Applied Energy 133 (2014) 298–307

Contents lists available at ScienceDirect

Applied Energy

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

Predicting winning and losing businesses when changing electricity tariffs

Ramon Granell [a], Colin J. Axon [b,1], David C.H. Wallom [a,*]

[a] Oxford e-Research Centre, University of Oxford, 7 Keble Road, Oxford OX1 3QG, United Kingdom
[b] School of Engineering and Design, Brunel University, Uxbridge, London UB8 3PH, United Kingdom

HIGHLIGHTS

- We have used a data set of 12,000 UK businesses representing 44 sectors.
- We used only 3 features to predict the winners and losers when switching tariffs.
- Machine learning classifiers need less data than regression models.
- Prediction accuracies of the winning and losing businesses of 80% were typical.
- We show how the accuracy varies with the amount of power demand data used.

ARTICLE INFO

Article history:
Received 4 April 2014
Received in revised form 24 July 2014
Accepted 25 July 2014
Available online 17 August 2014

Keywords:
Energy
Tariff switching
Classification
Neural Networks
Support Vector Machines
Regression models

ABSTRACT

By using smart meters, more data about how businesses use energy is becoming available to energy retailers (providers). This is enabling innovation in the structure and type of tariffs on offer in the energy market. We have applied Artificial Neural Networks, Support Vector Machines, and Naive Bayesian Classifiers to a data set of the electrical power use by 12,000 businesses (in 44 sectors) to investigate predicting which businesses will gain or lose by switching between tariffs (a two-classes problem). We have used only three features of each company: their business sector, load profile category, and mean power use. We are particularly interested in the switch between a static tariff (fixed price or time-of-use) and a dynamic tariff (half-hourly pricing). We have extended the two-classes problem to include a price elasticity factor (a three-classes problem). We show how the classification error for the two- and three-classes problems varies with the amount of available data. Furthermore, we used Ordinary Least Squares and Support Vector Regression models to compute the exact values of the amount gained or lost by a business if it switched tariff types. Our analysis suggests that the machine learning classifiers required less data to reach useful performance levels than the regression models.

1. Introduction

Uncertainty in global energy markets is leading to volatility of the prices that consumers pay for gas and electricity. Wholesale and retail energy prices have dropped recently in the USA, but are rising in many other nations [1]. For small and medium-sized businesses energy may form a significant cost, particularly in a recession. From the perspective of both individual businesses and energy providers (retailers), the ability to analyse energy use patterns (demand profiles) is important for economic and energy efficiency. For an individual business, the trade-off between cost and stability of price may be the most important factor. For the retailer, the ability to offer novel tariff structures to suit different types of organisations *e.g.* small shops or schools may be a way to differentiate themselves in a liberalised energy market [2]. Furthermore, different tariff structures may provide scope for improved network management *e.g.* load balancing by system operators [3–5]. The widespread deployment of cheap ICT for monitoring and sensing is making near-to-real-time data availability possible which is creating opportunities for machine learning and data mining techniques to be applied to this rich source of data. This is principally occurring in the electricity distribution sector. These factors are provoking interest in flexible tariffs.

There are a wide variety of tariff types used by electricity retailers [6]. We are examining three broad classes of tariff: fixed price, time-of-use, and real-time. A fixed price tariff (FPT) – the energy price is constant during all 24 h periods throughout the year. The

* Corresponding author. Tel.: +44 (0)1865 610601.
E-mail addresses: ramon.granell@oerc.ox.ac.uk (R. Granell), colin.axon@brunel.ac.uk (C.J. Axon), david.wallom@oerc.ox.ac.uk (D.C.H. Wallom).
[1] Tel.: +44 (0)1895 267932.

http://dx.doi.org/10.1016/j.apenergy.2014.07.098
0306-2619/

time-of-use tariff (TOUT) has different prices during some periods of the day (*e.g.* evening peak), but is the same for all days. The FPT and TOUT can be considered as static tariffs. A dynamic tariff, or real-time tariff (RTT) has a varying price on a basis of *e.g.* 30 or 60 min, with the price for each interval dependent on the demand expected and the availability of generators.

The consequence of a FPT is customers with demand when the electricity is cheaper subsidise customers with higher demand during peak periods. The RTTs will represent a more realistic pricing scheme. If switching to a RTT some customers would obtain benefit (be winners) whilst others would pay more (be losers) depending on their demand profile (a wealth transfer [7]). We have investigated how to predict which customers will win or lose when they change from a FPT or TOUT to a RTT based on their real behaviour. Businesses and light industries present highly heterogeneous energy consumption patterns, both within and between business sectors. The most frequent tariff change studied is from FPT to TOUT [7–11]. In [12], Norwegian houses are automatically assigned a critical peak tariff depending on outside temperature and their consumption pattern, and in [13] the longer term effects of households switching to TOUT have been studied. However, some analyse the change from static to dynamic tariffs [7,11]. These studies are usually performed using residential data; with only [7,8] using commercial data.

Our analysis goes beyond this to predict if a businesses is a winner or loser with the tariff change and by how much. The interest (and difficulty) in constructing this model lies in using only the basic pieces of information that are available in the electricity bill. This restriction is a significant constraint that has not been tackled previously due to the lack of (high resolution) electricity consumption data split by the type of business. We used machine learning techniques to perform experiments over an original data set of more than 12,000 UK businesses from 44 diverse commercial and industrial sectors.

Machine learning techniques have been applied in comparative tariff studies for some specific market such as insurances [14]. However, it is not common to apply machine learning to energy economics. In this area, [15] developed a tariff selection process algorithm (for FPT, RTT or TOUT) based on a Partially Observed Markov Decision Process and performed experiments over a 60 agent model simulating domestic customers. Another agent-based model to select the energy tariff that maximises savings for houses using Bayesian quadrature is developed by [16]. Our approach is different as we are not simulating the behaviour, but classifying it between winners and losers with the tariff changes using real data and employing Support Vector Machines, a Naive Bayes Classifier and Neural Networks. For predicting the quantity of the win or loss we used linear regression and Support Vector Regression models.

This article is structured in the following sections. First, we describe the data set and the pre-processing required to perform the experiments. Second, we define the different tariff schemes and the tariff switches that we investigate in Section 3. The prediction problems and the machine learning classifiers and regression models used to solve them are explained in Section 4. The experiments and their results are analysed in Section 5. The last section draws conclusions from our findings and proposes some ideas for future work.

2. The data set

The data set comprises half-hourly electricity use for 12,056 different UK businesses from 2006 to 2010. As almost all of the records have missing values or error signals due to loss of supply or other interruptions, we performed a pre-process to guarantee sufficient quality in the data set. The four stage process was:

1. Only readings from 2009 to 2010, where most of the businesses provide data were selected.
2. Readings whose values are less or equal to zero or with repeated time stamp were removed (around 11% of the readings).
3. For each business, readings whose values are higher than both the mean plus three times the standard deviation, and 10 kW h were purged (around 0.2% of the readings).
4. The businesses that do not contain at least ten different values in their readings are removed (1129 businesses were purged).

After this filtering process there were 10,926 businesses meeting our criteria. Subsequently, some of the businesses did not have sufficient readings to be considered representative. However, they were used for comparison. Subsets (of the full data set) were created using a threshold τ of the minimum number of readings available per business. Values of τ threshold range from half a month of readings ($48 * 30/2$) to 12 months of readings ($48 * 365$) creating different versions of the data set, removing the businesses with less than τ readings. These readings do not need to be consecutive, with some being spread during the two years period. A greater number of readings indicates a better representation of the energy behaviour of the business. Table 1 shows the averaged number of reading per business for different τ values.

The features that we are going to use are available on customer bills. From the data set, we are going to use the following set of features for each business:

Business Sector
There is a total of 44 different sectors of commercial and industrial activities. Although we used all of them for our experiments, we grouped them in five generic categories to preserve anonymity. Table 2 describes these sectors and groups. The percentage of businesses belonging to each category for the data set with different τ can be seen in Table 3 – Retail is the largest group and Social the smallest.

Mean of Energy-use
This is the mean for all the half hour readings of each business. As a reference, the average over the means for all the businesses of the data set with τ of half of month and one year are 2.87 kW h and 3.22 kW h respectively. For other values of τ, the mean is between these two values, increasing slightly with τ. The standard deviation is approximately 2 kW h.

Load Profile Category
This corresponds to one of the profile codes 05, 06, 07 and 08 that are the first two cyphers of the meter point administration number available in the standard British electricity bill. The meaning of these codes is shown in Table 4. Therefore, for computing the category of each business, first we need to calculate its load factor value using: $100 * (\text{mean energy use})/(\text{maximum energy use})$. The maximum energy use was computed by averaging the three maximum readings of each business. The percentage of businesses per load profile category are shown in Table 3; the distribution is quite even.

The business sector and load factor categories are discrete variables, whilst the mean of the energy use is continuous.

3. Addressing tariff changes

We have chosen three types of tariff for this study. Although many variations of these could be used, they represent the main broad classes of tariff. Moreover, they have relevance for the energy distribution network operators and electricity retailers.