Load forecasting for smart electricity distribution in SUNSEED project

Blaž Kažič
Jožef Stefan Institute
Jožef Stefan International
Postgraduate School
Ljubljana, Slovenia
blaz.kazic@ijs.si

Klemen Kenda
Jožef Stefan Institute
Jožef Stefan International
Postgraduate School
Ljubljana, Slovenia
klemen.kenda@ijs.si

Dunja Mladenić
Jožef Stefan Institute
Jožef Stefan International
Postgraduate School
Ljubljana, Slovenia
dunja.mladenic@ijs.si

ABSTRACT

One of the major goals of the SUNSEED European project is to improve the observability of the electricity distribution grid. The improved observability is one of the key challenges in the context of introducing smart grids. To avoid high investment costs into reinforced grid, the grid requires some intelligence, where all decisions are based on the in-real-time estimated state of the grid, as well as on the forecasted state of the grid. In this project showcase we present the short term load forecasting solutions developed during the project and its connectivity with the three phase state estimation module.

CCS CONCEPTS

• **Computing methodologies → Machine learning**; *Modeling and simulation*;

KEYWORDS

smart grid, short term load forecasting, state estimation

ACM Reference Format:
Blaž Kažič, Klemen Kenda, and Dunja Mladenić. 2018. Load forecasting for smart electricity distribution in SUNSEED project. In *Proceedings of KDD 2018 Conference (KDD2018).* ACM, New York, NY, USA, Article 4, 3 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

From the distribution system operators' perspective, observability of the electricity distribution grid is in particular important to allows higher penetration of distributed resources into the grid. By closely following the operation of the grid, potential disturbances can be identified and eliminated at the place of their origin. Furthermore, a forecasting module enables additional possibilities for control and future planning, offering applications, such as load switching and optimization, decision support for power systems operations, maintenance planning and trend detection.

KDD2018, August 2018, London, United Kingdom

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

SUSEED is an European FP7 EU project[1], actively running from February 2014 to August 2017. Project consortium consists of 9 partners from 6 different countries, including research organizations (Jozef Stefan Institute, Aalborg University, TNO) and partners from industry (Telekom Slovenije, Elektro Primorska, Elektroservisi, Gemalto, Gemalto M2M, Toshiba Research Europe).

In the scope of the project, several short term load forecasting models and three phase state estimation model were developed. We present some of them in the following two sections, followed by the modeling prototype architecture and use cases descriptions. We conclude with the discussion of closely coupling forecasting and state estimation components in order to improve the overall performance.

2 SHORT TERM LOAD FORECASTING

Forecasting module is one of the crucial building blocks in smart grids. In SUNSEED we have developed a decentralized, data-driven, streaming forecasting models, capable of processing large amount of real-time heterogeneous data streams and produce predictions on various nodes in the grid, from 5s to 24h into the future. In order to build the most effective model for the specific power grid (they differ in geographical properties, meteorological factors, grid topology, etc.), detailed exploratory data analysis was done. Beside the real-time grid measurements from the testbed (from AMI smart meters and WAMS meters), two additional external data-sources were taken into account: static date-time data (time of day, working day, holiday status, etc.) and weather data (current measurements and forecasts). For each data-source, additional streaming data aggregates were developed for the purpose of data cleaning, preprocessing and data enrichment (feature engineering).

In order to identify the most valuable and important features, importances were computed for all the extracted features for different prediction horizons and for different sensor nodes (with different levels of load aggregation). Analysis showed that beside the real time measurements data, static data such as *"date-time"* features and *"holidays"* are the most valuable features. Including weather forecasts might also improve accuracy, but not significantly. Results were also used for feature selection, in order to reduce model complexity and discard irrelevant or redundant features, which resulted in simpler models, shorter training times and reduced model overfitting.

[1]Partially funded by EU and EC under Grant agreement no: 619437 (http://sunseed-fp7.eu/)

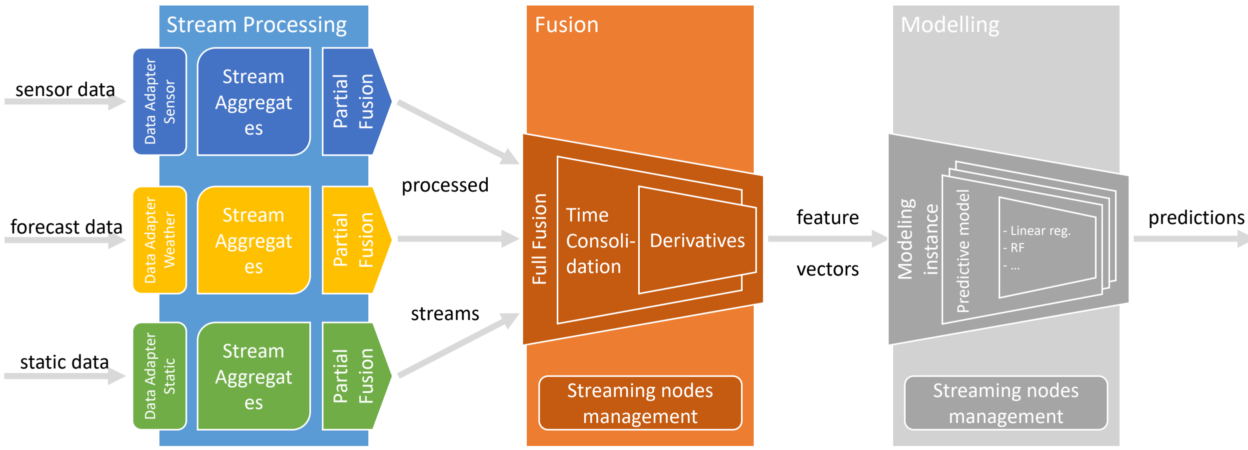

Figure 1: Data driven modeling architecture, capable of processing various real-time heterogeneous data streams.

Various data-driven machine learning models (i.e. Ridge Regression, K Nearest Neighbors, Random Forest Regressor and Neural Networks) were trained and evaluated with different feature sets in order to find the most appropriate combination of method and constructed feature set. Results show that Ridge Regression model and Random Forest models produce the most promising performance in terms of prediction accuracy, speed and model complexity. Further analysis also reveals that Random Forest models tends to use special features (such as holiday status) better, and therefore produces better forecasts for non-regular days (such as holidays seasons), but the model is more difficult to interpret. Evaluation analysis also show that the forecast accuracy decreases with forecasting horizon, meaning that the further into the future we are predicting - the worse the prediction, which is intuitive. And furthermore, lower aggregation nodes (approx. mean load 1kWh) performance at the individual level show much higher results (around 40 % MAPE), than with higher level nodes (approx. mean load 500kWh) (around 10 % MAPE or less), which is comparable to results in Sevlian et al. [4]. Extensive data analysis and model evaluation results are presented in the SUNSEED deliverable report [2] and will be published in an upcoming publication.

2.1 Use cases

Final forecasting prototype contains several decentralized models (for each sensor node from low and mid voltage grid), and supports three important project usecases:

- **Short Term Load Forecasting (STLF):** the main data-source for the STLF are smart meter (AMI) measurements (data is sent once per day, for 750 different nodes). Based on this data-source, the model is able to forecast the load for specific node for various prediction horizons: from 1h to 24h into the future. These predictions can be used by the distribution system operators, as decision support for power systems operations and as up-to-date estimations of the grid for state estimation module.
- **Very Short Term Load Forecasting (VSTLF):** the main data-source are WAMS (Wide Area Measurement System) measurement units, with a much higher streaming frequency than STLF (50Hz), which generate vast amounts of data (1GB per day per unit). In this case, the model is calculating very short term predictions - from 5s to 15min into the future, for 16 different nodes. Such extreme short term predictions are useful for autonomous load management applications and various control system modules.
- **User profiling and clustering:** the main data-source are smart meter (AMI) measurements. Weekly and daily consumption profiles (histograms) are calculated for all nodes once per day, which give an expert user additional introspection into the behavior of a certain consumer. This information is useful to grid operators for grid management, planning and maintenance. Consumption profiles also turned out to be very useful as feature vectors for clustering and anomaly detection.

3 THREE PHASE STATE ESTIMATION

During the project, three phase state estimation, based on a nonlinear Weighted Least Square (WLS) approach, was also developed. It supports a wide variety of measurement types and is capable of computing the system state in near real-time. Its performance depends on (i) the accuracy of the measurements, obtained by smart meters (AMI) and custom developed synchro-phasor measurement units (WAMS-SPMs), and (ii) on the parameters of the network model. Thus, for the DSO grid model series impedance and shunt admittance matrices were developed for all underground cables and overhead lines focusing to the specific testbed. The state estimation was successfully tested using real measurements from the project deployment site.

4 PROTOTYPE ARCHITECTURE

Final streaming prototype was built on the top of the QMiner[2] open-source framework [1] and supports processing a large-scale real-time data streams, from multiple sources in an online fashion. Prototype consists of three main streaming components (see Figure 1): (i) Stream processing (responsible for online handling data streams – cleaning, pre-processing), (ii) Fusion (includes mechanism for online merging and feature engineering), and (iii) Modeling component (includes online regression forecasting algorithms) [3].

To ensure the scalability of the system, the developed prototype due to its decentralized architecture enables parallelization (running different sets of models on different instances in parallel; in the current setup we are running one instance for 5 seconds event horizon and another for all the other models). Instances can be run anywhere within the closed network, where MQTT data is available.

5 DISCUSSION

During the project we have discovered that due to the nature of smart meter (AMI) measurements—which are primarily used for billing purposes and therefore obtained only once per day—the measurements are generally already obsolete at the time of the estimation of the grid state by the state estimation module, which needs as up-to-date measurements as possible. By closely coupling the state estimation module with the load forecasting module, forecasting models can be used to predict smart meter measurements for the time of interest and use them as up-to-date pseudo-measurements in the unified three-phase model for distribution network components, implemented in the power flow and state estimation solvers. On the other hand, state estimations' outcomes can be used by forecasting models, which gives the models additional information about the part of the electrical grid where there are no sensors installed, and potentially improve the overall forecasting performance.

REFERENCES

[1] Blaž Fortuna, Jan Rupnik, Janez Brank, Carolina Fortuna, Viktor Jovanoski, Mario Karlovčec, Blaž Kažič, Klemen Kenda, and Gregor Leban. 2014. QMiner: data analytics platform for processing streams of structured and unstructured data. In *Software Engineering for Machine Learning Workshop, Neural Information Processing Systems*.

[2] Blaž Kažič, Maja Škrjanc, Aleš Švigelj, Kemal Alič, Zhu Ziming, Silviu Nistor, Primož Škraba, and Dunja Mladenić. 2015. *SUNSEED deliverable report D4.2.1: Smart grid forecasting algorithms and mechanisms description.* Technical Report. Artificial Intelligence Laboratory, Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia. https://plus.si.cobiss.net/opac7/bib/29471015.

[3] Klemen Kenda, Blaž Kažič, Luka Stopar, Blaž Fortuna, Jan Rupnik, Maja Škrjanc, and Dunja Mladenić. 2018. Data Fusion Framework for Streaming Heterogeneous Data Sources. *Journal of Computer Science and Technology (submitted)* (2018).

[4] Raffi Sevlian and Ram Rajagopal. 2014. Short term electricity load forecasting on varying levels of aggregation. *arXiv preprint arXiv:1404.0058* (2014).

[2] https://qminer.github.io/