would also probably be needed, eventually moving into the realm of distributed computing.

## 6 Acknowledgements

This application is intended as a part of the project "Grid Operation and Distributed Energy Storage: Potential for improved grid-operation efficiency", funded by Agder Regional Research Fund, and the project "Electricity Usage in Smart Village Skarpnes", funded by the Norwegian Research Council. These projects are carried out together with Teknova, Agder Energi Nett and Eltek.

## 7 References

[1] Demonstrasjon og verifikasjon av intelligente distribusjonsnett (DeVID). `http://www.sintef.no/Projectweb/DeVID/Prosjektpartnere/`

[2] Sand, K., Foosnas, J., Nordgard, D.E., Kristoffersen, V., Solvang, T.B., Wage, D.: Experiences from norwegian smart grid pilot projects. In: Electricity Distribution (CIRED 2013), 22nd International Conference and Exhibition on, IET (2013) 1-4

[3] Goodwin. M, Yazidi. A: A Pattern Recognition Approach for Peak Prediction of Electricity Consumption. (2014). To be published in the proceedings of 10th International Conference on Artificial Intelligence Applications and Innovations.

[4] Feinberg, E., Genethliou, D.: Load forecasting. In Chow, J., Wu, F., Momoh, J., eds.: Applied Mathematics for Restructured Electric Power Systems. Power Electronics and Power Systems. Springer US (2005) 269-285

[5] Campbell, P.R., Adamson, K.: Methodologies for load forecasting. In: Intelligent Systems, 2006 3rd International IEEE Conference on, IEEE (2006) 800-806

[6] Weron, R.: Modeling and forecasting electricity loads and prices: A statistical approach. Volume 403. Wiley. com (2007)

[7] Mohsenian-Rad, A.H., Leon-Garcia, A.: Optimal residential load control with price prediction in real-time electricity pricing environments. Smart Grid, IEEE Transactions on 1 (2010) 120-133