Cluster Analysis of Smart Metering Data

An Implementation in Practice

Utilities and electricity retailers can benefit from the introduction of smart meter technology through process and service innovation. In order to offer customer specific services, smart meter mass data has to be analyzed. In the article we show how to integrate cluster analysis in a business Intelligence environment and apply cluster analysis to real smart meter data to identify detailed customer clusters.

DOI 10.1007/s12599-011-0201-5

The Authors

Dipl.-Wi.-Ing. Christoph Flath (✉)
Dipl.-Wi.-Ing. David Nicolay
Dr. Tobias Conte
PD Dr. Clemens van Dinther
Dr. Lilia Filipova-Neumann
Research Center for Information Technology
Haid-und-Neu-Str. 10–14
76131 Karlsruhe
Germany
flath@fzi.de
david.nicolay@energy4u.org
conte@fzi.de
Clemens.VanDinther@kit.edu
filipova@fzi.de

Received: 2011-03-01
Accepted: 2011-10-31
Accepted after two revisions by Prof. Dr. Terzidis.
Published online: 2012-01-12

This article is also available in German in print and via http://www.wirtschaftsinformatik.de: Flath C, Nicolay D, Conte T, van Dinther C, Filipova-Neumann L (2011) Clusteranalyse von Smart-Meter-Daten. Eine praxisorientierte Umsetzung. WIRTSCHAFTSINFORMATIK. doi: 10.1007/s11576-011-0309-8.

1 Introduction

Since the beginning of 2010, the installation of smart electricity meters is required by law (EnWG § 21c subpar. 1 a, b, c, d) in Germany. The introduction of smart metering poses significant economic and technical challenges for utilities companies. Smart meters are more costly than traditional meters and moreover require large investments for the necessary IT and communication infrastructure. Electricity companies, therefore, aim to recapture these investments through new services, improved processes enabled by smart meters. Faruqui et al. (2010) remark that utilities can profit from improved grid operations, lower costs for obtaining meter readings or faster identification of outages and interruptions. In addition, the implementation of dynamic electricity rates enables utilities to influence their customer's consumption behavior. This unlocks the potential to greatly reduce peak loads.

Both the design of segment-specific rates as well as the forecasting ability of electricity companies can be supported through targeted analysis of load data. Smart metering data can facilitate a customer segmentation based on dynamic load patterns instead of mere load totals.

In the present work we describe the implementation and evaluation of an IT-solution for cluster analysis of load data which we developed and realized in cooperation with ENERGY4U GmbH[1] and Allgäuer Überlandwerke[2] (AÜW). In the following sections we develop a software artifact following the design science approach as outlined by Hevner et al. (2004). We first address the problem of relevance and the related literature. Subsequently we describe the technical and methodological realization as well as the evaluation using real data. We then outline an application of the clustering results to the design of segment-specific electricity rates. The final section provides an outlook on further application scenarios as well as open research questions.

2 Problem Relevance and Related Work

New regulatory requirements, a changing public opinion regarding the orientation of the energy system, technological change, and increasing resource scarcity are the main drivers of changes in the energy sector. Responsibilities and tasks that were traditionally performed by a single integrated company are increasingly split up and distributed. This unbundling and the development of new market roles changes the electricity value chain and creates opportunities for new market actors. The introduction of smart meters may result in another paradigm change – especially in the areas of customer and portfolio management.

In Germany electricity metering using 15-minute intervals, the so-called "registrierende Leistungsmessung" (RLM) already is the standard protocol for large (industrial) customers with annual consumption levels exceeding 100,000 kWh. For smaller customers, however, electricity consumption was metered on an annual basis. Since the beginning of 2010

[1] Energy4U is an IT consulting firm for SAP solutions in the utilities sector (http://www.energy4u.org).
[2] Allgäuer Überlandwerke is a regional utilities company based in Kempten (http://www.auew.de).

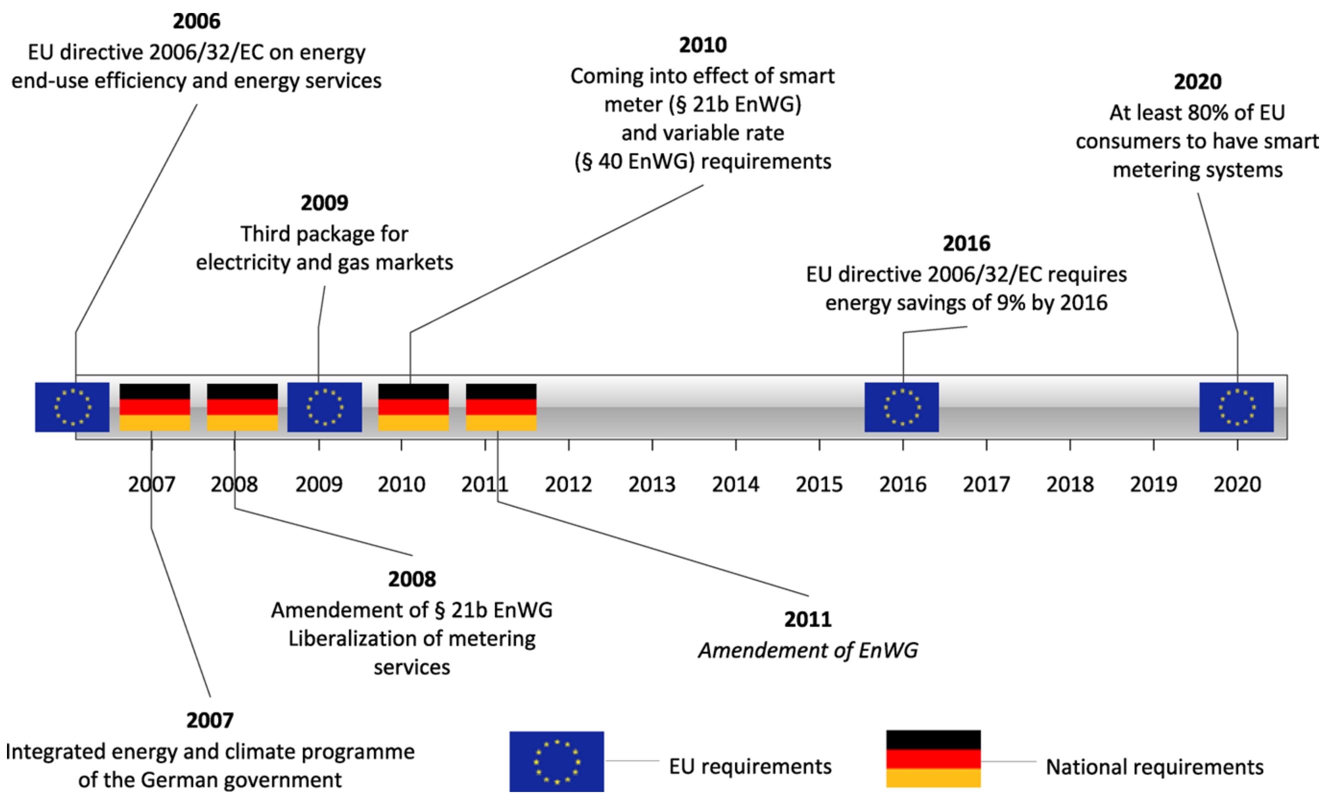

Fig. 1 European and German regulatory smart metering requirements

(cf. **Fig. 1**) the German Energy Industry Act (EnWG) requires metering service providers to install systems that are able to capture both the amount and the timing of electric energy consumption, i.e., smart meters. In its initial formulation this requirement applied only to new constructions and major renovations subject to technical and economic feasibility. The revised version of the EnWG from August 2011 extends this requirement to all buildings (§ 21c subpar. 1d EnWG). In addition, the revised law more explicitly defines the feasibility constraints (§ 21c subpar. 2 EnWG) as well as the required smart metering functionalities (§ 21d, e, i EnWG). These refinements help to meet prior criticism concerning the lack of clarity within the original formulations (cf. BNetzA 2010). Compared to the RLM-systems for large customers, the smart meters for small customers have to meet lower requirements concerning accuracy, data exchange frequencies, and completeness of measurement records (Nabe et al. 2009, p. 43).

Due to differing regulations on the national level, adoption of smart metering varies greatly within the EU. While Italy and Sweden already boast adoption rates of 86% respectively 59%, the introduction of smart meters in Germany is mostly confined to pilot regions (Jagstaidt et al. 2011).[3] The regulator's intention with the roll-out of smart meters and variable electricity rates is to achieve more transparency and efficiency in consumption.

Variable rates can be used to influence consumption behavior. This may provide utilities with significant potentials for cost reductions (Faruqui et al. 2010). To protect customers from intransparencies that may arise from the expected multitude of electricity rates, the EnWG strengthens customer rights: the change of suppliers is now to be completed within three weeks (§ 20a subpar. 2 EnWG), and § 40 requires electricity suppliers to provide their customers with concise billing documents. In particular suppliers are required to explain the relevant calculation factors and composition of prices in a lucid fashion, and the consumption profile of customers should be illustrated graphically. New data security and privacy requirements concerning collection, processing and usage of smart meter data are governed by § 21 g, h EnWG.

Given the very limited empirical values, the customer-specific design of such variable rates is a difficult challenge. The analysis of smart meter data is a potential key factor in successfully addressing this challenge. Systematical analysis requires the design and implementation of appropriate IT solutions. The aim of this paper is to develop a pilot instrument for customer clustering based on smart meter data. This could provide electricity suppliers with an in-depth overview of their customer portfolio and at the same time allow easy data interchange with other corporate IT systems like ERP or CRM solutions.

2.1 Business Intelligence and Knowledge Discovery

In practice, Business Intelligence (BI) has turned out to be the approach of choice to process large data sets into information and business-relevant knowledge (Cody et al. 2010). By means of IT-based data access and their IT-based analysis and processing, BI significantly sup-

[3]Darby (2006) as well as Burgess and Nye (2008) provide summaries of different international smart meter projects.

ports decision makers (Strauch and Winter 2002). BI itself is not bound to a specific system; in fact a diversity of information retrieval and analysis systems can be applied. Kemper et al. (2004) suggest a classification of these systems into ad hoc and model-driven systems. In this work, we focus on the model-driven analysis in line with Knowledge Discovery in Databases (KDD) as established by Fayyad et al. (1996).

KDD is the overall process consisting of data mining and the necessary preliminaries and reworking. Within this process, data mining is the actual tool with the aid of which previously unknown and potentially useful patterns can be retrieved and extracted (Han and Kamber 2006; Bissantz and Hagedorn 2009). The knowledge generated in this process can, for instance, be used in decision support systems (Kemper and Baars 2006). The Cross Industry Standard Process for Data Mining (CRISP-DM) is a procedural model that has established itself as the quasi standard for such tasks in the recent years (Kurgan and Musilek 2006; Shearer 2000). Here, the starting points for the data analysis are business problems and tasks which are systematically addressed in six process phases (Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment).

2.2 Data Mining in the Energy Industry

Since RLM is already common practice for major customers in the energy industry, the design of business processes for smaller smart metering customers and the usage of the associated large set of micro data represent a particular challenge. Performing a customer segmentation based on consumption data potentially reveals characteristic customer load profiles within the heterogeneous population.

The advancing liberalization of energy markets and the rising availability of consumption data have recently spotlighted data mining approaches to analyze load profile data. Pitt and Kitschen (1999) and Ramos and Vale (2008) study the applicability of different clustering techniques for the classification of large industrial clients' daily power consumption. Espinoza et al. (2005) consult the cluster analysis to scrutinize the daily and weekly load data delivered from transformer substations. Figueiredo et al. (2005) develop a procedural model to cluster and classify power customers which is then used for the analysis of customer data on the distribution network level. With respect to the consumption data, electricity providers can use the revealed information patterns to improve their business processes. For instance, Chicco et al. (2003) study the margin opportunity in the design of optimized tariffs for identified customer clusters. Espinoza et al. (2005) develop forecast models based on customer clusters on the distribution network.

Given the increasing adaption of smart grid technologies, it is hardly surprising that Keshav and Rosenberg (2011) as well as Ramos and Liu (2011) underline the central importance of data mining techniques for the transformation of the electricity system.

3 Implementation of the Cluster Analysis as a Design Science Artifact

As mentioned above, the use of data mining techniques to analyze load profiles is methodically sound and offers a variety of potentials for its application within the energy industry. In addition, its actual implementation within enterprises requires a suitable technical realization. In doing so, complex energy technology and utilization systems as well as the peculiarities of the smart metering landscape have to be taken into account. In the following, we will briefly describe the technical and methodical design decisions that underlie our cluster analysis.

3.1 System Environment Selection and Design of the Analysis Process

The cluster analysis was realized within the data warehouse software *SAP NetWeaverBusiness Intelligence* (SAP BI) because this system was already used by our project partner. However, this decision does not narrow the practical relevance of our work since SAP BI is applied in various mid- and large-sized enterprises in the electricity sector. SAP BI offers a platform to integrate, prepare, and analyze smart metering data. Moreover, the Analysis Process Designer (APD) and the Data Mining Workbench (DM Workbench) add up to a full data mining solution. The DM Workbench is utilized to create the models for the various data mining techniques (cluster analysis, ABC analysis, classification, etc.). In APD, the analysis process itself can be modeled and implemented.

In our case, a three-phase analysis process was used in accordance with the three to five phases of the CRISP-DM process model. As part of the data preparation, the input data are adjusted and converted into the appropriate format for subsequent analysis. In the subsequent analysis phase, a cluster analysis which is appropriate for the data is modeled and implemented. The results of the cluster analysis can then be evaluated with the help of an appropriate representation.

3.2 Data Preparation

Data preparation for the cluster analysis is a crucial step of the KDD process. First the integrity of the raw data has to be assured through filtering and substitute value creation. Thereafter, the data has to be transferred into the format that is needed for cluster analysis by means of suitable choice and normalization.

Data transmission in smart metering systems usually spans a multitude of technical components (e.g., meter, data concentrators, gateways, etc.). If one of these components fails, incorrect values or recording gaps are likely to occur. In order to interpolate time series of consumption data, literature proposes different approaches. For instance, Ramos and Vale (2008) consult artificial neural networks while Figueiredo et al. (2005) rely on a regression approach. Both approaches require comprehensive preliminary operations on the load data set which somewhat reduces their applicability to continuous data analyses. For the most part, the data set used in our work exhibits only singular incorrect values (the smallest time segment equals 15 minutes), therefore, we chose a simple linear interpolation approach for data cleansing. This lightweight approach achieves robust results for minor data gaps. However, it cannot reliably estimate major data gaps. Therefore, we discarded load data sets showing more than one hour of recording gaps.

In our work, the relative load development was determined as the basis for customer segmentation. Therefore, load profiles were normalized to a relative load profile since otherwise the absolute consumption level is likely to dominate the similarities of different load profiles. Based upon the normalized load

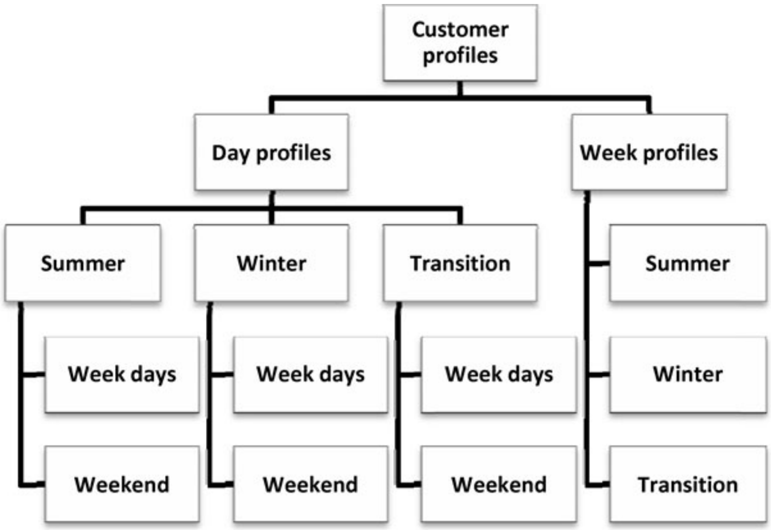

Fig. 2 Cluster analysis scenarios

curve, a characteristic load profile per time period (one day or one week) can be determined for each customer. The arithmetic means of the single time segments are the starting point for the formation of the characteristic load profiles.[4] In case of volatile consumption patterns, the arithmetic mean does not properly represent the customers' behavior. Academia proposes two approaches to cope with this issue. On the one hand, Räsänen and Kolehmainen (2009) extend the clustering objects by additional data points (e.g., standard deviation or skewness of the 15 minutes consumption). On the other hand, Ramos and Vale (2008) introduce a segmentation of the analysis days (weekday, season) which separates the raw data sets into more homogeneous subsets. As the load profiles with 15-minutes segments already give rise to very large data objects, we follow the segmentation approach proposed by Ramos and Vale (2008) in order to sustain scalability. For the data preparation, we identify the characteristic load profiles for nine "cluster scenarios", assembled from cluster type (week, weekday, weekend) and season (summer, winter, transitional). The described segmentation is illustrated in **Fig. 2**. This approach allows for a simple benchmarking with the standard load profiles utilized by the German energy industry which distinguish between similar day types.

To round off the data preparation phase, the clustering data can be enriched with further information available in other IT systems (e.g., ERP, CRM, or GIS data). Such a linkage of data can guide and support the subsequent analysis. In this work we do not go into details regarding this refining step.

3.3 Modeling and Implementation of the Cluster Analysis

The cluster analysis mainly aims at discovering structures in large data sets. According to Rodrigues et al. (2003), the *k*-means algorithm and a combination of *k*-means and artificial neural networks are suitable approaches for the clustering of load profiles. Both approaches achieve a similar clustering performance in the handling of customer load profiles. The *k*-means algorithm is considered to be the best known and most frequently applied partitioning clustering technique (Vercellis 2009) which is implemented as a standard in most of the established data mining software (also in SAP BI). Hence, we focus on the *k*-means algorithm in the following. The *k*-means algorithm works iteratively; it divides the data set into *k* clusters by minimizing the sum of all distances to the respective cluster centers. The algorithm does not guarantee to find a global optimum (Beringer 2008). Hence it is important to begin with various starting values for the cluster allocation in order to come up with a reasonably good result. Further, the choice of the number of clusters k is crucial for the algorithm's performance. The "right" number of clusters is not known ex ante; therefore, the cluster analysis is initially conducted with all cluster numbers to be considered. Subsequently, the results of the various starting values and the number of clusters are rated with respect to their performance in order to identify the best clustering.

In order to numerically rate the cluster quality, literature proposes, for instance, the Clustering Dispersion Indicator and the Mean Index Adequacy Indicator (Ramos and Vale 2008; Rodrigues et al. 2003). Both indicators have in common that they monotonically increase with the number of clusters. Thus, they are not suitable for a simultaneous determination of the optimal number of clusters. Davies and Bouldin (1979) introduced an index without this property: It rates the cluster quality based on the sum of the relations between the variance within two clusters and their centers' distances. The identification of local index minimums thus allows for the determination of a suitable number of clusters. For the search, only the basic domain needs to be indicated – which is typically given by the application context. That way, it is possible that different clustering can be used in different application contexts (tariff design, load forecast). In this work's technical realization, the Davies Bouldin Index has been modeled and integrated in APD.[5]

3.4 Evaluation and Visualization of Clustering Results

After selecting the final cluster specification, the obtained insights need to be processed for business applications. This is realized through appropriate reporting as well as automatic exchange with other information systems, e.g., decision support systems or CRM systems. We realize the visualization of results using the SAP analysis- and reporting tool *BEx Analyzer* which is embedded in Microsoft Excel. The *BEx analyzer* provides capabilities for both graphical and tabular reporting.

Figures 3 and **4** show that our clustering approach facilitates a homogeneous customer segmentation based on load

[4]In order to limit the size of the data objects, we enlarged the analysis interval to two hours for all week scenarios. Thus, for the latter, the characteristic load profiles are subdivided into two hour segments.

[5]Moreover, we included the Silhouette Index (Rousseeuw 1987) and the index proposed by Dunn (1974) as additional rating indicators.

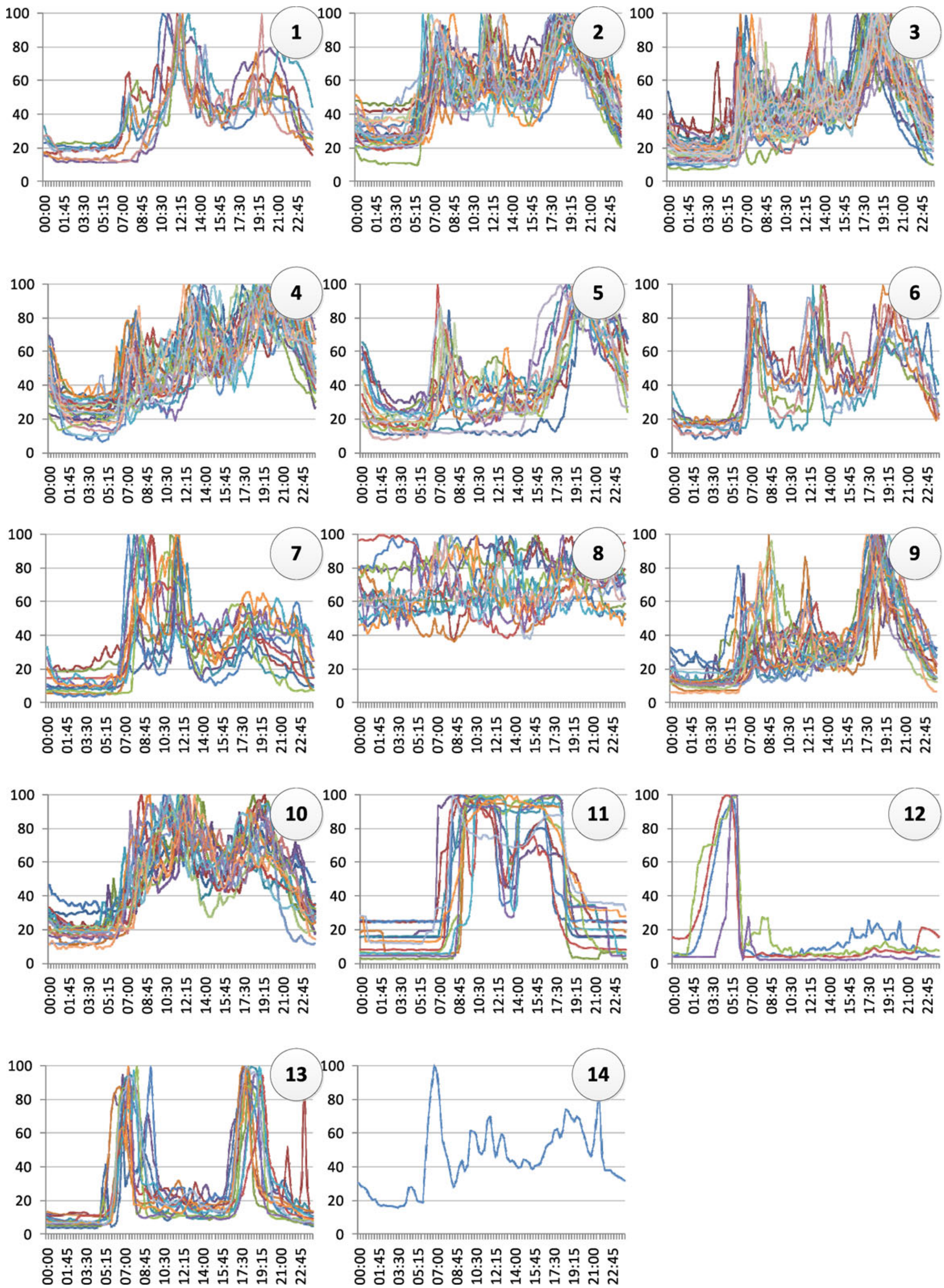

Fig. 3 Clustering results – day profiles (workdays, winter)

profiles. We can identify customer groups with distinct load profiles. In the analysis scenario weekdays (winter), we analyzed data from 215 customers which were segmented into 14 clusters. In contrast the scenario weekend days (winter) yielded only 10 clusters. The differing number of clusters indicates a lower diversity of weekend load profiles. Some of the profiles (e.g., cluster 11) only exist during the week which is typical for enterprises. Moreover, we can see a convergence of household consumption behavior on weekend days. This confirms the scenario-based clustering approach. Another characteristic difference between weekday and weekend clusters is the consumption distribution

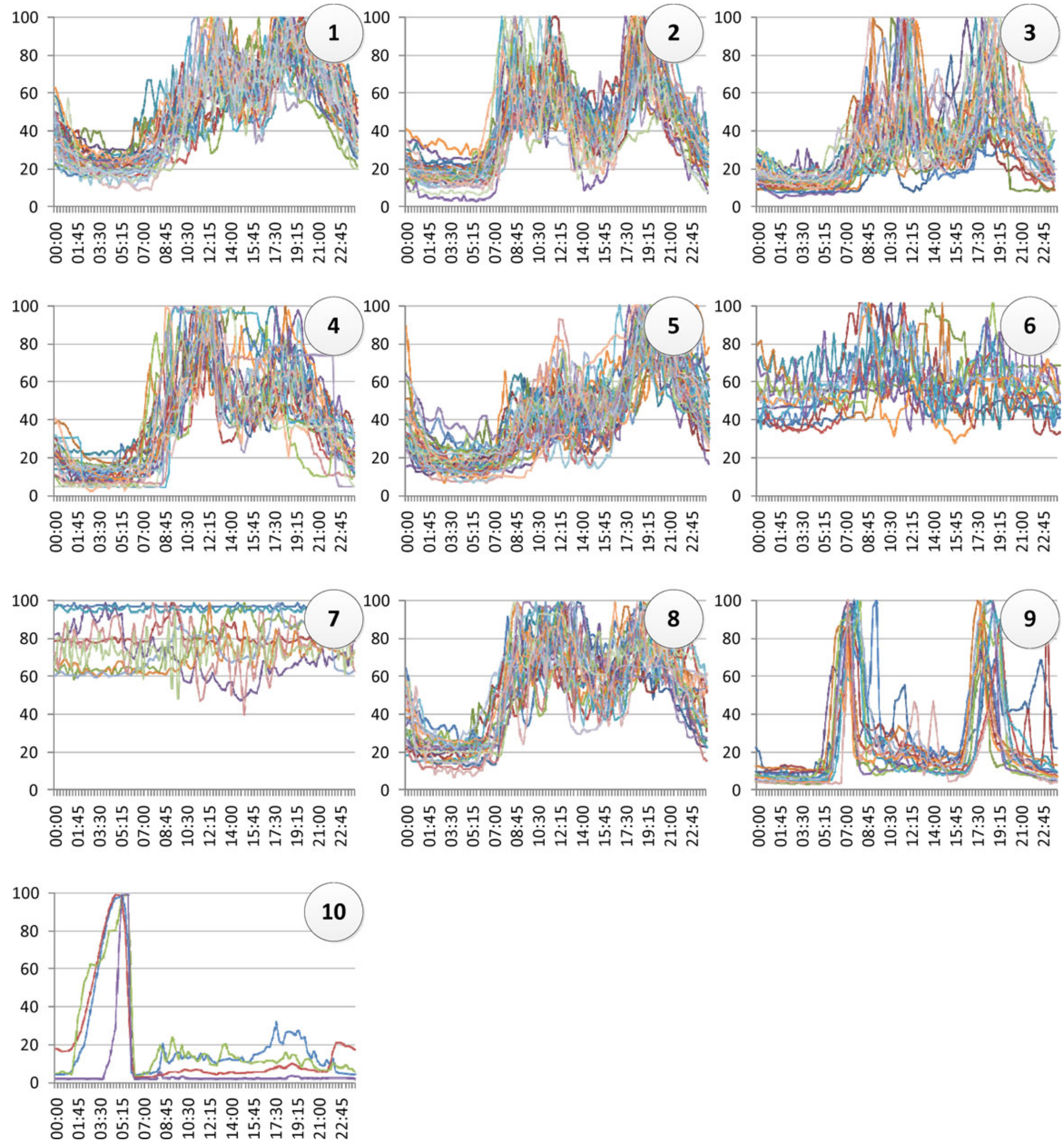

Fig. 4 Clustering results – day profiles (weekend, winter)

over the day. Weekdays exhibit a large share of consumption on mornings and evenings while weekends exhibit a more homogeneous consumption distribution.

As the clustering is solely based on consumption data, the question arises if conclusions regarding customer types are possible. The German standard load profiles[6] provide a natural benchmark for this type of evaluation. The clusters 11, 12, and 13 show distinct similarity with the standard load profiles for businesses (11), heat pumps (12), and farms (13). These assumptions are also supported by the absolute consumption levels. On the other hand the clusters 1–6, 10, and 14 with average annual consumption of about 3,600 kWh are most likely exclusively composed of private household customers. The random variation in the load profiles of these clusters provides further evidence for this assumption. It is the result of the aggregation of slightly differing daily load profiles to a representative household load profile, a characteristic, which can typically not be expected for industrial consumers (Müller 2010). The diversity of these household clusters and their large difference to the German standard household profile H0 show that cluster analysis of smart metering data can facilitate a more granular customer characterization. For instance cluster 5 exhibits a comparatively high consumption share during evening and early night hours. In view of the similarity with cluster 9, the challenge of determining a clear distinction of such micro-segments becomes clear. Adding information from other systems to the consumption data may enable an even better differentiation of these clusters. For clusters 7, 8, and 9 the type assessment is less clear, suggesting that they are most likely composed of different customer types.

[6] These profiles are available from the German Federal Association of Energy and Water Industries (BDEW).

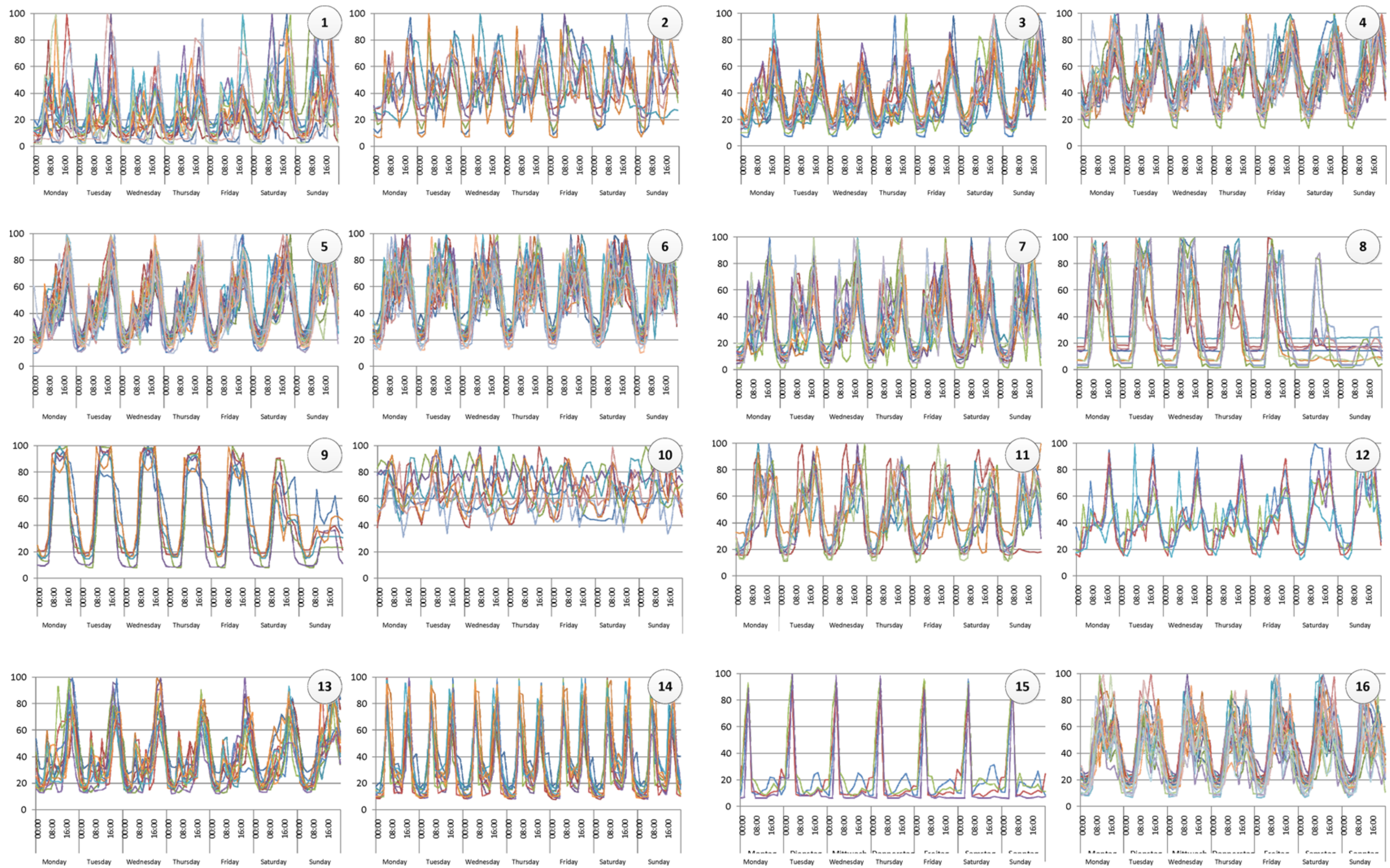

Fig. 5 Clustering results – week profiles (winter)

Abstract

Christoph Flath, David Nicolay, Tobias Conte, Clemens van Dinther, Lilia Filipova-Neumann

Cluster Analysis of Smart Metering Data

An Implementation in Practice

The introduction of smart meter technology is a great challenge for the German energy industry. It requires not only large investments in the communication and metering infrastructure, but also a redesign of traditional business processes. The newly incurring costs cannot be fully passed on to the end customers. One option to counterbalance these expenses is to exploit the newly generated smart metering data for the creation of new services and improved processes. For instance, performing a cluster analysis of smart metering data focused on the customers' time-based consumption behavior allows for a detailed customer segmentation. In the article we present a cluster analysis performed on real-world consumption data from a smart meter project conducted by a German regional utilities company. We show how to integrate a cluster analysis approach into a business intelligence environment and evaluate this artifact as defined by design science. We discuss the results of the cluster analysis and highlight options to apply them to segment-specific tariff design.

Keywords: Smart metering, Energy industry, Data mining, Business intelligence, Customer segmentation, Tariff design

The week (winter) load profile scenario yields a grouping with 16 clusters (cf. **Fig. 5**). Due to their repetitive structure, a graphical distinction between the different clusters is more difficult than in the case of single day load profiles. However, we can again spot distinct patterns (especially clusters 10, 13, and 15). At first glance the consumption ratios between different days as well as the periodicity of the load development seem fairly homogeneous. The main exceptions are clusters 8 and 9 which exhibit distinctly lower consumption on weekends with Sunday consumption levels being even lower than on Saturday.

4 Application of cluster results for segment-specific rate design

The economic benefits of time-variable electricity rates consist in the potential for utilities to improve price discrimination and to facilitate the reduction of peak loads. Their importance is described, among others, by Troxel (1938), Eckel (1985), Stephenson et al. (2001), and Parmesano (2007). However, there are only few approaches for segment-specific design based on customer load profile clusters of such rates (e.g., Chicco et al. 2003). Segment-specific rate design determines a time-variable rate for each individual customer segment. This allows consideration of consumption heterogeneity but also avoids the complexity of rate design on the individual customer level. The design of these rates requires setting the relevant endogenous parameters for each customer segment depending on its load curve. This encompasses the following aspects:

1. Determining the number of time zones:

   First the number of peaks and valleys in the load profile is determined. The time zones should correspond to the existing peaks and off-peak hours in the profile. Customer acceptance issues, which can possibly limit the number of time zones, will be considered for the decision.
2. Determining the start times of each time zones:

   Here, the difference between the average consumption in adjacent peaks and load time zones is to be maximized.
3. Determining the number of price zones:

   These will correspond to the number of time zones as long as it is relatively small. For acceptance reasons, it is conceivable to set the same price level for multiple time zones in cases with a relatively high number of time zones. Here, the maximum allowable number of price levels is to be determined applying the criterion of customer acceptance.
4. Modeling and maximizing supplier profit:

   The price elasticity values for the consumption are to be derived from recent field tests. Historical procurement costs in profit calculations are explicitly taken into account in order to consider any recurring peaks and valleys in the procurement cost.

While the introduction of different analysis scenarios in Sect. 3.2 increases cluster homogeneity, it also gives rise to a multitude of cluster combinations that may serve as a basis for rate design. Therefore, a central challenge in this process is the selection of an appropriate base scenario for determining the most relevant clustering. However, the cluster analysis in Sect. 3 indicated that across different analysis scenarios the different cluster populations are made up of varying customer compositions. Therefore, a stochastic assessment using cluster assignment probabilities is required. Initial results gathered in the context of this cluster analysis indicate that this is a promising approach that needs to be validated using a larger data set.

By influencing the demand side, segment-specific electricity rates can improve the energy system's efficiency. Moreover, time-differentiated billing allows greater transparency with respect to the impact of generation costs for the formation of retail electricity prices. Increased coupling between costs and prices increases the fairness of retail electricity pricing (Faruqui 2010).

5 Summary and Conclusion

The broad roll-out of smart meters poses significant challenges to electricity utilities. On the one hand, large investments in the metering infrastructure are required and on the other hand traditional business processes need to be redesigned. The processing and utilization of customer consumption data from smart meters may enable electricity companies to achieve large efficiency gains. In our work we have presented an implementation and evaluation of a cluster analysis approach for smart meter data within a business intelligence environment. The identified clusters are plausible and yield a true information gain for the utility. Given a direct integration with existing IT systems (e.g., ERP or CRM), this approach may be relevant for metering service companies as well as utilities. Based on the cluster analysis, metering service companies can offer innovative service products like energy management planning or regional load profiles. Suppliers can profit from the possibility of designing segment-specific rates which allow a better integration of the demand side into the control of the electricity system.

Our electricity load data analysis gives rise to subsequent research questions. We will need to validate the robustness of the identified clusters using a broader data set. Moreover, the integration of additional data sources, such as current rate information, industry code, or household properties such as demographic and socio-economic data can further support effective clustering. Following Newsham and Bowker (2010) there still is an insufficient understanding how the willingness of households to adapt their load behavior corresponds with their socio-economic properties. As this is a central question for future demand-centered control paradigms, research needs to focus on questions like assessing demand elasticity and load shifting potentials.

Acknowledgements

We especially want to thank Florian Schlüfter (Energy4U GmbH) and Stefan Mayer (Allgäuer Überlandwerke) for the excellent cooperation and support as well as for providing the smart grid data set.

References

BNetzA (2010) Bericht der Bundesnetzagentur zum Zähl- und Messwesen im Auftrag des Bundesministerium für Wirtschaft und Technologie. Wettbewerbliche Entwicklungen und Handlungsoptionen im Bereich Zähl- und Messwesen und bei variablen Tarifen

Beringer J (2008) Online-Data-Mining auf Datenströmen: methoden zur Clusteranalyse und Klassifikation. VDM, Saarbrücken

Bissantz N, Hagedorn J (2009) Data mining (Datenmustererkennung). WIRTSCHAFTSINFORMATIK 51(1):139–144

Burgess J, Nye M (2008) Re-materializing energy use through transparent monitoring systems. Energy Policy 36(12):4454–4459

Chicco G, Napoli R, Postolache P, Scutariu M, Toader C (2003) Customer characterization options for improving the tariff offer. IEEE Transactions on Power Systems 18(1):381–387

Cody W, Kreulen J, Krishna V, Spangler W (2010) The integration of business intelligence and knowledge management. IBM Systems Journal 41(4):697–713

Davies D, Bouldin D (1979) A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence 1(2):224–227

Darby S (2006) The effectiveness of feedback on energy consumption – a review for DEFRA of the literature on metering, billing and direct displays. Environmental Change Institute, University of Oxford

Dunn JC (1974) Well-separated clusters and optimal fuzzy partitions. Cybernetics and Systems 4(1):95–104

Eckel C (1985) A general model of customer-class pricing. Economics Letters 17(3):285–289

Espinoza M, Joye C, Belmans R, DeMoor B (2005) Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. IEEE Transactions on Power Systems 20(3):1622–1630

Faruqui A (2010) The ethics of dynamic pricing. The Electricity Journal 23(6):13–27

Faruqui A, Harris D, Hledik R (2010) Unlocking the €53 billion savings from smart meters in the EU: how increasing the adoption of dynamic tariffs could make or break the EU's smart grid investment. Energy Policy 38(10):6222–6231

Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (1996) Advances in knowledge discovery and data mining. AAAI Press/MIT Press, Menlo Park/Cambridge

Figueiredo V, Rodrigues F, Vale Z, Gouveia J (2005) An electric energy consumer characterization framework based on data mining techniques. IEEE Transactions on Power Systems 20(2):596–602

Han J, Kamber M (2006) Data mining: concepts and techniques. Morgan Kaufmann, San Francisco

Hevner A, March S, Park J, Ram S (2004) Design science in information systems research. MIS Quarterly 28(1):75–105

Jagstaidt UCC, Kossahl J, Kolbe LM (2011) Smart metering information management. WIRTSCHAFTSINFORMATIK 53(5):313–317

Kemper H, Baars H (2006) Business Intelligence und Competitive Intelligence. HMD – Praxis der Wirtschaftsinformatik 43(247):7–20

Kemper H, Mehanna W, Unger C (2004) Business Intelligence. Grundlagen und praktische Anwendungen. Eine Einführung in die IT-basierte Managementunterstützung. Vieweg, Wiesbaden

Keshav S, Rosenberg C (2011) How internet concepts and technologies can help green and smarten the electrical grid. ACM SIGCOMM Computer Communication Review 41(1):109–114

Kurgan La, Musilek P (2006) A survey of knowledge discovery and data mining process models. The Knowledge Engineering Review 21(1):1–24

Müller K (2010) Gewinnung von Verhaltensprofilen am intelligenten Stromzähler. Datenschutz und Datensicherheit – DuD 34(6):359–364

Nabe C, Beyer C, Brodersen N, Schäffler H, Adam D, Heinemann C, Tusch T, Eder J, de Wyl C, Vom Wege J, Mühe, S (2009) Ökonomische und technische Aspekte eines flächendeckenden Rollouts intelligenter Zähler, Ecofys im Auftrag der Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen

Newsham G, Bowker B (2010) The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: a review. Energy Policy 38(7):3289–3296

Parmesano H (2007) Rate design is the No 1 energy efficiency tool. The Electricity Journal 20(6):18–25

Pitt BD, Kitschen DS (1999) Application of data mining techniques to load profiling. In: Proc 21st IEEE international conference on power industry computer applications (PICA '99), pp 131–135

Ramos C, Liu CC (2011) AI in power systems and energy markets. IEEE Intelligent Systems 26(2):5–8

Ramos S, Vale Z (2008) Data mining techniques to support the classification of MV electricity customers. In: IEEE power and energy society general meeting – conversion and delivery of electrical energy in the 21st century, pp 1–7

Räsänen T, Kolehmainen M (2009) Feature-based clustering for electricity use time series data. In: Proc adaptive and natural computing algorithms, pp 401–412

Rodrigues F, Duarte J, Figueiredo V, Vale Z, Cordeiro M (2003) A comparative analysis of clustering algorithms applied to load profiling. In: Proc of MLDM 2003, Leipzig, Germany, pp 73–85

Rousseeuw P (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20(11):53–65

Shearer C (2000) The CRISP-DM model: The new blueprint for data mining. Journal of Data Warehousing 5(4):13–22

Stephenson P, Lungu I, Paun M, Silvas I, Tupu G (2001) Tariff development for consumer groups in internal European electricity markets. In: Electricity Distribution, 2001. Part 1: Contributions. CIRED. 16th international conference and exhibition on electricity distribution (IEE Conf Publ No 482). IET

Strauch B, Winter R (2002) Vorgehensmodell für die Informationsbedarfsanalyse im Data Warehousing. In: Maur E, Winter R (eds) Vom Data Warehouse zum Corporate Knowledge Center. Physica, Heidelberg, pp 359–378

Troxel C (1938) Class prices for gas and electricity. The American Economic Review 28(2):275–289

Vercellis C (2009) Business intelligence: data mining and optimization for decision making. Wiley, New York