Applied Energy 177 (2016) 435–448

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Predictive segmentation of energy consumers

Adrian Albert *, Mehdi Maasoumy

C3 IoT, Inc., 1300 Seaport Blvd, Redwood City, CA 94063, United States

HIGHLIGHTS

- An algorithm is proposed for building optimal, predictive consumer segmentations.
- The method structures existing qualitative knowledge using machine learning.
- It extracts predictive rules from data and optimally combines them into segments.
- The method is applied to identify predictive segments in a sample of 1*M* users.
- It uncovers homogeneous segments that are 2–3 times more effective for targeting.

ARTICLE INFO

Article history:
Received 19 January 2016
Received in revised form 30 April 2016
Accepted 21 May 2016
Available online 28 May 2016

Keywords:
Segmentation
Targeting
Energy efficiency

ABSTRACT

This paper proposes a *predictive segmentation* technique for identifying sub-groups in a large population that are both *homogeneous* with respect to certain patterns in customer attributes, and *predictive* with respect to a desired outcome. Our motivation is creating a highly-interpretable and intuitive segmentation and targeting process for customers of energy utility companies that is also optimal in some sense. In this setting, the energy utility wants to design a small number of message types to be sent to appropriately-chosen customers who are most likely to respond to different types of communications. The proposed method uses consumption, demographics, and program enrollment data to extract basic predictive patterns using standard machine learning techniques. We next define a feasible potential assignment of patterns to a small number of segments described by expert guidelines and hypotheses about consumer characteristics, which are available from prior behavioral research. The algorithm then identifies an optimal allocation of patterns to segments that is feasible and maximizes predictive power. This is formulated as maximizing the minimum enrollment rate from across the segments, which is then expressed as solving a mixed-integer linear-fractional program. We propose a bisection-based method to quickly solve this program by means of identifying feasible sets. We exemplify the methodology on a large-scale dataset from a leading U.S. energy utility, and obtain segments of customers whose likelihood of enrollment is more than twice larger than that of the average population, and that are described by a small number of simple, intuitive rules. The segments designed this way achieve a 2–3× improvement in the probability of enrollment over the overall population.

1. Introduction

In recent years energy utility companies have become ever keener on improving their relationship with a customer base that has traditionally been disengaged with their electricity provider. In the past, both energy companies and their consumers have understood the role of a utility as "keeping the lights on"; however, current technology trends and shifting customer attitudes, particularly fueled by the rise in consumer-facing Internet companies that excel at understanding and anticipating the preferences of their customers, have lead to an increased interest at utilities to engage with their customers. Compounding these trends are the increase in data availability, both high-granularity consumption data collected through sensing infrastructure such as smart meters and other "meta-data" on the consumers themselves, and in computational methods (e.g., [1,2]) to process this data. As such, energy utilities increasingly rely on analytic techniques that may provide them with ways to increase their customer satisfaction and engagement, as well as participation in environmentally-friendly programs within their customer base.

* Corresponding author.
E-mail addresses: adrian.albert@c3iot.com (A. Albert), mehdi.maasoumy@c3iot.com (M. Maasoumy).
URL: http://www.adrianalbert.info (A. Albert).

http://dx.doi.org/10.1016/j.apenergy.2016.05.128

Customer segmentation is a cornerstone of the marketing toolbox of small and large organizations, as a technique for understanding customers and for identifying ways to act upon that understanding. It is used heavily in marketing [3], online ads [4], or e-commerce [5], to name a few applications. As utilities strive to develop a more personal and modern relationship with their customers, they've enthusiastically embraced segmentation as a means to tailor their communications about efficiency measures and other programs to increase participation and engagement.

Most market segmentation techniques used in practice focus on the application of fixed rule-sets. For example, consumers who live in large homes and have children are assigned to a "high consumption" category, whereas those who subscribe to environmentalist magazines are ascribed to the "green advocates" group. Typically, these rules stem from counter-factual or anecdotal experience, behavioral studies, or small-scale psychology experiments, and are seen as "accepted facts" in practice. Being the result of distilled domain knowledge, such segmentation strategies are certainly valuable and should inform theory and practice; however, current literature in the field is silent about the extent to which performance of such qualitative approaches may be quantified and, it is hoped, improved upon.

In this paper, we propose a novel methodology for extracting *predictive* segments of energy utility customers from the individual household-level consumption, building characteristics, demographics, and program participation data that is becoming ever more common at utility companies. We are interested in uncovering *interpretable* segments that are both *homogeneous*, i.e., customers in those segments share certain demographics, building characteristics, or consumption characteristics, and *predictive*, in that consumers in certain segments have a higher probability of enrollment in efficiency programs than the population as a whole. For this, we first extract predictive patterns from the raw data that are characterized by their *support*, to how many consumers the patterns apply to, and by their *effectiveness*, the rate at which consumers covered by that pattern enroll in efficiency programs as compared to the overall population. These patterns may be associated with a small number of consumer typologies, as suggested by prior behavioral research, e.g., high income, educated consumers may be assumed to also have environmentally-friendly attitudes. Then, we build segments of consumers by combining appropriately-chosen patterns from the respective subsets associated with the segments, to maximize effectiveness while maintaining desired levels of support. We exemplify this approach by extracting predictive segments from nearly one million customers of a large U.S. utility.

We specifically consider an operational setting of identifying marketing leads and running mailing (paper or email) campaigns aimed at enrolling consumers into Energy Efficiency (EE) programs. From conversations with utility company partners, we found that this is currently the most widely used methodology for customer upsell and consumer acquisition for enrollment into demand-side management programs in the industry, as opposed to, e.g., running ad campaigns on online search engines or social networks. In this situation, marketing program managers prefer intuitive, simply-defined segments that they can understand from a marketing perspective, in order to craft appropriate messages to each segment.

Apart from the clear advantages, in terms of cost effectiveness and customer-relationship management to the utility company, a more accurate targeting for EE programs would be of significant interest to the consumers themselves. In particular, those consumers who lean towards enrolling in a specific program, but lack either the appropriately-packaged information presented from a viewpoint that may convince them to take action, or need additional nudging to make them decide to enroll, would benefit from a targeted approach. Another important result of improved, targeted communication is an increased satisfaction of the customer with the service provided by the utility company. Furthermore, by using mainly survey data as input, the methodology is highly amenable to situations where highly detailed sensor (smart meter) data is not present, so its adoption does not require prior investment into monitoring infrastructure.

1.1. Related literature

Customer segmentation and targeting for energy programs has recently received attention from seemingly disparate literature in engineering and computer science, operations management, and marketing.

Engineering research on demand-side management has been motivated recently by the availability of detailed customer data, including fine-grained consumption readings and socio-demographic information. It has typically focused on a few main areas:

(*i*) using whole-home data (either from smart meters or from custom instrumentation) to model building energy consumption behavior and describe consumption patterns of populations of users with the goal of informing programs such as tailored time-of-day pricing or smart thermostat controls [6,7], or design automated supply-following control algorithms for buildings with particular load profile [8];

(*ii*) collecting both whole-home *and* individual-appliance experimental data to reconstruct separate end-use consumption signals from an aggregate signal [9,10];

(*iii*) studying average effects of different internal and external factors, in particular occupancy, weather, building characteristics, on building energy consumption [11–13] and developing appropriate techniques for modeling and control of building energy consumption [14].

Other work investigates the relationship between patterns in consumption and consumer attributes [15,16], that can inform EE program targeting; however there is no immediate, actionable connection between the ability to infer consumer attributes and whether those consumers are likely to enroll in specific programs. Typically, these approaches have been motivated by the customer information obtained from online portals, and availability of high-frequency (interval) consumption data that is not yet available in the majority of utilities. Therefore, most of this work does not attempt to incorporate, test, or structure in any way the qualitative knowledge that currently exists in utilities regarding their customers, but proposes new, data-intensive techniques that are largely tangential to the current state of practice.

Most recent literature on energy analytics is concerned with characterizing consumption patterns (*load profiling*) in an extension of traditional demand-management practices at utilities that use aggregate demand profiles to inform programs. A segmentation strategy of consumers by the cost that their consumption behavior poses to the grid has been proposed in [17] as a way to target those groups of consumers who contribute most to the volatility in demand. A popular topic of study is the heterogeneity in typical daily load profiles (which typically entails clustering daily user consumption load shapes using off-the-shelf unsupervised algorithms such as `K-Means`) that can later be used for interventions such as differential pricing or incentives to reduce energy. This approach is taken in e.g., [18–23]. Other variations on segmenting load profiles based on first learning generative models of consumption, then clustering the obtained models have been discussed in [16,24]. This line of research is however largely descriptive in nature, as typically no clear use case is provided for the identified load patterns – and few programs at utilities currently exist that can incorporate such information. Moreover, these approaches also require a high amount of granular smart meter data from many consumers, which may not be practical for regions or countries where smart metering has not yet achieved scale.

Another set of recent literature in the area of energy analytics has focused on modeling energy consumption of buildings sector, as the most significant energy consumer, with energy consumption larger than transportation and industry sectors [25]. Randomness of internal effects such as occupancy and user behavior, as well as external effects such as climate variations, affect both transient and steady state thermal response of building models. Robust building modeling and robust control of energy consumption of Heating, Ventilation and Air Conditioning (HVAC) systems of buildings are crucial to effective and optimal design and operation of such systems. Buso et al. [26] proposes a robust building design, where buildings' performances show little variations with alternating occupant behavior patterns. Other studies such as [27] have attempted to model the stochastic behavior of electrical loads in residential buildings. Alternative approaches to address randomness and unpredictability of buildings, are discussed in [28] which proposes a controller for building HVAC systems that is robust against both internal and external random factors affecting building consumption, and in [29] which proposes a guideline for choosing model predictive control versus robust model predictive control, versus a rule-based controller based on the level of model uncertainty.

On the other hand, the operations management and marketing literature have seen a growing interest in applications to energy over the past several years. This may have been influenced by the fact that, at many utility companies, the department that is concerned with allocating, enrolling, and targeting consumers with efficiency programs has traditionally been either Operations or Marketing. In these fields, researchers are primarily concerned with qualitative studies in the context of segmentation and targeting, not necessarily applied to energy. A popular type of behavioral and psychology-inspired techniques employs a "psychographic segmentation" [30], which is based in qualitative surveys and concepts from marketing and behavioral psychology. Similar approaches have been developed, e.g., in [31,32]. Such qualitative work has been influential in the development and deployment of segmentation and targeting for EE programs at utility companies, and has guided the discussion around the topic until recently.

Our work contributes to the larger discussion in the fields of engineering, computer science, operations management and marketing, by providing a simple and transparent methodology that produces interpretable segments building on existing domain knowledge at operations and marketing departments at energy utilities. It fills the widening gap between the qualitative segmentation and targeting approaches that were widely in use in past decades and are still in use in many energy companies and the modern, data-driven techniques that have gained popularity in the last few years. As our proposed methodology does not rely on still-scarce granular smart meter data, although it can readily incorporate it if available, it is immediately applicable at organizations that do not have access to large quantities of interval consumption data.

The remainder of this paper is organized as follows. In Section 2 we formally introduce the predictive segmentation problem. Then, in Section 3 we develop the computational methods used to extract predictive segments from real-world customer data. Section 4 discusses the data that we used to illustrate our technique. Section 5 presents experimental results. We conclude in Section 6.

2. Predictive segmentation

For the setting we consider in this paper, a transparent and useful segmentation strategy should achieve the following:

1. internalize existing, valuable domain knowledge and best practices so that practitioners can easily relate to and adopt them;
2. be cost-aware, interpretable, and intuitive for non-technical program administrators at energy utilities, as well as useful for crafting marketing communications;
3. offer certain optimality guarantees in terms of effectiveness, i.e., be highly discriminative with respect to its purpose of identifying sub-groups whose members will be more likely to take action than consumers taken at random from the population at large.

As a precursory step in the process of obtaining effective segments algorithmically, we also identify those (measured) characteristics of the consumers that provide the best predictive power for the outcome of interest, the likelihood of enrolling in EE programs.

To the first point above, much expertise and practical experience exists at energy utilities that allows them to put forth hypotheses about certain high-level types of customers that they wish to identify among their base. For example, most experienced program administrators would agree that "Green Advocate" consumers respond to other types of communications, emphasizing environmental impact, than those consumers who are more "Cost Conscious", who may be responsive to arguments about monetary savings.

To the second point, we start from *existing* domain knowledge that associates certain variables with each given segment (e.g., "Green Advocates" might be defined by their income, household type, and level of education), and identify simple logical rules involving those variables that lead to the most effective segmentation strategy. Such intuitive segments should allow crafting appropriate messaging strategies. For example, consumers in the "Green Advocates" groups will receive messages that emphasize the environmental aspects of energy savings, while those consumers in the "high consumption" category will be informed about ways in which they could reduce their energy bills. A consequence of this approach is cost-awareness of the resulting model – by defining a small number of predictive segments defined by simple rules that involve only a handful of variables, we effectively select 15–20% of the variables available for analysis.

The challenge then becomes (as presented in the third point above) to develop an algorithmic segmentation methodology that internalizes the desiderata of points 1 and 2 while ensuring useful properties of the resulting segments, as well as achieving the best possible segmentation satisfying the imposed structure. We consider that the desired outcome is to maximize the impact of the marketing communications on EE program enrollment, i.e., target those customers that are more likely to enroll. As both tailoring communications and managing campaigns are costly, there is a real incentive to create messages for small number of segments, and to have those segments include consumers who are most likely to take action.

2.1. Problem setting

A population $\mathcal{X}$ consisting of N consumers is serviced by an operator (an energy utility company); for each consumer the utility observes a number of M features $x \in \mathbb{R}^M$ that comprise of both consumption and customer characteristics such as socio-demographic and physical building attributes; the features across all consumers are stored in a matrix $X \in \mathbb{R}^{N \times M}$. The utility also observes for each consumer i, whether he has enrolled in any program in the past year, which is encoded as a binary variable $y : y_i = 1$ iff *customer* i has enrolled.

The utility wishes to use the data (X, y) to identify K segments within the population that are "homogeneous" with respect to the attributes X, with the purpose of informing, simplifying, and increasing the effectiveness of targeted communications for demand-side efficiency program enrollment. Based on prior marketing research, the utility may have certain hypotheses as to what "types" of customers it services. We assume that this prior knowledge is of the form:

> *"Green Advocates" have a relatively high income or at least a college degree.*
> *"Home Improvers" are home owners or own a large equity share on their home.*
> ⋮

Then the data (X, y) can be used to make these hypotheses specific by extracting a set $\mathcal{P}$ of V patterns, $\mathcal{P} \equiv \{P_1, \ldots, P_V\}$, that are both *descriptive*, in that the characteristics of the consumers they refer to exhibit these patterns, and *predictive*, in that the consumers who fall in a certain pattern are more likely to enroll than a consumer selected at random from the entire population. We thus define a *pattern* to be a logical expression of the form:

$$P = \{x \in \mathcal{X} | r_1(x) \wedge r_2(x) \wedge \ldots\}, \tag{1}$$

where the P's are *base rules* (logical statements). Hence a pattern is defined as a succession of conjunctions. Interchangeably we refer to the pattern as the set of consumers that follow the logical definition of the pattern. We consider the base rules to be of the form:

$$r_j(x) := x_j \leqslant t_j \text{ or } r_j(x) := x_j \geqslant t_j. \tag{2}$$

As such, a base rule is defined by the variable x_j (the j-th variable in x) it refers to, a *direction* (either "$\geqslant$" or "$\leqslant$"), and a *threshold* t_j learned from data. We consider a rule $P_j(\cdot)$ to be *consistent* with a hypothesis if both the variable and the direction that define that rule match the hypothesis. Similarly, we define a pattern P to be δ-*consistent* with a hypothesis if it contains at least $\delta \geqslant 1$ rules that are consistent with the hypothesis.

It is useful to define a *coverage matrix* C that summarizes the extent to which an item i is covered by pattern m:

$$c_{im} = \begin{cases} 1 & \text{if pattern } P_m \text{ covers consumer } i \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The *effectiveness* of a pattern P may be computed as the (empirical) enrollment probability of consumers covered by that pattern:

$$q(P) = \frac{\sum_{i \in P} \mathbb{1}\{y_i = 1\}}{|P|}. \tag{4}$$

With the setup above, we define K segments as collections of patterns, $S_k \subset \mathcal{P}(\mathcal{P})$, such that every pattern in each segment is δ-consistent with the hypotheses that define that segment. Let $B \in \mathbb{R}^{M \times K}$ define the (known) consistency matrix that describes the allowed relationship between segments and patterns:

$$b_{mk} = \begin{cases} 1 & \text{if pattern } P_m \text{ can be included in segment } S_k \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Finally, a *segmentation* is defined as the set of individual segments

$$\mathcal{S} \equiv \{\mathcal{S}_1, \ldots, \mathcal{S}_K\} \tag{6}$$

2.2. Effective segmentation

Here we consider a segmentation strategy to be *effective* if it is able to discriminate between consumer segments with respect to the rate of enrollment. That is, a good strategy (on K segments) will identify those segments in the population that enroll with probabilities $q_k,\ k = 1, \ldots, K$ that are very different from (either smaller or greater than) the overall rate q observed in the entire population. For example, if our segmentation consists of $K = 2$ groups A and B, it is perfectly effective if all consumers in A enroll, but no consumer in B enrolls (so $q_A = 1$ and $q_B = 0$). A perfectly ineffective segmentation is one where consumers in A enroll at the same rate as consumers in B (so $q_A = q_B$). Of course, one could always group consumers into two segments by having all those who have enrolled in EE programs be in one of the segments; however the challenge is to identify patterns in the consumer characteristics X that lead to interpretable, intuitive definitions of segments that are also predictive of enrollment.

The *effectiveness* of each segment may be computed in a similar way to the effectiveness of a pattern as the (empirical) enrollment probability of consumers in that segment:

$$q(S) = \frac{\sum_{i \in S} \mathbb{1}\{y_i = 1\}}{|S|}. \tag{7}$$

A segment is thus a good proxy for enrollment if $|q_k - q| \gg 0$, where $q = \frac{\sum_{i \in \Omega} \mathbb{1}\{y_i = 1\}}{N}$ is the rate of enrollment in the overall population. The problem we want to solve is to allocate at least $\underline{\pi}$ and at most $\overline{\pi}$ patterns to each segment such that the resulting segments have desirable effectiveness properties, for example:

- maximize the minimum effectiveness:
$$\max_{S_1, \ldots, S_K} \min_k q(S_k) \tag{8}$$
- ensure an appropriate balance of effectiveness across segments:
$$\max \theta_1 q(S_1) + \cdots + \theta_K q(S_K), \tag{9}$$
where θ is a given weights vector.

For this, define the decision variables z_{mk} ($Z \in \mathbb{R}^{M \times K}$) such that

$$z_{mk} = \begin{cases} 1 & \text{if pattern } P_m \text{ is included in segment } S_k \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where a segment k is defined as:

$$S_k = \cup_{m: z_{mk} = 1} P_m. \tag{11}$$

Then our problem becomes finding the values of z_{mk} such that one of the objectives (8) and (9) is maximized, and the following *feasibility constraints* are satisfied:

$$\begin{array}{ll} \text{Include patterns only in allowed segments} & z_{mk} \leqslant b_{mk}, \quad \forall m, k \\ \text{Limit number of patterns per segment} & \underline{\pi} \leqslant \sum_m z_{mk} \leqslant \overline{\pi}, \quad \forall k \\ \text{A pattern can only belong to one segment} & \sum_k z_{mk} \leqslant 1, \quad \forall k \\ \text{Either select a pattern or not} & z_{mk} \in \{0, 1\}, \quad \forall m, k \end{array} \tag{F$_0$}$$

2.3. Overlapping patterns and segments

There may be many patterns that are feasible for a given segment, i.e., $|\{P_m | b_{mk} > 0\}| > 1$; moreover patterns may overlap (that is, the sets of consumers they define are not disjoint, $\exists\ m, m',\ P_m \cup P_{m'} \neq \varnothing$). Then the segments in $\mathcal{S}$ may overlap as well, if they happen to contain patterns that overlap in the customers they describe. This imposes an additional complication to appropriately formulating an optimization problem that addresses (8) and (9) and satisfies the constraints (F_0).

Were the patterns not overlapping, the segment effectiveness could be written as:

$$q_k = \frac{\sum_{i,m} y_i c_{im} z_{mk}}{\sum_{i,m} c_{im} z_{mk}} = \frac{y^T C z_k}{\mathbf{1}^T C z_k} = \frac{a^T z_k}{d^T z_k}, \tag{12}$$

where $a \equiv C^T y$ and $d \equiv C^T \mathbf{1}$.

However, since pattern overlap can be substantial, the above expression overcounts the consumers that fall into multiple patterns. One simplification we adopt to address this issue is to relax the definition of the coverage matrix C, noting that a consumer who is covered by n different patterns may be considered as having a fractional coverage of $\frac{1}{n}$ on each pattern. This translates to a modified coverage matrix $\widetilde{C}$:

$$\tilde{c}_{im} = \frac{c_{im}}{\sum_m c_{im}} \tag{13}$$

As such the modified coverage matrix assigns a *weight* to each consumer i that indicates the fractional coverage of a single pattern, giving equal importance to each pattern. For simplicity we refer to this modified matrix still by C.

3. Computing predictive segments

The design of an algorithm to compute predictive segments will be determined by the specific form that the objective function takes, given that the constraints are mixed-integer linear ones. Here we focus on the situation where the objective is to allocate allowable patterns to segments such as to maximize the minimum effectiveness across the K segments – see Eq. (8). This is a natural requirement for a program administrator that wishes to have guarantees on the minimum effectiveness of his targeted communications strategy.

3.1. Maximizing the minimum effectiveness

The above formulation in Eq. (12) makes use of K vectors z_k that encode the decision variables for each segment. In order to express the objective and constraints in the more familiar affine form using a single decision variable vector we can employ the following notation:

$$\mathbf{v}_k \equiv \left(\mathbf{0}^T\ \mathbf{0}^T \dots \underbrace{v^T}_{kth\ \text{position}} \dots \mathbf{0}^T \right) \tag{14}$$

$$\tilde{\mathbf{1}}_m \equiv \left((0 \dots \underbrace{1}_{mth\ \text{position}} \dots 0) \dots (0 \dots \underbrace{1}_{mth\ \text{position}} \dots 0) \right) \tag{15}$$

$$\mathbf{z} \equiv \left(z_1^T\ z_2^T \dots\ z_k^T \dots z_K^T \right), \tag{16}$$

with $\mathbf{z}$, $\mathbf{v}_k$, and $\tilde{\mathbf{1}}_m \in \mathbb{R}^{1 \times MK}$. The effectiveness can then be expressed as:

$$q_k = \frac{\mathbf{a}_k^T \mathbf{z}}{\mathbf{d}_k^T \mathbf{z}}, \tag{17}$$

and the feasibility conditions in (F_0) can be expressed as:

$$\begin{aligned} &\mathbf{z} \leqslant \mathrm{vec}(B) \\ &\mathbf{1}_k^T \mathbf{z} \leqslant \overline{\pi},\ \forall k = 1, \dots, K \\ &\tilde{\mathbf{1}}_m^T \mathbf{z} \leqslant 1,\ \forall m = 1, \dots, M \\ &z_{mk} \in \{0, 1\} \end{aligned} \tag{F}$$

In the case of the max–min objective function (8), the optimization problem tries to increase as much as possible the lower bound on the effectiveness across the segments. This results in a relatively homogeneous distribution of the q_ks. This situation may be desirable, e.g., when action will be taken on each of the segments. In this case the optimization problem may be expressed as:

$$\begin{aligned} \max_{\mathbf{z}}\ \min_{1 \leqslant k \leqslant K}\ & \frac{\mathbf{a}_k^T \mathbf{z}}{\mathbf{d}_k^T \mathbf{z}} \\ \text{subject to}\ & \mathbf{z} \leqslant \mathrm{vec}(B) \\ & \mathbf{1}_k^T \mathbf{z} \leqslant \overline{\pi},\ \forall k \\ & \mathbf{1}_k^T \mathbf{z} \geqslant \underline{\pi},\ \forall k\ k = 1, \dots, K \\ & \tilde{\mathbf{1}}_m^T \mathbf{z} \leqslant 1,\ \forall m = 1, \dots, M \\ & z_{mk} \in \{0, 1\} \end{aligned} \tag{LFIP}$$

Problem (LFIP) is a *generalized (max–min) linear-fractional integer program* with linear constraints. This class of problems has been extensively studied in the literature (see e.g., [33–35]).

Following [36], we propose an equivalent formulation of (LFIP) as a linear-integer programming feasibility problem (LFIP-F):

$$\begin{aligned} \max_{\mathbf{z}}\ & \lambda \\ \text{subject to}\ & (\mathbf{A} - \lambda \mathbf{D}) \mathbf{z} \geqslant \mathbf{0} \\ & \mathbf{z} - \mathrm{vec}(B) \leqslant \mathbf{0} \\ & \mathbf{1}_k^T \mathbf{z} - \overline{\pi} \leqslant 0,\ \forall k = 1, \dots, K \\ & \underline{\pi} - \mathbf{1}_k^T \mathbf{z} \leqslant 0,\ \forall k = 1, \dots, K \\ & \tilde{\mathbf{1}}_m^T \mathbf{z} - 1 \leqslant 0,\ \forall m = 1, \dots, M \\ & z_{mk} \in \{0, 1\} \end{aligned} \tag{LFIP-F}$$

where $\mathbf{A}$ is a matrix with rows $\mathbf{a}_k^T$ and $\mathbf{D}$ is a matrix with rows $\mathbf{d}_k^T$, for $k = 1, \dots, K$.

For a given value of λ, the above feasibility problem (LFIP-F) can be solved using standard mixed-integer programming packages. Although the initial customer characteristics data can be quite large (here $N \approx 1M$ consumers), the number of patterns is expected to be much smaller ($M \sim 1000$), as is the number of segments (here $K = 5$). Then a standard package can offer an excellent out-of-the-box performance.

Then a maximum $\lambda^* \equiv \max_{\mathbf{z}} \lambda$ with a corresponding optimum $\mathbf{z}$ can be found efficiently using an iterative bisection Algorithm 1 that solves a feasibility problem (LFIP-F) at each step. Starting with a large interval $[l_0, u_0]$ in which the optimum λ^* is guaranteed to be (here $[0, 1]$), the algorithm successively narrows down the interval $[l, u]$, at every step ensuring that $\lambda^* \in [u, b]$. This is outlined in Lemma 1 below which builds upon [37].

Algorithm 1. Bisection algorithm for solving Problem LFIP-F.

Input: Interval $[l, u]$ that contains the optimum λ^*; tolerance parameter ϵ.

```
while l < u and |u − l| ⩾ ϵ do
    λ ← (u+l)/2
    z ← satisfies LFIP-F(λ)    ▷ Solve using a standard MIP solver such as GUROBI
    if λ feasible then
        u ← λ
    else
        l ← λ
return z
```

Lemma 1. *The output of Algorithm 1 is an optimal $\mathbf{z}^*$ corresponding to λ^*, the maximum value of λ, within a tolerance ϵ, and within $\log_2\left(\frac{\epsilon_0}{\epsilon}\right)$ iterations.*

To prove Lemma 1 we must show that the Algorithm 1 will find a unique value λ^* that is the maximum feasible value that λ can take. For this, define the feasible set as:

$$\Lambda \equiv \Big\{ \lambda \mid \exists \mathbf{z} \in \{0,1\}^{MK},\ (\mathbf{A} - \lambda \mathbf{D})\mathbf{z} \geqslant \mathbf{0},\ \mathbf{z} \leqslant \mathrm{vec}(B), \quad \mathbf{1}_k^T \mathbf{z} \leqslant \overline{\pi},\ \mathbf{1}_k^T \mathbf{z} \geqslant \underline{\pi},\ \tilde{\mathbf{1}}_m^T \mathbf{z} \leqslant 1 \Big\} \quad (18)$$

With this notation we have $\lambda^* \equiv \sup\{\lambda \in \Lambda\}$, and the optimal pattern allocation to segment $\mathbf{z}^*$ corresponds to λ^*. By definition the optimum λ^* is the (upper) transition point between the feasible set Λ and the unfeasible set $\bar{\Lambda} \equiv \{\lambda \mid \lambda \notin \Lambda\}$, so the following must hold for a tolerance parameter $\epsilon > 0$ (small):

- $\lambda \in \Lambda \Rightarrow \lambda - \epsilon \in \Lambda$
- $\lambda \notin \Lambda \Rightarrow \lambda + \epsilon \notin \Lambda$

To prove that Algorithm 1 will find the optimum λ^* we need to show that it satisfies the above conditions. We focus solely on the term containing λ in our analysis.

To prove the first condition, we take $\lambda \in \Lambda$, and we must prove that $\lambda - \epsilon \in \Lambda$. The fact that $\lambda \in \Lambda$ implies that $\exists \mathbf{z}_\lambda$, s.t. $(\mathbf{A} - \lambda\mathbf{D})\mathbf{z}_\lambda \geqslant \mathbf{0}$. Then for $\lambda + \epsilon$ and the same $\mathbf{z}_\lambda$, we have

$$(\mathbf{A} - (\lambda - \epsilon)\mathbf{D})\mathbf{z}_\lambda = (\mathbf{A} - \lambda\mathbf{D})\mathbf{z}_\lambda + \epsilon\mathbf{D}\mathbf{z}_\lambda \geqslant \mathbf{0}.$$

The second term above is positive since $\epsilon > 0$ and both $\mathbf{D}$ and $\mathbf{z}_\lambda$ have only non-zero entries.

To prove the second condition, fix a value $\lambda \notin \Lambda$; then we wish to show that $\lambda + \epsilon \notin \Lambda$ for $\epsilon > 0$. The fact that $\lambda \notin \Lambda$ implies that $\nexists \mathbf{z}$, s.t. $(\mathbf{A} - \lambda\mathbf{D})\mathbf{z} \geqslant \mathbf{0}$; as such we must have $(\mathbf{A} - \lambda\mathbf{D})\mathbf{z} < \mathbf{0}$, $\forall \mathbf{z}$ for the given value of λ. Let $\mathbf{z}_\lambda$ be the decision variable vector corresponding to λ that produces the largest value of $(\mathbf{A} - \lambda\mathbf{D})\mathbf{z}$ and satisfies all the other conditions that define the feasibility set Λ. Then $(\mathbf{A} - \lambda\mathbf{D})\mathbf{z}_\lambda \geqslant (\mathbf{A} - \lambda\mathbf{D})\mathbf{z}$, $\forall \mathbf{z} \in \{0,1\}^{MK}$. From the infeasibility of λ we further have $(\mathbf{A} - \lambda\mathbf{D})\mathbf{z}_\lambda < 0$. Then for $\lambda + \epsilon$ take a decision vector $\mathbf{z}_{\lambda+\epsilon}$ that produces the largest value of $(\mathbf{A} - (\lambda + \epsilon)\mathbf{D})\mathbf{z}$. But from before we have $(\mathbf{A} - \lambda\mathbf{D})\mathbf{z}_\lambda \geqslant (\mathbf{A} - \lambda\mathbf{D})\mathbf{z}$, $\forall \mathbf{z}$, including $\mathbf{z}_{\lambda+\epsilon}$. Then we have for $\mathbf{z}_{\lambda+\epsilon}$:

$$(\mathbf{A} - (\lambda + \epsilon)\mathbf{D})\mathbf{z}_{\lambda+\epsilon} = (\mathbf{A} - \lambda\mathbf{D})\mathbf{z}_{\lambda+\epsilon} - \epsilon\mathbf{D}\mathbf{z}_{\lambda+\epsilon} \leqslant (\mathbf{A} - \lambda\mathbf{D})\mathbf{z}_\lambda - \epsilon\mathbf{D}\mathbf{z}_{\lambda+\epsilon} < \mathbf{0}. \quad (19)$$

Then since $(\mathbf{A} - (\lambda + \epsilon)\mathbf{D})\mathbf{z}_{\lambda+\epsilon} < \mathbf{0}$ we conclude that $\lambda + \epsilon \notin \Lambda$. As such, the Algorithm 1 will always find a maximally feasible λ^* corresponding to an optimum allocation vector $\mathbf{z}^*$. Moreover, since with each step the algorithm halves the search interval $[l, u]$, it takes at most $\log_2\left(\frac{u_0 - l_0}{u^* - l^*}\right) \leqslant \log_2\left(\frac{\epsilon_0}{\epsilon}\right)$ steps to reach the completion condition of $|u - l| < \epsilon$ $\square$.

3.2. Extracting predictive patterns from data

Given a set of observations encoded as the feature matrix X and the binary response (enrollment) vector y, we wish to extract patterns P that are highly effective ($q \gg q_0$). For this we adopt the following approach:

1. Use an ensemble method such as Random Forests or AdaBoost [38] having classification trees as base learner to generate many decision trees of varying depths (here we generated trees of up to depth 5). This step allows us to construct a list P_0 of initial patterns that we obtain by traversing the decision tree to each leaf. Depending on the level of the trees used as base classifier in the boosted ensemble, these rules can take varying forms of complexity, from single statements (trees of depth 1, or *decision stumps*) to conjunctions of multiple base rules.
2. Prune the patterns list $\mathcal{P}_0$ to eliminate those rules that do not correspond to some set criteria of "quality". For this purpose we shall consider a pattern $P \in \mathcal{P}_0$ as "effective" if it meets both of these criteria:
   - minimum support: $|P| > \eta$. Here we used $\eta = 500$, which was a number of customers that our utility company partners considered significant enough to be worth to potentially craft a targeted communication to.
   - minimum effectiveness: $q(P) > \zeta q_0$. Here we used $\zeta = 2$. This choice for ζ was motivated by the desire to ensure that only highly-predictive patterns would be used in designing segments.
3. Further remove patterns that overlap more than v % (here $v = 70\%$) with other patterns and have a lower effectiveness q. This step in effect ensures a level of (indirect) control of multicollinearity in the predictors, as we prune out those patterns that largely describe the same consumers that are also described by more predictive patterns.

This procedure results in a pruned set of patterns $\mathcal{P}$.

4. Experimental setup

4.1. The customer characteristics data

The data that we used in this paper was obtained from a large energy company in the U.S. and was comprised of ~100 socio-demographic and building characteristics, as well as of monthly energy consumption readings across two years for $N = 957,150$ consumers. After standard data cleaning procedures we selected 43 variables of interest that had at least 80% valid entries across the entire population. Out of those, 19 variables were categorical variables, whereas 24 were numerical variables. Converting the categorical variables to binary dummy variables, we obtained the final dataset of $P = 304$ variables. Overall, 48,310 consumers, corresponding to a fraction $q_0 = 4.9\%$, had enrolled in any energy efficiency program in the two years prior to the data collection.

Table 1 describes several categorical variables of interest. A large majority of consumers (~80%) own their homes, and only ~16% rent. The education levels reflect society at large, with a

Table 1
Example categorical-valued customer characteristics.

	Level	Percentage
Variable GreenAware	Behavioral Greens	52
	Think Greens	21
	Potential Greens	14
	True Browns	12
Combined homeowner	Homeowner	80
	Renter	20
Education	High school diploma	25
	Bachelor degree	25
	Graduate degree	23
	Some college	17
	Less than high school diploma	8
Marital status	Single	54
	Married	46
Home Heat Ind	Hot water	75
	Furnace	20
	Electric&Other	4
Spouse gender code	Female	75
	Male	25
Presence of child age 0–18	Inferred no children present	72
	Confirmed presence of children	27
	No adult in household	1

Table 2
Example numerical-valued socio-demographic and building-related customer characteristics.

	Mean	Std	Min	25%	50%	75%	Max
Length of residence	12.3	11.9	0.0	3.0	8.0	19.0	64.0
Number of adults in household	2.2	1.4	0.0	1.0	2.0	3.0	8.0
Birth year	1957	13.0	1880	1951	1960	1963	1995
Year built	1937	38.0	1900	1900	1925	1973	2014
Home total rooms	6.8	2.7	1.0	6.0	6.0	7.0	41.0
kW h annual	8661.6	8659.4	0.0	3952.5	6848.6	11,160.0	1,103,400.0

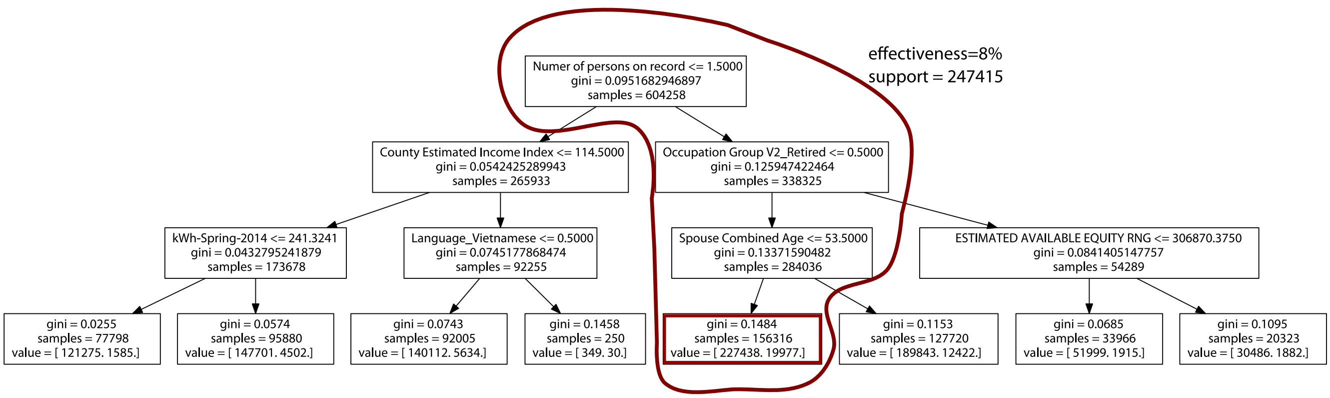

Fig. 1. Example decision tree of height 3 extracted from data. The highlighted pattern is a path in the decision tree starting from the root for which the effectiveness (proportion of positive samples) is 8%.

quarter of consumers having each college degrees and graduate degrees, while half of the consumers have a high school diploma or less. The "Green Aware" variable summarizes the result of a third-party analysis that takes into account factors such as magazine subscriptions, community involvement, political leaning, and affiliations to different organizations to result in an inferred level of interest in environmental matters. The purpose of the variable is to offer a single descriptor as a proxy to the likelihood that a person (or a household) is likely to adopt environmentally-friendly actions or be inclined to change behavior to favor a positive environmental outcome. As such, "Behavioral Greens" are consumers who both think and act in an environmentally-friendly way; "Think Greens" have a positive outlook towards environmental matters, but are less likely to act on that attitude if it involves effort on their part; "Potential Greens" are consumers who may be swayed to adopt environmentally-friendly actions if that is in their interest; and, finally, "True Browns" are consumers who do not have any interest in environmental and energy-efficiency matters. We unfortunately did not have access to the details of how this analysis was performed.

Table 2 summarizes several more numerical variables of interest. The average birth year is 1957, which suggests a baby-boomer demographic. The average family in our sample lives in a large home (6 rooms) with a tenure of more than 12 years. Ideally, we would have liked to have data of building envelope characteristics and the thermo-physical features of the building stock. Building location proximity features can be used as an approximation for building envelope characteristics, as buildings in the same neighborhoods are often built from similar material with similar thermo-physical characteristics. However, we did not have access to such data. Instead, we approximate these features by building age (function of Year Built).

4.2. Predictive patterns extracted from data

We extracted predictive rules from the data as described above in Section 3. An example of such a rule extracted from a decision tree of depth 3 is given in Fig. 1. There, a rule of a higher-than-average effectiveness (8% vs 5% for the average population) is highlighted in red. After pruning, the list of predictive patterns, whose effectiveness was at least $2 \times q_0 \approx 0.10$, and that had a support of at least $\eta = 500$, contained $M_0 = 2965$ patterns of up to 5 base rules each (1852 patterns with 5 base rules, 963 patterns with 4 base rules, 143 patterns with 3 base rules, and 7 patterns with 2 base rules). Fig. 2 illustrates the distribution of pattern effectiveness $q(R)$ for pattern of different complexities (2—5 base rules). As expected, for each of the levels of pattern complexity considered, the distribution exhibits an exponential behavior, with a many patterns of lower effectiveness, and fewer highly-effective patterns. Lower-complexity patterns, with 3 base rules, as displayed in blue in the figure, will also be less effective (lower q) than more complex patterns (with 4 or 5 base rules, displayed in green and red in the figure). However, there are highly effective 5-rule patterns ($q > 0.3$), and so it is expected that our algorithm will select some of those to design segments. Almost regardless of

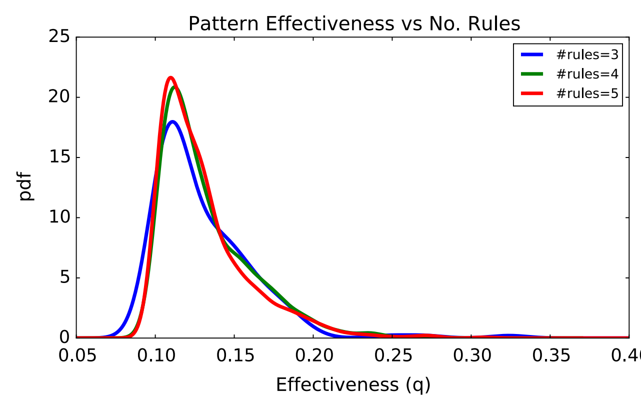

Fig. 2. Distribution of pattern effectiveness for the $M_0 = 2965$ patterns extracted from data. Different complexities (2–5 base rules) display a similar exponential drop-off in the number of highly predictive patterns.

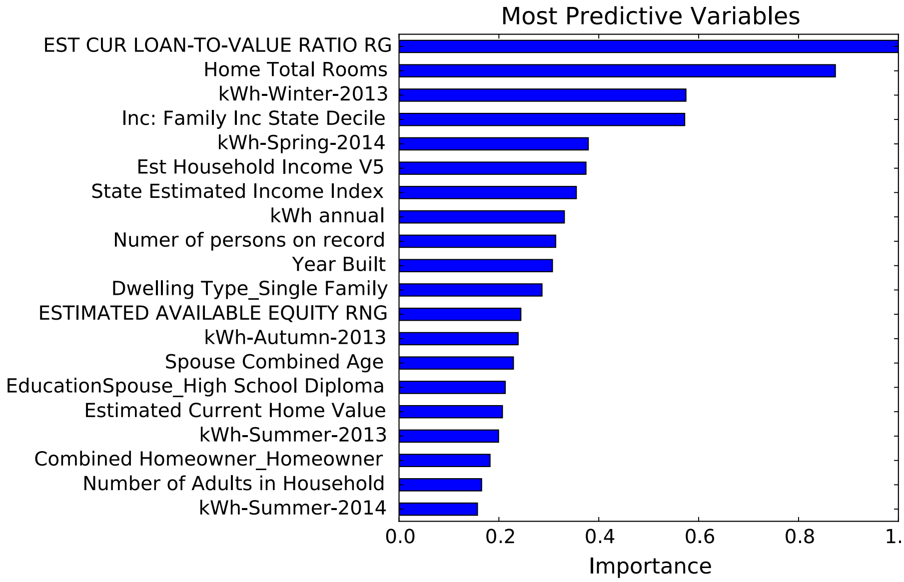

Fig. 3. Top 20 most important variables that explain enrollment.

complexity, the average effectiveness of the patterns we have used is ≈ 0.12, a $2.5\times$ improvement over the average probability to enroll in the overall population (which stands at 4.9%).

The top 20 most important variables for predicting enrollment are listed in Fig. 3. These include the amount of ownership on the house such as loan to value ratio and available equity, the size of the house and of the family living there, and measures of family income among others. This suggests that enrollment depends on the perception of financial commitment and ability as pertains to improvements to the house. Note that the interpretation of the relative magnitudes of relative "predictive importance" in this figure is to be understood from the perspective of how an ensemble classifier works: shown is the relative frequency with which variables were selected by the algorithm to grow the weak classifiers (decision trees) in the ensemble classifier.

The emergence of financial variables related to the equity owned on the house (shown in all-caps letters in Fig. 3) as key predictors was a confirmation to our utility partners of their existing knowledge about who enrolls in EE programs. Their intuition was that, the more invested a family was in their home, and the more of an owner they were, the more likely they were to want to improve the that home for the long-term. Other financial variables are related to the economic situation of the family (income, decile relative to the income distribution in the state), which corroborates further the knowledge at the partner utility that families are likelier to enroll in home-improvement or cost-reduction programs depending on what the impact of these measures is on their financial situation.

Physical building characteristics such as building age ("Year Built"), size ("Home Total Rooms"), and type (single family home) are also among the factors that are most predictive of the likelihood to enroll in EE programs. This is again in line with the current knowledge at the utility partner, and was information that was already being used in their segmentation design. What this analysis provides, at a minimum, is a better characterization of, e.g., what an "old" or "new" building is, or what a "large" or "small" house is. This is expressed as logical patters (see Section 2) that are based off numeric thresholds computed from the data. Similarly, some of the measured characteristics of the family (number of persons and the number of adults in the household) were relevant in predicting enrollment, which again was not a surprising finding.

The consumption data that was available for this study was at the level of a monthly billing cycle; as such, we have only computed measures of average consumption overall and by season as input features into the ensemble classifier that we used to generate predictive rules. Average consumption in the winter and spring are identified as more important predictors than consumption in the summer, which may be explained by the fact that, geographically, the consumers in our sample were located in a region (northeastern U.S.) where heating usage is more prominent. This suggests that consumers in that particular geography may be easier to swayed to enroll in an EE program related to their heating appliance or usage than to one related to their air conditioning. We expect that with richer consumption data from which actual usage patterns could be extracted beyond seasonal averages, more refined assertions could be made about the determinants of enrollment.

The present analysis only considered enrollment into *any* energy efficiency programs; it is likely that analyzing specific programs geared towards more specific types of consumers will yield more refined distinctions in the important variables (such as rebates for insulation as opposed to efficient appliances).

4.3. Associating patterns to segments

To define segments we have used results of prior behavioral research and extensive interaction with the energy utility that provided the data. The utility wished to identify consumers falling into a small number of segments that it had already defined based on its own internal expertise and research, as well as independent third-party behavioral and marketing studies such as [39]. As described in Section 2 above, the purpose of the segments was twofold: (i) crafting a small number of marketing communications such as standardized emails with appropriate information and framing for each segment, and (ii) identifying consumers corresponding to each segment that were likely to enroll in an energy efficiency program.

Based on this prior art, the utility believed that consumers fall into $K = 5$ segments: "Green Advocates", "High Consumption", "Home Improvers", "Cost Conscious", and "Cultural Drivers". The segment meaning that encode this hypothesis are summarized in Table 3. There, listed are the number of variables (out of the P total) corresponding to each of the segments, as well as the number of all the patterns extracted from the data that satisfy the requirements. We also list the average pattern effectiveness q, as well as the coverage (number of consumers) for each of the segments. Here, the effectiveness and coverage correspond to the entire set of possible patterns that could potentially be assigned to each of the segments; the algorithm presented in this paper will select a subset of patterns for each segment that improves q while ensuring a large enough segment coverage and minimizing overlap with other segments. Note, however, that there is overlap between segments, as discussed in Section 5 below.

Given these segment definitions, we associated potential patterns P from $\mathcal{P}_0$ to different segments by ensuring that each pattern P was δ-consistent (see Section 2) with the hypothesis about the meaning of the respective segment. That is, for a given segment S we found those rules $P \in \mathcal{P}_0$ that contained at least δ base rules $P_j \in R$ that matched both in the variable j and in the direction (either greater than or smaller than a threshold learned from the data). The resulting set of patterns $\mathcal{P}$ contained $M = 219$ patterns. Not all consumers were covered by the reduced set of patterns $\mathcal{P}$, with $N = |\cup_{P \in \mathcal{P}} P| = 614,830$ (64% of the original sample), but 89% of the enrolled consumers were included in the reduced set.

The number of patterns obtained for each segment, as well as their coverage (number of consumers in the pattern) are also listed in Table 3. While the number of patterns is smaller than the initial ~3000, it is still a non-trivial task to select a small enough number that maximizes effectiveness. The association matrix B that encodes the pattern-to-segment assignment feasibility is illustrated in Fig. 4. Some patterns may belong to multiple segments, as illustrated in Fig. 5. There, the distribution of the number of patterns that cover users is plotted. Most users are covered by a small number of patterns; however there are a small number of users that simultaneously fall into more than 50 patterns.

Two examples of rules extracted from the data and assigned to segments “High Consumption” and “Cost Conscious” are displayed in Fig. 6. The patterns assigned to “High Consumption” contain at least $\delta = 1$ base rules that involves a condition that consumption be greater than a given threshold value.

A summary is presented in Table 4 of the thresholds extracted from the data that we used to define patterns using rules associated with the $K = 5$ segments considered. There, we have listed the variables that form base rules, along with the direction (either “High” or “Low”) of the rule, and the average values and standard deviations of the thresholds obtained from data. We have only listed the variables that were associated with segment hypotheses, according to the statements defining the segments, see Section 2. The interpretation of the high and low values listed in the table is that, in order for a given variable to be considered as having a “high” (respectively, “low”) value, it has to be greater (lower) than the corresponding *High* (*Low*) threshold. For example, for our segmentation design, the patterns that contain rules related to the total number of rooms in the house indicate that, for predicting EE program enrollment, a high number of rooms is more than 6. Similarly, the current loan-to-value ratio is considered to be high from a predictive standpoint if it is greater than 9.8.

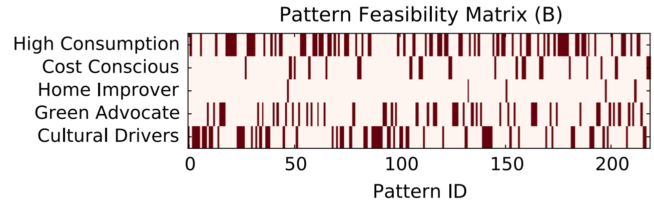

Fig. 4. Pattern feasibility matrix *B*.

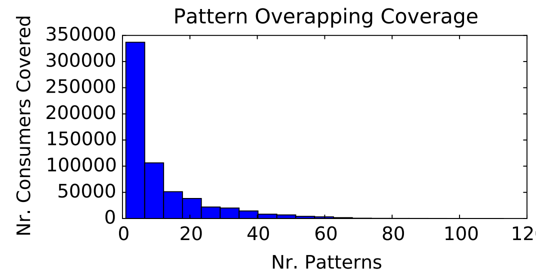

Fig. 5. Pattern overlap distribution.

5. Results

We used Algorithm 1 to obtain an optimally feasible allocation of patterns to segments Z in the case where $(\overline{\pi} = 5,\ \underline{\pi} = 1)$. The algorithm narrows the search region from $[0, 1]$ (of width ϵ_0) down until convergence in 14 iterations, when $|u - l| < \epsilon = 10^{-14}$. The bisection search process is illustrated in Fig. 7. The resulting optimal allocation matrix is displayed in Fig. 8. There, the horizontal axis orders patterns by an arbitrary ID number in the same format as that used in Fig. 4 to represent the allowable assignment matrix B. The algorithm has selected a small number of patterns with the best effectiveness properties that satisfy the constraints in (F_0).

The optimal solution contains 10 patterns spread out across the 5 segments. Table 5 summarizes the effectiveness and size of the

Table 3
Segment definition and statistical description of the associated patterns extracted from data.

Segment	Meaning	# Variables	# Patterns	q (%)	Coverage
High consumption	Large annual or monthly kW h True Brown	7	79	8.2	416,018
Cost conscious	Small home Low income Non-professional Home ownership is relevant Marital status is relevant	20	24	8.3	431,020
Home improver	Home owner Large financial stake in home Long-term occupant	8	5	10.2	81,340
Green advocate	Educated (college or above) High income Professional occupation	17	50	8.3	373,834
Cultural drivers	Ethnicity is relevant Language is relevant Religion is relevant	90	62	9.1	308,146

```
Example patterns: High Consumption
    Spouse Combined Age less than 53.50  &
      kWh-Spring-2014 less than 277.19  &
        Home Total Rooms greater than 6.50  &
          kWh annual greater than 5593.38  &
              *** Rate: 0.104; Support:   2134 ***
    kWh annual greater than 6009.88  &
      kWh-Summer-2013 greater than 726.16  &
        Birth Year greater than 1962  &
          Presence of Child Age 0-18 V3 is Confirmed &
            Year Built greater than 1907  &
                *** Rate: 0.112; Support:  56398 ***
```

```
Example patterns: Cost Conscious
    Home Total Rooms greater than 6.50  &
      Dwelling Type is Single Family  &
        Birth Year less than 1968  &
          Home Total Rooms less than 9.50  &
              *** Rate: 0.129; Support:  21641 ***
    Spouse Gender Code is Female  &
      Dwelling Type is Single Family  &
        Person Type is Other  &
          Current Home Value greater than $269,015  &
              *** Rate: 0.103; Support:   6228 ***
```

Fig. 6. Example patterns associated with two segments, “High Consumption” and “Cost Conscious”.

Table 4
Summary of threshold values used for base decision rules. Here, in order for a given variable to be considered as having a “high” value, it has to be greater than the corresponding *High* threshold; conversely, a variable has a “low” value if it is lower than the *Low* threshold.

Variable	Direction	Avg.	Std.
Birth year	High	1969.5	0.0
	Low	1959.5	0.0
Combined age	High	53.5	0.0
	Low	49.5	5.7
County estimated income index	High	90.5	0.0
County income percentile	High	56.0	12.5
EST CUR LOAN-TO-VALUE RATIO RG	High	9.8	25.3
ESTIMATED AVAILABLE EQUITY RNG	High	308,746.9	38,549.4
	Low	2625.0	0.0
Est Household Income V5	High	103,571.4	36,421.6
	Low	750.0	0.0
Estimated current home value	High	330,438.5	0.0
Home total rooms	High	6.5	0.0
Inc: Family Inc State Decile	High	5.2	0.4
Number of adults in household	High	4.5	0.0
	Low	1.5	0.0
Number of persons on record	High	1.5	0.0
Spouse combined age	Low	76.0	17.1
Spouse ethnicity	Low	4.8	1.7
State estimated income index	High	125.5	42.2
Year Built	High	1985.5	2.0
	Low	1926.0	19.5
kW h annual	High	6218.7	1626.0
kW h-Autumn-2013	High	458.9	121.9
kW h-Spring-2014	High	373.5	117.3
	Low	236.8	106.2
kW h-Summer-2013	High	508.6	149.7
kW h-Summer-2014	High	228.6	177.9
	Low	319.3	82.9
kW h-Winter-2013	High	386.1	79.8
	Low	1643.4	0.0

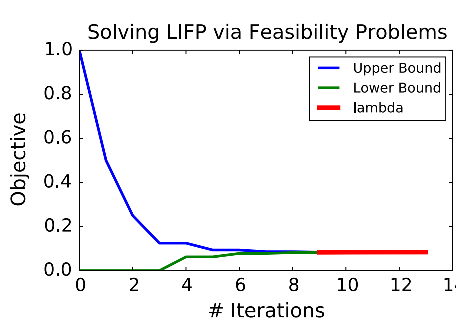

Fig. 7. Finding a maximum lower bound λ on segment effectiveness by iteratively solving a feasibility problem (LFIP) with $\overline{\pi} = 5$ and $\underline{\pi} = 1$.

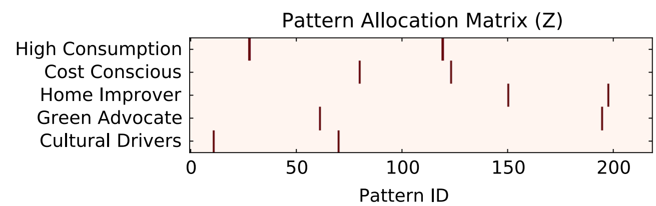

Fig. 8. Pattern-to-segment allocation matrix Z for $\overline{\pi} = 5,\ \underline{\pi} = 1$.

Table 5
Caption.

	High consumption	Cost conscious	Home improver	Green advocate	Cultural drivers
$q(S_k)$	0.101	0.120	0.123	0.104	0.145
$\lvert S_k \rvert$	119,520	113,310	63,768	141,304	7881

resulting segments. The final effectiveness numbers are all greater than $2 \times q_0$, with consumers assigned to one segment (“Cultural Drivers”) enrolling at almost three times the rate in the overall population.

Fig. 9 shows examples of the overlap between segments. This overlap is induced because the patterns themselves that make up the segments may and do overlap in the customers they cover. Segment overlap is however a natural concept in reality, as consumers may have certain traits that may ascribe them to one segment (such as “Cost Conscious”), while other traits are shared with consumers in a different segment (such as “Home Improver”). Our segmentation technique transparently accounts for this situation. A more exhaustive view of segment overlap is presented in Fig. 10 as a network plot. There, each segment is represented as a node of a size proportional to the number of customers in that segment; the weight of the links between the segments represents the pairwise overlap of the segments. As the constraints are changed from $(\overline{\pi} = 4,\ \underline{\pi} = 1)$ (left panel) to $(\overline{\pi} = 5,\ \underline{\pi} = 2)$ (right panel), the structure of the segmentation changes as more patterns are used to construct some of the segments.

Note that “segments” are an abstract concept that are defined by the program administrator as to aid with creating and managing communications that differentiate among consumers to some extent while keeping operational cost and complexity low. They uncover some heterogeneity, but at the same time do not allow for fully tailoring an intervention down to the individual. Imposing that every consumer belongs to one segment only imposes unrealistic assumptions, which our approach circumvents.

In Fig. 11 we list the patterns defining the segments that correspond to the optimal pattern allocation for $(\overline{\pi} = 5,\ \underline{\pi} = 2)$. The hypotheses in Table 3 about the meaning of each segment are enriched with specific information such as thresholds t_j (defining precisely what “high” and “low” mean) and additional base rules.

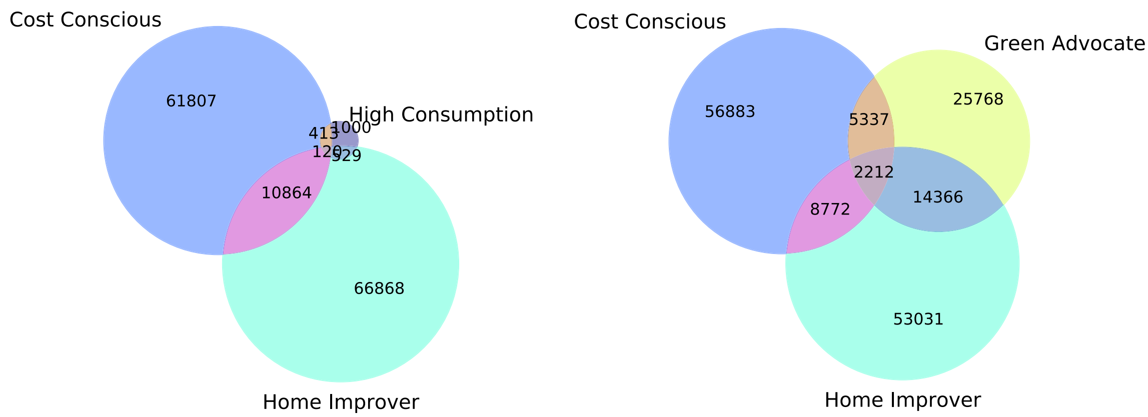

Fig. 9. Segment overlap examples. There are customers in the sample that are both "Cost Conscious" and "High Consumers", or "Green Advocates" and "Cost Conscious".

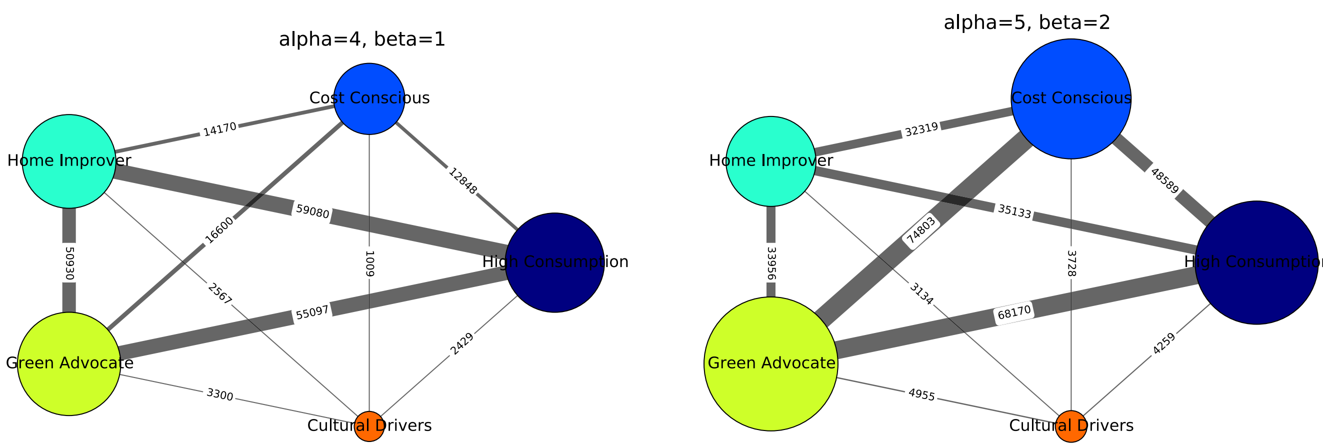

Fig. 10. Segment size and overlap. *Left:* ($\overline{\pi} = 4,\ \underline{\pi} = 1$); *Right:* ($\overline{\pi} = 5,\ \underline{\pi} = 2$).

For example, one type of "Home Improvers" who enroll in energy efficiency programs at a high rate are South Asians who earn more than $75,000 a year, and who own an equity on their house of more than $306,870. Similarly, one type of "Green Advocates" are families that earn more than $75,000 a year, making at least two and a half times the average income level for their state, who have children, and don't live in multi-family accommodations. The patterns in each segment may then be used to design marketing communications specific to that segment, as to include elements which consumers in that segment are seen to be responsive to. Moreover, the specificity of the patterns (in terms of thresholds learned from data) allows to target those consumers that are most likely to enroll.

From the discussion above it is clearly apparent that the structure of the segmentation obtained depends strongly on the nature of the constraints, specifically on the values of $\overline{\pi}$ and $\underline{\pi}$. To study this dependence, we ran Algorithm 1 for a grid $(\overline{\pi},\ \underline{\pi})$ where $1 \leqslant \overline{\pi} \leqslant 9$ and $1 \leqslant \underline{\pi} \leqslant \overline{\pi}$. The optimum value of the objective $\lambda^*(\overline{\pi}, \underline{\pi})$ and the number of patterns selected for the segmentation are illustrated in Fig. 12. The best results are obtained when $\underline{\pi} = 1$ (so the algorithm does not force more patterns into segments than necessary). Good results ($\lambda^* \approx 12\%$) are shown to be obtained for moderate to large values of $\overline{\pi}$ (4–8) and low values of $\underline{\pi}$ (1–3). These maps thus offer a guideline of how to trade off model complexity and segmentation effectiveness.

For a given value of $\underline{\pi}$, we study how the objective λ^* and the individual segment effectiveness values q_k $k = 1, \ldots, K$, vary with $\overline{\pi}$. Then this can serve as tuning parameter for the complexity of the resulting segmentation, which can be designed to accommodate desired effectiveness values of individual segments of interest. This is illustrated in Fig. 13 for a value of $\underline{\pi} = 2$. For example, if the emphasis falls on "Cultural Drivers", a segmentation with $\overline{\pi} \in \{3, 4, 5\}$ is preferred. Note that for all values of k, q_k is distinctly greater than λ^*.

Lastly, we illustrate the dependency of individual segment effectiveness on the segmentation complexity (i.e., total number of patterns selected across segments) in Fig. 14. This highlights the best possible effectiveness values that can be achieved for a fixed, given value of segmentation complexity. For example, if the efficiency program manager wishes to select a total number of patterns between 20 and 25, he/she can expect the optimum effectiveness of the "Cultural Drivers" segment to always be greater than that of the "Cost Conscious" segment. For that range of $\overline{\pi}$, the "Home Improvers", "Green Advocates" and "Cultural Drivers" have all an effectiveness value around 11%.

6. Conclusions

Designing and running energy efficiency programs has developed into a key component of the environmental and financial strategies of energy utility companies. Pushed to innovate by ever more demanding environmental regulations, utility companies are increasingly looking to use data analytics to understand how to improve key metrics of performance, such as customer engagement with energy programs, while more wisely spending their operational budgets. Through extensive discussions with our utility company collaborators we learned that "black-box" algorithms for predicting the likelihood of adopting energy

```
======== High Consumption ========
Rate: 0.199; Support: 2062
kWh-Summer-2014 less than 410  &
  kWh annual greater than 4420  &
    kWh-Autumn-2013 greater than 553  &
      kWh-Summer-2013 greater than 917  &
        EST CUR LOAN-TO-VALUE RATIO RG greater than 2.50  &
            *** Rate: 0.277; Support:    786 ***
--- OR ---
Presence of Child Age 0-18 is Confirmed  &
  Spouse Religion is not Catholic  &
    kWh-Spring-2014 greater than 275  &
      Language is Hindi  &
        kWh-Summer-2013 greater than 526.50  &
            *** Rate: 0.187; Support:   1334 ***
```

```
======== Cultural Drivers ========
Rate: 0.147; Support: 3971
Spouse Combined Age less than 53.50  &
  Ethnicity is not Hispanic  &
    kWh annual greater than 4530  &
      General Ethnicity is South Asian  &
        Language is not English  &
            *** Rate: 0.148; Support:   3110 ***
--- OR ---
Marital Status is not Single  &
  Home Heat Ind is not No heat  &
    County Income Percentile greater than 68.50  &
      Spouse Language is Hindi  &
        Language is not Urdu  &
            *** Rate: 0.156; Support:   1652 ***
```

```
======== Home Improver ========
Rate: 0.102; Support: 78381
ESTIMATED AVAILABLE EQUITY greater than $262,500  &
  ESTIMATED AVAILABLE EQUITY greater than $306,870  &
    Est Household Income V5 greater than $75,000  &
      General Ethnicity is South Asian  &
        kWh-Winter-2013 less than 1643.43  &
            *** Rate: 0.160; Support:   1755 ***
--- OR ---
ESTIMATED AVAILABLE EQUITY greater than $262,500  &
  ESTIMATED AVAILABLE EQUITY greater than $306,870  &
    kWh-Autumn-2013 greater than 367  &
      Home Total Rooms greater than 6.50  &
        State Estimated Income Index greater than 112.50  &
            *** Rate: 0.102; Support:  77723 ***
```

```
======== Green Advocate ========
Rate: 0.119; Support: 47683
Est Household Income V5 greater than $75,000  &
  kWh-Autumn-2013 greater than 358.66  &
    Presence of Child Age 0-18 is Confirmed  &
      State Estimated Income Index greater than 259  &
        Dwelling Type is not Multi-Family & Condominiums  &
            *** Rate: 0.129; Support:  36024 ***
--- OR ---
EducationSpouse is High School Diploma  &
  Est Household Income V5 greater than $75,000  &
    County Income Percentile greater than 52.50  &
      ESTIMATED AVAILABLE EQUITY less than $262,500  &
        Birth Year greater than 1967  &
            *** Rate: 0.102; Support:  12708 ***
```

```
======== Cost Conscious ========
Rate: 0.124; Support: 73204
Inc: Family Inc State Decile greater than 5.50  &
  Dwelling Type is not Marginal Multi-Family  &
    Dwelling Type is Single Family  &
      Spouse Combined Age less than 53.50  &
        Combined Age less than 41.50  &
            *** Rate: 0.176; Support:  26720 ***
--- OR ---
Est Household Income greater than $75,000  &
  Occupation Group is not Engineering/Computers/Math  &
    Est Household Income less than $100,000  &
      Numer of persons on record greater than 1.50  &
        County Estimated Income Index less than 121.50  &
            *** Rate: 0.101; Support:  50326 ***
```

Fig. 11. Segment definitions corresponding to the optimal pattern allocation.

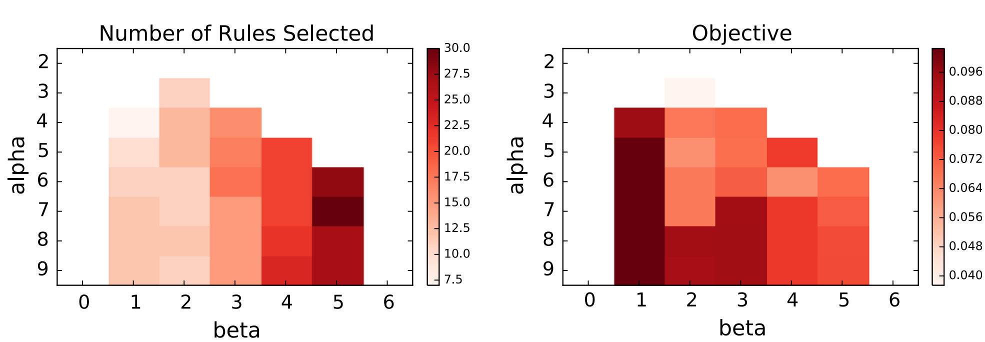

Fig. 12. Sensitivity analysis: dependence of objective λ (*left*) and number of selected patterns (*right*) on $\overline{\pi}$ and $\underline{\pi}$.

conservation measures were not transparent to program and marketing managers, who most of the times are not trained in digesting and understanding machine learning concepts. As a result, anecdotal evidence suggests that such methods typically end up not being used at all. Instead, we found that current practice in market segmentation at utilities is often to segment consumers according to certain characteristics (home square footage, income group, etc.) using simple logical rules that encode existing knowledge (hypotheses based on best-practices and experience) about customer types. We used this intuitive understanding of customer

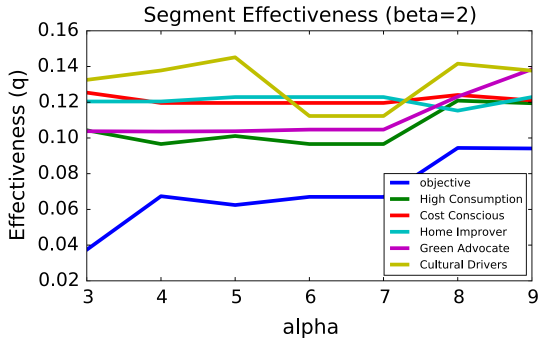

Fig. 13. Segment effectiveness for different values of $\overline{\pi}$ and a fixed $\underline{\pi} = 2$. Generally the optimum effectiveness stays the same or increases when allowing more patterns into segments.

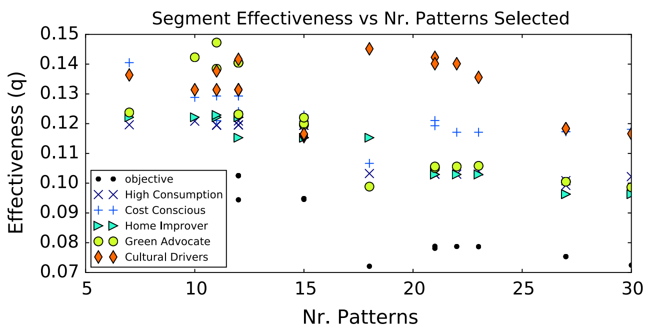

Fig. 14. Segment effectiveness as function of segmentation complexity (number of patterns allocated overall into segments).

types to develop our methodology. In addition to the transparent definition of segments, our approach offers guarantees of optimality (with respect to the empirical probability to enroll) of the obtained segments.

This paper introduced a methodology for programatically constructing interpretable, predictive segmentation of energy consumers. We formulated the *predictive segmentation problem* that is based on first extracting predictive patterns (conjunctions) from data, then optimally allocating the patterns to segments. The segments were defined using prior behavioral and marketing research at an energy utility. We formulated the optimal allocation as solving a generalized (max–min) linear-fractional integer program with linear constraints. To solve this program we proposed an efficient bisection algorithm. We illustrated the method on identifying optimally predictive segments in a population of $\sim 1M$ electricity consumers of a large U.S. energy utility. These segments represent consumers that the utility may craft appropriate messages to, and for which are more effective and economical to target. Generally, the segments identified were defined with a small number of patterns, as the model was able to select optimum pattern allocations to segments from the many possible combinations. In communicating these results to the energy utility company we have obtained the data from, we found that program managers particularly appreciated the ability to construct their own "custom" segments from among the different predictive patterns extracted from the data, and to compare them against the ones that were optimally computed using our algorithm.

6.1. Practical application scenario

A typical scenario in which we envision this method to be used in practice involves an off-line marketing campaign where a reaction from the customer is not easy to obtain instantly (as in, e.g., running online ads). To design an effective segmentation campaign, the utility would go through the following steps:

1. Identify several typologies of people that the Marketing (or Operations) organization believes that their customers follow, based on previous practice and behavioral research studies. Encode this information as a set of statements for each segment, of the type "Income is high" or "House size is small". For the example considered in this paper, the statements that define the 5 segments considered are defined in Table 3.
2. Decide on operational parameters such as the minimum size of a subset of customer that may constitute a segment and the complexity of the segments obtained (the minimum and maximum number of patterns to be included).
3. Gather a dataset of customer attributes (including consumption patterns, socio-demographic information, house characteristics), as well as enrollment information (yes/no) for each consumer, and apply the algorithm in Section 3 to obtain optimal segment definitions.
4. Develop communications and messages specific for each segment, by using both the existing knowledge of the segment meaning, as well as the variables identified as relevant for that segment.

While we have had energy conservation at utility companies as motivating application for our predictive segmentation methodology, the algorithms developed in this paper are certainly more generally applicable. This includes natural extensions to other customer-relationship management (CRM) applications (e.g., for call-centers or sales leads) as well as to purely engineering applications reducible to constructing optimal sets from overlapping subsets (e.g., sensor placement).

References

[1] Li Z, Yang F. Governing millions meters data. In: Power energy society general meeting, 2015 IEEE; 2015. p. 1–5. http://dx.doi.org/10.1109/PESGM.2015.7286131.

[2] Liu X, Nielsen PS. Streamlining smart meter data analytics. In: Proceedings of the 10th conference on sustainable development of energy, water and environment systems.

[3] I.R.M. Association. Marketing and consumer behavior: concepts, methodologies, tools, and applications. IGI Global; 2014. Available from: <https://books.google.com/books?id=RVwFogEACAAJ>.

[4] Yan J, Liu N, Wang G, Zhang W, Jiang Y, Chen Z. How much can behavioral targeting help online advertising? In: Proceedings of the 18th international conference on World wide web. ACM; 2009. p. 261–70.

[5] Bhatnagar A, Ghose S. A latent class segmentation analysis of e-shoppers. J Bus Res 2004;57(7):758–67. Available from: <http://EconPapers.repec.org/RePEc:eee:jbrese:v:57:y:2004:i:7:p:758-767>.

[6] Kwac J, Flora J, Rajagopal R. Household energy consumption lifestyle segmentation using hourly data. IEEE Trans Smart Grid; 2015.

[7] Albert A, Rajagopal R. Thermal profiling of residential energy use. IEEE Trans Power Syst 2015;30(2):602–11. http://dx.doi.org/10.1109/TPWRS.2014.2329485.

[8] Maasoumy M, Rosenberg C, Sangiovanni-Vincentelli A, Callaway DS. Model predictive control approach to online computation of demand-side flexibility of commercial buildings hvac systems for supply following. In: American Control Conference (ACC), 2014. IEEE; 2014. p. 1082–9.

[9] Armel K Carrie, Gupta A, Shrimali G, Albert A. Is disaggregation the holy grail of energy efficiency? the case of electricity. Energy Policy 2013;52(C):213–34. Available from: <http://EconPapers.repec.org/RePEc:eee:enepol:v:52:y:2013:i:c:p:213-234>.

[10] Kolter JZ, Jaakkola T. Approximate inference in additive factorial hmms with application to energy disaggregation. J Mach Learn Res – Proc Track 2012;22:1472–82.

[11] Houde S, Todd A, Sudarshan A, Flora J, Armel KC. Real-time feedback and electricity consumption: a field experiment assessing the potential for savings and persistence. Energy Policy. http://dx.doi.org/10.1109/TCE.2011.5735484.

[12] Kavousian A, Rajagopal R, Fischer M. Determinants of residential electricity consumption: using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. Energy 2013;55:184–94.

[13] Kavousian A, Rajagopal R, Fischer M. Ranking appliance energy efficiency in households: utilizing smart meter data and energy efficiency frontiers to

estimate and identify the determinants of appliance energy efficiency in residential buildings. Energy Build 2015;99:220–30.

[14] Maasoumy M, Sangiovanni-Vincentelli A. Total and peak energy consumption minimization of building hvac systems using model predictive control. IEEE Des Test Comput 2012;29(4).

[15] Beckel C, Sadamori L, Santini S. Automatic socio-economic classification of households using electricity consumption data. In: Proceedings of the 4th international conference on future energy systems (ACM e-Energy '13). Berkeley (CA), USA: ACM; 2013. p. 75–86. note: This paper provides the results of a preliminary analysis. To see (or cite) the full results of our household classification analysis, refer to the paper above with the title Revealing Household Characteristics from Smart Meter Data published in the journal Energy.

[16] Albert A, Rajagopal R. Smart meter driven segmentation: what your consumption says about you. IEEE TransPower Syst 2013;28(4):4019–30. http://dx.doi.org/10.1109/TPWRS.2013.2266122.

[17] Albert A, Rajagopal R. Cost-of-service segmentation of energy consumers. IEEE Trans Power Syst 2014;29(6):2795–803. http://dx.doi.org/10.1109/TPWRS.2014.2312721.

[18] Flath C, Nicolay D, Conte T, van Dinther C, Filipova-Neumann L. Cluster analysis of smart metering data – an implementation in practice. Bus Inform Syst Eng 2012;4(1).

[19] Rasanen T, Kolehmainen M. Feature-based clustering for electricity use time series data. In: Kolehmainen M, Toivanen P, Beliczynski B, editors. Adaptive and natural computing algorithms. Lect notes comput sci, vol. 5495. Berlin/Heidelberg: Springer; 2009. p. 401–12.

[20] Figueiredo V, Rodrigues F, Vale Z, Gouveia J. An electric energy consumer characterization framework based on data mining techniques. IEEE Trans Power Syst, 20(2). http://dx.doi.org/10.1109/TPWRS.2005.846234.

[21] Smith BA, Wong J, Rajagopal R. A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting. In: ACEEE summer study on energy efficiency in buildings; 2012.

[22] Tsekouras G, Hatziargyriou N, Dialynas E. Two-stage pattern recognition of load curves for classification of electricity customers. IEEE Trans Power Syst 2007;22:1120–8.

[23] Espinoza M, Joye C, Belmans R, DeMoor B. Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. IEEE Trans Power Syst 2005;20(3):1622–30. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1490617>.

[24] Alzate C, Espinoza M, Moor B, Suykens JA. Identifying customer profiles in power load time series using spectral clustering. In: Proceedings of the 19th international conference on artificial neural networks: Part II. ICANN '09. Berlin, Heidelberg: Springer-Verlag; 2009. p. 315–24.

[25] Maasoumy M, Sangiovanni-Vincentelli A. Smart connected buildings design automation: foundations and trends. Found Trends Electr Des Autom 2016;10(1–2):1–143. http://dx.doi.org/10.1561/1000000043. Available from: <http://dx.doi.org/10.1561/1000000043>.

[26] Buso T, Fabi V, Andersen RK, Corgnati SP. Occupant behaviour and robustness of building design. Build Environ 2015;94(Part 2):694–703. http://dx.doi.org/10.1016/j.buildenv.2015.11.003. Available from: <http://www.sciencedirect.com/science/article/pii/S0360132315301669>.

[27] Ortiz J, Guarino F, Salom J, Corchero C, Cellura M. Stochastic model for electrical loads in mediterranean residential buildings: validation and applications. Energy Build 2014;80:23–36. http://dx.doi.org/10.1016/j.enbuild.2014.04.053. Available from: <http://www.sciencedirect.com/science/article/pii/S037877881400379X>.

[28] Maasoumy M, Sangiovanni-Vincentelli A. Optimal control of hvac systems in the presence of imperfect predictions. In: Dynamic system control conference (DSCC). ASME; 2014.

[29] Maasoumy M, Razmara M, Shahbakhti M, Vincentelli AS. Handling model uncertainty in model predictive control for energy efficient buildings. Energy Build 2014;77:377–92.

[30] Pedersen M. Market segmentation and energy efficiency program design. In: ACEEE summer study on energy efficiency in buildings.

[31] Moss SJ, Cubed M. Segmenting residential customers: energy and conservation behaviors. CIEE Behavior and Energy Program.

[32] Dethman L, Thomley D. Comparison of segmentation plans for residential energy consumers. Energy Trust of Oregon, Inc.; 2010.

[33] Horst R, Pardalos PM, editors. Handbook of global optimization. Nonconvex optimization and its applications. Dordrecht, Boston: Kluwer Academic Publishers; 1995. Available from: <http://opac.inria.fr/record=b1088635>.

[34] Feng Q, Jiao H, Mao H, Chen Y. A deterministic algorithm for min–max and max–min linear fractional programming problems 2011;4:134–41. http://dx.doi.org/10.1080/18756891.2011.9727770.

[35] Schaible S, Shi J. Recent developments in fractional programming: single ratio and max–min case; 2004.

[36] Boyd S, Vandenberghe L. Convex optimization. New York (NY), USA: Cambridge University Press; 2004.

[37] Patel S, Sevlian R, Zhang B, Rajagopal R. Pricing residential electricity based on individual consumption behaviors. Available from: arXiv:1312.1243.

[38] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer series in statistics. Springer; 2009. Available from: <https://books.google.com/books?id=tVIjmNS3Ob8C>.

[39] Frankel D, Tai H, Heck S. Using a consumer-segmentation approach to make energy-efficiency gains in the residential market, McKinsey and Co. Research Study; 2013.