Pilotage de la consommation électrique par envoi d'incitations tarifaires

Target Tracking for Contextual Bandits: Application to Demand Side Management

Gilles Stoltz

(Membre du bureau de l'Agence maths–entreprises, pour les Pays de la Loire)

ICML 2019 paper with: Margaux Brégère (PhD student), Pierre Gaillard (Inria Paris), Yannig Goude (EDF R&D)

Motivation

Aim: maintain balance between production and consumption

Current solution: forecast consumption and adapt production

Prospective solution: encourage/discourage consumption
by dynamically setting prices

Bandit monitoring: trade-off between

- Learning behaviors of customers (= exploration)
- Optimizing incentives sent (= exploitation)

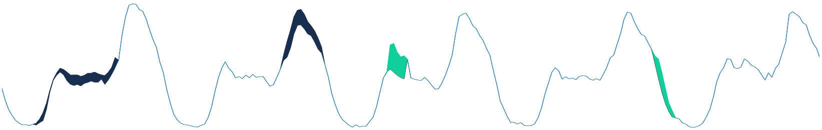
$\longrightarrow$ Stochastic bandit theory should be applicable...

Motivation: data set

"SmartMeter Energy Consumption Data in London Households"
Public dataset – by UK Power Networks

Individual consumptions at half-hourly frequency in year 2013

About 1,000 customers with tariff incentives



K=3 tariffs: Low (L), Normal (N), High (H)

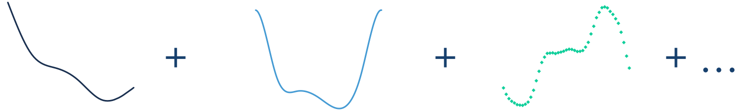
Modeling (1/3): consumption → known and effective methodology

Population assumed to be homogeneous (as a first approach)

(Mean) consumption Y depends on context $x_t \in \mathbb{R}^d$: temperature, season, day of the week, hour of the day, etc.

Also depends on tariff $k \in \{1, \dots, K\}$

$$\mathrm{Y_t} = \mathrm{f_1(temperature)} + \mathrm{f_2(position\ in\ the\ year)} + \mathrm{f_3(hour)} + \mathrm{f_4(tariff)} + \dots + \mathrm{noise}$$



Temperature Position in the year Hour

If single tariff k picked: $$Y_{t,k} = \gamma_k + \sum_{i=1}^{d} f_i(x_{t,i}) + \text{noise}$$

Modeling (1/3): consumption → known and effective methodology

Context $x_t \in \mathbb{R}^d$ and tariff k: $$Y_{t,k} = \gamma_k + \sum_{i=1}^{d} f_i(x_{t,i}) + \text{noise}$$

Generalized additive model → effective modeling

(Wood, 2006; Goude et al., 2014; Gaillard et al., 2016)

- The f_i are cubic splines
- We fix the number q_i of knots and their location
- There exists a basis $b_1^{(i)}, \dots, b_{q_i}^{(i)}$
- We write $f_i = \sum_{1 \leqslant j \leqslant q_i} \beta_j^{(i)} b_j^{(i)}$ for each i

Summary: For x_t and k,

$$Y_{t,k} = \beta^{\mathsf{T}} \varphi(x_t) + \gamma_k + \varepsilon_{t,k}$$

where β and γ_k are unknown, but $\varphi(x_t)$ is known

Modeling (1/3): consumption → extension to various tariffs

If tariffs $\{1, \ldots, K\}$ are distributed in shares $p = (p_1, \ldots, p_K)$

Then (cf. homogeneous population), mean consumption:

$$Y_{t,p} = \sum_{k=1}^{K} p_k Y_{t,k} = \sum_{k=1}^{K} p_k \big(\beta^{\mathsf{T}} \varphi(x_t) + \gamma_k + \varepsilon_{t,k}\big)$$
$$= \theta^{\mathsf{T}} \phi(x_t, p) + p^{\mathsf{T}} \varepsilon_t$$

with θ unknown, but $\phi(x_t, p)$ is known (and linear in p)

Noise: ε_t iid, $\mathbb{E}[\varepsilon_t] = 0$, sub-Gaussian, $\Gamma = \mathrm{Var}(\varepsilon_t)$

We use this model as a data generator to test our bandit strategies

(cf. impossible on historical data!)

Modeling (2/3): target tracking for contextual bandits

Known parameters

- K tariffs
- Context set $\mathcal{X}$
- Transfer function $\phi : \mathcal{X} \times \mathcal{P} \to \mathbb{R}^m$
- Bound C on consumptions Y

Unknown parameters

- Coefficients $\theta \in \mathbb{R}^m$
- Covariance matrix $\Gamma = \mathrm{Var}(\varepsilon_t)$

For each round $t = 1, 2, \dots$

1. Observe a context $x_t \in \mathcal{X}$ and a target $c_t \in [0, C]$
2. Choose an allocation of tariffs $p_t \in \mathcal{P}$
3. Observe a mean consumption $Y_{t,p_t} = \theta^\top \phi(x_t, p_t) + p_t^\top \varepsilon_t$
4. Encounter an error $(Y_{t,p_t} - c_t)^2$

Modeling (2–3/3): target tracking – evaluation

Context x_t and target c_t: allocation p_t and consumption $Y_{t,p_t} = \theta^{\mathsf{T}}\phi(x_t, p_t) + p_t^{\mathsf{T}}\varepsilon_t$

Not maximizing some sum of rewards (as in classical contextual bandits) but minimizing some sum of errors

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

By concentration: $$\approx \sum_{t=1}^{T} \mathbb{E}\big[(Y_{t,p_t} - c_t)^2 \,\big|\, \mathcal{F}_{t-1}\big]$$

where $\mathcal{F}_{t-1}$: information available at the beginning of round t

Modeling (3/3): evaluation

Cumulative error: $$\approx \sum_{t=1}^{T} \mathbb{E}\big[(Y_{t,p_t} - c_t)^2 \,\big|\, \mathcal{F}_{t-1}\big]$$

p_t is $\mathcal{F}_{t-1}$–measurable, $\mathbb{E}[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$ and $\mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1}) = \Gamma$

thus $$\mathbb{E}\big[(Y_{t,p_t} - c_t)^2\big] = \mathbb{E}\Big[\big(\theta^\top \phi(x_t, p_t) + p_t^\top \varepsilon_t - c_t\big)^2\Big]$$
$$= \dots = \underbrace{\big(\theta^\top \phi(x_t, p_t) - c_t\big)^2}_{\text{bias}} + \underbrace{p_t^\top \Gamma p_t}_{\text{variance}}$$

Hence the (conditional) regret:

$$\overline{R}_T = \sum_{t=1}^{T} \big(\theta^\top \phi(x_t, p_t) - c_t\big)^2 + p_t^\top \Gamma p_t$$
$$- \sum_{t=1}^{T} \min_{p \in \mathcal{P}} \Big\{ \big(\theta^\top \phi(x_t, p) - c_t\big)^2 + p^\top \Gamma p \Big\}$$

Minimize cumulative error $\longleftrightarrow$ Minimize regret

Classical approach in contextual bandits

Aim: maximize

$$\sum_{t=1}^{T} Y_{t,p_t} = \sum_{t=1}^{T} \theta^{\mathsf{T}} \phi(x_t, p_t) + p_t^{\mathsf{T}} \varepsilon_t \approx \sum_{t=1}^{T} \theta^{\mathsf{T}} \phi(x_t, p_t)$$

that is, maximize $\displaystyle \sum_{t=1}^{T} \theta^{\mathsf{T}} \phi(x_t, p_t) - \min_{p \in \mathcal{P}} \sum_{t=1}^{T} \theta^{\mathsf{T}} \phi(x_t, p)$

Algorithm LinUCB → optimistic algorithm

(Li et al., 2010; Chu et al., 2011; Abbasi-Yadkori et al., 2011)

- Estimates θ by a confidence region centered at $\widehat{\theta}_{t-1}$ while picking the p_t
- Gets confidence intervals $\widehat{\theta}_{t-1}^{\mathsf{T}} \phi(x_t, p) \pm a_{t,p}$ on the $\theta^{\mathsf{T}} \phi(x_t, p)$
- Picks $\displaystyle \arg\max_{p \in \mathcal{P}} \Big\{ \widehat{\theta}_{t-1}^{\mathsf{T}} \phi(x_t, p) + a_{t,p} \Big\}$

Guarantees a $\widetilde{\mathcal{O}}(\sqrt{T})$ regret bound

Classical approach in contextual bandits

We will re-use the estimation of θ, performed while picking the p_t

For some $\lambda > 0$: at the beginning of round $t \geqslant 2$,

$$\widehat{\theta}_{t-1} \in \mathop{\arg\min}_{\widetilde{\theta} \in \mathbb{R}^m} \left\{ \lambda \left\| \widetilde{\theta} \right\| + \sum_{s=1}^{t-1} \left(Y_{s,p_s} - \widetilde{\theta}^{\mathsf{T}} \phi(x_s, p_s) \right)^2 \right\}$$

(there exists a closed-form expression, cf. regularized OLS)

Under some natural normalization assumptions

namely, $\|\phi\|_\infty \leqslant 1$ as well as $\|\theta\|_\infty \leqslant C$ so that $\phi^{\mathsf{T}}\theta \in [0, C]$

w.p. $1-\delta$, $$\left\| V_{t-1}^{1/2} \left(\widehat{\theta}_{t-1} - \theta \right) \right\| \leqslant \sqrt{\lambda m}\, C + \rho \sqrt{2 \ln \frac{1}{\delta} + m \ln \left(1 + \frac{t-1}{\lambda} \right)}$$

where the Gram matrix $V_t = \lambda \mathrm{I}_d + \sum_{s=1}^{t-1} \phi(x_s, p_s)\, \phi(x_s, p_s)^{\mathsf{T}}$

and where ρ is the sub-Gaussian parameter of the $\varepsilon_1, \varepsilon_2, \ldots$

Our algorithm for tracking with contextual bandits

- Performs initial estimation of $\Gamma = \mathrm{Var}(\epsilon)$ for $\sim T^{2/3}$ rounds getting a confidence region centered at $\widehat{\Gamma}$
- Estimates θ with confidence regions while picking the p_t (exactly like LinUCB proceeds)
- Gets confidence intervals $\Bigl(\bigl[\widehat{\theta}_{t-1}^{\mathsf{T}} \phi(x_t, p)\bigr]_C - c_t\Bigr)^2 + p^{\mathsf{T}} \widehat{\Gamma} p \pm \alpha_{t,p}$ on the conditional errors $\bigl(\theta^{\mathsf{T}} \phi(x_t, p) - c_t\bigr)^2 + p^{\mathsf{T}} \Gamma p$
- Picks $\operatorname*{arg\,min}_{p \in \mathcal{P}} \Bigl\{ \bigl(\widehat{\theta}_{t-1}^{\mathsf{T}} \phi(x_t, p) - c_t\bigr)^2 + p^{\mathsf{T}} \widehat{\Gamma} p - \alpha_{t,p} \Bigr\}$ i.e., plays optimistically

Not satisfactory yet: wish to estimate Γ while picking the p_t

Issue: argmin over a non-convex function of p

$\rightarrow$ restrict $\mathcal{P}$ to a grid in the simulations

Steps 2–4 are straightforward adaptations, but Step 1 was a bit more challenging

Analysis

Mimics the one of LinUCB for steps 2–4

(Li et al., 2010; Chu et al., 2011; Abbasi-Yadkori et al., 2011; Lattimore and Szepesvari, 2019 + blog)

Salient and challenging part: estimation of Γ (step 1)

Iterate for $n \sim T^{2/3}$ rounds over p_t of the form "same tariff for all" or 50%–50% allocation of two tariffs

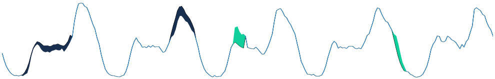
$$\hat{\Gamma}_n \in \operatorname*{arg\,min}_{\hat{\Gamma} \in \mathcal{M}_K(\mathbb{R})} \sum_{s=1}^{n} \left(\left(Y_{s,p_s} - \left[\hat{\theta}_n^{\mathsf{T}} \phi(x_s, p_s) \right]_C \right)^2 - p_s^{\mathsf{T}} \hat{\Gamma} p_s \right)^2$$

We obtain: w.p. $1 - \delta$

$$\begin{aligned} \overline{R}_T = & \sum_{t=1}^{T} \left(\theta^{\mathsf{T}} \phi(x_t, p_t) - c_t \right)^2 + p_t^{\mathsf{T}} \Gamma p_t \\ & - \sum_{t=1}^{T} \min_{p \in \mathcal{P}} \left\{ \left(\theta^{\mathsf{T}} \phi(x_t, p) - c_t \right)^2 + p^{\mathsf{T}} \Gamma p \right\} \quad \leqslant \widetilde{\mathcal{O}}(T^{2/3}) \end{aligned}$$

Realistic simulations

Reminder: we have a historical data set with $K = 3$ tariffs and p_t chosen as Dirac masses



K=3 tariffs: Low (L), Normal (N), High (H)

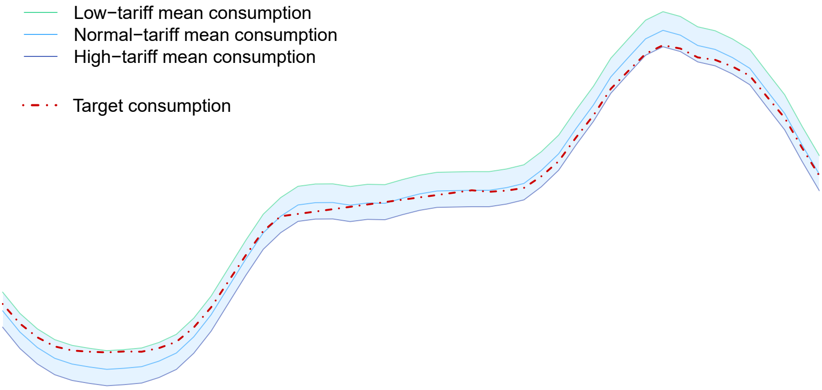
We are bound to generate new data

We estimate our GAM model with 1 year of such data

We then generate new consumption data
based on this model + historical weather variables

And also, we create attainable targets: $\theta^{\mathsf{T}}\phi(x_t, 1) \leqslant c_t \leqslant \theta^{\mathsf{T}}\phi(x_t, 3)$

Realistic simulations

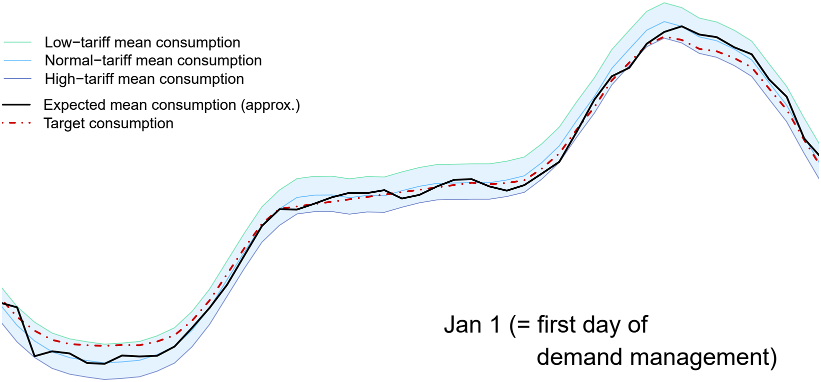


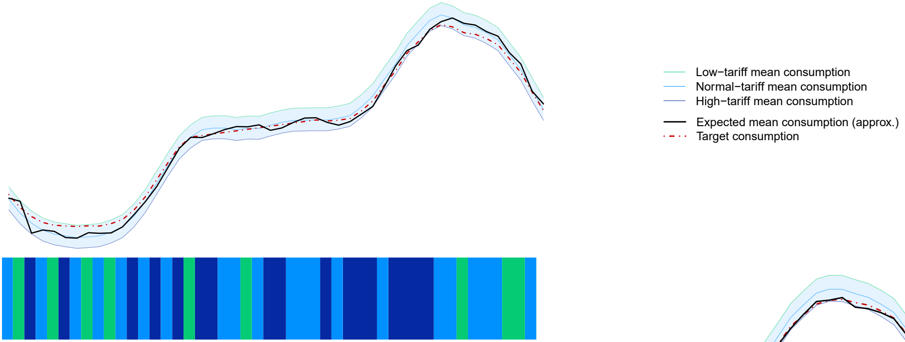
Aim: smooth out consumption

Experiment design → provider changing its policy

- Pick the "normal" tariff for 1 year, i.e., $p_t = (0, 1, 0)$
- Then start picking different allocations with at most 2 tariffs (either 1+2 or 2+3)

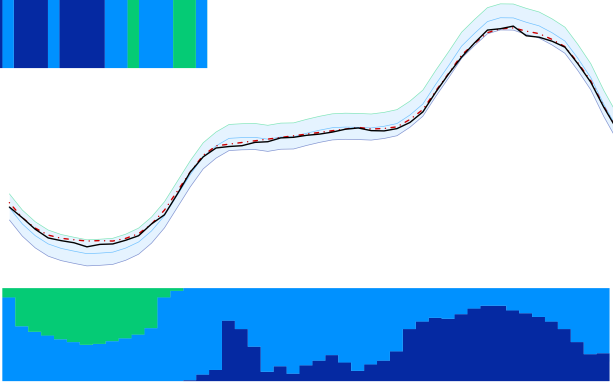
Repeat this 200 times





Top: January 1

Bottom tariff allocations based on a single run



Bottom: January 30

What's next?

The case of inhomogeneous consumers

- Create clusters of clients according to their profiles
- Tailor allocations picked to each cluster

Agence maths–entreprises

Soutient les collaborations recherche entre un laboratoire et une entreprise (en général, TPE–PME–ETI)

Offre phare #1: abondement au contrat de collaboration

Offre phare #2: organisation de semaines d'études pour doctorant.e.s