Machine Learning in Wind Energy Information Systems

Oliver Kramer, Nils André Treiber, Fabian Gieseke

Department of Computing Science
University of Oldenburg
Uhlhornsweg 84
26111 Oldenburg Germany

1 Introduction

For the integration of wind and photovoltaic power, a precise forecast of energy has an important part to play. Up to now, the integration of decentralized energy into the electricity grid is often ignored. Further, it is estimated that the grid gets unstable, if the amount of integrated renewable energy exceeds about 15 to 20%. Since the amount of integrated wind and solar energy resources is steadily increasing, a precise prediction for subhourly scheduling becomes necessary for a successful integration of the resources into the grid. Effective forecast systems will allow balancing and integrating multiple volatile power sources at all levels of the transmission and distribution grid [4]. The increasing number of sensors installed in wind turbines and wind farms allows their detailed observation and monitoring. The new data can be used for short-term prediction of wind energy. However, the efficient extraction of meaningful patterns and the learning process of large wind energy data sets requires the application of powerful data analysis tools.

Among the most prominent techniques are the so-called support vector machines [6, 5], which depict one of the state-of-the-art schemes for learning tasks like classification (i. e., automatic partition of objects into classes) and regression tasks (i. e., the prediction of a real value) [1]. The goal of the corresponding schemes is to generate high-quality models that can make reasonable predictions for new patterns, based on the knowledge given by the observed training data. In this paper, we consider support vector machines to address the task of predicting the wind energy of single wind turbines based on time-series data that is available for neighbored turbines and provide a detailed analysis of several learning scenarios, thus extending our previous work [2, 3] in this context. Our main concern are short-term prediction systems that can provide valuable estimates for time-horizons up to 20, 60 and 120 minutes, respectively.

2 Support Vector Regression Models

One of the most popular tools in the field of machine learning are support vector machines, which can be used for classification, regression and a variety of other learning settings [1, 5, 6]. In general, the basis for training appropriate models is a set $T = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\} \subset \mathbb{R}^d \times \mathcal{Y}$ consisting of labeled patterns. For classification settings, the space $\mathcal{Y}$ of labels is discrete (e.g., $\mathcal{Y} = \{-1, +1\}$ for the binary case). For regression scenarios, the space $\mathcal{Y}$ is given by $\mathbb{R}$; here, the goal of the learning process consists in finding a prediction function $f : \mathcal{X} \to \mathbb{R}$ that maps unseen patterns $\mathbf{x} \in \mathcal{X}$ to reasonable real-valued labels. These models can be seen as a special instance of problems having the form

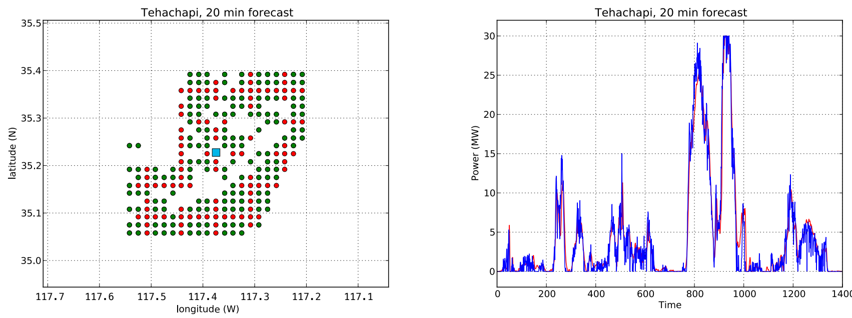
$$\inf_{f \in \mathcal{H}, b \in \mathbb{R}} \frac{1}{N} \sum_{i=1}^{n} L\big(y_i, f(\mathbf{x}_i + b)\big) + \lambda ||f||_{\mathcal{H}}^2, \tag{1}$$

where $\lambda > 0$ is a user-defined real-valued parameter, $L : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ a loss function and $||f||_{\mathcal{H}}^2$ the squared norm in a so-called *reproducing kernel Hilbert space* $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}} = \{f : \mathcal{X} \to \mathbb{R}\}$ induced by an associated *kernel function* $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. The space $\mathcal{H}$ contains all considered models and the term $||f||_{\mathcal{H}}^2$ is a measure for the "complexity" of a particular model f [1, 6, 5]. Ideally, one would like to generate models that represent the training data well and that are, at the same time, not too complex to avoid overfitting. This idea is reflected by the two objectives present in the above optimization task: The first term measures how well the model f fits to the data (according to the definition of the loss function L), whereas the second term measures the complexity of the model. The parameter λ is called *regularization parameter* and determines the trade-off between these two objectives.

3 Experimental Analysis

We formulate the wind forecasting task as regression problem and assume that a time series $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^d$ of N wind measurements of K wind grid points and corresponding measurements $y_1, \ldots, y_N \in \mathbb{R}$ of wind energy production of a target point is given. The task is to predict the production $\mathbf{y}_t$ at time $t = t_i + \theta$ with $\theta \in \mathbb{N}$ based on past wind measurements at time $t_i - 1, t_i - 2, \ldots, t_i - \mu$, with $\mu \in \mathbb{N}$ past observations. We investigate the questions, if prediction of wind energy can exclusively be based on the existing infrastructure of wind turbines for three different time horizons, i.e., 20 minutes, one hour and two hours.

The question is how far we can look into the future and how much information from the past is necessary for a reliable forecast. After parameter tuning and training of the SVR with optimal settings on the first nine months of 2006, we evaluated the SVR models on randomly chosen time series intervals of the last quarter of 2006. The intervals contain two major peaks (ramp events), which are very relevant for balancing the grid. Such events can be caused by storm fronts passing. Each evaluation interval consist of 1,400 time steps that correspond to approximately ten days. Figure 1 shows the experimental results of a 20-minute ahead prediction. On the left, the environment of the Tehachapi wind park is visualized with the target wind grid point in the center, training input grids points are shown in green. Figure 1(b) compares the 20-minute SVR prediction (red) with the actual wind energy (blue).



(a) (b)

Fig. 1. 20-minute forecast of a wind grid point in Tehachapi: (a) selection of input grid points (green), (b) comparison of SVR prediction (red) with actual wind energy (blue). Prediction and target wind energy curve match perfectly.

We can observe that the precision of the 20-minute ahead forecast is very high. The two peaks in the second half of the interval almost perfectly match. For larger time horizons, grid points spread in a larger neighborhood were necessary for optimal predictions. Although the accuracy is relative high, small deviations from the target curve can be observed. This observations is even more evident in case of the 2-hour forecast with a tendency towards a temporal shift. The predictions will be evaluated w.r.t. statistical properties like RMSE, maximum and minim deviation from the target values.

4 Conclusions

Wind energy prediction is an important aspect for a stable energy grid. The integrity and stability can be improved with a precise forecast of the volatile energy sources. We have demonstrated that SVR is a successful method for the prediction of wind energy production only based on wind measurements from wind turbines, in particular without further meteorological data or weather simulation models. SVR turns out to be a fast and robust time series prediction technique. For the wind resource scenarios we have found recommendable parameters for the ε-loss and typical bounds for optimal kernel parameters. The lattice introduced in this paper is a new contribution and is a convenient way to select input grid points for the learning process. Our experiments have shown that a reliable forecast on the level of wind grid points is possible for the 20-minute and 2-hour ahead prediction. On the level of a whole wind park, the results have shown that even a reasonable six-hour forecast is possible [3].

References

1. T. Hastie, R. Tibshirani and J. Friedman. *The Elements of Statistical Learning*. Springer, 2 edition, 2009.
2. O. Kramer and F. Gieseke. Analysis of wind energy time series with kernel methods and neural networks. In *Proceedings of the 7th International Conference on Natural Computation*, pages 2381–2385, 2011.
3. O. Kramer and F. Gieseke. Short-term wind energy forecasting using support vector regression. In *Proceedings of the International Conference on Soft Computing Models in Industrial and Environmental Applications*, Advances in Intelligent and Soft Computing, pages 271–280, 2011.
4. M. Milligan, K. Porter, E. DeMeo, P. Denholm, H. Holttinen, B. Kirby, N. Mille, A. Mills, M. O'Malley, M. Schuerger and L. Soder. Wind power myths debunked. *IEEE Power and Energy Society*, February, 2009.
5. B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, Cambridge, MA, USA, 2001.
6. A. J. Smola and B. Schölkopf. A tutorial on support vector regression, 1998.