Is Unsupervised Ensemble Learning Useful for Aggregated or Clustered Load Forecasting?

Peter Laurinec and Mária Lucká

Faculty of Informatics and Information Technologies,
Slovak University of Technology in Bratislava,
Ilkovičova 2, 842 16 Bratislava, Slovak Republic
{peter.laurinec,maria.lucka}@stuba.sk

Abstract. This paper presents a comparison of the impact of various unsupervised ensemble learning methods on electricity load forecasting. The electricity load from consumers was simply aggregated or optimally clustered to more predictable groups by cluster analysis. The clustering approach consists of efficient preprocessing of data gained from smart meters by a model-based representation and the K-means method. Two types of ensemble learning methods were implemented to investigate the performance of forecasting on clustered or simply aggregated load: bootstrap aggregating based and the newly proposed clustering based. The smart meter datasets used in our experiments come from Ireland and Slovakia, where data from more than 3600 consumers were available in both cases. The achieved results proved that unsupervised ensemble learning for forecasting aggregated and clustered load improves accuracy.

Keywords: load forecasting, clustering, bagging, ensemble learning

1 Introduction

Modern information technologies produce a large amount of data that can be used for further analysis, giving important insights into data that supports making informed decisions. One important source of data is smart meters - sensors measuring electricity consumption or production. A smart grid is an ecosystem created from smart meters that can control or monitor electricity load (of consumers) or both load and production (of prosumers), collecting a large amount of data and making interventions when needed. To make these interventions or decisions useful, they must be supported by information provided by smart meter data. One of the key tasks in a smart grid is electricity load forecasting, which is essential for energy distribution and utility companies, salesmen and for end users. Developing more sophisticated and accurate forecasting methods is important and for these purposes, data mining and machine learning methods are developed and adapted.

There are several suitable methods for load forecasting such as time series analysis and regression methods. Both types of them have their limitations such as an inability of adaptation to sudden changes (concept drift) and the noisy behaviour of time series. Therefore to find and choose the most suitable forecasting method is difficult. Besides already existing forecasting methods, there is a promising approach using the proper combination of various methods overcoming limitations of particular ones. This method is called ensemble learning. Moreover, forecasting methods themselves can be tuned by the simplest of all ensemble methods - bootstrap aggregating (bagging).

We have evaluated two methods of bagging: a) moving block bootstrap in combination with time series analysis methods; b) and a method based on classical sampling with replacement combined with randomized values of hyperparameters for regression trees. We have proposed several new ideas to unsupervised ensemble learning approaches relying on a proper combination of multiple bootstrap forecasts. Their advantages and disadvantages were discussed in our work.

Another very important but totally different approach to optimising forecast accuracy is based on advanced time series data mining methods. This includes cluster analysis that is used for consumer segmentation according to their consumption patterns, so more predictable groups of consumers are created. As we showed in our previous works [1, 2], this approach has the promising improvement of forecasting accuracy.

The aim of this paper is to compare the combination of the time series data mining approach with the newly proposed ensemble learning methods for improving forecasting accuracy.

This paper is structured as follows: Section 2 contains an introduction with related works, while in Section 3 the datasets used in our experiments are described. Section 4 presents a description of our approach together with the methods used for time series processing, cluster analysis, forecasting and ensemble learning. Section 5 presents the description and the evaluation of performed experiments and the paper concludes with Section 6.

2 Related Work

Electricity load forecasting is a highly discussed research area due to the interesting character of data coming from smart meters. The time series of electricity consumption can have various patterns and are affected by multiple seasonalities (daily, weekly and yearly), weather, holidays and other unexpected changes. For this reason, sophisticated machine learning methods are applied to tackle challenges linked with smart meter data.

Ensemble learning in load forecasting is a highly used method in solving the above-mentioned problems. Adhikari et al. [3] proposed a ranking based ensemble approach to incorporate only the best models to the final ensemble forecast. Shen et al. [4] proposed a pattern forecasting ensemble model, which combines forecasts created by clustering algorithms. Grmanova et al. [5] used median-based approach to optimise weights in the incremental heterogeneous ensemble learning model for consumption forecasting.

The usage of cluster analysis for more accurate forecasts of aggregated load is noted in the work of Shahzadeh et al. [6]. This paper deal with the clustering of consumers in three different ways of feature extraction from a time series. As a clustering method, K-means was used and the neural network has been applied as a forecast method. Wijaya et al. [7] used correlation-based feature selection as a representation of consumers in clustering, and linear regression, multi-layer perceptron and support vector regression were used as forecasting methods. In our previous work [2], four different representations of time series and ten forecasting methods were evaluated, in order to verify their suitability for the forecasting of clustered load. We have proved that optimised clustering of consumers significantly improves the accuracy of forecasts in combination with triple exponential smoothing, ARIMA, Random Forests and bagging.

Until now, the combination of clustering of consumers and ensemble learning has not been explored and evaluated. Therefore in the proposed paper we will a) evaluate from two different types of bagging on basic forecasting methods and examine their behaviour on clustered load, b) design several new unsupervised clustering ensemble learning approaches for forecasting, c) for clustering electricity consumers, propose efficient preprocessing through model-based representation of time series based on K-means clustering.

3 Smart Meter Data

For verification of our approach, we have used in our experiments two different datasets, comprising of data from smart meters. This data includes Irish and Slovak electricity consumption. The Irish data was collected by the Irish Commission for Energy Regulation (CER) and available from ISSDA[1] (Irish Social Science Data Archive). This data contains three different types of customers: residential, SMEs and others. The largest group is residential, where after removing consumers with missing data, we have 3639 residential consumers left. The frequency of data measurements was on a half-hour basis, during a day 48 measurements were performed. Slovak data was collected within the project International Centre of Excellence for Research of Intelligent and Secure Information-Communication Technologies and Systems. These measurements were obtained from Slovak enterprises, having a completely different nature than the Irish data. After removing consumers with missing data, those with zero consumption and consumption higher than 42 kW, the dataset comprised 3630 consumers. The frequency of data measurements was on a quarter-hour basis, so daily 96 measurements were performed. The frequency of data measurements was transformed to half-hourly in order to make it comparable with the Irish data.

The difference between the residential and enterprise data is significant. The amount of consumption in residences was low and not regular, as opposed to the enterprise, where the amount of consumption was very high and mostly regular during the week and irregular during the year (i.e., holidays for whole factory or in school, the period of year when central heating is turned on). Therefore different evaluation results for these two datasets could be expected.

[1] `http://www.ucd.ie/issda/data/commissionforenergyregulationcer/`

4 Proposed Approach

Two approaches for the aggregation of electricity load were compared: based on clustering of consumers and based on simple aggregation. Moreover, three types of forecasting methods were compared: six basic methods, six basic methods with bagging and six ensemble methods. The clustering approach consists of these phases:

1. Normalisation of time series by z-score and calculation of a model-based representation of time series (estimation of regression coefficients). Extracted representations then enter a clustering method.
2. Calculation of an optimal number of clusters for given representations of a time series by DB-index. The actual clustering of consumers is followed by the K-means method.
3. Aggregation of consumption within the clusters and the application of a forecast model on training data. The forecast for the next period is calculated and aggregated. Finally it is compared with the real consumption.

This process is repeated incrementally until new data is available. The sliding window has length of 21 days and it is shifted by one day. It means that the oldest day from the window is removed and the data from a new day are added.

4.1 Clustering of Electricity Consumers

The first necessary step is the normalisation of the times series of electricity consumption by the z-score because we want to cluster similar patterns and not the time series according to the amount of energy consumption.

The next is the computation of the time series representation, which is an input to the clustering algorithm. The modification of the time series to its representation is performed by a suitable transformation. The main reason for using representations of time series is the pursuit of more effective and easier work with time series, depending on the application. Using time series representations is appropriate because by reducing the dimension, it will reduce memory requirements and computational complexity, and it implicitly removes noise and emphasizes the essential characteristics of data. We conducted from our previous work that model-based representations are highly appropriate for seasonal time series [1]. For a model, multiple linear regression is used for extraction of regression coefficients of two seasonalities (daily and weekly). Formally, the model can be written as follows:

$$x_t = \beta_{d1}u_{td1} + \beta_{d2}u_{td2} + \cdots + \beta_{ds}u_{tds} + \beta_{w1}u_{tw1} + \cdots + \beta_{w6}u_{tw6} + \varepsilon_t,$$

for $t = 1, \dots, n$, where x_t is the $t-th$ electricity consumption, $\beta_{d1}, \dots, \beta_{ds}$ are regression coefficients for daily season, s is the length of period of one day, $\beta_{w1}, \dots, \beta_{w6}$ are regression coefficients for a weekly season. Weekly regression coefficients are just six, not seven, because of prevention from singularity of the model. The $u_{td1}, \dots, u_{tdseas}$, u_{tw1}, $\dots$, u_{tw6} are independent binary (dummy) variables representing the sequence numbers in the regression model. They are equal to 1 in the case when they point to the $j - th$ value of the season, $j = 1, 2, \ldots, s$, in case of a daily season and $j = 1, 2, \ldots, 6$ in case of a weekly season. The ε_t are random errors having the normal distribution of $N(0, \sigma^2)$ that are for different t mutually independent. The most widespread method for obtaining an estimate of the vector $\boldsymbol{\beta} = (\beta_{d1}, \ldots, \beta_{ds}, \beta_{w1}, \ldots, \beta_{w6})$ is the Ordinary Least Squares method. In Fig. 1 the transformation of time series of length of three weeks $(48 \times 21 = 1008)$ to the model-based representation of length 54 is shown.

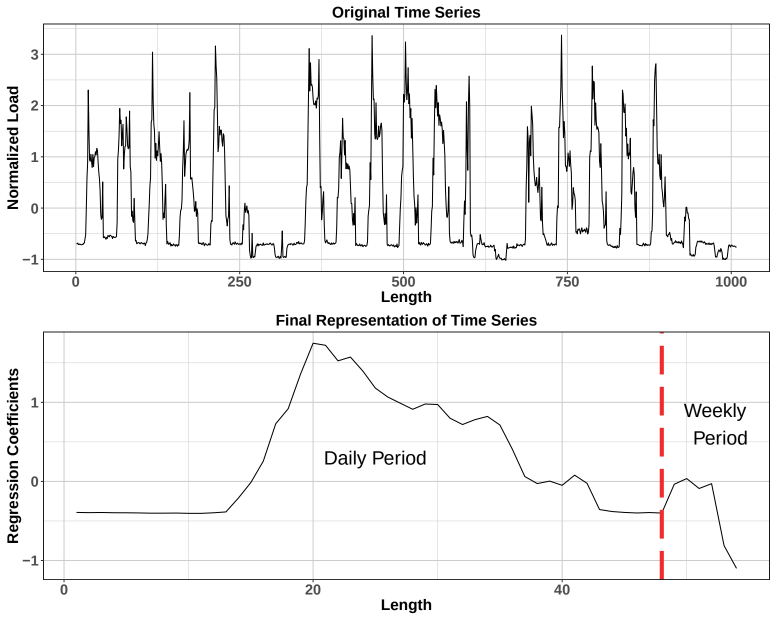

Fig. 1. Proposed representation of time series of a randomly picked Slovak consumer.

For grouping consumers into clusters, the centroid-based clustering method K-means with centroids initialization K-means++ [8] was used. The advantage over conventional K-means is based on carefully seeding of initial centroids, which improves the speed and accuracy of clustering.

In each iteration of a batch processing, we have automatically determined the optimal number of clusters to K using the internal validation rate Davies-Bouldin index [9]. The optimal number of clusters ranged from 8 to 18.

In Fig. 2 clustered time series representations of consumers from Slovakia are shown. We can see that the clusters 1, 2 and 8 have a similar daily pattern, but the weekly pattern is remarkably different, so our clustering approach is working correctly. As is apparent, other clusters are visibly different from each other.

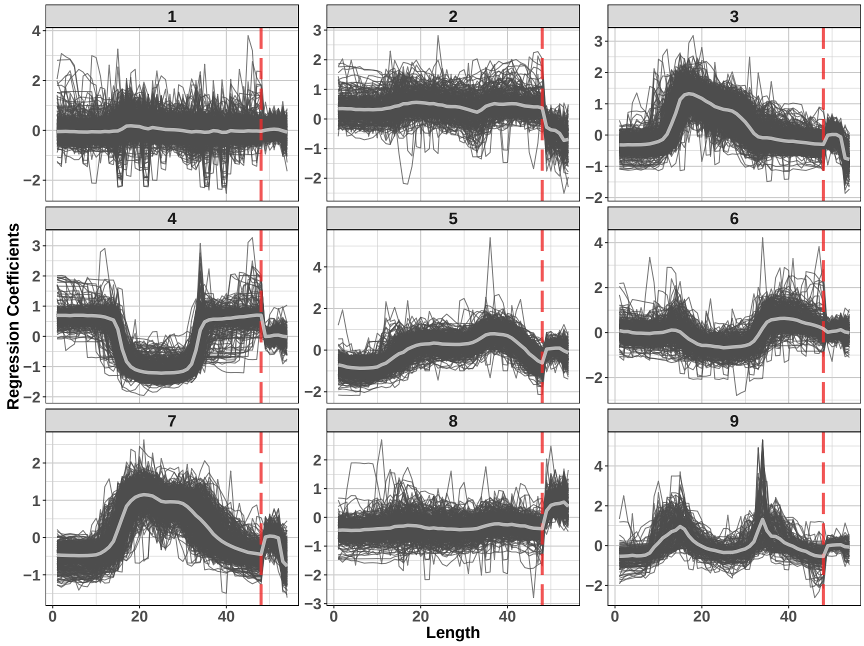

Fig. 2. Nine clusters of Slovak consumers. Grey line represents the centroid of a cluster.

4.2 Forecasting Methods

Basic Forecasting Methods We have compared six basic forecasting methods to investigate their relevance in combination with bootstrapping, and to see if it benefits from clustering.

Seasonal decomposition of time series by Loess (STL) is a method that decomposes a seasonal time series into three parts: trend, seasonal and remaining [12]. For the resulting three time series, the result is used separately for the forecast with Holt-Winters exponential smoothing and ARIMA model (STL+ARIMA). The ARIMA model has been introduced by Box and Jenkins [13] and is one of the most popular approaches in forecasting. The Holt-Winters exponential smoothing [14] is a forecasting method applied to a time series, whereby past observations are not weighted equally, but the weights decrease exponentially with time. The exponential smoothing method was used with STL decomposition (STL+EXP) and also stand-alone (EXP).

Recursive partitioning regression trees that belong to Classification and Regression Trees methods (CART) search over all possible splits by maximising an information measure of node impurity, selecting the covariate showing the best split [10]. The most important hyperparameters that must be tuned are the minimum number of observations in needed in node to split (set to 2), maximal depth of a tree (set to 30) and the complexity parameter (cp). The last parameter cp is a threshold deciding if each branch fulfils conditions for further processing (only nodes with fitness larger than factor cp = 1E-6 are processed).

The attributes to the CART model are daily and weekly seasonal vectors. We considered the size of a daily period $s = 48$ and weekly period $w = 7$. Daily seasonal vector has the form $\mathbf{day}_j = (1, 2, \dots, s)$, $j = 1, 2, \dots, d$, where d is the number of days in the training window and $\mathbf{day} = (\mathbf{day}_1, \dots, \mathbf{day}_d)$. Let $\bar{\mathbf{i}} = (i, \dots, i)$ is a vector of dimension s. Then the weekly seasonal vector has the form $\mathbf{week} = (\bar{\mathbf{1}}, \bar{\mathbf{2}}, \dots, \bar{\mathbf{w}}, \bar{\mathbf{1}}, \dots)$ and has the dimension of $s * d$.

Conditional inference trees (CTREE) is a statistical approach to recursive partitioning, which takes into account the distributional properties of the measurements [11]. Here the important hyperparameter for tuning is the minimal criterion that must be exceeded in order to implement a split (set to 0.95).

Two variants of CTREE method based on different attributes entering to the model were evaluated. The first one (CTREE.lag) has four seasonal attributes for daily and weekly periods in the sinus and cosinus form: $(\sin(2\pi\frac{\boldsymbol{day}}{s}) + 1)/2$ resp. $(\cos(2\pi\frac{\boldsymbol{day}}{s}) + 1)/2$, and $(\sin(2\pi\frac{\boldsymbol{week}}{7}) + 1)/2$ resp. $(\cos(2\pi\frac{\boldsymbol{week}}{7}) + 1)/2$. Another attribute for the model is the seasonal component of STL decomposition with a one day lag. The second one (CTREE.dft) uses as attributes two seasonal Fourier terms. As we found experimentally the best results were achieved with six terms for daily period $(\sin(\frac{2\pi jt}{48}),\ \cos(\frac{2\pi jt}{48}))_{j=1}^{6}$, and twelve pairs of terms for weekly seasonality $(\sin(\frac{2\pi jt}{7}), \cos(\frac{2\pi jt}{7}))_{j=1}^{12}$, where $t = (1, \dots, n)$.

Bootstrap Aggregating Methods Bootstrap aggregating (bagging) is an ensemble meta-algorithm [15], which creates multiple versions of a learning set to produce a multiple number of predictors. These predictors are then aggregated, for example by arithmetic mean. We have implemented two different bagging methods in order to adapt to two different types of forecasting methods: regression trees and time series analysis methods.

Classical bagging proposed by Breiman [15], generates multiple training sets by uniformly sampling the original one with replacement. In our approach, the sample ratio and hyperparameters mentioned in the previous section concerning regression trees were also randomised. The sample ratio was randomly sampled in the range of $0.7 - 0.9$. The CART hyperparameters were sampled this way: maximal depth in range of $26 - 30$, minimal split $2 - 3$ and cp 9E-7 − 1E-5. The CTREE hyperparameter minimal criterion is sampled $0.88 - 0.97$. Each time 150 trees were created and the resulting forecasts were aggregated by median.

For the time series analysis methods (STL+ARIMA, STL+EXP, EXP), the bagging proposed by Bergmeir et al. [16] was used. At first a Box-Cox transformation to the data was applied, then the series was decomposed into three components by STL. The remainder component is then bootstrapped using the moving block bootstrap (mbb), and to every created bootstrap version of the remainder, the trend and seasonal components are added, and then the Box-Cox transformation is inverted. So a random pool of similar bootstrapped time series is generated (in our case 150). After applying the forecasting method to each time series, the forecasts are aggregated by median. In Fig. 3 the results of mbb method applied on the Irish and Slovak data are shown. As it can be seen the method produces very noisy time series when the original data is also noisy (Slovak data).

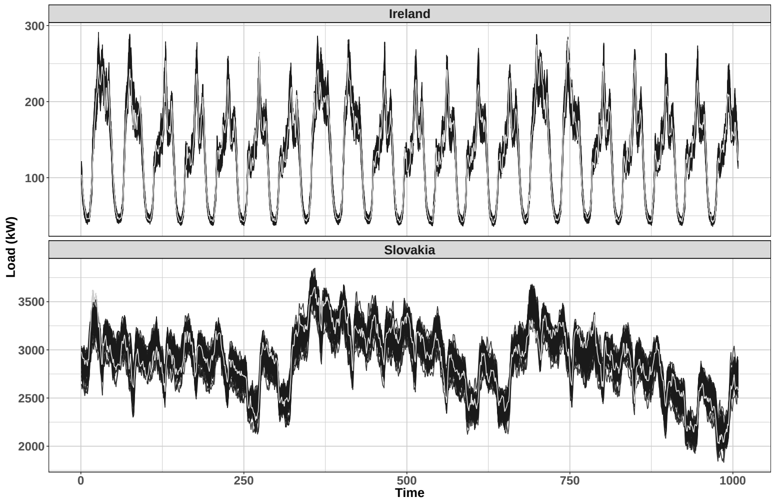

Fig. 3. Mbb method used on two different time series of length three weeks from Ireland and Slovakia. The original time series is illustrated with the light grey colour.

Ensemble Learning We have implemented six different ensemble learning methods that we can divide into these three groups:

a) Simple aggregation based - average and median,
b) Naive cluster based - average of medians of methods,
c) Cluster based - K-means based, DBSCAN based and OPTICS based.

There are other widely used ensemble learning methods, which are error based, so they weight each prediction method by its performance. We proposed ensemble learning methods, which are structure based, so it uses unsupervised learning to create a final ensemble forecast. As we have found and experimentally proven, only unsupervised approaches are suitable for time series created by clustering, which are newly generated in each data window. Reason of this claim is that each created clustered time series needs to apply different forecasting method.

Average and median ensemble simply aggregates all available forecasts, in our case, there are $6 \times 150 = 900$ forecasts, which were produced by bagging.

The first cluster-based method uses priori information on which forecasting method was used. Each method creates a median. After this, a set of medians (in this case having a set length of 6) is averaged to a final ensemble forecast.

The next three methods are cluster-based. Before using a clustering algorithm on a dataset of forecasts (matrix of dimension 900×48), the Principal Component Analysis is used to extract just the first three principal components in order to reduce noise. The K-means based procedure (Section 4.1) was used to create clusters of forecasts. First, DB-indexes were computed and an optimal number of clusters in the range of $3-8$ was found. Next K-means produced clusters with corresponding centroids, which were averaged to the final ensemble forecast.

All ensemble methods mentioned above used all forecasts to produce the final one, even when anomalous, which could cause loss of forecasting accuracy. For this reason, density-based clustering methods which can automatically filter noisy objects were implemented. First of them, DBSCAN [17] (Density-Based Spatial Clustering of Applications with Noise) clustering algorithm was used. It requires two parameters: ϵ (set to 1.45) and $minPts$ (set to 8). Ensemble forecast is created by the average of medians of clusters. The biggest drawback of this approach is that these parameters in the whole process of evaluation are set statically and not dynamically. However, deviations are reduced by principal components normalisation, which guarantees stability in the range of data values.

For producing density-based clustering that automatically adapts to the shape of objects, OPTICS [18] (Ordering Points To Identify the Clustering Structure) algorithm with automatic ξ-cluster procedure was implemented. ξ defines the degree of steepness (set to 0.045), which is applied in the so-called reachability plot of distances. The final ensemble forecast is the median of medians of clusters. The accuracy of the load forecast was measured by MAPE (Mean Absolute Percentage Error). MAPE is defined as $100 \times \frac{1}{n}\sum_{t=1}^{n}\frac{|x_t-\overline{x}_t|}{x_t}$, where x_t is a real consumption, $\overline{x}_t$ is the forecasted load and n is a length of data.

5 Experiments

We have performed several experiments to evaluate our proposed approach. The Ireland testing dataset contains 3 months of measurements from the year 2010 (1.2.2010 − 28.2.2010, 1.5.2010 − 31.5.2010 and 1.8.2010 − 31.8.2010), comprising 90 days. The Slovak testing dataset contains also 3 months of measurements from years 2013 and 2014 (23.9.2013 − 26.10.2013, 10.2.2014 − 11.3.2014 and 2.6.2014 − 1.7.2014), comprising 94 days. Moreover we had additional data coming from 21 days before each of the six tested periods that were used in clustering and train forecasting methods.

The first comparison of forecasting methods is shown in Table 1, where all ensemble methods are compared on aggregation and clustering approaches on both datasets. Besides the comparison of average values of MAPE, the p-values of Wilcoxon rank sum test are also shown. They show whether forecast errors with the clustering approach have significantly lower values than simple aggregation.

CTREE.bagg.dft method was best basic bagged method on Irish data and CTREE.bagg.lag on Slovak data. The best ensemble method on Irish data was simple median. On the simple aggregated Slovak data the method of average of medians had lowest MAPE. The lowest MAPE on clustered Slovak dataset has been achieved by the DBSCAN-based ensemble method. A significant improvement of results with the clustering approach on Irish data was attained using two methods: CTREE.bagg.lag and K-means. On the other hand, clustering helped in 8 of the 12 cases on Slovak data. The significance of the best ensemble approach against best bagged method is shown in Table 2.

Table 1. Average daily MAPE (%) of 12 forecasting methods evaluated on two datasets and two types of aggregation. Agg. represents simple aggregation of consumption and Clust. clustering approach. Bold values represent the lowest MAPE among bagged basic and among ensemble methods. P-vales less than 0.05 are bold.

	Agg.-Irel.	Clust.-Irel.	p-value	Agg.-Slov.	Clust.-Slov.	p-value
CART.bagg	3.7908	3.7964	0.4872	3.1561	3.0993	**0.0092**
CTREE.bagg.lag	3.8081	3.7599	**0.0301**	**2.9568**	**2.8730**	**0.0012**
CTREE.bagg.dft	**3.6746**	**3.7103**	0.9018	3.0080	2.9341	**0.0018**
STL+ARIMA.mbb	3.9344	3.9085	0.0935	3.0325	2.9993	**0.0090**
STL+EXP.mbb	3.9901	4.0221	0.9881	3.0306	3.0021	0.1472
EXP.mbb	4.0565	4.0723	0.5923	2.9760	2.9446	0.1282
Average	3.7034	3.6970	0.2717	2.8312	2.8086	0.0533
Median	**3.6103**	**3.6046**	0.3363	2.8329	2.7980	0.0832
AveMedians	3.6704	3.6771	0.8054	**2.8179**	2.7901	**0.0318**
K-means	4.3018	4.0189	**0.0140**	2.9715	3.0916	**0.0105**
DBSCAN	3.9752	3.7985	0.2625	2.9352	**2.7532**	**0.0103**
OPTICS	3.7482	3.7239	0.4710	2.9253	2.7982	**0.0003**

Table 2. P-values from hyphothesis if ensemble method is better than forecasting method with bagging.

Agg.-Irel.	Clust.-Irel.	Agg.-Slov.	Clust.-Slov.
Median-CTREE.dft	Median-CTREE.dft	AveMed.-CTREE.lag	DBSCAN-CTREE.lag
0.0011	<**0.0001**	0.1379	0.2415

The results of simple median ensemble method were significantly better on both approaches on Irish data, however on Slovak data ensembles were not significantly better, even though their MAPE was lower than by using the bagged basic method.

Table 3 summarises the results of basic methods used without bagging. Corresponding p-values shows that Irish data clustering has a significant effect on the accuracy of forecasts. On Slovak data, significant improvements include the regression tree methods, but time series analysis methods have failed testing.

Table 4 shows p-values of comparison (differences) of Tables 1 and 3, where it is tested if the bagging basic forecasting methods improve forecasting accuracy significantly. On the simple aggregated Irish data, all methods achieved statistically significant results, except the CART.bagg method. On the Slovak data, the bagging regression trees had significant results, but time series analysis methods had failed. This is caused by the mbb method that is not adaptable on noisy and fluctuated time series which is highly present in Slovak data.

6 Conclusion

In our paper we have proposed and tested two techniques for forecasting electricity load. We have compared and implemented various ensemble learning methods

Table 3. Average daily MAPE (%) of 6 forecasting methods evaluated on two datasets and two types of aggregation. Agg. represents simple aggregation and Clust. clustering approach. Bold values represent lowest MAPE. P-values less than 0.05 are bold.

	Agg.-Irel.	Clust.-Irel.	p-value	Agg.-Slov.	Clust.-Slov.	p-value
CART	**3.8570**	3.8502	**0.0171**	3.1921	3.1416	**0.0107**
CTREE.lag	3.9203	**3.7523**	<**0.0001**	2.9950	2.8954	**0.0008**
CTREE.dft	4.0849	3.9214	<**0.0001**	3.1944	3.0096	<**0.0001**
STL+ARIMA	4.0718	3.8943	**0.0247**	2.7567	2.7404	0.0738
STL+EXP	4.2750	4.1866	0.5560	2.6887	2.6424	0.1748
EXP	4.8000	4.2219	<**0.0001**	**2.3957**	**2.4672**	0.9928

Table 4. P-values of testing if bagging six basic forecasting methods improves forecasting accuracy significantly.

	Agg.-Irel.	Clust.-Irel.	Agg.-Slov.	Clust.-Slov.
CART.bagg	0.1113	0.0760	**0.0075**	**0.0002**
CTREE.bagg.lag	**0.0001**	0.4807	**0.0004**	**0.0036**
CTREE.bagg.dft	<**0.0001**	<**0.0001**	<**0.0001**	**0.0002**
STL+ARIMA.mbb	**0.0178**	0.0592	0.9875	0.9996
STL+EXP.mbb	**0.0380**	**0.0330**	0.9980	0.9995
EXP.mbb	<**0.0001**	0.2559	1.0000	1.0000

in order to compare its forecasting accuracy using aggregated and clustered electricity load. Six base forecasting methods combined with two different bagging methods and six unsupervised ensemble approaches, combining all available forecasts created by bootstrap methods, were evaluated. A new approach based on unsupervised ensemble learning in combination with clustered load was proposed and evaluated. The clustering of consumers was performed by efficient preprocessing using estimated regression coefficients as a representation of time series, K-means method and optimally finding a number of clusters by DB-index. We have proven that the bagging of regression trees significantly improves accuracy on both aggregated and clustered load. On the other hand, bagging of time series analysis methods can be unreliable, because of weak adaptivity to noisy and fluctuated data. Simple median ensemble approach performed significantly better than any other forecasting methods on Irish smart meter data. The best forecasting method on aggregated Slovak data was the average of medians method and on clustered data the DBSCAN-based method. However, these two methods did not gain significantly better results when compared with the bagged basic models. The simple ensemble method - median of all forecasts, was better (in mean of MAPE) in all cases when compared with the best bagged basic methods. For this reason we conclude that ensemble learning is suitable for aggregated and also clustered load forecasting. The clustering of consumers stably improved forecasting accuracy of all methods, except for the time series analysis methods on Slovak data.

Smart meter data are often noisy and fluctuated. They force us to develop more robust methods to the detect trend shift (concept drift) and handle the noisy character of data. In future work, we want to focus on ensemble and clustering methods that are more adaptable for the aforementioned problems.

Acknowledgments. This work was partially supported by the Scientific Grant Agency of The Slovak Republic, Grant No. VG 1/0752/14 and STU Grant scheme for Support of Young Researchers.

References

1. Laurinec, P., Lucká, M.: Comparison of Representations of Time Series for Clustering Smart Meter Data. In: Proceedings of WCECS, pp. 458–463 (2016)
2. Laurinec, P., Lóderer, M., Vrablecová, P., Lucká, M., Rozinajová, V., Ezzeddine, A. B.: Adaptive Time Series Forecasting of Energy Consumption Using Optimized Cluster Analysis. In: Proceedings of IEEE ICDMW, pp. 398–405 (2016)
3. Adhikari, R., Verma, G., Khandelwal, I.: A Model Ranking Based Selective Ensemble Approach for Time Series Forecasting. Proc. Comp. Scie. 48, pp. 14–21 (2015)
4. Shen, W., Babushkin, V., Aung, Z., Woon, W. L.: An ensemble model for day-ahead electricity demand time series forecasting. In: e-Energy '13, ACM, pp. 51–62 (2013)
5. Grmanová, G., Laurinec, P., Rozinajová, V., Ezzedine, A. B., Lucká, M., Lacko, P., Vrablecová, P., Návrat, P.: Incremental Ensemble Learning for Electricity Load Forecasting. Acta Polytechnica Hungarica 13(2) (2016)
6. Shahzadeh, A., Khosravi, A., Nahavandi, S.: Improving load forecast accuracy by clustering consumers using smart meter data. In: Proceedings of IJCNN, (2015)
7. Wijaya, T. K., Vasirani, M., Humeau, S., Aberer, K.: Cluster-based Aggregate Forecasting for Residential Electricity Demand using Smart Meter Data. In: Proceedings of IEEE International Conference on Big Data, pp. 879–887 (2015)
8. Arthur, D., Vassilvitskii, S.: K-means++: The Advantages of Careful Seeding. In: SODA '07, pp. 1027–035 (2007)
9. Davies, D. L., Bouldin, D. W.: A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence 2, pp. 224–227 (1979)
10. Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J: Classification and Regression Trees. Chapman and Hall/CRC, Wadsworth (1984)
11. Strasser, H., Weber, Ch.: On the asymptotic theory of permutation statistics. Mathematical Methods of Statistics 8, pp. 220–250 (1999)
12. Cleveland, R. B. et al.: Seasonal-Trend Decomposition Procedure based on LOESS. J. Official Stat. 6, pp. 3–73 (1990)
13. Box, G. E. P., Jenkins, G. M.: Time Series Analysis: Forecasting and Control, Holden-Day (1970)
14. Holt, C. C.: Forecasting seasonals and trends by exponentially weighted moving averages. ONR Research Memorandum, Carnegie Inst. of Tech. 52, (1957)
15. Breiman, L.: Bagging Predictors. Machine Learning 24(2), pp. 123–140 (1996)
16. Bergmeir, Ch., Hyndman, R. J., Bentez, J. M.: Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation. International Journal of Forecasting 32(2), pp. 303–312 (2016)
17. Ester, M., Kriegel, H. P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the KDD, pp. 226–231 (1996)
18. Ankerst, M., Breunig, M. M., Kriegel, H. P., Sander, J.: OPTICS: ordering points to identify the clustering structure. In: Proc. of ACM SIGMOD, pp. 49–60 (1999)