

Improved Short Term Energy Load Forecasting Using Web-Based Social Networks

Mehmed Kantardzic¹, Haris Gavranovic², Nedim Gavranovic³, Izudin Dzafic², Hanqing Hu¹

¹J. B. Speed School of Engineering, University of Louisville, Louisville, KY, USA

²Faculty of Engineering and Sciences, International University of Sarajevo, Sarajevo, Bosnia and Herzegovina

³College of Economics, American University in Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina

Email: mehmed.kantardzic@louisville.edu

Received 14 October 2015; accepted 27 October 2015; published 30 October 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this article, we are initiating the hypothesis that improvements in short term energy load forecasting may rely on inclusion of data from new information sources generated outside the power grid and weather related systems. Other relevant domains of data include scheduled activities on a grid, large events and conventions in the area, equipment duty cycle schedule, data from call centers, real-time traffic, Facebook, Twitter, and other social networks feeds, and variety of city or region websites. All these distributed data sources pose information collection, integration and analysis challenges. Our approach is concentrated on complex non-cyclic events detection where detected events have a human crowd magnitude that is influencing power requirements. The proposed methodology deals with computation, transformation, modeling, and patterns detection over large volumes of partially ordered, internet based streaming multimedia signals or text messages. We are claiming that traditional approaches can be complemented and enhanced by new streaming data inclusion and analyses, where complex event detection combined with Web-based technologies improves short term load forecasting. Some preliminary experimental results, using Gowalla social network dataset, confirmed our hypothesis as a proof-of-concept, and they paved the way for further improvements by giving new dimensions of short term load forecasting process in a smart grid.

Keywords

Short Term Energy Load Forecasting, Smart Grid, Social Networks, Event Detection

1. Introduction

A smart grid is an advanced electricity transmission and distribution network that utilizes information, communication and control technologies to improve economy, efficiency, and security of the grid [1] [2]. We were witnesses of tremendous development and improvement of electrical distribution systems during the last decade. The governments as well as electricity distribution companies are keen to improve the automatic control and to provide the seamless integration of the various new elements in the smart grid. Distributed generation of electricity is replacing the traditional centralized electricity generation [3]. This trend is enhanced in particular by ever increasing ecological awareness in the modern society and recent technological advancement in the electricity generation systems. The internal combustion engines, gas turbines, micro-turbines, photovoltaic, fuel cells, and wind power turbines can be found all together in a relatively small geographical region (such as micro-grid on **Figure 1**) to be used as a complement to the classical electrical energy supply system.

In the **Figure 1**, one can see the graphical presentation of the complexity of the hypothetical islanded smart grid. On the left side of **Figure 1**, different consumers are presented together with the smart meters and distributed intelligence (DI) needed for the collection of the appropriate data from the customers. In the middle part, distributed generation is presented together with the low voltage transformer substations and the customers like electric vehicles who are not fixed to the single position or to the one single area. On the right side, Control Microgrid 1 is presented where the data for that particular micro-grid are processed for optimal match of the distributed generation and consumption in the islanded smart micro-grid. Control Microgrid 2 is needed as well as a control of more than one islanded micro-grids connected into the smart grid.

We can talk about the concept of smart grid only if the management of the grid is supported by advanced automated solutions using information technology supported by applied math. Data generation in smart grids has grown exponentially, and new challenges in organizing and analyzing these big data are to be faced to efficiently optimize smart grids. These data include production levels, loads, price information and others, and they are collected at regular short time intervals (usually up to half an hour intervals). Traditional energy forecasting systems dispose with other types of data which are influencing consumption or production of electrical energy. Such data include outdoor temperature, humidity, social events, geographical differences, demographic information, traffic intensity and day of the week. Also, it is clear, based on previous analyses, that the load demands have multiple patterns including seasonal, daily and weekly periodicity. Besides that, the load level during the

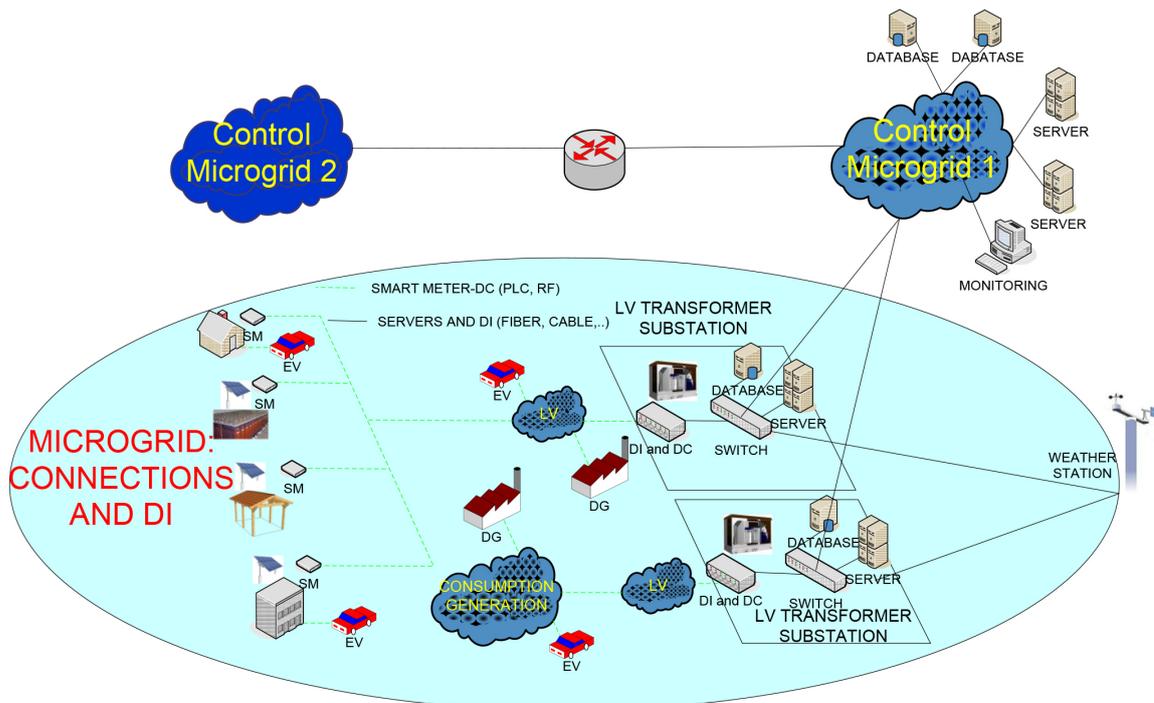


Figure 1. Islanded smart micro-grid example [3].

weekends and holidays is lower than that of the working days, which leads to conclusion that the load demands are highly influenced by the calendar days [4].

In general, forecast relies on historical consumption data to determine how much power a consumer (or group of customers) may require. Forecasting on a level of smaller geo-locations is obviously more difficult due to variability of human behavior. Because of a complexity and the amount of data, forecasting problem may be classified as big data problem requiring new approaches to computing [5]. The techniques from different fields, such as artificial intelligence, high performance computing, data mining, simulation and stochastic modeling, are employed for the development of forecasting subsystems on a smart grid.

Accurate load forecasting enables a utility provider to plan the resources and control actions balancing the supply and the demand of electricity. However, short term load forecasting (STLF) is a difficult task because energy consumption is influenced by many factors such as weather conditions, daily, weekly or seasonal cyclic characteristics, special events, economy status, patterns of behavior for different consumer types, and habits of individual customers. Current short term forecasting models for smart grid primarily combine factors such as power usage data, weather forecast, time-related cyclic patterns, and some demographics data. A variety of modeling techniques have been applied with different levels of success including autoregressive integrated moving average (ARIMA), fuzzy-logic and neuro-fuzzy methods, artificial neural networks such as radial basis function neural network (RBFN), and recently support vector regression methodology (SVR).

One-size-fits-all models are gone for the utility forecaster. To improve forecast performance, reduce uncertainties and generate value in the new data-intensive environment, forecasters must be able to decide which specific factors affect the load, and what are their characteristics in time and space? For example, spatiotemporal load profiles are analyzed in Finland, where characteristics for different regions in the city of Helsinki are obtained, and one illustrative example is given in the [Figure 2](#) [6]. Graphs show that power demand peaks in the city center are at noon or in the early afternoon, whereas in the outer suburbs of the city with high share of residential buildings, the peak occurs around 7-8 pm.

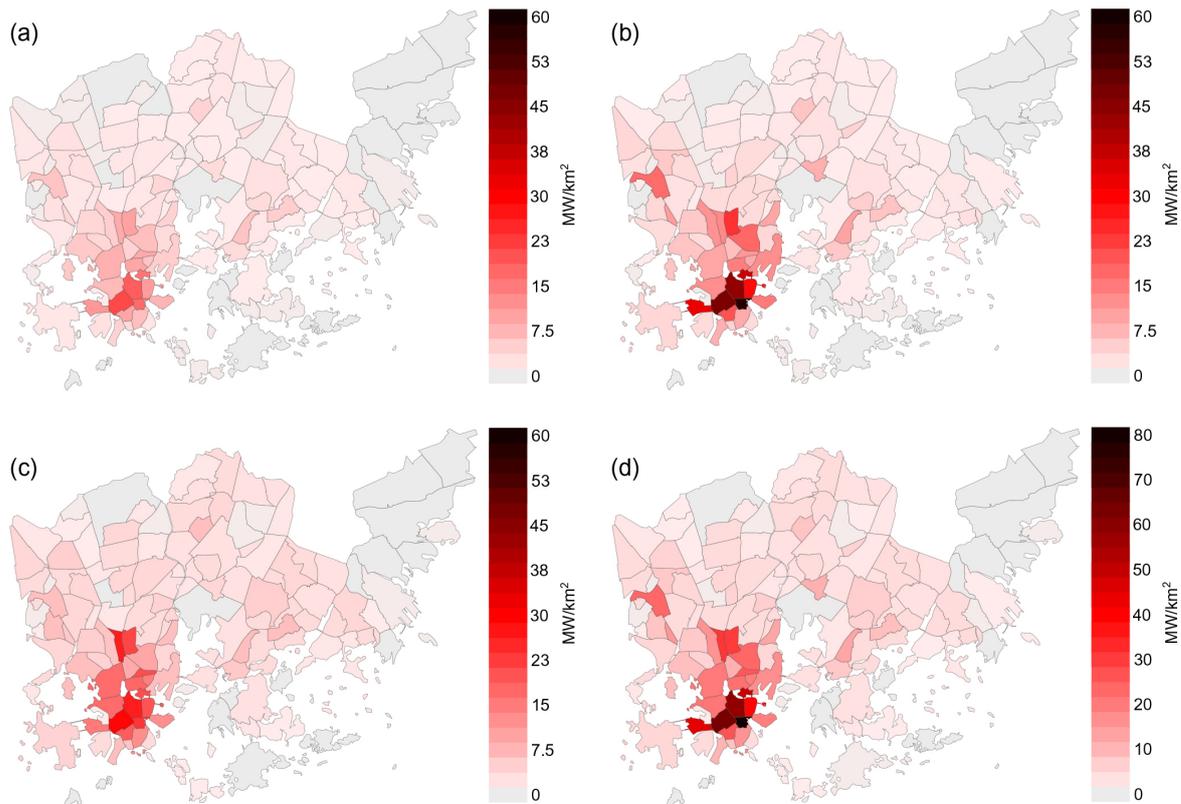


Figure 2. Spatiotemporal load for Helsinki in early February: (a) 2-3 am; (b) 11-12 am; (c) 7-8 pm; (d) Peak load for the day [6].

In this article, we present a new approach for predicting spatiotemporal, short-term power demands in a form of function $P(x, y, t)$, where (x, y) are coordinates of the micro-grid or a smaller location while t is a discrete time [6]. The power demand P is affected by a large set of:

- “traditional” factors, such as type of consumer, human behavior, cultural factors, time of year and weather conditions, but also
- “non-traditional” factors, such as social and sport events in the region, outage logs, etc.

Social networks on the Web and the other Web-based apps may be a primary source of data related to “non-traditional” factors. They could significantly improve short term load forecasting especially in the situation where the emphasis is on peak load detection.

2. Related Work

Traditional approaches in STLF are based on so-called “the similar day method”, where the task is to find a day in the history that is similar to the forecasted day. The similarity is usually based on day of the week, season of a year, and the weather patterns. While the similar day method is one of the earliest methods being applied to load forecasting, even today, many system operators are still having the load and temperature profiles of the representative days hanging on the wall of the operations room [7]. Modern similar day method is often implemented using some clustering techniques. Instead of one similar day, the algorithms may identify several similar days or similar segments of a day, and then combine them to obtain the forecasted load profile [7] [8].

New load forecasting models in the smart grid era take advantages of modern automatic metering infrastructure, information technology, and advancements in atmosphere science. Most of load forecasts were made using one of three categories of methods: trending, simulation, or hybrid approaches [9]. Trending methods use a mathematical function to fit the past load growth and then extrapolate to the future load. Many research studies assume the electricity loads as the time series. Even the naïve, direct approaches, analyzing trends, cycles and errors of the time series, are considered as a part of a trending class [7]. More complex methods use the Artificial Neural Networks [3] [10] [11], and specifically developed statistical methods [12] [13]. Recently, in [14], the authors are using statistical learning theory, and they develop the prediction method based on the support vector regression. More complex methods show better performances, but at the same time they are often more time-consuming. Today, the most common approaches include multiple regression, ARIMA technique, support vector machines, fuzzy logic systems, or artificial neural networks, in which the load history is made a function of trend terms [7] [10] [11]. The advantages of machine learning based methods include ease of use, simplicity, and a short-range response to recent load-growth trends. But, many utilities are simply not comfortable or not allowed to use these black-box models for forecasts in everyday practice. Despite how fancy the models are, they are not useful for businesses because the explanations are weak and often not available [9].

New extended data sources support improved load forecasting models using machine learning techniques. These sources generate data such as: hourly/daily load history at end user level, hourly weather history at weather stations, demographic and economy information, industry code mapping, outage logs, and energy system loss information [5] [15]. Additional attributes may include the effects of varying sunrise and sunset times which determine when domestic and street lighting are used. Weather data may be extended with cloud cover, humidity, visibility, and precipitation [16]. Modeling repetitive patterns and some aspects of seasonality, includes techniques which determine how loads during different holidays and their surrounding days are being altered to better reflect the holiday load profiles. For example, in New England area Independence Day may be modeled as typical Sunday, while Thanksgiving Day is most similar to Saturday [9]. A comprehensive short term load model should make a distinction between three classes of electricity customers: residential, commercial, and industrial customers [5] [8] [9]. Residential customers have the most weather responsive electricity consumption behavior. Although the load of an individual residential customer can be quite stochastic (**Figure 3**), residential load aggregated in a micro-grid is more predictable than the loads of other customer classes.

Load forecasting for small commercial customers is similar to residential load forecasting, because small commercial customers usually have close response to weather. In addition to weather, most large commercial loads are significantly affected by the business schedules. Also, some of them have strong seasonal patterns. For instance, the load of hotels in a vacation area is mainly affected by tourism demand during holiday seasons, while the load of hotels in central business districts responds to the local economy and major conferences [9]. Additional illustrative example is the load of education institutions which also has its own characteristics and

patterns. For instance, many students stay up late studying. Therefore, daily peaks of student dorms often occur close to midnight. At annual level, education institutions follow academic calendars. The load level is recognizable low during academic holidays.

Most of the previous methodologies assume that the data, directly originated from the smart grid, is complement with the outdoor temperature and sometimes with the data about the humidity, or type of the day, and they are used together for load forecasting. For example, in [12] there is a use of humidity, sunrise and sunset times beside the temperature. In [14] the authors used the temperature and humidity, while Figure 4 clearly shows the correlation and interdependence between the electrical load and the outside temperature over the period of 300 days. The authors in [13] use price-incentive signals together with weather projections for STLF (in the presence of wind turbine which efficiency can be forecast for few hours to few days). Lately, the promising results are achieved with a kind of disaggregation of the load forecasts among costumers, and their later recombination at several levels: forecast the integrated loads for a day and thereafter predict hourly consumption [11]. All data necessary for STLF are not collected only by energy industry. Similar to other utilities, STLF system can “borrow” data from external sources to perform the analyses. Typical sources include national survey conducted by

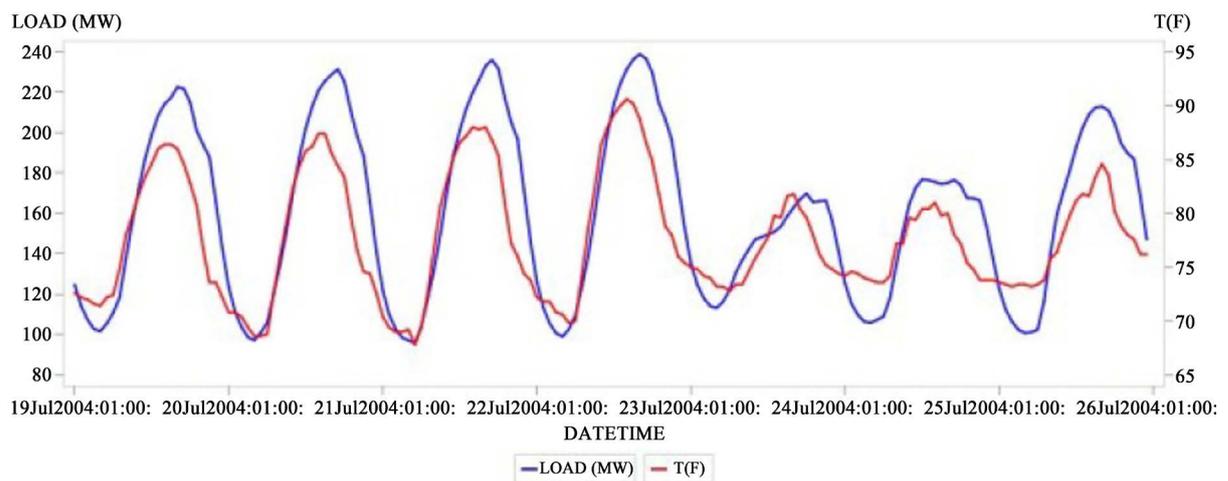


Figure 3. Diverse daily load profiles of residential customers [8].

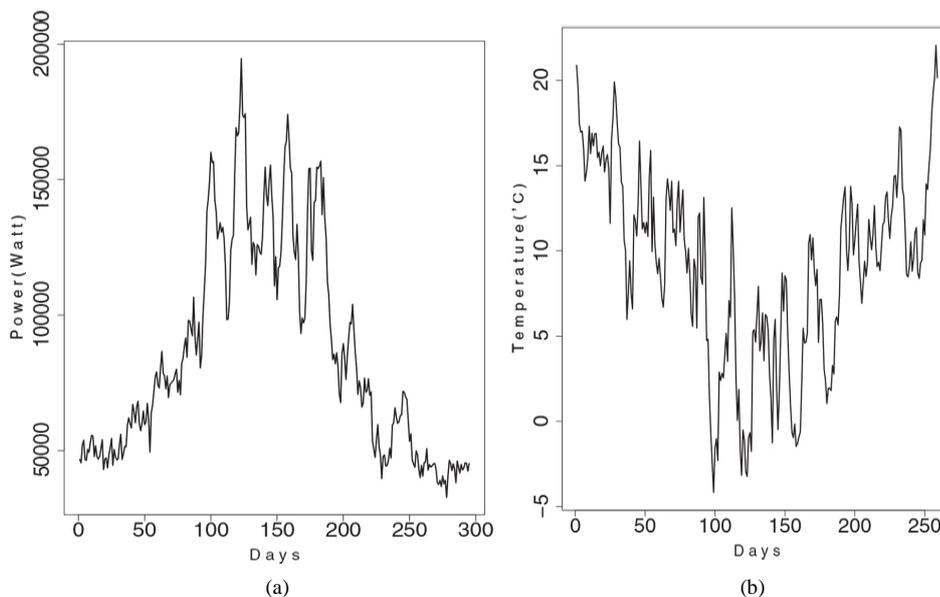


Figure 4. Correlation between load and temperature is very high: 0.83 [7]. (a) Load data during 10 month period; (b) Temperature during the same 10 month period.

the government, case studies performed by consulting firms, public reports from other utilities, and academic publications [8]. Many other outside factors such as price of energy, social and other events, in the stronger or weaker manner influence the electrical consumption, and can be used for the prediction of electrical loads.

The impact of human activities to energy consumption can be realized in several aspects. In the hourly resolution, the impact varies over the calendar variables including day of the week, and/or month of the year. In the long-term monthly or quarterly resolution, the impact varies on different economic conditions. For instance, during the year of 2009, which is the early part of a recession in US, the energy consumption of US is lower than that of 2008, because people were using power more conservatively, and lots of businesses were closed [17].

A STLF model very often includes the loads of some preceding hours as part of the inputs, which captures the autocorrelation of the current hour load and the preceding hour loads. But the problem is that there are some outliers/peaks in the load, which are not detectable by standard approaches [17]. Outliers' removal by averaging peaks is not improving short term forecasting. On the contrary, losing the information about short term local peaks, the predictive system is not performing one of its main STLF tasks; peak control. Our research and the proposed methodology make emphasis on a detection and successful forecasting of these short term, irregular peaks.

3. Short Term Load Forecasting Using Web-Based Social Network and Event Schedule Information

Recent activities and publications show that a great amount of effort is invested in the research of load forecasting [4] [8] [10] [15] [18]. Main results include some applications targeted the load forecast for big, national electric energy network where the aggregation of the load is based on over millions of users. The appearance of the smart grid, where the islanded mode of operation is favored, changed completely the perspective, and directions of research and applications [10]. The interest is shifted towards the sub-grid network of at most several thousand consumers (industrial, commercial, residential, apartment buildings, sport venues etc.), where different types of aggregation came to the stage. **Figure 1** presents one such islanded smart micro-grid where the production, the consumption, and the control and management are localized to the narrow geographical area. Only a fraction of consumed energy is supplied from outside a micro-grid. Reduction in space, is also followed by reduction in time. Long-term forecasting methodologies are followed with more medium-term, and recently even short-term load forecasting approaches (STLF—for the length of period going maximum up to one day).

This research is primary interested in the STLF methodology for the islanded smart grid with up to several thousands of users. Short term load forecasting (STLF) in that framework has several characteristics we have to take care when we are developing a new modeling methodology [15] [18]:

- Large volume of data, because many regions today have been storing hourly load for years.
- Requirements for high quality of data due to the maturity of the metering technology. Most outliers are physically explainable, such as system outages, sudden demand response activities, etc.
- Multiple patterns of seasonality, such as: hours in a day, days of week, and months of a year.
- Load is highly dependent on several input variables, including high correlation to the local temperature due to space heating and cooling needs.
- High accuracy requirements are established for STLF system applicability. For example, it is expected that MAPE (Mean Absolute Percentage Error) for hourly forecast in one day is <5%.
- Societal necessity support—improvements in STLF quality can help system reliability and energy efficiency, which means less outages, reduction of cost, and in general greener environment.

The data-rich smart grids offer new opportunities to optimize their management more efficiently. There is a huge set of new, innovative input variables influencing the smart grid model, sometimes in a very complex way. The importance of better short term load forecasting is recognized, and it should include new real time data sources supported by new methodologies, and sometimes even totally new approaches. Regardless of the method applied, one important issue of STLF is the selection of input features from a large pool of candidates. Many input features, such as historical loads with different time lags, meteorological factors, and calendar information, have been widely examined in the load forecasting literature [4]. Comprehensive analyses show that some of them might be redundant or even irrelevant to a specific STLF problem. Therefore, an effective and efficient feature selection approach that is able to identify the best predictive subset of features by eliminating

noisy, irrelevant, and redundant features without degrading the performance of the model is highly needed.

The days of one-size-fits-all models are gone. To improve forecast performance, reduce uncertainties and generate value in the new data-intensive environment, forecasters must be able to decide which model, or combinations of models, is best. In addition, they must be able to determine relevant indicators that affect load [4]. Many energy forecasting papers reported amazingly low errors, but failed miserably in practice [9]. To further advance our predictive system, we have to take an interdisciplinary approach by involving various communities, such as statistical forecasting, artificial intelligence, meteorological science, and power engineering, but also social science which is becoming more important with the development of internet-based social networks.

We are proposing an integrated forecasting framework with the concentration on the short term load forecasting (STLF) engine that can easily be linked and extended to various other forecasts. One of the most important goals of short term energy demand forecasting is to accurately predict the occurrence of peak loads—a situation where the demand for power goes beyond the current power generation capacity. Traditionally energy demand has been often cyclical, with load patterns observed across a 24 hour period for different types of users (commercial, industrial, and residential), and easy recognizable seasonal patterns seen across a calendar year. The current STLF models for Smart Grid primarily combine power usage, weather forecast and sometimes demographics data where numerous sensors in smart appliances and smart meters are under deployment to monitor continuously power use activities [1]. Improvements to power information systems rely on data from existing and new information sources generated outside the power grid from other relevant domains—weather, real-time traffic, schedule of large events and conventions, equipment duty cycle schedule, data from call centers, Facebook and Twitter feeds, city websites, other social networks, and so on [19]. All these distributed data sources pose information integration and analysis challenges. The problem space is affected by the dynamic nature of information present within this environment. For example, as people migrate within the city because of some important events, prior load forecast models will need to adapt. We are exploring two important aspects related Smart Grid data:

- 1) streaming data collection, processing and analysis, and
- 2) complex event detection,

both applicable in the field of short-term energy demand forecasting. This means that both, the algorithms used for data mining, as well as the information sources used, will need to change and adapt over time. The ability of the system to continuously learn and rapidly incorporate new information sources and predictors will be essential. Predictive models should be responsible for handling different types of irregularities in the input data. The problem is that some input data which are source of irregularities are not used to build the model. Usually, in current models, these effects are interpreted as outliers or random peaks, which are removed to “improve the model” [5]. Therefore, we have to revisit the issue of outlier removal which are correct measurements, with useful information that we wrongly excluded [16].

The technologies that will enable these tasks assume scalable data mining and complex event processing systems that accept continuously data from distributed sources and detect emergency situations based on defined policies. The system should be smart enough to filter out noise from distributed multimedia data and present only information and patterns that are relevant to the current energy consumption. A variety of distributed streaming data, collected from numerous of traditional or non-traditional sources, should be included in the context of energy demands analysis and prediction [1]:

- Real Time Consumption—data collected from smart meters and other currently available Smart Grid measurement equipment in the area,
- Infrastructure Information—information about the power grid network interconnectivity, together with building structures, building orientation, and equipment installation,
- Natural Conditions—weather and seasonal changes at different spatial and temporal scale,
- Customer Behavior—collected data that helps in understanding customer power usage patterns including analysis of customer demographics, social networks roles, adopters of new tools, person’s action influence, etc.,
- Schedule Information—scheduling information provides knowledge about a future occurrence of differ events ahead of time. This information enables estimation of the demand at the particular region based on the type of event scheduled, as well as on the number of people (the size of a crowd!) expected to turn up for the event. Schedule information about individual people as well as facilities may be useful, but they require much more complex information infrastructure to collect and process all data [1] [2].

Complex event detection deals with computation, transformation and pattern detection in time and space over large volumes of partially ordered streaming multimedia signals, structured records, or text messages [20]. This approach has been used successfully in financial services industry to detect stock trading patterns or as a support for supply chain management systems. Several vendors such as Oracle, Microsoft have developed integrated software products to support event detection [20] [21]. The main idea is that traditional approaches can be complemented by data driven dynamic algorithms that integrate multiple streaming data, locate patterns among a large class of real-time information, and predict usage trends and peak occurrences. We are proposing a novel dynamic prediction technique which should be implemented as additional modules in the integrated short term forecasting system. This extension will enable a finer control of power use. For example, announcing and scheduling of a massive convention in the area, blog postings about the concert of a famous entertainer in a local concert hall, or referencing a public holiday, all can affect power usage in a city, a region, or a micro-grid. Prediction algorithms should recognize and locate patterns using a large class of historical and real-time information, and adjust dynamically models used to predict power usage [22]. Complex event detection combined with Web-based technologies is a promising solution to dynamic demand forecasting and our proposal for modified architecture of a forecasting system from [2] is presented in Figure 5.

Figure 5 summarizes and schematically presents proposed approach where the social network data and event schedule information are used to adjust and refine the forecast for the peak loads. The main novelty in the proposed research is the proof of the ability to improve STLF system by integrating text-based information available on the Web with a streaming real-time data about energy consumption and weather condition. Using machine learning techniques in the modeling phase we expect to prove the advantages in use of dynamic, Web-based, multimedia data for modeling in load forecasting.

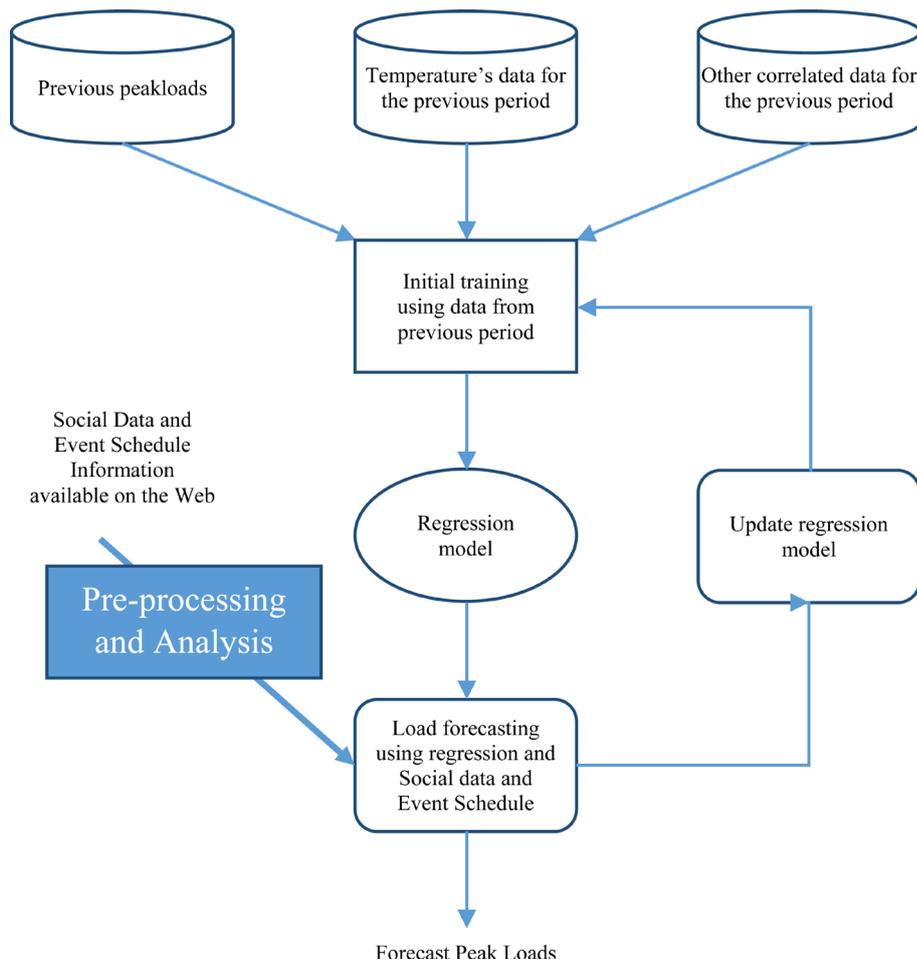


Figure 5. Overview of the proposed load forecasting scheme (modified from [2]).

Highlighting and better understanding the process of peak load detection may significantly improve prediction results. Let's illustrate the problems with a simple illustrative example. Standard average value as a smoothed replacement for peak occurred in **Figure 6** part F1, may not be adequate measure for assessing a forecast in time. We have to take care about regular peaks, which are relatively easy detectible, but also about irregular behavior which is most difficult to detect. The hypothetical forecasts, illustrated with dashed lines, consist of a flat forecast (F1) (corresponding to the average usage) and a single peak forecast centered around three different times (F2-F4), with the correct background usage. In the context of using the forecasts to reduce the peak demand, F2 is a very good forecast, F3 is reasonable, and both F1 and F4 are poor.

Any successful forecast method requires a degree of flexibility in the spatial/temporal positioning of the peaks [23]. Selection of input features plays an important role in developing models for short-term load forecasting (STLF) which includes irregular peak detection. Since electric power loads often exhibit nonlinear and non-stationary dynamics over time, various factors such as climate factors, social activities and seasonal factors should be explored for accurate electricity load forecasting [24].

4. Load-Related Events Detection Using Social Networks Data and Other Web-Based Resources

This section is presenting a proof-of-concept that Web-based resources may give enough spatio-temporal information about human crowd in specific locations. While this crowd may cause load peaks in energy consumption, the available information is important step in improvements of short term load prediction system. To experimentally validate the proposed approach in irregular peak detection using social networks and other Web-based tools, we used the following resources:

- 1) Gowalla Social Network,
- 2) Google Map, and
- 3) Louisville City Events Website.

Gowalla is a location based social network website. Users check-in at locations, and share their checked-in locations within the network. Gowalla records participating locations, and assigns a location ID and a static GPS coordinate to each location ID. For any users checking in at the same location, the GPS coordinates of that check-in instance are always the same. The dataset is separated into user friendship network data and location data. The user network data contains pairs of user ID with a total of 950,327 pairs. Each pair means the two users are "friend" within the Gowalla social network. The location data contains all location check-in data with 6,442,892 instances, and **Table 1** shows a sample data. The Gowalla check-in dataset is obtained from Stanford Network Analysis Project (SNAP) website (<http://snap.stanford.edu/data/loc-gowalla.html>).

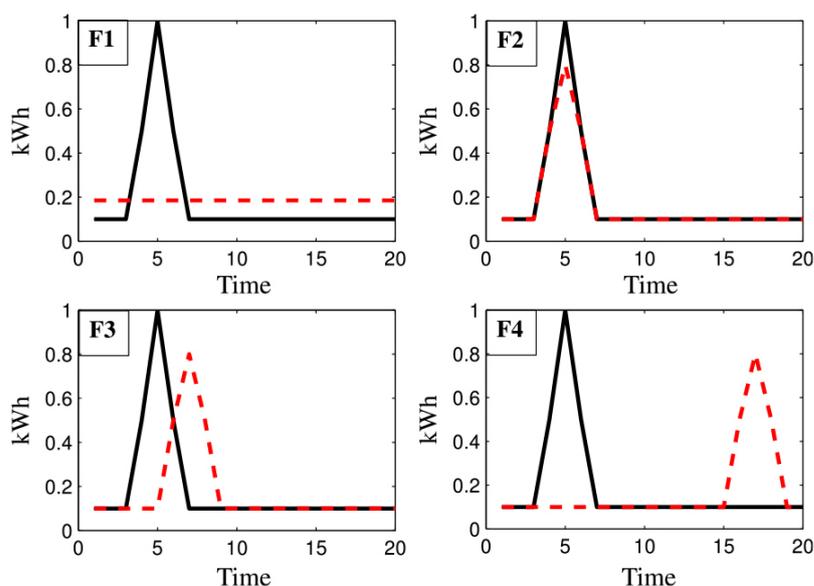


Figure 6. Four “forecasts”, F1, F2, F3 and F4 (dashed lines), together with the actual data (solid lines) [23].

We focus on the location check-in data in the crowd-based event detection experiment. An event is defined as a set of check-ins of users at the same location with the number above a specified threshold value, and within a short time span. To detect a possible event in Gowalla database, our approach looks for check-ins at a given location within a 15 minutes timespan. We applied a grid approach with two dimensions: longitude and latitude, to count the number of check-ins at each location. After analyzing the entire dataset, we count the number of check-ins in each grid and look for grid with above a threshold value density. In our experiment we found several locations that contains more than 9 check-ins in less than 15 minutes. **Table 2** illustrates some of the top check-in locations.

We cross-reference the longitude and latitude for each location with Google Map available information, and we are able to confirm the character of human gathering on these locations. The result is shown in **Table 3**.

From **Table 3** we are able to detect multiple types of social crowd: restaurant gathering, in Austin, TX, school activities and church event in Lille, France, and finally late hour's crowd in one of Austin's hotels. As shown in **Table 3** there are different types of event occurring around the world. We are also interested in looking for time dynamics of these check-in event patterns. Temporal distributions of check-ins for all four events from **Table 3**, with the one minute grid interval, is visualized on **Figure 7**.

Presented patterns in time domain are very simple, and they are only proof-of-concept. In the cases of large human crowd, with longer term duration of an event, we may use these temporal diagrams for better prediction of event dynamics through the crowd size. When, using real time data, the beginning part of the diagram is recognized as initiation of an event, the ending part of a pattern will give the best estimation of the trend in the crowd and duration of an event. These real world large events, with an influence on load peaks, we didn't detect in our relatively small Gowalla dataset. While the other more popular social networks may enable this detection and further analyses, the main advantage of Gowalla social network is free availability of social network data.

Second source of Web data are specific Web sites which are announcing scheduled large events in the area. Extracting the data from these sites, and including this information in our predictive system, may also significantly improve short term load predictive model, and in general may influence on local energy consumption. Almost every larger city has this kind of Web site, where the data on events, if available, are already preprocessed and prepared for inclusion in the proposed STLF system. As an illustrative example, we visited the Louisville City Events Web site, extracted events in the spring 2015, and segment of preprocessed results in a form of table is presented in **Figure 8**. Efficient implementation of social network event data collection and

Table 1. A sample of Gowalla location check-in data. Source: (SNAP) website. (<http://snap.stanford.edu/data/loc-gowalla.html>)

User ID	Check-in time	Latitude	Longitude	Location ID
196514	2010-07-24T13:45:06Z	53.3648119	-2.272346583	145064
196514	2010-07-24T13:44:58Z	53.36051123	-2.276369017	1275991

Table 2. Selected locations with high density of check-in. Source: (SNAP) website. (<http://snap.stanford.edu/data/loc-gowalla.html>)

Time	Location ID	Longitude	Latitude	# of check-ins
3/15/2009 18:07	9250	30.26713388	-97.74451911	19
10/15/2009 10:11	27090	50.61957326	3.048555851	28
10/15/2009 13:50	27426	50.63573561	3.067001402	12
3/14/2010 3:08	9222	30.25801062	-97.73973942	9

Table 3. Google map location of event. Maps source: <https://www.google.ba/maps/>.

Time	Location ID	Name	Type	City	Country
3/15/2009 15:00	9250	tenOak: Bourbon House and Lounge	Restaurant	Austin	US
10/15/2009 10:00	27,090	Lycée International Montebello	School	Lille	France
10/15/2009 13:45	27,426	Paroisse Saint Maurice	Church	Lille	France
3/16/2009 22:15	9222	The Driskill	Hotel	Austin	US



Figure 7. Check-in pattern in time for locations from Table 3.

Louisville Upcoming Events	
Thu, Jul 30, 2015	7:00 PM
Cracker	Cracker is an American alternative rock band
Headliners Music Hall	
Thu, Aug 6, 2015	12:00 AM
Down	Down is an American heavy metal super-group
The Mercury Ballroom	
Mon, Aug 3, 2015	7:00 PM
Gavin DeGraw	Gavin Shane DeGraw is an American musician
KFC Yum! Center	
Fri, March 24, 2016	7:00 PM
NCAA Division I Men's Basketball Regional Championships	
KFC Yum! Center	
....	

Figure 8. Louisville city events (Source: <https://louisvilleky.gov/events>).

analysis with data from scheduled events Web sites, and integration with other modules of forecasting system, should represent an important direction of smart micro-grid short term load forecasting.

The results presented in this Section clearly show that complex events can often be detected using social networks. The first technique is using the activities recorded on social networks, analysis of these big data and partly or completely automatic recognition of the complex events. The second technique is more naïve and relies on the data provided for announcing scheduled large events in the area. Both techniques have their importance because the first one detects the complex events when they are already in progress and the second one detects

them in advance. Proper approach is the combination of both complex events detection techniques.

Once these techniques are available for detecting complex events, short term energy load forecasting can be improved because complex events are usually connected with a specific energy load for micro-grid. We know that events where many people gather together on the small area are related to much higher energy consumption at the venue of the event and up to certain point lower energy consumption at home or business premises for example.

We believe to be the first one to investigate the idea of complex event detection application into the problem of STLF for micro-grid. Our approach improves STLF because one important information can be included in the context of energy demand analysis and prediction. The results in this Section are the starting point for improved STLF model. Such model is improvement of the current models because it will include an information with very specific features. Improved STLF model for micro-grid control will handle energy consumption peaks on the small area. This control is important for STLF because the energy supply disruption are to be expected as a results of unusual events or behavior of many individuals in the area. Complex events that we can detect and recognize are in this group of events and incorporating the technique of complex events detections will provide important information to be used for improvements of the STLF models.

5. Conclusions

We initiated our research with the hypothesis that improvements in short term load forecasting may rely on inclusion of data from new information sources generated outside the power grid and weather related systems. Relevant domains of data may include scheduled activities on a smart grid, large events and conventions in the area, equipment duty cycle schedule, data from call centers, real-time traffic, social networks feeds, and variety of city or region websites. All these distributed data sources pose information collection, integration, and analysis challenges. Our approach is concentrated on complex non-cyclic events detection, where detected events have a human crowd magnitude that is influencing power requirements in the area. We are claiming that traditional approaches can be complemented and enhanced by new streaming data inclusion and analyses, where complex event detection is enhanced by Web-based technologies, and that may improve short term load forecasting. Preliminary experimental results, using Gowalla social network, Google maps, and Louisville city event website, show directions how Web-based resources may be used as a part of enhanced forecasting system.

We are planning to study, design, develop, and experimentally test the proposed approach, summarized in **Figure 5**, within two complementary real-world frameworks:

- Considering the entire islanded smart micro-grid with up to several thousand users and taking perspective of aggregated load.
- Classify all nodes in the smart grid with respect to their nature (industrial, commercial, residential, etc.), and build the enhanced model using Web data for the local short term prediction on each node, or on each class of nodes.

Our final goal is to optimize energy production such that consumption is better predicted because complex events are recognized in advance and may be followed in time. It was possible to evaluate possible cost reduction under the assumption that complex events were recognized compared to the environment where complex events were not detected in advance. Results presented in this article are just initial ideas and some preliminary experiments performed as a proof-of-concept. If this approach is successfully implemented and achieved expected goals, it will also pave the way for further improvements of smart micro-grids.

Acknowledgements

The author Haris Gavranovic would like to acknowledge the contribution of the COST Action TD1207.

References

- [1] Zhou, O., *et al.* (2012) Semantic Information Modeling for Emerging Applications in Smart Grid. *Proceedings of the 2012 Ninth International Conference on Information Technology—New Generations, ITNG'12, IEEE Computer Society*, Washington DC, 775-782. <http://dx.doi.org/10.1109/ITNG.2012.150>
- [2] Aung, Z., Toukhy, M., Williams, J.R., Sanchez, A. and Herrero, S. (2012) Towards Accurate Electricity Load Forecasting in Smart Grids. *The Fourth International Conference on Advances in Databases, Knowledge, and Data Application, DBKDA*.

- [3] Hernandez, L., Baladrón, C., Aguiar, J.M., Carro, B., Sanchez-Esguevillas, A.J. and Lloret, J. (2013) Short-Term Load Forecasting for Microgrids Based on Artificial Neural Networks. *Energies*, **6**, 1385-1408. <http://dx.doi.org/10.3390/en6031385>
- [4] Lai, S.-H. and Hong, T. (2014) When One Size No Longer Fits All—Electric Load Forecasting with a Geographic Hierarchy. White Paper, SAS. <http://assets.fiercemarkets.com/public/sites/energy/reports/electricloadforecasting.pdf>
- [5] Day, P., et al. (2014) Residential Power Load Forecasting. *Proceeding of the CSER 2014 Conference*, Redondo Beach, March 2014, 457-464. <http://dx.doi.org/10.1016/j.procs.2014.03.056>
- [6] Mikkola, J. and Lund, P.D. (2014) Models for Generating Place and Time Dependent Urban Energy Demand Profiles. *Applied Energy*, **130**, 256-264. <http://dx.doi.org/10.1016/j.apenergy.2014.05.039>
- [7] Ding, N., Besanger, Y., Wurtz, F., Antoine, G. and Deschamps, P. (2011) Time Series Method for Short-Term Load Forecasting Using Smart Metering in Distribution Systems. *IEEE Trondheim Power Tech*, Trondheim, 19-23 June 2011, 1-6. <http://dx.doi.org/10.1109/ptc.2011.6019331>
- [8] Hong, T. and Shahidehpour, M. (2015) Load Forecasting Case Study. EISPC, U.S. Department of Energy.
- [9] Hong, T. (2014) Energy Forecasting: Past, Present, and Future. *FORESIGHT*, Winter 2014, 43-48.
- [10] Amjady, N., Keynia, F. and Zareipour, H. (2010) Short-Term Load Forecast of Microgrids by a New Bilevel Prediction Strategy. *IEEE Transactions on Smart Grid*, **1**, 286-294. <http://dx.doi.org/10.1109/TSG.2010.2078842>
- [11] Ilić, S.I., Vukmirović, S.M., Erdeljan, A.M. and Kulić, F.J. (2012) Hybrid Artificial Neural Network System for Short-Term Load Forecasting. *Thermal Science*, **16**, 215-224. <http://dx.doi.org/10.2298/TSCI120130073I>
- [12] Feinberg, E.A., Fei, J., Hajagos, J.T. and Rossin, R.J. (2011) Smart Grid Software Applications for Distribution Network Load Forecasting. *Proceedings of the First International Conference on Smart Grids, Green Communications and IT Energy-Aware Technologies*, Venice, 22-27 May 2011.
- [13] Hosking, G. and Zhang, N.S. (2013) Short-Term Forecasting on the Daily Load Curve for Residential Electricity Usage in the Smart Grid. IBM Research Report.
- [14] Jain, A. and Jain, M.B. (2013) Fuzzy Modeling and Similarity Based Short Term Load Forecasting Using Swarm Intelligence—A Step towards Smart Grid. In: Bansal, J.C., et al., Eds., *Proceedings of the Seventh International Conference on Bio-Inspired Computing: Theories and Applications*, Springer India, New Delhi, 15-27.
- [15] Serres, R. (2014) Short-Term Load Forecasting in New York State: NYISO Method and Support Vector Regression. Power System Analysis Project, Columbia University, New York.
- [16] Charlton, N. and Singleton, C. (2014) A Refined Parametric Model for Short Term Load Forecasting. *International Journal of Forecasting*, **30**, 364-368. <http://dx.doi.org/10.1016/j.ijforecast.2013.07.003>
- [17] Hong, T. (2010) Short Term Electric Load Forecasting. PhD Dissertation, North Carolina State University, Raleigh.
- [18] Hong, T., Wang, P. and White, L. (2015) Weather Station Selection for Electric Load Forecasting. *International Journal of Forecasting*, **31**, 286-295. <http://dx.doi.org/10.1016/j.ijforecat.2014.07.001>
- [19] Simmhan, Y., et al. (2011) An Informatics Approach to Demand Response Optimization in Smart Grids. Technical Report, USC.
- [20] Microsoft Corp (2013) Smart Energy Reference Architecture V2.0. Technical Report.
- [21] Oracle Corp (2010) Los Angeles Department of Water & Power Selects Oracle. Press Release.
- [22] Luckham, D.C. (2007) A Brief Overview of the Concepts of CEP. Technical Report.
- [23] Haben, S., Ward, J., Greetham, D.V., Singleton, C. and Grindrod, P. (2014) A New Error Measure for Forecasts of Household-Level, High Resolution Electrical Energy Consumption. *International Journal of Forecasting*, **30**, 246-256. <http://dx.doi.org/10.1016/j.ijforecast.2013.08.002>
- [24] Hu, Z.Y., Bao, Y.K., Xiong, T. and Chiong, R. (2015) Hybrid Filter—Wrapper Feature Selection for Short-Term Load Forecasting. *Engineering Applications of Artificial Intelligence*, **40**, 17-27. <http://dx.doi.org/10.1016/j.engappai.2014.12.014>