Target Tracking for Contextual Bandits: Application to Demand Side Management

Margaux Brégère[1,2,3], Pierre Gaillard[3], Yannig Goude[1,2], Gilles Stoltz[2]

[1]EDF R&D, [2]Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, CNRS, [3]Inria

Introduction

Motivation

As electricity is hard to store, balance between production and demand must be maintained at any time.

Current solution.
Forecast consumption and adapt production accordingly

- As renewable energies are subject to climate, the production becomes hard to adjust
- New communication tools (smart meters) will provide to data access and instantaneous communication

Prospective solution.
Send incentive signals (electricity tariff variations) to manage demand response

Problem

How to optimize these signals learning from clients behaviors?
Learn from clients behaviors & Optimize tariffs sending
Exploration - Exploitation trade-off

Idea

Apply contextual-bandit theory to demand side management by offering price incentives

Setting and model

Modeling of the electricity consumption

Some exogenous factors (temperature, day, etc.) form, at an instance t, a context vector $x_t \in \mathcal{X}$. The individual consumption of a customer getting the price level $j \in \{1, \dots, K\}$ is $\psi(x_t, j)$ + white noise. A share $p_{t,j}$ of customers receives tariff j at t. The mean consumption observed for an homogeneous population, with $p_t \in \mathcal{P} \subset \{(p_1, \dots, p_K) \in [0,1]^K \,|\, \sum_k p_k = 1\}$, equals

$$Y_{t,p_t} = \sum_{j=1}^{K} p_{t,j}\psi(x_t, j) + \text{noise}.$$

Assumption.
There is an unknown vector θ and a known function ϕ such that $\sum_{j=1}^{K} p_{t,j}\psi(x_t, j) = \phi(x_t, p_t)^T\theta$.

Model 1. Tariff-dependent noise

$$Y_{t,p_t} = \phi(x_t, p_t)^T\theta + p_t^T\varepsilon_t$$

Sub-Gaussian i.i.d noise vectors $\varepsilon_1, \dots, \varepsilon_t$ with $\mathbb{E}[\varepsilon_1] = (0, \dots, 0)^T$ and $\mathrm{Var}(\varepsilon_1) = \Gamma$

Model 2. Global noise

$$Y_{t,p_t} = \phi(x_t, p_t)^T\theta + e_t$$

Sub-Gaussian i.i.d scalar noises $e_1, \dots, e_t$ with $\mathbb{E}[e_1] = 0$ and $\mathrm{Var}(e_1) = \sigma^2$

Target Tracking for Contextual Bandits

Inputs
Parametric context set $\mathcal{X}$
Set of legible convex weights $\mathcal{P}$
Transfer function ϕ: $\mathcal{X} \times \mathcal{P} \to \mathbb{R}^d$
Bound on mean consumptions C

Unknown parameters
Transfer parameter $\theta \in \mathbb{R}^d$
Covariance matrix $\Gamma \in \mathcal{M}_K(\mathbb{R})$ (Model 1) or Variance σ^2 (Model 2)

At t=1,2,...
- Observe a context $x_t \in \mathcal{X}$ and a target $c_t \in (0, C)$
- Choose an allocation of price levels $p_t \in \mathcal{P}$
- Observe a resulting mean consumption $Y_{t,p_t} \in (0, C)$
- Suffer a loss $\ell_t = (Y_{t,p_t} - c_t)^2$

Aim. Minimize the cumulative loss

$$L_T = \sum_{t=1}^{T} (Y_{t,p_t} - c_t)^2$$

Results

With $\ell_{t,p} = \mathbb{E}\left[(Y_{t,p} - c_t)^2 \,|\, \text{past}\right]$,
the regret R_T is considered $R_T = \sum_{t=1}^{T} \ell_{t,p_t} - \sum_{t=1}^{T} \min_{p\in\mathcal{P}} \ell_{t,p}$

Model 1	Model 2
$\ell_{t,p} = (\phi(x_t, p)^T\theta - c_t)^2 + p^T\Gamma p$	$\ell_{t,p} = (\phi(x_t, p)^T\theta - c_t)^2 + \sigma^2$

For both models
- Estimate parameter θ with a Ridge regression $\hat{\theta}_{t-1} = \arg\min_{\tilde{\theta}} \sum_{s=1}^{t-1}(Y_{s,p_s} - \phi(x_s, p_s)^T\tilde{\theta})^2 + \lambda\|\tilde{\theta}\|^2$
  thus $\hat{\theta}_{t-1} = V_{t-1}^{-1}\sum_{s=1}^{t-1} Y_{s,p_s}\phi(x_s, p_s)$ with $V_{t-1} = \lambda I_d + \sum_{s=1}^{t-1}\phi(x_s, p_s)\phi(x_s, p_s)^T$
- Create a confidence set $\|\hat{\theta}_{t-1} - \theta\|_{V_{t-1}} \le B_t$

Estimate Γ on the first n rounds

$$\hat{\Gamma}_n = \arg\min_{\tilde{\Gamma}} \sum_{s=1}^{n} (\hat{Z}_t^2 - p_t^T\tilde{\Gamma})^2$$

with $\hat{Z}_t = Y_{t,p_t} - [\phi(x_t, p_t)^T\hat{\theta}_n]_C$
For for $p_1, \dots p_n$ well chosen, $\|\hat{\Gamma}_n - \Gamma\|_\infty \le \gamma_n$

Assumption. Attainability

$$\forall t \ge 1, \exists p_t^\star \in \mathcal{P}, \phi(x_t, p_t^\star)^T\theta = c_t$$

and thus $\min_{p\in\mathcal{P}} \ell_{t,p} = \ell_{t,p_t^\star} = \sigma^2$

- Estimate losses and get a confidence bound for each p thanks to B_t and γ_n

$$\hat{\ell}_{t,p} = \left(\left[\phi(x_t, p)^T\hat{\theta}_{t-1}\right]_C - c_t\right)^2 + p^T\hat{\Gamma}_n p \quad \text{and } \|\hat{\ell}_{t,p} - \ell_{t,p}\| \le \alpha_{t,p}$$

$$\tilde{\ell}_{t,p} = \left(\phi(x_t, p)^T\hat{\theta}_{t-1} - c_t\right)^2 \quad \text{and } \|\tilde{\ell}_{t,p} - \ell_{t,p}\| \le \beta_{t,p}$$

- Select price level optimistically

$$p_t \in \arg\min_{p\in\mathcal{P}}\{\hat{\ell}_{t,p} - \alpha_{t,p}\} \qquad p_t \in \arg\min_{p\in\mathcal{P}}\{\tilde{\ell}_{t,p} - \beta_{t,p}\}$$

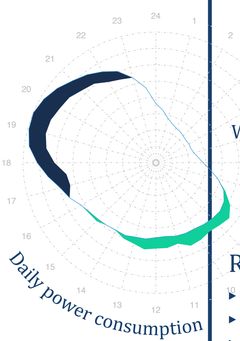
Theorem 1. For proper choices of confidence levels $\alpha_{t,p}$, B_t, γ_n, with $n = o(T^{2/3})$ with probability at least $1-\delta$ the regret is upper bounded as
$R_T \lesssim T^{\frac{2}{3}}\ln^2 T/\delta\sqrt{\ln 1/\delta}$

Theorem 2. For proper choices of confidence levels $\beta_{t,p}$, B_t and regularisation λ, with probability at least $1-\delta$ the regret is upper bounded as
$R_T \lesssim \ln^2(T)$

- Efficient algorithm: sub-linear regret
- Hard computation: non convex minimization problem

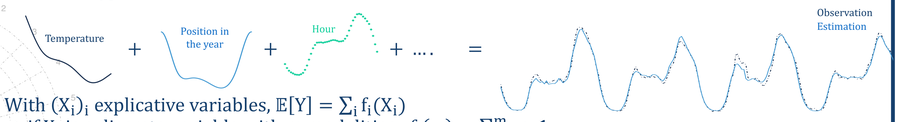
Data*

Consumption at half-an-hour intervals of
1 100 clients subjected to Dynamic Time of Use energy prices
three tariffs: Low, Normal, High



A generalized additive model for consumption

$$Y_{t,p_t} = f_1(\text{temperature}) + f_2(\text{position in the year}) + f_3(\text{hour}) + f_4(p_t) + \dots + \text{noise}$$

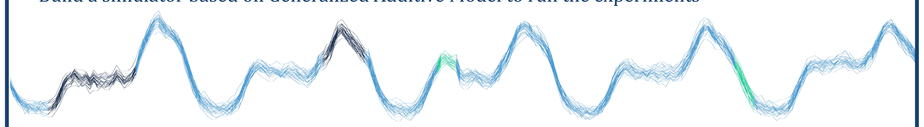


With $(X_i)_i$ explicative variables, $\mathbb{E}[Y] = \sum_i f_i(X_i)$
- if X_i is a discrete variable with m modalities : $f_i(x_i) = \sum_{j=1}^{m} \alpha_j \mathbf{1}_{x_i=j}$
- if X_i is a continuous variable, f_i is a spline:C^2-function defined piecewise by polynomials

There is a known transfer function ϕ and an unknown parameter θ such that $\mathbb{E}[Y] = \phi(X)^T\theta$

Realistic simulator. Context + Price level → Mean consumption

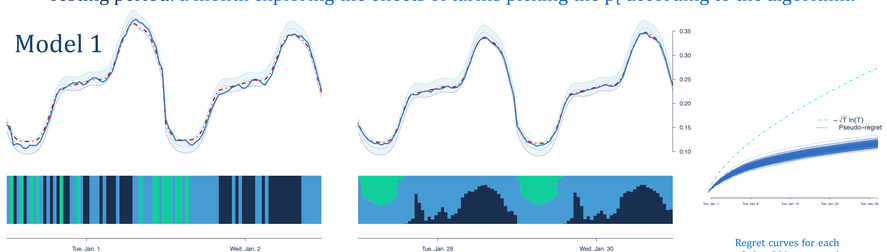
- Select customers with more than 95% of data available and consider their mean consumption
- Estimate covariance matrix Γ
- Build a simulator based on Generalized Additive Model to run the experiments



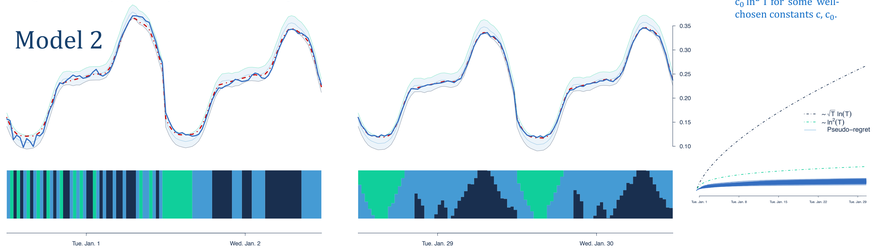
*Low Carbon London - UK Power Networks"Smart Meter Energy Consumption Data in London Households"

Experiments

$\mathcal{P}$ restriction: only two tariffs at the same round. Population split into 100 equal parts. Target creation. smooth attainable targets.
Training period: a year of data using historical contexts with only Normal tariffs picked.
Testing period: a month exploring the effects of tariffs picking the p_t according to the algorithm.



Top: average of mean consumptions over 200 runs (full blue line); target consumption (dashed red line) and mean consumption associated with each tariff (pastel). Bottom: proportions p_t used over time for a single run.



Regret curves for each of the 200 runs and plots of $c\sqrt{T}\ln T$ and $c_0\ln^2 T$ for some well-chosen constants c, c_0.

References

- Auer, P. Using confidence bounds for exploitation- exploration trade-offs, 2002.
- Chu, W., Li, L., Reyzin, L., and Schapire, R. Contextual bandits with linear payoff functions., 2011.
- Goude, Y., Nedellec, R., and Kong, N. Local short and middle term electricity load forecasting with semi-parametric additive models, 2014.
- Lattimore, T. and Szepesvári, C. Bandit algorithms, 2018.
- Li, L., Chu, W., Langford, J., and Schapire, R. A contextual- bandit approach to personalized news article recommendation, 2010.
- Schofield, J., Carmichael, R., Tindemans, S., Woolf, M., Bilton, M., and Strbac, G. Residential consumer responsiveness to time-varying pricing, 2014.
- Wood, S. Generalized Additive Models: An Introduction with R, 2006.