Politechnika Gdańska

Wydział Elektrotechniki i Automatyki

Gdańsk 2015

PREDICTION OF CONSUMER ELECTRICITY NEEDS BASED ON INTERNET WEATHER FORECASTS

Artur OPALINSKI

Politechnika Gdańska, Wydział Elektrotechniki i Automatyki
tel.: 58 347 1347, e-mail: Artur.Opalinski@pg.gda.pl

Abstract: Electrical energy is considered both as an important driver for producing and transporting goods in companies, as well as a good in itself which requires planning and management for generating and delivering it to consumers in proper time and amounts. Weather information can be considered to convey part of the data on energy delivery needs of consumers. Free meteorological data sources on the Web do not offer consistent data to automate energy consumption forecasts. The paper identifies and addresses these inconsistencies to provide for automatic data gathering and supply to a demand forecasting model.

Keywords: demand prediction, electrical energy, weather forecasts.

1. INTRODUCTION

Short-term load forecasting (STLF), with lead times ranging from a few hours to several days ahead, helps electrical grid operators to make cost effective scheduling of resources, purchase of energy, and maintenance and security analysis studies [1][2].

SLTF can be based on meteorological parameters from weather forecasts. For selected USA locations the range of annual energy consumption due to actual weather variations can be as much as +7.0% or -11.0% from long-term average weather patterns [3]. Annual peak electrical demand variation for the same locations is -4.7% to 4.9% [3]. In Poland, the most meteorological influence on demand is attributed to the temperature [4], as it is in other parts of the world [5] Other meteorological conditions are considered less frequently, but nevertheless are also reported as influencing energy use in specific cases [6][7]: sun light, relative humidity, wind direction [8] and speed [8][9], barometric pressure [8].

Recent opening of electricity markets in many countries complicates weather - based predictions, as customer locations are no longer geographically constrained to a continuous, neighboring region. Additional data may therefore be needed for some STLF, including weather forecast data for diverse customer locations.

The question of weather data sources and data acquisition methods for commercial, non-academic use of the existing models remains in shadow. It can be guessed that data originating from Institute of Meteorology and Water Management (Instytut Meteorologii i Gospodarki Wodnej, IMGW) is primarily used by commercial companies in Poland, and data from similar, relevant meteorological institutions is used thorough the world (ECOMET institutions in European countries). Meteorological data is expensive for commercial use in volume and in digital form when provided by IMGW, or when provided by other ECOMET institutions, and hard to obtain without their involvement [10].

Meteorological institutions, including IMGW, are also required by law to publish free daily weather forecasts. For a long time this forecast data could only be manually extracted from IMGW publication archives or from popular newspaper archives for free; the daily forecasts could also be sourced form TV or radio broadcasts. Due to the recent technological advancements, a lot of useful information exists on the World Wide Web, including free weather forecasts. Sourcing data from the Web allows to get the most current data by anyone, at any time, and for diverse short-term planning horizons and geographical locations.

The goal of this paper it to present and to overcome challenges which arise when data useful for existing electricity demand prediction models is collected from the Web.

2. STATE OF THE ART

Weather forecast public data is made available on the Web in a variety of formats, using one of two mechanisms:

- Weather API - conceived for machine-processing by Web mashups, provide data mainly in XML format. There is no standard of data representation and also different communications methods are used to request the data, resulting in unique APIs used on different sites [11].
- Human-ready HTML presentations – very popular on the Web for weather forecasts, including also the sites providing Weather API. HTML is oriented on human presentation and thus data embedded therein is difficult for other machine processing. Web data is mixed together with formatting code. Actual HTML files downloaded from the web contain many formatting inconsistencies, which require special handling. Weather forecast data is often intertwined with irrelevant data, e.g. advertisements. As the result of site software upgrades or reconfigurations, unexpected and unannounced changes to the presentation format may occur. While the changes often go unnoticed by humans, they mostly require

changes to the machine processing of the weather forecasts.

Weather APIs, while promising, are still too immature due to the lack of standard. Therefore accessing data in the human-readable form is further investigated in this paper, as providing broader range of data sources to choose from.

Existing HTML data extraction techniques are based on wrappers and can be divided into two groups. Position-based solutions, currently employing DOM and XPath to describe the location (coordinates) of information in the document structure. Position-based solution provide very high accuracy when wrappers are generated by humans, but separate wrappers must be provided for different document structures used by Web sites considered as sources [12]. Ontology-based solutions, which offer the possibility to extract data from previously unseen web pages by employing domain knowledge and data labeling (if available). These methods allow to harvest data from many sources without growing human labor, but their accuracy is lower [13].

Due to the need of correct input data for predictions, this paper is pursuing the position-based approach. The arising issues of different HTML document structures at different Web sites and the problem of document format changes are addressed in the following section.

3. PROBLEM FORMULATION

STLF models are algorithms which produce their output based on a predefined set of inputs. Inputs are usually introduced in the form of time series, to represent parameter changes in the past, and require input data to be uniform in:

- **Scope**, i.e. it is required that the same parameters are available for each time instant considered by model.
- **Units**, i.e. t is required that the same parameters are expressed in the same measurement units for every time instant.
- **Time resolution**, i.e. t is required that time series contain parameter values taken at common time instants.

To the contrary to the above requirements, a short review of random, popular meteorological data sources on the Web (see Table 1) reveals that they are not uniform in respect to the above. Additionally, meteorological data sources on the Web differ by the following aspects vital to automatic data gathering:

- Inconsistent data **format**. Data is usually expressed in HTML tables, but these tables are organized very differently by different sources. Additionally sources present only textual data, some present only graphical data (pictures), and some provide meteorological information in mixed form. Thus localizing and perceiving data is intuitive for humans, but hard for automatic processing.
- Inconsistent data **values** for same time and geographical location, resulting from using different weather models and different input data for generating the weather forecasts. Web sources generally do not disclose models nor provenience of data they rely on.
- Inconsistent **time range** of published data, both in regard to history and to future.

4. SOLUTION

To solve all the above identified issues (data inconsistencies in scope, units, time resolution, format, values, and time range, availability), system architecture has been proposed as in Figure 1.

Acquisition Agents (AAs) are configured with Web URL and with the name of the geographic location(s) to collect data for. AAs are the most variable part of the system, as they must be programmed to parse data format provided by Web source to extract all currently available data, map graphical symbols to numerical values, and convert values to use a unified set of units (SI unit system is used further down processing). AAs are called at predefined instants of time by external mechanisms to collect data, preprocess it, and to send it to store. When data source is unavailable, AAs send reserved values (NULL) to the store to indicate lack of data for a specified time instant Thus AA role is to remove the limitations in data format, data units, and to address partially data availability, subject to Meteorological Data Store cooperation.

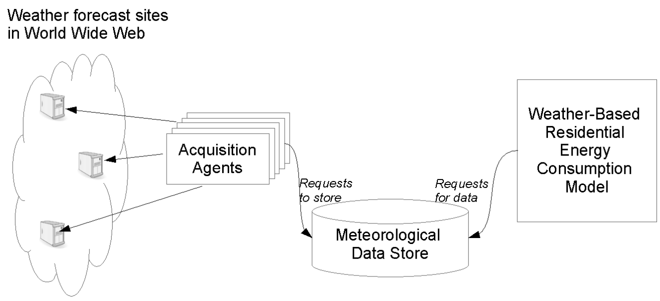

Fig. 1. Software architecture for automated import of Internet weather forecasts

Main task of Meteorological Data Store (MDR) is to accept data from AAs and to store it in database, to make it available for the Weather-Based Residential Energy Consumption Model (WBRECM). For each stored piece of data, also data source and time related to this data is stored. This additional information is checked against data requests from WBRECM to provide parameters from indicated sources and for indicated time range. If the requested range contains NULL values, a WBRECM request for this data is declined.

MDR uses a common time resolution of 1 hour for data from all sources. If a source of data offers lower time resolution of forecast data, MDR interpolates missing values. The simplest form of interpolation may be repeating the last value from this source. MDR is also responsible for removing duplicate data entries, should some agents be called too often, i.e. before their Web sources offered refreshed data. Duplicates are prevented by always accepting the newest data values from AAs for a given time instant, unless these values are NULL.

The role of MDR is therefore to remove the limitations caused by inconsistencies in data time resolution, data scope, and data availability. It also partially addresses data scope and data time range, subject to WBRECM requests.

WBRECM sends requests for data to MDR, indicating data time range, data parameters, and data sources for each of these parameters. Thus the user of the model is able to express his preferences over selection of values from specific data sources. This completes resolution of data time range, and helps to solve the issue of selecting data sources, i.e. of selecting between different data values obtained from different sources.

Tab. 1. Inconsistencies in five aspects of selected free meteorological Web data sources

	Scope [absolute temperature, temperature 'feels like', precipitation, pressure, humidity , wind speed, wind direction, cloudiness, sunrise/sunset]	**Units** [metric/US+ remarks]	**Time resolution** [h]	**Data format** [graphic/textual /mixed]	**Forecast time range** [h]
meteo.pl	temperature, precipitation, pressure, wind speed, wind direction, cloudiness	Metric; cloudiness in octants, precipitation as mm/h	1	graphic	60/72/80
yr.no	temperature, precipitation, wind speed, wind direction, cloudiness, sunrise/sunset	Metric; cloudiness as graphics, precipitation as mm/h	4	mixed	66
mojapogoda.pl (meteogroup.com)	temperature, precipitation, wind speed, wind direction, cloudiness	Metric; cloudiness as graphics, precipitation as mm/h and probability	4	graphic	168
storm247.com	absolute temperature, temperature 'feels like', precipitation, pressure wind speed, wind, humidity, direction, cloudiness, sunrise/sunset]	Metric; precipitation as mm/h, wind direction as graphics	6	mostly textual	216
weather.com	Scope [absolute temperature, temperature 'feels like', precipitation, humidity , wind speed, wind direction, cloudiness, sunrise/sunset]	US, precipitation as mm/h, cloudiness as graphics and in descriptive form, wind direction in compass bearings	1	mixed	8/72/120/240
weather.msn.com	absolute temperature, precipitation, cloudiness	US, precipitation as probability, cloudiness as graphics	1	mixed	144+
twojapogoda.pl	absolute temperature, temperature 'feels like', precipitation, pressure, humidity , wind speed, wind direction, cloudiness	Metric, precipitation as mm/h, cloudiness in percent, wind direction as graphics	1	Mostly textual	384
pogodynka.pl	absolute temperature, precipitation, wind speed, wind direction, cloudiness	Metric, precipitation and cloudiness as graphics, wind direction as graphics	12/24	mixed	144

5. VERIFICATION

Software components in C and in Python have been developed for verification, including weather-based STLF residential energy consumption model (WBRECM). It captures possibly many meteorological parameters, being mentioned as having potential significance in different models referenced before; it is not specific to any geographic region. The model reflects: heating and cooling needs, based on temperature and wind, with a 6h lag,, lightning needs, based on: sun radiation and twilight times for GMT+1 timezone, and weekly electricity usage periodicity.

To prove that the proposed solution overcomes the identified inconsistencies of data scope, units, time resolution, format, values, and time range, two verification cases have been used:

Case 1: two Web data sources (A: pogodynka.pl and B: twojapogoda.pl) for the same geographic location, providing the same data scope and units, but differ in time resolution, data format, data values, and time range.

Case 2: for the same geographic location part of data (temperature) acquired from source A , and part of data (wind speed and Sun radiation) from source B. This was to simulate overcoming a limited scope of data in a single Web data source.

Results are depicted in Figure 2. Data set A and data set B come from two Web sources and differ in data format, units (coding), and time resolution. Despite that, model is able to generate predictions for both. Differences in predictions result from differences in weather forecast data; decision on preferring one source over another one is left to the user.

The combined input set A+B demonstrates that in case of data unavailability or inconsistencies in data scope, modeling is still possible by combining data from selected Web sources. Such forecast will of course differ from forecasts made using only data set A or only data set B. Selecting the Web sources is left to the user to mitigate issues of data quality as required.

6. CONCLUSION

Predicting electricity consumption is important for balancing its generation and for better distribution. Freely available weather forecasts can be used to automatically generate information on future demand, after overcoming data issues specific in such usage scenario.

This paper presents how to automatically and reliably collect and use data from different Web sources. The following data inconsistencies have been identified: scope, units, time resolution, format, values time range and availability. The paper proves that by carefully designing a distributed system architecture, is is possible to fully address all identified data inconsistencies. The most decisive factors are left to the user, as user is involved in selecting data sources for each parameter, and data ranges to match his actual modeling needs.

a)

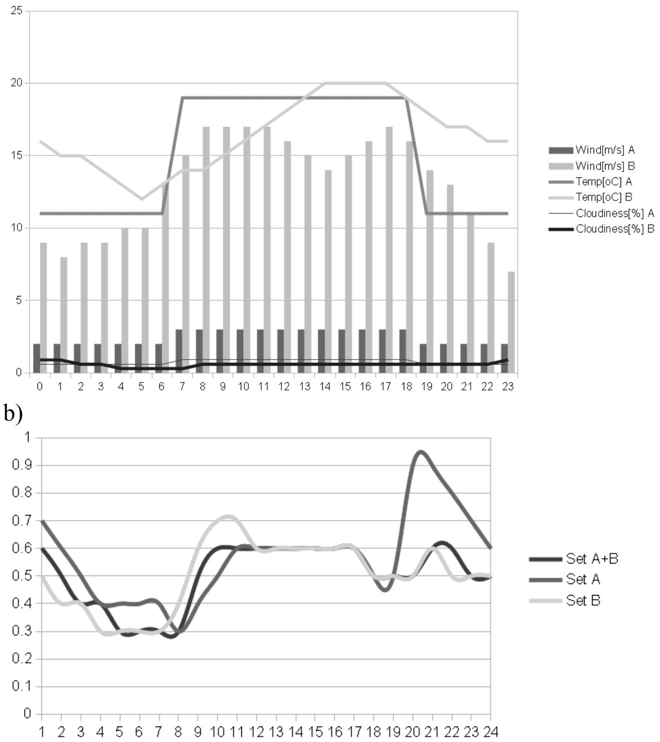

Fig. 2. a) Model input meteorological data from sources A and B, for 24 consecutive hours, on Thursday, in Summer, same geographic location, b) Predicted electricity consumption values for scenario Case 1 (Set A and set B) and for scenario Case 2 (set A+B).

7. REFERENCES

1. Hayes, B., Gruber, J., Prodanovic, M.: Short-Term Load Forecasting at the local level using smart meter data, in Proc of IEEE PowerTech, 2015.
2. Wu, X. ; Wang, X. ; Bi, S., Zhang, P.: Short-term load forecasting based on big data technologies, CSEE Journal of Power and Energy Systems, Vol.1, Issue:3, pp.59-67, 2015.
3. Drury B. Crawley, Which Weather Data Should You Use for Energy Simulations of Commercial Buildings?, ASHRAE Transactions, vol.104 part 2, 1998.
4. Łyp, J., Popławski, T., Starczynowska, E.: The comprehensive analyze of influencing meteorological factors on load variation in the National Power System, Electrotechnicla Review, R.87, No.2, pp.97-100, 2011.
5. Eugenio F. Sanchez-Ubeda, Ana Berzosa: New variables to improve electricity and natural gas consumption forecasting: Dynamic degree-days, Avances in Inteligencia Artificial, Vol.1, 2011.
6. Ziser, C. J., Dong, Z. Y., Saha, T. K.: Investigation of Weather Dependency and Load Diversity on Queensland Electricity Demand, Proc. AU Power Engineering Conf, pp.457-462, 2005.
7. Beccali M., Cellura M., Brano V. L., Marvuglia A.: A neural network model to forecast urban electricity consumption from weather data, World Renewable Energy Congress VIII Book of Abstracts, USA, 2004.
8. Drury B. Crawley, Which Weather Data Should You Use for Energy Simulations of Commercial Buildings?, ASHRAE Transactions, vol.104 part 2, 1998.
9. Włodarczyk, A., Zawada, M.: Modeling the impact of the weather factors on the electrical energy consumption in one of the regions in the southern Poland, Proc. of 7th International Conference on the European Energy Market (EEM), 2010.
10. Maj M.: IMGW Data – not for free, and the price may be prohibitive, Internet Users's Journal, Business and Law, 2012, available online, URL:http://di.com.pl/news/46012,0,Dane_IMGW_-_nie_dostaniemy_ich_darmo_a_ceny_bywaja_zaporowe.html (DOA: 1.08.2015).
11. John Musser, 5 Weather APIs – From WeatherBug to Weather Channel, web blog entry. available online, URL: http://blog.programmableweb.com/2009/04/15/5-weather-apis-from-weatherbug-to-weather-channel/ (DOA: 1.07.2015).
12. Alberto H. F. Laender, Berthier A. RibeiroNeto, Altigran S. da Silva, Juliana S. Teixeira, A Brief Survey of Web Data Extraction Tools, ACM SIGMOD, Vol.31 Issue 2, pp.84-93, 2002.
13. Ferrara E., De Meo P., Fiumara G., Baumgartner R.: Web Data Extraction, Applications and Techniques: A Survey, available online, URL: http://www.emilio.ferrara.name/ wpcontent/uploads/2011/07/survey-csur.pdf, (DOA: 1.08.2015).

PRZEWIDYWANIE ZAPOTRZEBOWANIA NA ENERGIĘ ELEKTRYCZNĄ NA PODSTAWIE INTERNETOWYCH PROGNOZ POGODY

Energia elektryczna jest istotna w produkcji i transporcie, zaś w pewnych gałęziach gopodarki również jako dobro samo w sobie, wymagające planowanie i zarządzania generacją oraz dystrybucją, w celu dostarczenia do klienta w odpowiednim czasie i ilości. Przyjmuje się, że informacje pogodowe niosą dane na temat zapotrzebowania energetycznego konsumentów. Darmowe źródła pogodowe w Internecie oferują dane niespójne, które nie pozwalają na automatyzację procesu prognozowania zapotrzebowania. Artykuł wskazuje te niespójności, oraz wskazuje sposoby ich zniwelowania, w celu zapewnienia automatycznego zbierania danych oraz generowania prognoz zapotrzebowania.

Słowa kluczowe: planowanie zapotrzebowania, energia elektryczna, prognozy pogody.