

A Pattern Recognition Approach for Peak Prediction of Electrical Consumption

Morten Goodwin, Anis Yazidi

► **To cite this version:**

Morten Goodwin, Anis Yazidi. A Pattern Recognition Approach for Peak Prediction of Electrical Consumption. 10th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Sep 2014, Rhodes, Greece. pp.265-275, 10.1007/978-3-662-44654-6_26 . hal-01391323

HAL Id: hal-01391323

<https://hal.inria.fr/hal-01391323>

Submitted on 3 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



A Pattern Recognition Approach for Peak Prediction of Electrical Consumption

Morten Goodwin^{1,2} and Anis Yazidi³

¹ Department of ICT, University of Agder, Grimstad, Norway

² Teknova AS, Grimstad, Norway

³ Institute for ICT, Oslo and Akershus University College of Applied Sciences, Norway

Abstract. Predicting and mitigating demand peaks in electrical networks has become a prevalent research topic. Demand peaks pose a particular challenge to energy companies because these are difficult to foresee and require the net to support abnormally high consumption levels. In smart energy grids, time-differentiated pricing policies that increase the energy cost for the consumers during peak periods, and load balancing are examples of simple techniques for peak regulation. In this paper, we tackle the task of predicting power peaks prior to their actual occurrence in the context of a pilot Norwegian smart grid network.

While most legacy studies formulate the problem as time-series-based estimation problem, we take a radically different approach and map it to a classical pattern recognition problem using a simple but subtle formulations. Among the key findings of this study is the ability of the algorithms to accurately detect 80% of energy consumption peaks up to one week ahead of time. Further, different classification methods have been rigorously tested and applied on real-life data from a Norwegian smart grid pilot project.

Keywords: Peak Prediction, Energy Consumption, Classification

1 Introduction

Changes in the consumers electrical power consumption influences demand peaks, which are difficult to predict due to their seemingly random occurrence. This is because consumption peaks are consequences of aggregated and collective behaviour of several customers, and are influenced by many external factors. A simple example of typical consumer's behavioral influence is when people turn on their heater when the weather is cold. When multiple consumers do this at the same time it, which is natural since a drop in temperature affects them all, the aggregated behavior causes a peak in the overall electrical consumption. However, since there are many factors other than temperature that influences a user electrical consumption, it is far from trivial to foresee what the consumption will be and in turn predict high loads.

Both consumption peaks and peak supplies cause challenges for electrical service providers because they need to design their grid for the maximum potential load. This is crucial since energy scarcity has severe consequences such as power outages. An alternative approach to over dimensioning electrical grids for the absolute maximum potential consumption peaks is evening out the peaks with peak controlling methods. The difficulty lies in that to control consumption, smart grid systems need know peaks prior to their occurrence, and there is no existing method that can accurately do so. There is however no doubt that the existence of such a method would help power companies, such as Agder Energy, to design intelligent strategies for reducing the overall consumption. The overall impact of successful peak prediction is enabling for electrical grids with less load and in turn collectively less energy usage.

This paper addresses predicting electrical power consumption peaks in small communities where there are no external data available, such as weather predictions or detailed customer data. This contributes to a crucial part of larger smart grid energy system which performs load balancing and avoids energy over-supply to keep grid voltage at an acceptable level. The peak prediction algorithms are meant to be used in a smart grid installation to: (1) Automatically turn off smart electrical appliances. (2) Automatically ask consumer power cells to be turned on. (3) Carry out local level load balancing. Hence, a potential peak at one customer household will avoid contributing to an overall demand peak in the system. This is a realistic scenario and such algorithms are planned to be part of Smart Village neighbourhood in Arendal, Norway.

This is particularly difficult as peaks are influenced by complex behaviour and typically occur in abnormal situations. Further, since this is planned to be installed in a new Smart Village neighbourhood, there is no external data available. We do not have access to other data such customer behavior, and since this is a small relatively secluded area weather reports is expected to have an equal impact on all customers and is therefore not particularly valuable⁴. The latter is different most approaches in the literature [1].

Our main contributions in this paper are: In contrast to most other approaches, mapping peak detection into a two-labelled classification problem using statistical terms. We further test it out with data from a real Norwegian pilot smart grid project, and are able to get results in line with state-of-the-art with relatively simple classification algorithms.

This paper is organised as following. Section 2 formally defines the problem to be solved as a two-labelled classification problem. Section 3 continues with the most relevant development in the area from two research areas: peak prediction/detection and rare item classification. The data used in these experiments are described in section 4 and the methods used for peak predictions are described in section 5. Section 6 continues with the results and findings from these methods. Finally, conclusion and future work is outlined in section 7.

⁴ The authors acknowledge that a change in temperature will influence electricity usage and in turn yield peaks. However, if there is an over all energy usage increase for all customers, techniques such as local load balancing will have limited value.

2 Problem setting

The goal of this research is to predict load peaks prior to their occurrence. Formally, given a time sequence \mathbf{t} let $p(t_i, n)$ be a statistical function asserting whether t_i is a peak where $t_i \in \mathbf{t}$ and n tells the extremity of the peak as following:

$$p(t_i, n) = \begin{cases} \text{peak if } t_i > \mu + n * \sigma \\ \text{not peak otherwise} \end{cases} \quad (1)$$

where μ and σ are the arithmetic average and standard deviation of \mathbf{t} , and n is a number defined as the extremity of the peak. Thus, a value is defined as a peak if it has a value more than n times the standard deviation above the mean value. Is it possible to predict the output from $p(t_i, n)$ using only part of the sequence occurring prior to t_i ($t_{<i} \in \mathbf{t}$)? Hence, the peak detection problem is deducted to a standard two-labelled classification problem. This is less trivial than most two-labelled classification problems because the peaks are rare and “abnormal”. Rare abnormalities are particularly hard to predict.

In layman’s terms: Is it possible to predict future load peaks using only past data on electrical consumptions?⁵

3 State-of-the-art

This section presents the most relevant research in the area of peak prediction. Section 3.1 presents the research on prediction electrical loads, and section 3.2 presents research on classifications when there is a skewed division of labelled data — i.e. classification when the positive labels are so few that it has a significant negative impact on the classification results.

3.1 Electrical consumption prediction

Many methods for load forecasting is available in the literature [1, 3–5]. Common for many of these is a reliance on some sort of third-party data for predictions and/or try to predict the actual power load. There are obvious similarities between peak prediction and load forecasting, but the objective is notably different.

A common approach is to use linear regression models in combination with other factors such as heat data. This is shown very useful to predict short-term peak loads [6, 7] and annual loads[8]. This research uses various functional linear regression models and support vector regression for the actual prediction, and clusters the data and assign one regression model per cluster.

Another common approach of load forecasting is to find a similar day based on the available data, such as weather reports, and assume that similar days

⁵ Note that is very different from the seemingly similar peak detection (positive outlier detection). For outlier detection it is possible calculate whether t_i is an outlier using t_i itself.[2]

have similar loads. Most relevant for this work is hierarchical two-step classifiers used for short-term load forecasting [9] which first predicts the “type of day” then an hourly forecast.

Some work exists without reliance on third party data with the assumption that the data has some internal structure than can be learnt.⁶ Perhaps most well known is the use of evolutionary algorithms and fuzzy logic approach to predict hourly peaks from one to seven days ahead [10], wavelet transformation techniques to accurately forecast power consumptions about 4 hours ahead of time [11] and short-term predictions using exponential smoothing methods [12]. Further, other machine learning techniques been attempted such as non-linear curve-fitting [13, 14] and Support Vector Machines (SVM) [15, 16].

3.2 Rare item classification

Rare item classification is a field within pattern recognition where the majority of the data is impertinent, typically the normal situation, and consequently the interesting data appears much rarer. Often times even the labels to be classified are not known in advance. Further, there is an imbalance in the data available for training (a priori). Exhaustive labelling is commonly required to build up a proper training database. Both labelling and corresponding classification is inevitably costly. Much research exists on this, but, to the best of our knowledge, none of these directly address the peak prediction problem.

An effective technique when working with rare item classification is combination methods, such as hierarchical classifiers with more than one classification technique [17, 18]. Generally this means organising local classifiers in an hierarchical structure and defining rules giving a global consensus. Based on techniques such as majority voting, the classifiers yields significantly better results for rare item classification than comparable algorithms, even with a flat structure [19]. Others use clever re-sampling, such as random under- and oversampling [17].

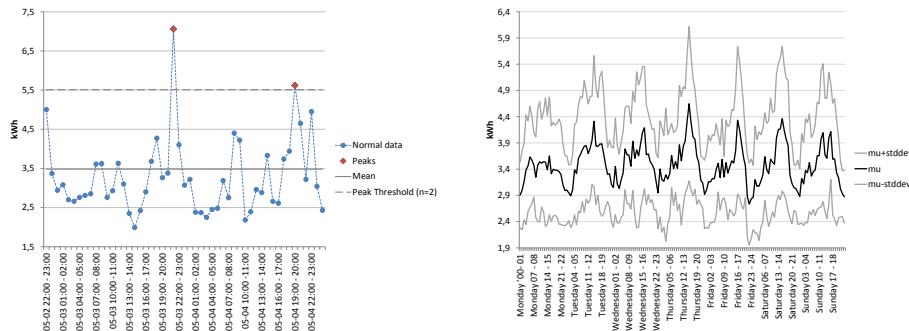
4 Data with electrical consumption peaks

The data used in this project is from a summer cabin area in Hvaler, Norway for six weeks in mid 2012 [20]. In total there are more than 7900 installations with varying degree of activity — making the data very vacillated. From each installation, accumulated hourly energy consumptions (kWh) was collected throughout the six weeks.

An examples of electrical consumption for one installation is presented in Figure 1(b). This shows averaged values of 7 days — displaying some clear trends: A rise and fall of the consumption with its highest around Norwegian dinner time and lowest at night, repeated nightly.

Figure 1(a) shows an example of consumption with peaks for a specific 24 hour period. The threshold level, n , for the peak data in Figure 1(a) is set to 2

⁶ This is the same assumption we have in this paper.



(a) Example of peak in the consumption data with n set to 2 (b) Consumption averaged per 7 days

Fig. 1: Consumption patterns and peaks

yielding two peaks: around 05-03 22:00 and 05-04 16:00. Overall, by varying n in the equation 1, the percentage of peaks varies from 15% for $n = 1$ down to 0.8% for $n = 3$.

5 Approach

This sections presents approaches for to predict $p(t_i, n)$. The first approach is a straightforward solution introduced for comparison purposes. This is continued with standard classification techniques inspired by similar work on load forecasting [6, 15, 16, 21]. Lastly, we present hierarchical classification solutions inspired by techniques in the literature on rate item classification[17–19]. All classifiers aim at predicting $p'(t_{<i}, n)$ as close to $p(t_i, n)$ as possible.

Majority voting(c_m):Firstly, a simple majority voting classifier was implemented that counts the number of peaks in the training data and checks is there are more peaks the previous day than could be expected based on simple statistics.

Consistent peaks(c_c):Secondly, another classifier was implemented that always predict the peak equal to the previous peak value as following:

$$c_c(t_{<i}, d, n) = p'(t_{<i}, n) = p(t_{i-d}, n). \tag{2}$$

where t_i is the data to predict and d is how many hours in advance t_i is.

SVM(c_l, c_p):The purpose of the SVM is to create a function, linear(c_l) or polynomial(c_p), that separates data that results in peaks from the data that does not. Upon predicting $p(t_i, n)$ the vector space consists of a defined number of points prior to $t_i \in \mathbf{t}$, namely $t_{<i}$. Each individual point in $t_{<i}$ is then a separate dimension. Two variants of SVM are implemented; The linear SVM approach is defined as $c_l(t_{<i}, d, n)$, while the polynomial is defined as $c_p(t_{<i}, d, n)$,

GMM(c_g):GMM is a classifier known for handling outliers well [22]. It has, to the best of our knowledge, not been used for load or peak prediction previously.

It has however been used for many other classification applications with success, and is specially suited for rare item classification. Similar, to the SVM classifiers, when classification of t_i with GMM we populate the vector space with points prior to t_i ($t_{<i}$). This classifier is defined as $c_g(t_{<i}, n)$.

Two variants of hierarchical classifiers are presented: Hierarchical and Hierarchical with Consistent Peaks. They both use a flat approach similar to variants in the literature [19] — namely straightforward combinations of the other classifiers.

Hierarchical (c_{h1}): The hierarchical classifier is simply a combination of the linear SVM and GMM similar to the common practice in the literature as following:[23, 24].⁷

$$c_{h1}(t_{<i}, d, n) = c_l(t_{<i}d, n) | c_g(t_{<i}d, n) \quad (3)$$

Hierarchical with Consistent Peaks (c_{h2}): An additional classifier was created utilizing the data from the consistent peaks (see equation 2). This extends equation 3 as following.

$$c_{h2}(t_{<i}, d, n) = c_l(t_{<i}, d, n) | c_g(t_{<i}, d, n) | c_c(t_{<i}, d, n) \quad (4)$$

6 Experiments

This section presents empirical results from running the predictive algorithms introduced in section 5. In line with common practice in the field [10], we present predictions starting from one hour into the future and up to 7 days (168 hours).

Section 6.1 presents a comparative experiments when n set to 1 — predicting moderate peaks. Additionally, section 6.2 shows the behaviour when varying n — making the peaks to detect more extreme — influences the algorithms.

All results presented in this chapter use a vector size of 24.⁸

Further, several metrics exists to evaluate information retrieval approaches. This paper promotes classification methods that favour false positives over false negatives. This is because a false positive, predicting a peak that is not present, has significantly less consequences than not predicting peaks which are present. Thus, high recall is much more sought after than high precision.

6.1 Predicting small consumption peaks

This section presents peak predictions with n in equation in $p(t_i, n)$ (equation 1) set to 1. This is the most straightforward approach where peaks are defined as simple values larger than the standard deviation.

Figure 2(a) and 2(b) show the recall and precision of the applied algorithms. These figures all show predictions 1 to 168 (7 days) into the future with n set to

⁷ Note that the decision rule is simpler than most other setups.

⁸ Several experiments were carried out with various vector sizes concluding with a vector size of 24 units. These results are not crucial to the understanding of the approach and therefore omitted.

1. This means that each time instance in the data, the algorithms try to predict from 1 to 168 hours ahead of time and the average is shown in the graph.

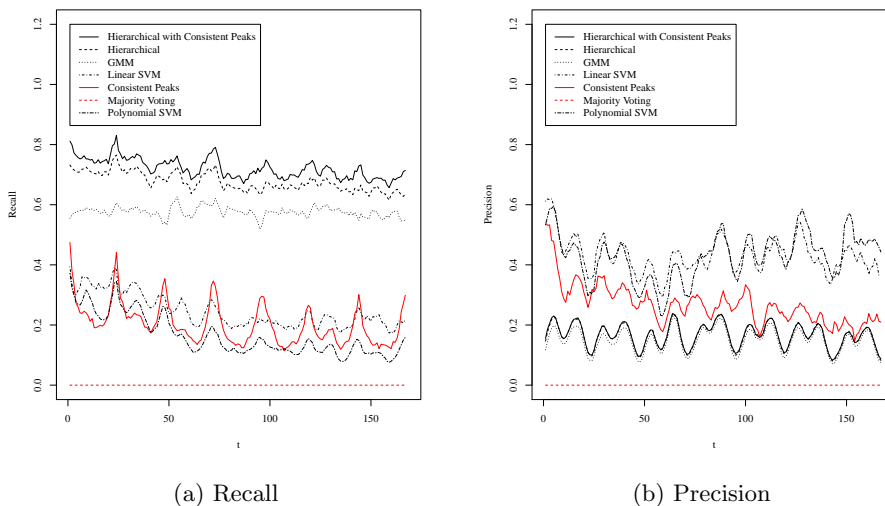
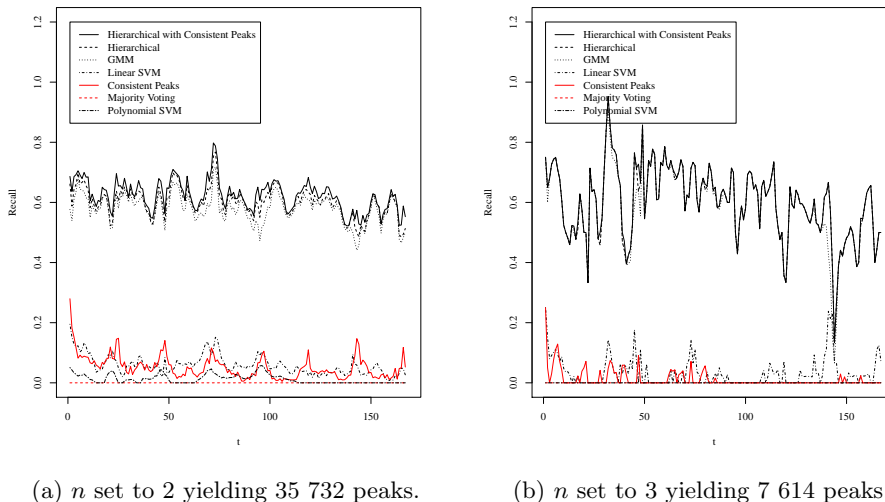


Fig. 2: Recall and precision for predicting 1 to 168 hours into the future with n set to 1

All pattern recognition algorithms, SVM, GMM and Hierarchical has a decrease in recall as the t increases. This means that, in line with intuition, the difficulty increases the further into the future the algorithm try to predict peaks. However, when t reaches 24 the quality of the measure increases. Hence, it is easier to predict 24 hours into the future than it is to predict other values. This suggests that people behave similarly in 24 hour cyclic intervals (cook dinner, eat, sleep at roughly the same intervals). This is supported by “Consistent peaks”-classifier is surprisingly accurate.

The figure shows that the hierarchical classifiers have an excellent recall (Figure 2(a)). Thus, these algorithms are able to detect about 80% of the peaks in the data. Further, the recall decreases as t increases for all algorithms. Hence, prediction becomes more difficult the further into the future the algorithms aim at predict. This is however less true for hierarchical classifiers and GMM than for comparative algorithms.

On the other hand, the hierarchical classifiers have a low precision (Figure 2(b)). The algorithms have many false positives (close to 80%). We can say that an hierarchical classifier is able to predict 80% of the peaks, and whenever it predicts a peak it is a 20% chance that it is an actual peak. Keep in mind that the objective of the algorithms is to favor false positives compared to false negatives having less impact in electrical power grids.

Fig. 3: Recall predicting 1 to 168 hours into the future varying n

6.2 Predicting larger consumption peaks

This section presents results upon varying n in $p(t_i, n)$ (equation 1). By increasing n the peaks become more extreme and (presumably) more difficult to predict.

Figure 3(a) and 3(b) show recall over time when n is set to 2 and 3. This is same setup as shown in figure 2(a) with n set to 1, but the behaviour is notably different. First and foremost, all classifiers have a decreased recall. Most notably, the SVM classifiers drop fast to a recall close to 0. In this scenario, SVM is only able to predict close to 24 hours into the future, and not very well. Consequently, the hierarchical classifiers are only slightly better than GMM — which means that SVM does not contribute much in an increased recall in the hierarchical classifiers.

The trend in Figure 2(a) ($n=1$) and 3(a) ($n=2$) is continued in figure 3(b) ($n=3$), but is more extreme. The SVM classifiers more quickly drops to 0 and the difference between the hierarchical and GMM classifiers are much less. Still, the best algorithms are able to reach a recall more than 70%.⁹

Another trend which is more visible is that the results are more chaotic which may be caused by only some of these extreme peaks are predictable, while others are not. With peaks as n increase this becomes more apparent in the data rendering more challenging results.

A conclusion to be drawn from this is that an increase in the extremities of the peaks makes the classification more difficult. However, GMM and corresponding

⁹ The trend continues when increasing n even further. This is deliberately left out of the paper since this does not contribute to further insight of the algorithms.

hierarchical classifiers are still able to detect more than 70% of the peaks. All other classifiers fall short with a recall close to 0.

7 Conclusion and future work

This paper deals with predicting electrical consumption peaks as input to load balancing and/or smart pricing strategies. This is done in a completely new way by mapping the prediction activities into a two-labelled classification problem. Further, in contrast to most existing approaches, the features are based solely on previous consumptions, avoiding reliance on third party data for the predictions.

Several classification algorithms are implemented and successfully applied on real-life data from a Norwegian smart-grid pilot project. The most promising results are produced by a simple hierarchical classifier combining linear support vector machines, Gaussian mixture models and deterministic assumption of consistent peaks. This solution is able to detect 80% of consumption peaks up to one week ahead of time. This promising result shows the usefulness of our approach.

This paper is part of an ongoing research project aiming at a developing smart energy housing technology with peak prediction as an essential component. We plan to carry out the following research: Comparison with additional classifiers, such as graph based classification approaches and more advanced hierarchical classifiers. We also plan to include weather information data as commonly done in the literature in order to improve the prediction results. Most importantly, this work is planned to be used in the scheduled Smart Village Skarpnes with 40 passive houses.

8 Acknowledgements

This project is partially funded by Agder Regional Research Fund with the project “Grid Operation and Distributed Energy Storage: Potential for improved grid-operation efficiency” and the Norwegian Research Council funded project “Electricity Usage in Smart Village Skarpnes”. These projects are carried out together with University of Agder, Teknova, Agder Energi Nett and Eltek. Data is collected through the external DeVID project.¹⁰

References

1. Feinberg, E., Genethliou, D.: Load forecasting. In Chow, J., Wu, F., Momoh, J., eds.: Applied Mathematics for Restructured Electric Power Systems. Power Electronics and Power Systems. Springer US (2005) 269–285
2. Hodge, V., Austin, J.: A survey of outlier detection methodologies. *Artificial Intelligence Review* **22**(2) (2004) 85–126
3. Campbell, P.R., Adamson, K.: Methodologies for load forecasting. In: Intelligent Systems, 2006 3rd International IEEE Conference on, IEEE (2006) 800–806

¹⁰ <http://www.sintef.no/Projectweb/DeVID/Prosjektpartnere/>

4. Weron, R.: Modeling and forecasting electricity loads and prices: A statistical approach. Volume 403. Wiley. com (2007)
5. Mohsenian-Rad, A.H., Leon-Garcia, A.: Optimal residential load control with price prediction in real-time electricity pricing environments. *Smart Grid, IEEE Transactions on* **1**(2) (2010) 120–133
6. Goia, A., May, C., Fusai, G.: Functional clustering and linear regression for peak load forecasting. *International Journal of Forecasting* **26**(4) (2010) 700 – 711
7. Pardo, A., Meneu, V., Valor, E.: Temperature and seasonality influences on spanish electricity load. *Energy Economics* **24**(1) (2002) 55 – 70
8. Wang, J., Li, L., Niu, D., Tan, Z.: An annual load forecasting model based on support vector regression with differential evolution algorithm. *Applied Energy* **94**(0) (2012) 65 – 70
9. Ilić, S.A., Vukmirović, S.M., Erdeljan, A.M., Kulić, F.J.: Hybrid artificial neural network system for short-term load forecasting. *Thermal Science* **16**(suppl. 1) (2012) 215–224
10. Yang, H.T., Huang, C.M.: A new short-term load forecasting approach using self-organizing fuzzy armax models. *Power Systems, IEEE Transactions on* **13**(1) (1998) 217–225
11. Annamareddi, S., Gopinathan, S., Dora, B.: A simple hybrid model for short-term load forecasting. *Journal of Engineering* **2013** (2013)
12. Taylor, J.W.: An evaluation of methods for very short-term load forecasting using minute-by-minute british data. *International Journal of Forecasting* **24**(4) (2008) 645–658
13. Zhang, G., Eddy Patuwo, B., Y Hu, M.: Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting* **14**(1) (1998) 35–62
14. Suganthi, L., Samuel, A.A.: Energy models for demand forecasting a review. *Renewable and Sustainable Energy Reviews* **16**(2) (2012) 1223–1240
15. Mohandes, M.: Support vector machines for short-term electrical load forecasting. *International Journal of Energy Research* **26**(4) (2002) 335–345
16. LI, Y.c., FANG, T.j., YU, E.k.: Study of support vector machines for short-term load forecasting [j]. *Proceedings of the Csee* **6** (2003) 010
17. Kotsiantis, S., Kanellopoulos, D., Pintelas, P., et al.: Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering* **30**(1) (2006) 25–36
18. Silla Jr, C.N., Freitas, A.A.: A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery* **22**(1-2) (2011) 31–72
19. Xiao, Z., Dellandrea, E., Dou, W., Chen, L.: Hierarchical classification of emotional speech. *IEEE Transactions on Multimedia* (2007)
20. Sand, K., Foosnas, J., Nordgard, D.E., Kristoffersen, V., Solvang, T.B., Wage, D.: Experiences from norwegian smart grid pilot projects. In: *Electricity Distribution (CIRED 2013), 22nd International Conference and Exhibition on, IET* (2013) 1–4
21. Chen, B.J., Chang, M.W., et al.: Load forecasting using support vector machines: A study on eunite competition 2001. *Power Systems, IEEE Transactions on* **19**(4) (2004) 1821–1830
22. Jain, A.K., Duin, R.P.W., Mao, J.: Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **22**(1) (2000) 4–37
23. Uguroglu, S.: *Robust Learning with Highly Skewed Category Distributions*. PhD thesis, Carnegie Mellon University (2013)
24. Huang, H., He, Q., Chiew, K., Qian, F., Ma, L.: Clover: a faster prior-free approach to rare-category detection. *Knowledge and Information Systems* **35**(3) (2013) 713–736