Study Committee D2

Information systems and Telecommunication

Paper 10619_2022

APPLICATION OF MODERN TIME-SERIES ANALYTICS TOOLS TO IMPROVE PEAK LOAD MANAGEMENT AND PLANNING THE EPU DEVELOPMENT

Pavel Litvinov, Sergey Nesterov

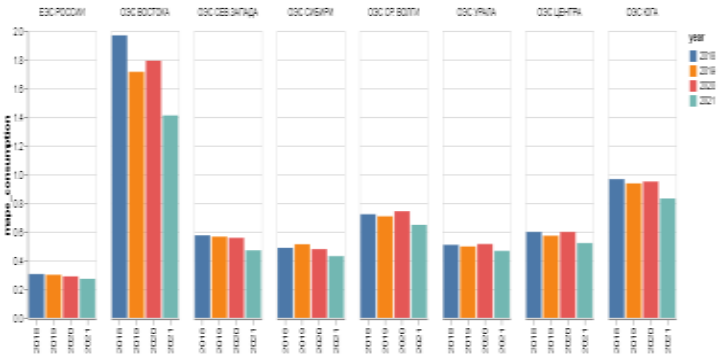
INTELAB, LLC

Motivation

- The solution of the key tasks of the electric power industry - maintaining the balance of generation and consumption in the power system, taking into account the existing restrictions, is becoming an increasingly difficult task. One of the objective reasons for this is the development of weather-dependent renewable energy.
- Long-term planning of consumption is also becoming a challenge - the trend towards energy saving opposed to the development of industry, electric transport, an increase in the number of devices using batteries, etc.
- The authors decided to jointly test artificial intelligence methods and simulation tools to solve a complex multifactorial and multiparametric planning tasks and forecasting problem.

Datasets

- 300,000 hourly values of electricity consumption
- 7 regions (Energy Systems)
- 4+ years



RME of consumption forecast by regions of Russia

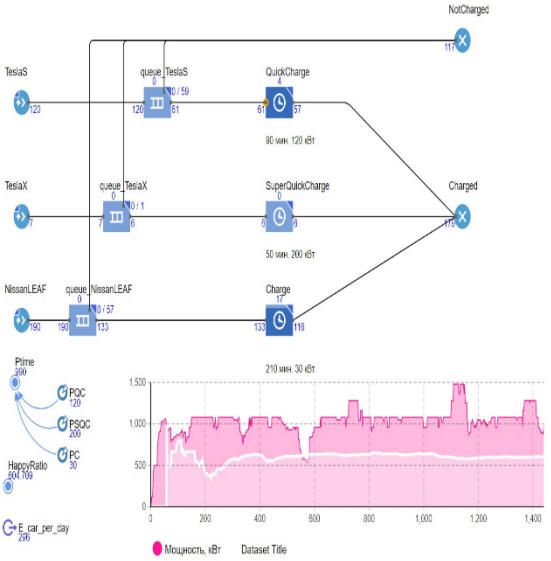
Example of visual display of peak load hours by region

Time series in-memory processing and visual displaying

- For the processing and visual display of time series, was used the methodology, the new thing of which consists in the preparation, transformation and processing of data exclusively in RAM, with guaranteed preservation of the row data.
- The popular open source Grafana was chosen to display the time series.
- Abnormal values are marked with red dots on the line chart. Temperature maps and histograms charts visually show the probability distribution of peak hours separately for the last five years and within the selected interval, which allows you to visually assess the differences between the specified period and the average values.

Study Committee D2
Information systems and Telecommunication
Paper 10619_2022

APPLICATION OF MODERN TIME-SERIES ANALYTICS TOOLS TO IMPROVE PEAK LOAD MANAGEMENT AND PLANNING THE EPU DEVELOPMENT continued



Simulation model of the charging station

Simulation models

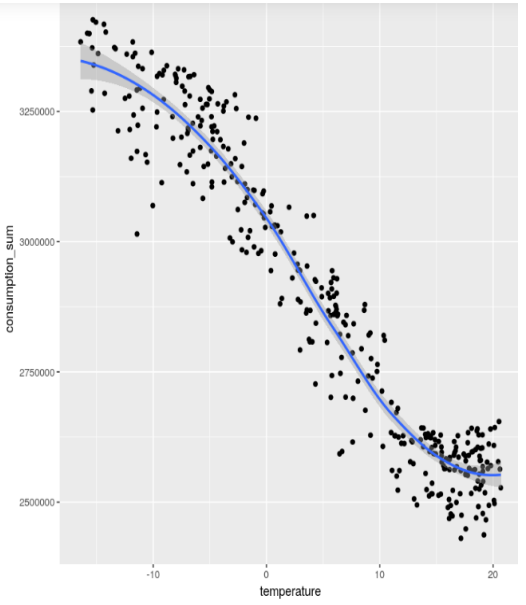
- For long-term forecasts, when there was no historical data, simulation modeling was applied.
- An example of such data for Russia is the consumption profile of charging stations for electric vehicles.
- The rapid growth in demand for charging infrastructure must be taken into account when planning the development of distribution networks, especially in large cities.
- To produce reasoned recommendations for planning the development of power grid companies in case of unavailable actual data, AnyLogic software simulation models have been applied.
- To build long-term multivariate forecasts, a system dynamics model was used.

Machine Learning

- As a key factor, the consumption forecast was selected, followed by the prediction of peak load hours.
- On a small number of parameters and values, preliminary estimates of the effectiveness of using various methods were made:

Random forest	0.441
Gradient Boosted Trees	0.389
Ridge (L2) regression	0.054
XGBoost	0.414
Decision Tree	0.452

- All algorithms except Ridge regression showed comparable accuracy results. The authors chose the XGBoost as the main, since one is easy to adapt and configure with a subsequent increase in the number of parameters.
- Historical data integrally take into account the influence of many parameters. Prediction results significantly improve additional data enrichment:
  – weather;
  – holidays;
  – geographical coordinates of the area, influencing the time of sunset and sunrise;
  – calendar (days of the week, month and season);
  – air temperature forecast.



Correlation between temperature and energy consumption

Study Committee D2

Information systems and Telecommunication

Paper 10619_2022

APPLICATION OF MODERN TIME-SERIES ANALYTICS TOOLS TO IMPROVE PEAK LOAD MANAGEMENT AND PLANNING THE EPU DEVELOPMENT continued

An example of a consumption forecast built using a neural network

Discussion

- Difficulties in data collection:
  - not all EPU collect and store in their databases data of the required quantity and quality databases. The term "necessary" includes sufficient depth of history, synchronized time stamps, comparability of measurement methods, the availability of service information explaining the history of deviations, etc.;
  - some data has commercial value and is not provided for analysis, for fear of being used by competitors;
  - there is no market for selling data that has importance for the forecast, except for weather information;
  - there are no motivation for mutually beneficial exchange of data, methods and forecasting results.
- All participants are interested in improving the forecasting accuracy, but the actual accuracy especially for some regions, remains rather low. The most modern prediction methods, including machine learning, do not increase one too much. There are several objective reasons:
  - neural networks training requires large amounts of labeled data. An increase in the number of years included in the training sample, predictably, first improves the forecast accuracy, and then begins to worsen it;
  - the degree of influence of various factors changes over time. For example, the correlation with daylight hours becomes weaker as the use of LED lamps expands, and as the number of electric vehicles increases, the evening peak will increase;
  - ongoing climatic changes, in particular, sharp fluctuations in temperature or atypical values for the region;
  - researchers have serious difficulties in obtaining historical datasets containing the necessary features: actual consumption volumes, the factory production program, planned switchings and outages, new consumers, etc. Moreover, such information is not available in real time;
  - most of information the from the public domain and open data is published on web pages, which requires the development of tools for extracting, normalizing and loading it into the database.

Used software and libraries for the R and Python programming languages

Domain	Open source Software or Library	
	R	Python
Web scraping	rvest	Selenium, BS
Data preparation	dplyr	Pandas
Database	MySQL	
Statistics	stats	Statsmodels
Machine Learning	–	Sklearn
Neural Network	–	TensorFlow
Charts and Diagrams	ggplot2	Pyplot, seaborn
Reporting	R Markdown	–
Interactive Computing Platform and IDE	RStudio Server, Jupyter Notebook	Jupyter Notebook
Time series visualization	Grafana	
Dashboard	Apache Superset	
Simulation	AnyLogic	

Conclusion

- One of the important results can be considered the conclusion that the selected set of libraries for R and Python can be effectively used for processing time series and solving forecasting problems.
- The forecast precision was improved due to the preliminary removal of abnormal values from the training sample. The best results were shown with the use of Isolation Forest algorithm.
- The need to add time series with the air temperature in the region to all data sets was confirmed. In Russia, there is a strong correlation between temperature and energy consumption. The correlation calculated by different methods was: Pearson correlation coefficient - 0.961, Spearman's rank correlation coefficient - 0.936.
- The measurements of the execution time of various operations confirmed the advisability of using RAM to speed up calculations. For example, loading a time series of 30,000 values took 34 ms. The following transformations, including filtering, grouping, and statistical functions, took 7ms. This opens up the possibility of implementing automatic control or creating systems of the Intelligence Amplification that function as a guide in real-time.
- It is shown that it is impossible to highly improve the accuracy of forecasting without using additional information from consumers, for example, planned dates of production interruption, changes in work schedules or production programs, installation of local generation, etc. As a solution, it is planned to create a pilot service that allows consumers to send such information in a suitable form, which after anonymization will be used to improve the accuracy of forecasts in the energy district.