
Clustering-Ansatz zur Erstellung von Lastprofilen zur Vorhersage des Stromverbrauchs

Ein Modell zur Erstellung von Lastprofilen durch Anwendung von Fuzzy-C-Means auf Smart-Meter-Zeitreihen

Christian Bock

Heinrich-Heine-Universität
Institut für Informatik
Universitätsstraße 1
40225 Düsseldorf, Deutschland
bock@cs.uni-duesseldorf.de

BTU EVU
Beratung GmbH
Leostraße 31
40545 Düsseldorf, Deutschland

ABSTRACT

In dieser Arbeit stellen wir einen praxisorientierten Ansatz zur Erstellung von *Lastprofilen* zur Vorhersage des Energieverbrauchs anhand von *Smart-Metering-Zeitreihen* vor. Die prinzipielle Idee ist es, historische Verbrauchszeitreihen mittels *Fuzzy-Clustering* in Gruppen zu unterteilen. Diese Einteilung in Gruppen ermöglicht es Energiewirtschaftsunternehmen Kunden mit jeweils ähnlichem Verbrauchsverhalten zu identifizieren und darauf basierend den Einkauf von Energiemengen entsprechend zu planen. Die Besonderheit dieses Ansatzes ist es, dass auch energiewirtschaftspezifisches Anwendungswissen bei der Erstellung und Auswertung der gewonnenen Lastprofile berücksichtigt wird; so ist beispielsweise das Konsumverhalten von Endverbrauchern an lokalen Feiertagen unter Umständen gänzlich anders als an Werktagen. Der hier vorgestellte Ansatz identifiziert daher nicht nur unterschiedliche Kundengruppen sowie deren typisches Verbrauchsverhalten, sondern unterteilt das Verbrauchsverhalten darüber hinaus auch zeitlich. Die durch diese Unterteilung gewonnenen Lastprofile haben dadurch ein Format, das sich stark an den bei den Stadtwerken eingesetzten Lastprofilen orientiert und eignen sich daher besonders gut für den Einsatz in deren Geschäftsprozessen oder denen entsprechender Dienstleister.

Categories and Subject Descriptors

G.3 [Probability And Statistics]: Time series analysis; H.2.8 [Database Management]: Database Applications—*Data mining*; H.3.3 [Information Storage And Retrieval]: Information Search and Retrieval—*Clustering*; H.4 [Information Systems Applications]: Miscellaneous; J.1 [Administrative Data Processing]: Business

Keywords

Big Data, Data Mining, Knowledge Discovery, Clustering, Time Series, Smart Metering, Databases

1. EINLEITUNG

Für Betreiber von Stadtwerken ist die korrekte Bilanzierung von Energiedaten von essentieller Wichtigkeit für die Stabilität der Energieversorgung. Die Herausforderung von Stadtwerken besteht dabei in der Notwendigkeit, den Energieverbrauch ihrer Kunden bereits im Voraus zu kennen oder zumindest abschätzen zu können; dies ermöglicht es den Stadtwerken, den Strombedarf frühzeitig bei Stromerzeugern anzumelden, wodurch diese die erforderlichen Kapazitäten vorhalten können, sobald diese benötigt werden. Die Abschätzung des Stromverbrauchs erfolgt dabei mittels so genannter *Lastprofile* (siehe Abschnitt 3). Üblicherweise setzen Stadtwerke hierbei die *Standardlastprofile* ein, welche vom *Bundesverband der Energie- und Wasserwirtschaft (BDEW)* bereitgestellt werden. Diese Standardlastprofile wurden in den 1990er Jahren anhand von Messungen des damaligen *Verbands der Elektrizitätswirtschaft (VDEW)* erzeugt. Hieraus ergibt sich jedoch das Problem, dass die Profile aufgrund mangelnder Anpassung an aktuelle Entwicklungen zunehmend schlechter als Modell zur Prognose des Energieverbrauchs geeignet sind [27]. Durch die Liberalisierung des Strom- und Gasmarktes und dem dadurch entstandenen Konkurrenzdruck zwischen den Energieversorgungsunternehmen, sowie durch den zunehmenden Einsatz von Technologien wie *Smart Metering*, wurden Innovationen, welche eine höhere Effizienz bei gleichzeitig niedrigerem Risiko erlauben, sowohl ermöglicht als auch notwendig [15]. Im Nachfolgenden stellen wir daher einen Ansatz vor, welcher *Fuzzy-Clustering* [6] verwendet, um optimale Lastprofile anhand von *Smart-Metering-Zeitreihen* zu gewinnen.

2. VERWANDTE ARBEITEN

Mit dem zunehmenden Umfang, in dem im Verlauf der letzten Jahrzehnte Daten auch im Bereich der Energiewirtschaft computergestützt erhoben, verarbeitet und ausgewertet werden können, haben parallel dazu auch entsprechende Forschungsarbeiten an Bedeutung gewonnen [18, 24, 9]. [1, 21] zeigen beispielsweise Methoden zur Verbrauchsprognose auf. Weitere Anwendungsgebiete beinhalten unter Anderem

^{28th} GI-Workshop on Foundations of Databases (Grundlagen von Datenbanken), 24.05.2016 - 27.05.2016, Nörten-Hardenberg, Germany. Copyright is held by the author/owner(s).

Ausreißererkennung [29] sowie Marketing und Tarifoptimierung [8, 20]. In vielen Fällen, in denen dabei Clustering zur Datenanalyse zum Einsatz kommt, wird dabei auf den *K-Means*-Algorithmus [4, 11, 16, 25, 22] oder darauf aufbauende Verfahren [19] zurückgegriffen; zunehmend kommt aber auch *Fuzzy-Clustering* zur Anwendung [29, 18, 26]. Neue Möglichkeiten der Datenanalyse ergeben sich zudem durch den stetig steigenden Ausbau des *Smart Metering* [3, 12].

In dem in dieser Arbeit vorgestellten Verfahren verwenden wir einen *Fuzzy-Clustering*-Ansatz [6], um *Lastprofile* aus einem Datensatz von *Smart-Metering*-Zeitreihen zu extrahieren. Diese wiederum eignen sich als Modell zur Prognose des Energieverbrauchs. Als Neuerung werden hierbei jedoch nicht nur *Kundengruppen* identifiziert, sondern auch *Typtage*. Dies ermöglicht nicht nur eine praxisbezogenere Darstellung der Lastprofile, da die Kombination aus Kundengruppe und Typtag sehr gut von den heutigen IT-Systemen der Stadtwerke verarbeitbar ist, sondern auch eine realistischere Modellierung des Verbrauchsverhaltens.

3. STRUKTURELLER AUFBAU VON LASTPROFILIEN

Die Prognose des Energiebedarfs wird von Stadtwerken typischerweise anhand von *Lastprofilen* durchgeführt: hierzu werden zunächst alle Kalendertage in eine Menge von *Typtagen* unterteilt. Ein *Typtag* im Sinne der Energiewirtschaft ist hierbei eine Menge von Kalendertagen, an denen sich das Verbrauchsverhalten der Kunden signifikant von dem anderer Typtage unterscheidet. Die von der *BDEW* empfohlene Einteilung besteht dabei aus 9 Typtagen, nämlich jeweils *Werktag*, *Samstag* sowie *Sonn- und Feiertag* für drei festgelegte Jahreszeiten *Sommer*, *Winter* und *Übergangszeit*. Diese Einteilung lässt sich leicht als Modell eines typischen Wochenablaufs nachvollziehen: so entspricht es beispielsweise dem intuitiven Empfinden, dass ein durchschnittlicher Haushalt tagsüber an einem Werktag aufgrund von Berufstätigkeit ein grundsätzlichlich anderes Verbrauchsverhalten aufweist gegenüber einem Wochenendtag.

Darüber hinaus werden die einzelnen Kunden in Kundengruppen eingeteilt, beispielsweise *Haushalt*, *Landwirtschaftsbetriebe allgemein*, *Wochenendbetrieb* und weitere. Jeder solchen Kundengruppe wird für jeden Typtag ein Verbrauchsmuster in Form einer Zeitreihe zugewiesen, welche das für die jeweilige Kundengruppe typischste Verbrauchsverhalten an dem jeweiligen Typtag beschreiben. Als Beispiel zeigt Abbildung 1 den Lastprofilverlauf des Haushaltsprofils (*H0*) für die Sommersaison. Die jeweiligen Verbrauchsmuster bieten dabei eine zeitliche Auflösung von 15 Minuten und sind normiert, das heißt sie stellen lediglich den relativen Energieverbrauch der einzelnen Kunden dar. Mithilfe des bei den Endverbrauchern im Rahmen der jährlichen Stromzählerablesung erhobenen Jahresverbrauchs, welcher die Grundlage für die *Jahresverbrauchsprognose (JVP)* des Kunden für das Folgejahr bildet, und des für das jeweilige Profil spezifischen *Normierungsfaktors*, lassen sich die Lastprofile für jeden Kunden individuell skalieren. Somit ermöglicht die Kenntnis der Jahresverbrauchsprognose mithilfe der Lastprofile für jeden Verbraucher eine viertelstundenscharfe Prognose des Energiebedarfs für das komplette Jahr.

Aufgrund der Unterteilung der Energieverbräuche in zuvor festgelegte *Typtage* und *Kundengruppen* stellen Lastprofile ein sehr einfaches Prognosemodell dar. Je zuver-

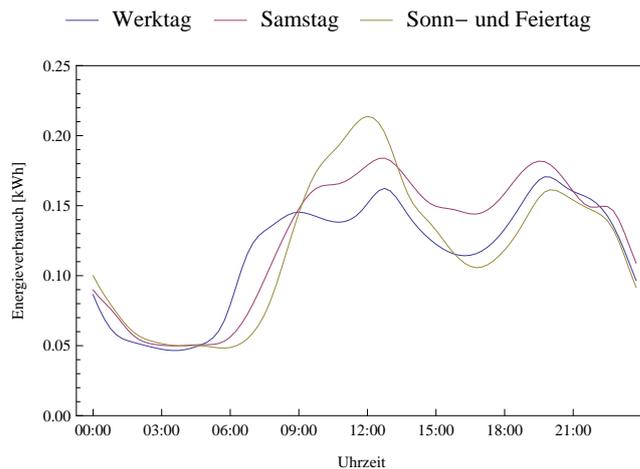


Figure 1: Übersicht des VDEW-Standardlastprofils *H0* (Haushaltsprofil), gültig für Sommermonate. Die Zeitreihen sind normiert und stellen das repräsentative Verbrauchsverhalten jeweils für die Typtage *Werktag*, *Samstag* sowie *Sonn- und Feiertag* dar.

lässiger und präziser die verwendeten Lastprofile dabei jedoch den tatsächlichen Energiebedarf der Kunden modellieren, desto höher ist die Planungssicherheit der Stadtwerke beim Einkauf des Stroms von den Kraftwerken. Hierbei wird nicht vorausgesetzt, dass sämtliche einer bestimmten Kundengruppe zugewiesenen Verbraucher an gleichen Typtagen stets das gleiche Verbrauchsmuster aufweisen. Dies ist beispielsweise aufgrund von für Stadtwerke unvorhersehbaren, individuellen Urlaubsreisen einzelner Endverbraucher nicht realisierbar. Stattdessen nimmt man an, dass die der jeweiligen Kundengruppe zugewiesenen Verbraucher als Ganzes besagtes Verbrauchsmuster aufweisen. Aufgrund dieser Eigenschaft eignen sich *Clustering*-Ansätze, um Lastprofile basierend auf bekannten Energieverbrauchszeitreihen zu erzeugen. Im nachfolgenden Abschnitt stellen wir ein solches Verfahren vor.

4. ANSATZ ZUM ERZEUGEN VON LASTPROFILIEN

Für die Zusammenstellung der Lastprofile aus *Smart-Meter*-Zeitreihen verwenden wir ein dreistufiges Verfahren:

- Im ersten Schritt bestimmen wir die optimale Anzahl an Typtagen sowie deren Aufteilung auf die einzelnen Kalendertage
- Im zweiten Schritt bestimmen wir für jeden Typtag die optimale Anzahl an Verbrauchsmustern sowie deren konkrete Ausprägungen
- Im dritten Schritt werden die Lastprofile aus den zuvor ermittelten Daten zusammengestellt und die Messstellen einem Profil zugewiesen

4.1 Bestimmung der Typtage

Zur Bestimmung von Typtagen ist es wichtig zu wissen, ob und wie stark sich die Energieverbräuche pro Tag als Ganzes voneinander unterscheiden. Diese Betrachtung muss folgende Eigenschaften aufweisen:

- Sie muss unabhängig von der Anzahl der Messstellen sein
- Sie muss die *Form* der Verbrauchszeitreihe berücksichtigen, nicht aber die *Menge* der Energie

Der erste Punkt ist intuitiv leicht verständlich. Beim zweiten Punkt hingegen fließt energiewirtschaftsspezifisches Anwendungswissen mit ein: so werden Lastprofile seitens Stadtwerken ausschließlich dazu verwendet, um zwischen verschiedenen Arten von Verbrauchsmustern bei Kunden zu unterscheiden; eine Skalierung des vom Lastprofil prognostizierten Tagesverbrauchs erfolgt anhand der *Jahresverbrauchsprognose* des jeweiligen Kunden. Diese errechnet sich aus der jährlich von den Stadtwerken durchgeführten Ablesung des Stromzählers und wird hier nicht weiter betrachtet.

Um die oben genannten Punkte bei der Identifikation der Typtage zu berücksichtigen, konstruieren wir zunächst aus den Smart-Meter-Zeitreihen $S_i, 1 \leq i \leq N$ aller N Kunden mit den Messwerten $s_{i,j}$ für die Zeitpunkte $t_j, 1 \leq j \leq T$ eine neue Zeitreihe $X = \{x_1, x_2, \dots, x_T\}$ wie folgt:

$$x_j = \frac{1}{N} \sum_{i=1}^N \frac{s_{i,j}}{\sum_{j'=1}^T s_{i,j'}} \quad (1)$$

Hierbei ist $\sum_{j'=1}^T s_{i,j'}$ ein zeitreihenspezifischer Normierungsfaktor. Dieser ermöglicht es beim Clustering ausschließlich die *Form* des Verbraucherverhaltens zu berücksichtigen, wodurch Gering- und Vielverbraucher, deren Konsumverhalten sich bis auf einen skalaren Faktor sehr ähneln, nach Abschluss unseres Verfahrens denselben Clustern zugeordnet wurden. Anschaulich entspricht die Zeitreihe X somit einer *Mittelwertzeitreihe* aus allen normierten Smart-Meter-Zeitreihen. Durch die Mittelwertbildung wird eine Unabhängigkeit in Bezug auf die Anzahl der Messstellen erreicht. Im Anschluss konstruieren wir aus X den Datensatz D mit den Elementen d_l wie folgt:

$$D = \left\{ d_l = (x_j, \dots, x_{j+m}) \mid \begin{array}{l} \forall a \text{ mit} \\ 1 \leq j \leq a \leq j+m \leq T : \\ t_a \text{ gehört zum} \\ l\text{-ten Kalendertag} \end{array} \right\} \quad (2)$$

$(m+1)$ entspricht hierbei der Anzahl der gemessenen Werte pro Tag. In der Praxis häufig anzutreffende Messintervalle für Smart-Meter-Zeitreihen sind 1 Messung alle 15, 30 oder 60 Minuten, wodurch die d_l jeweils einem 96-, 48- beziehungsweise 24-Tupel entsprechen. Diesen Datensatz D segmentieren wir mittels *Fuzzy-C-Means* [6] mit verschiedenen Werten für die Anzahl der Cluster (c); der optimale Wert für c und die optimale Clustereinteilung werden anschließend durch die Auswertung verschiedener *Cluster-Validity-Indizes* [7] bestimmt. Mit dieser Vorgehensweise ist es nicht nur möglich die optimale Anzahl an Typtagen auf Basis der Smart-Meter-Zeitreihen zu bestimmen; aufgrund der Kenntnis, welche d_l denselben *Cluster-Zentroiden* zugeordnet wurden und welchem Kalendertag sie jeweils entsprechen, ist es möglich zu bestimmen welche Kalendertage zu einem Typtag zusammengefasst werden. Ein Analyst begutachtet die Kategorisierung der Kalendertage, arbeitet Regelmäßigkeiten heraus und überträgt sie auf zukünftige Tage.

4.2 Bestimmung typischer Verbrauchsmuster

Um die Anzahl und Ausprägungen der Verbrauchsmuster in Hinblick auf die finalen Kundengruppen herauszuar-

beiten, unterteilen wir die zur Verfügung stehenden Smart-Meter-Messwerte zunächst in disjunkte Mengen P_1, \dots, P_L ein. Seien hierzu $K_n, 1 \leq n \leq L$ die in Abschnitt 4.1 bestimmten Typtage, welche jeweils die zum jeweiligen Typtag zugeordneten t_j enthalten. Dann konstruieren wir $P_n, 1 \leq n \leq L$ mit den Elementen $p_{e,n}$ wie folgt:

$$P_n = \left\{ (y_{i,j}, \dots, y_{i,j+m}) \mid \begin{array}{l} \forall a, b \text{ mit} \\ 1 \leq j \leq a, b \leq j+m \leq T : \\ t_a, t_b \in K_n \text{ und} \\ y_{i,a}, y_{i,b} \text{ gehören zu} \\ \text{demselben Kalendertag} \end{array} \right\} \quad (3)$$

mit $y_{i,j} = \frac{s_{i,j}}{\sum_{j'=1}^T s_{i,j'}}$

Anschaulich entspricht P_n ähnlich wie D in Abschnitt 4.1 einem Datensatz, bei dem Messwerte eines Kalendertages zu jeweils einem $(m+1)$ -Tupel zusammengefasst wurden. Anders als D basiert P_n jedoch auf den normierten Messwerten $y_{i,j}$ und enthält nur die Daten der Kalendertage, die dem Typtag K_n zugeordnet wurden. Jeden der Datensätze P_n segmentieren wir anschließend unter Verwendung von *Fuzzy-C-Means*. Analog zur Bestimmung der Typtage segmentieren wir auch in diesem Fall die P_n mit verschiedenen Werten für die Anzahl der Cluster c und bestimmen die optimale Clustereinteilung mithilfe von *Cluster-Validity-Indizes*; die daraus resultierenden für den jeweiligen P_n -Datensatz als optimal bewerteten *Cluster-Zentroiden* $C_{q,n}, 1 \leq q \leq c_{n,optimal}$ entsprechen den gesuchten typischsten Verbrauchsmustern für den Typtag K_n .

4.3 Bestimmung der Lastprofile

Lastprofile im Sinne von Abschnitt 3 kann man als eine Menge von L -Tupeln darstellen, bei denen der n -te Eintrag eines solchen Tupels dem Verbrauchsmuster an zum Typtag K_n gehörigen Kalendertagen beschreibt. Jedem Kunden wird anschließend ein solches Lastprofil zugewiesen. Das Erzeugen der Menge von Lastprofilen G ist somit prinzipiell durch Bildung des *kartesischen Produktes* aus den in Abschnitt 4.2 gewonnenen Verbrauchsmustern $C_{q,n}$, gruppiert anhand der in Abschnitt 4.1 ermittelten Typtage K_n , möglich:

$$G = \prod_{n=1}^L W_n \quad \text{mit} \quad W_n = \{C_{q,n} \mid 1 \leq q \leq c_{n,optimal}\} \quad (4)$$

Bei dieser Vorgehensweise fehlen allerdings noch die zur praktischen Verwendung der Profile notwendigen *Profilzuordnungen* Z ; diese beschreiben, welcher Kunde welchem Lastprofil zugewiesen wurde, und somit welche Verbrauchsmuster als Grundlage für die Bilanzierung des zu erwartenden Energieverbrauchs dieses Kunden dienen. Darüber hinaus ist nicht garantiert, dass jedes durch das kartesische Produkt gebildete Lastprofil einen Kunden zugewiesen bekommt. Aus diesem Grund schlagen wir vor, für jeden Kunden ein individuelles Profil zu berechnen; dabei sagen wir, dass ein Kunde, hier repräsentiert durch die entsprechende Smart-Meter-Zeitreihe S_i , am Typtag K_n dem Verbrauchsmuster $C_{q,n}$ zugewiesen wurde, wenn die auf S_i basierenden Tupel im Datensatz P_n beim Clustering am häufigsten den höchsten Zugehörigkeitsgrad zum Verbrauchsmuster $C_{q,n}$ aufweisen.

Algorithmus 1 beschreibt diese Vorgehensweise in Pseudocode. Die Hilfsvariable H stellt hierbei ein Lastprofil in

Algorithmus 1 Bestimmung der Lastprofile

Eingabe: $S_i, P_n, K_n, C_{q,n}, U_n$
Ausgabe: Menge aller Lastprofile G , Menge der Profilverordnungen Z

```

1:  $G \leftarrow \emptyset$ 
2:  $Z \leftarrow \emptyset$ 
3: for  $i = 1$  to  $N$  do
4:   for  $n = 1$  to  $L$  do
5:      $H[n] \leftarrow C_{q,n}$  mit
            $q = \arg \max_{q'} \left\{ \left. \begin{array}{l} p_{e,n} \\ \exists j : (y_{i,j}, \dots, y_{i,j+m}) = p_{e,n} \\ \wedge \nexists q'' : u_{q'',e,n} > u_{q',e,n} \end{array} \right\} \right.$ 
6:   end for
7:    $G \leftarrow G \cup H$ 
8:    $Z \leftarrow Z \cup (S_i, H)$ 
9: end for
10: return  $G, Z$ 
    
```

Form eines L -dimensionalen Arrays dar, bei dem $H[n]$ dem Verbrauchsmuster am Typtag K_n entspricht. Die Profilverordnungen Z werden durch eine Menge von 2-Tupeln beschrieben, wobei der jeweils erste Eintrag den Kunden beziehungsweise seine Smart-Meter-Zeitreihe S_i und der zweite Eintrag das ihm zugewiesene Lastprofil enthält. Die in Algorithmus 1 verwendeten $u_{q,e,n} \in U_n$ entsprechen den in Abschnitt 4.2 mittels *Fuzzy-C-Means* bestimmten Zugehörigkeitsgraden des Datensatz-Tupels $p_{e,n}$ zum Verbrauchsmuster $C_{q,n}$.

5. EXPERIMENTELLE UNTERSUCHUNG

5.1 Beschreibung des Datensatzes

Zur Evaluation des in Abschnitt 4 vorgestellten Verfahrens steht uns ein realer Smart-Meter-Datensatz zur Verfügung. Die darin enthaltenen Zeitreihen entsprechen den im Bilanzierungsgebiet eines Stadtwerks registrierten Kunden, darunter beispielsweise Haushaltskunden, Gewerbekunden und Landwirtschaftsbetriebe. In unserem konkreten Fall entspricht dies 7668 Messstellen, jeweils mit Daten für einen Zeitraum vom 26 Monaten, bei einer zeitlichen Auflösung von 1 Messung pro Stunde. Aufgrund der Tatsache, dass es sich hierbei um ein Gebiet mit einem vollständigen Rollout von Smart-Meter-Zählern handelt, wird es uns hierdurch ermöglicht, das vorgestellte Verfahren unter realistischen Bedingungen zu bewerten. Bei der Verarbeitung der Daten wurden ausschließlich die Zeitreihenmesswerte berücksichtigt; Informationen über gegebenenfalls aktuell den Messstellen zugeordneten Kundengruppen blieb außer Acht.

5.2 Auswertung

Um die optimale Einteilung in Typtage zu bestimmen, wurden die zur Verfügung stehenden Smart-Meter-Zeitreihen gemäß Formel 2 zu einem 24-dimensionalen Datensatz aufbereitet und anschließend mittels *Fuzzy-C-Means* mit verschiedenen Werten für c geclustert ($2 \leq c \leq 25$). Als Distanzfunktion wurde die euklidische Distanz verwendet. Um verfälschten Ergebnissen aufgrund der Initialisierung der Cluster mit Zufallskoordinaten vorzubeugen, wurde für jeden Wert für c das entsprechende Clustering-Experiment 100-mal unabhängig voneinander wiederholt und jeweils mit *Cluster-Validity-Indizes (CVI)* bewertet.

Tabelle 1 zeigt eine Übersicht der Bewertung der Clustereinteilungen anhand verschiedener, in der Literatur weit

verbreiteter, Cluster-Validity-Indizes. Die einzelnen Zellen der Tabelle enthalten hierbei die Häufigkeit, wie oft von den 100 Iterationen der jeweilige Wert für c vom entsprechenden CVI als der beste bewertet wurde. Auf den ersten Blick ist hierbei erkennbar, dass die CVIs V_{PC} [5], V_{NPC} [2, 23] und V_{XIE} [28] eine kleine Anzahl an Clustern ($c \leq 5$) deutlich bevorzugen. Dem gegenüber stehen die CVIs V_{FH} , V_{PD} und V_{APD} [13]; diese bewerten große Werte für die Anzahl an Clustern ($c \geq 5$) als sehr gut, darunter insbesondere den Bereich $5 \leq c \leq 10$, jedoch mit einer auffallend negativen Bewertung des Wertes $c = 6$. Einzig der Index V_{BWS} bewertet den Bereich $5 \leq c \leq 7$ durchgehend positiv.

Die Tendenz der CVIs V_{PC} , V_{NPC} und V_{XIE} zu einer eher geringen Anzahl an Clustern ist hierbei wenig überraschend. Im Falle von V_{XIE} belegen zahlreiche Untersuchungen die monoton abfallenden Werte des Bewertungsmaßes für steigende c [17]; V_{PC} beziehungsweise V_{NPC} liefern nur bei klar getrennten, sphärischen Clustern gute Ergebnisse und unterschätzen in sonstigen Fällen den besten Wert für c sehr häufig [7, 17]. Die CVIs V_{FH} , V_{PD} und V_{APD} hingegen weisen eine zu große Streuung der Ergebnisse auf, als dass diese für eine aussagekräftige Schlussfolgerung herangezogen werden können. V_{BWS} konnte sich in experimentellen Untersuchungen sowohl bei synthetischen als auch realen Datensätzen als hervorragendes Bewertungsmaß profilieren [7, 17]. In unseren Experimenten hat V_{BWS} die Clustereinteilungen im Bereich $5 \leq c \leq 7$ als sehr gut eingestuft; für weiterführende Experimente zur Bestimmung typischer Verbrauchsmuster und Erzeugung der Lastprofile erachten wir daher die Zugrundelegung der Typtagunterteilungen mit 5 bis 7 Typtagen für sinnvoll.

6. AUSBLICK

In dieser Arbeit haben wir einen Ansatz zur Generierung von Lastprofilen aus Smart-Meter-Zeitreihen vorgestellt. Aufgrund der Tatsache, dass wir bei diesem Verfahren weder die Anzahl der Typtage, die Anzahl der Kundengruppen, noch deren konkrete Ausprägungen fest vorgeben, bilden die finalen Lastprofile das Verbrauchsverhalten der Kunden optimal ab. Durch Herausarbeiten von Regelmäßigkeiten in dem Auftreten der ermittelten Typtage und deren Übertragung auf zukünftige Tage wird es den Stadtwerken somit ermöglicht, das zukünftige Konsumverhalten von Verbrauchern zu prognostizieren und so die Menge der einzukaufenden Energie besser an den tatsächlichen Bedarf anzupassen.

Die weiterführende experimentelle Untersuchung unseres Verfahrens hinsichtlich der Identifizierung von Verbrauchsmustern, der Erzeugung der Lastprofile, deren Vergleich zu den BDEW-Standardlastprofilen und den Einfluss von Ausreißern auf die Qualität der Ergebnisse steht derzeit noch aus; diese werden wir in zukünftigen Veröffentlichungen publizieren.

In dieser Arbeit haben wir zur Bestimmung der Typtage und Verbrauchsmuster *Fuzzy-C-Means* verwendet. Prinzipiell wären hierfür auch andere zentroidbasierte, partitionierende Clustering-Verfahren wie *Gustafson-Kessel* [14] oder *Fuzzy Maximum Likelihood Estimation (FMLE)* [13] denkbar. Während bei der Bestimmung der Typtage auch dichte-basierte Clustering-Verfahren wie *DBSCAN* [10] zum Einsatz kommen können, ist dies bei der Erkennung der Verbrauchsmuster nach unserem Ansatz nicht möglich, da die bei dichte-basierten Clustering-Verfahren fehlenden Zentro-

Table 1: Clusterbewertung für die Bestimmung der Anzahl und Ausprägung der Typtage anhand der Cluster-Validity-Indizes *Partition Coefficient* (V_{PC}), *Normalized Partition Coefficient* (V_{NPC}), *Compactness & Separation by Xie and Beni* (V_{XIE}), *Compactness & Separation by Bouguessa, Wang and Sun* (V_{BWS}), *Fuzzy Hypervolume* (V_{FH}), *Partition Density* (V_{PD}) sowie *Average Partition Density* (V_{APD}). Für jeden Wert von c wurde das entsprechende Clustering-Experiment 100-mal wiederholt. Die Werte in den Zellen beschreiben daher, wie oft der jeweilige Cluster-Validity-Index den entsprechenden Wert für c als den besten bewertet hat.

| c | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|-----------|----|----|---|----|----|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| V_{PC} | 79 | 15 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V_{NPC} | 79 | 14 | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V_{XIE} | 79 | 14 | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V_{BWS} | 0 | 0 | 0 | 26 | 36 | 23 | 8 | 4 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V_{FH} | 0 | 0 | 0 | 8 | 1 | 15 | 7 | 9 | 10 | 5 | 6 | 2 | 6 | 7 | 4 | 1 | 4 | 4 | 3 | 4 | 0 | 1 | 2 | 1 |
| V_{PD} | 0 | 0 | 0 | 5 | 0 | 15 | 8 | 7 | 6 | 4 | 6 | 4 | 6 | 7 | 5 | 2 | 5 | 5 | 3 | 4 | 2 | 4 | 1 | 1 |
| V_{APD} | 0 | 0 | 0 | 22 | 0 | 16 | 7 | 12 | 9 | 6 | 4 | 1 | 5 | 4 | 4 | 1 | 2 | 2 | 1 | 1 | 0 | 1 | 1 | 1 |

iden die typischsten Verbrauchsmuster darstellen, welche gemäß Algorithmus 1 Teil der Lastprofile sind. In weiterführenden Arbeiten planen wir auch den Vergleich der Qualität der durch die Verwendung verschiedener zentroidbasierten, partitionierenden Clustering-Verfahren gewonnenen Lastprofile.

Bei den von uns durchgeführten Experimenten konnten die Typtage von einem handelsüblichen Desktop-Computer (Intel Core i7-4790 @ 3,6 GHz) innerhalb von 30 Minuten gefunden werden. Da jedoch unser Verfahren die Lastprofile, insbesondere die Verbrauchsmuster, anhand der Messwerte aller Smart-Meter-Zeitreihen zusammenstellt ist zu erwarten, dass die benötigte Rechenzeit durch zukünftig gemessene Smart-Meter-Daten bei einer Neuerstellung der Lastprofile signifikant zunimmt. Dies betrifft speziell Stadtwerke mit vielen Kunden. Ein möglicher Ansatz für weiterführende Arbeiten bestünde daher darin zu bestimmen, welchen Zeitraum die Smart-Meter-Daten für eine sinnvolle Erstellung der Lastprofile mindestens abdecken müssen. Wäre dies bekannt, könnten Stadtwerke ihre Datenbestände entsprechend filtern, sodass nur die aktuellsten für das Clustering notwendigen Messwerte herangezogen werden; dies würde den Rechenaufwand auch für große Stadtwerke in einem angemessenen Rahmen halten. Eine alternative Vorgehensweise wäre es, neu hinzukommende Smart-Meter-Messwerte inkrementell in existierende Lastprofile zu integrieren und die Qualität der auf diese Weise erzeugten Profile mit denen einer Neuberechnung zu vergleichen.

7. ACKNOWLEDGEMENT

Diese Arbeit entstand mit freundlicher Unterstützung der BTU EVU Beratung GmbH.

8. REFERENCES

- [1] F. Andersen, H. Larsen, and T. Boomsma. Long-term forecasting of hourly electricity load: Identification of consumption profiles and segmentation of customers. *Energy Conversion and Management*, 68:244 – 252, 2013.
- [2] E. Backer. *Cluster analysis by optimal decomposition of induced fuzzy sets*. PhD thesis, TU Delft, Delft University of Technology, 1978.
- [3] C. Beckel, L. Sadamori, T. Staake, and S. Santini. Revealing household characteristics from smart meter data. *Energy*, 78:397 – 410, 2014.
- [4] I. Benítez, A. Quijano, J.-L. Díez, and I. Delgado. Dynamic clustering segmentation applied to load profiles of energy consumption from spanish customers. *International Journal of Electrical Power & Energy Systems*, 55:437 – 448, 2014.
- [5] J. C. Bezdek. Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology*, 1(1):57–71.
- [6] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer US, 1981.
- [7] M. Bouguessa, S. Wang, and H. Sun. An objective approach to cluster validation. *Pattern Recognition Letters*, 27(13):1419 – 1430, 2006.
- [8] G. Chicco and I. S. Ilie. Support vector clustering of electrical load pattern data. *IEEE Transactions on Power Systems*, 24(3):1619–1628, Aug 2009.
- [9] P. D. Diamantoulakis, V. M. Kapinas, and G. K. Karagiannidis. Big data analytics for dynamic energy management in smart grids. *CoRR*, abs/1504.02424, 2015.
- [10] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.
- [11] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia. An electric energy consumer characterization framework based on data mining techniques. *IEEE Transactions on Power Systems*, 20(2):596–602, May 2005.
- [12] F. Fusco, M. Wurst, and J. Yoon. Mining residential household information from low-resolution smart meter data. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3545–3548, Nov 2012.
- [13] I. Gath and A. Geva. Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:773–780, 1989.
- [14] D. Gustafson and W. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on*, volume 17, pages 761–766, 1978.
- [15] M. Hayn, V. Bertsch, and W. Fichtner. Electricity load profiles in europe: The importance of household segmentation. *Energy Research & Social Science*, 3:30 – 45, 2014.

- [16] L. Hernández, C. Baladrón, J. M. Aguiar, B. Carro, and A. Sánchez-Esguevillas. Classification and clustering of electricity demand patterns in industrial parks. *Energies*, 5(12):5215, 2012.
- [17] L. Himmelspach. *Fuzzy Clustering of Incomplete Data*. PhD thesis, Heinrich-Heine-Universität Düsseldorf, 2016.
- [18] K. le Zhou, S. lin Yang, and C. Shen. A review of electric load classification in smart grid environment. *Renewable and Sustainable Energy Reviews*, 24:103 – 110, 2013.
- [19] J. J. López, J. A. Aguado, F. Martín, F. Muñoz, A. Rodríguez, and J. E. Ruiz. Hopfield-k-means clustering algorithm: A proposal for the segmentation of electricity customers. *Electric Power Systems Research*, 81(2):716 – 724, 2011.
- [20] N. Mahmoudi-Kohan, M. P. Moghaddam, and M. Sheikh-El-Eslami. An annual framework for clustering-based pricing for an electricity retailer. *Electric Power Systems Research*, 80(9):1042 – 1048, 2010.
- [21] M. Misiti, Y. Misiti, G. Oppenheim, J.-M. Poggi, et al. Optimized clusters for disaggregated electricity load forecasting. *Revstat*, 8(2):105–124, 2010.
- [22] F. Rodrigues, J. Duarte, V. Figueiredo, Z. Vale, and M. Cordeiro. *Machine Learning and Data Mining in Pattern Recognition: Third International Conference, MLDM 2003 Leipzig, Germany, July 5–7, 2003 Proceedings*, chapter A Comparative Analysis of Clustering Algorithms Applied to Load Profiling, pages 73–85. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.
- [23] M. Roubens. Pattern classification problems and fuzzy sets. *Fuzzy Sets and Systems*, 1(4):239 – 253, 1978.
- [24] C. Rudin, D. Waltz, R. N. Anderson, A. Boulanger, A. Salleb-Aouissi, M. Chow, H. Dutta, P. N. Gross, B. Huang, S. Ierome, D. F. Isaac, A. Kressner, R. J. Passonneau, A. Radeva, and L. Wu. Machine learning for the new york city power grid. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(2):328–345, Feb 2012.
- [25] T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Applied Energy*, 87(11):3538 – 3545, 2010.
- [26] J. P. L. Viegas, S. M. Vieira, and J. M. C. Sousa. Fuzzy clustering and prediction of electricity demand based on household characteristics. In J. M. Alonso, H. Bustince, and M. Reformat, editors, *2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT-15)*, Gijón, Spain., June 30, 2015. Atlantis Press, 2015.
- [27] D.-I. S. von Roon, D.-I. T. Gobmaier, K. Wachinger, and M. Hinterstocker. Statusbericht zum Standardlastprofilverfahren Gas. <https://www.ffegmbh.de/kompetenzen/system-markt-analysen/508-statusbericht-standardlastprofile-gas>, 11 2014.
- [28] X. L. Xie and G. Beni. A validity measure for fuzzy clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(8):841–847, Aug. 1991.
- [29] X. Zhang and C. Sun. Dynamic intelligent cleaning model of dirty electric load data. *Energy Conversion and Management*, 49(4):564 – 569, 2008.