0911 논문 리뷰

- 연구 목표

$$\mathbb{E}\left[\sum_{t=1}^{\mathcal{T}} \sum_{x \in \mathcal{X}} \sum_{e=1}^{\mathcal{E}} T_{xe}(t) R_x(t)\right].$$

1. **Non-stationarity:** The prevalence $R_x(t)$ is time-varying.
2. **Imbalanced Data:** On average, only 2 in 1000 passengers tests positive, meaning the data $\{P_x(t'), N_x(t')\}_{x \in \mathcal{X}, t' < t-2}$ is very imbalanced with mostly negative tests.
3. **High-Dimensionality:** The number of possible features $\mathcal{X}$ is very large.
4. **Batched decision-making:** All testing allocations for a day must be determined at the start of the day (batch) for each point of entry.
5. **Delayed Feedback:** Labs may take up to 48 hours to return testing results.
6. **Constraints:** Each point of entry is subject to its own testing budget and arrival mix.

- 추정

$$P_k = \sum_{t'=t-16}^{t-3} P_k(t'), \qquad N_k = \sum_{t'=t-16}^{t-3} N_k(t'),$$

과거 16일 ~ 3일 전(딜레이: 2일) 테스트 결과를 참고

- 그룹

L : 위험도를 나타내는 집단(블랙 리스트, 화이트 리스트, 그레이 리스트)

나라 리스트

K : 나라

[승객] C [K] C [L]

→ 현재 클릭 사기 테스트 알고리즘에 대응:

L : 전체를 하나의 그룹으로 취급

K : Kmeans로 20개의 집단 세팅

1 IP → 1 유저

$$\hat{\alpha}_l = \frac{M_{1l}^2(1 - M_{1l})}{M_{2l}^2 - M_{1l}^2} - M_{1l}$$

$$\hat{\beta}_l = \hat{\alpha}_l \frac{(1 - M_{1l})}{M_{1l}}$$

$$M_{1l} = \frac{1}{|\mathcal{L}_l|} \sum_{k \in \mathcal{L}_l} \frac{P_k}{T_k}$$

$$M_{2l} = \frac{1}{|\mathcal{L}_l|} \sum_{k \in \mathcal{L}_l} \frac{P_k\,(P_k - 1)}{T_k(T_k - 1)}$$

$$\widehat{\alpha}_k = \widehat{\alpha}_l + P_k \text{ and } \hat{\beta}_k = \hat{\beta}_l + N_k$$

Subroutine 1: Estimating Prevalence via Empirical Bayes

Input: *# of positives P_k, negatives N_k and tests T_k for each type k in the past 14 days of test results.*

for each color designation of countries $l \in \{\mathcal{W}_t, \mathcal{G}_t, \mathfrak{B}_t\}$

Compute M_{1l}, M_{2l} from Eq. (2) and (3)

$\widehat{\alpha}_l \leftarrow \frac{M_{1l}^2(1 - M_{1l})}{M_{2l}^2 - M_{1l}^2} - M_{1l}, \widehat{\beta}_l \leftarrow \widehat{\alpha}_l \frac{(1 - M_{1l})}{M_{1l}}$

for all types $k \in \mathcal{L}_l$ whose origin country is in l:

$\widehat{\alpha}_k \leftarrow \widehat{\alpha}_l + P_k$ and $\hat{\beta}_k \leftarrow \hat{\beta}_l + N_k$

end

end

Al 을 계산할 때 M2l - M1l**2 대신에 M1l - M2l

- Optimistic Gittins Indices

$$\lambda_k = \frac{\alpha_k}{\alpha_k + \beta_k}\left(1 - \gamma F_{\alpha_k+1,\beta_k}(\lambda_k)\right) + \gamma\lambda_k\left(1 - F_{\alpha_k,\beta_k}(\lambda_k)\right).$$

- 테스트 할당 알고리즘
  - 제한

    Ce : 미리 정해진 공항 e에 제한된 테스트 횟수

    → 클릭 사기 알고리즘에서의 제한: 바로 직전 시간 IP 유저 개수 * 비율(ex: 0.05)로 제한

    Yk : 타입 k승객에 아직 테스트가 안된 승객 명 수

  - 규칙

기튼 지수 람다가 가장 높은 타입 k에 테스트 하나 추가 → 이후 타입 k 파라미터 업데이트

Subroutine 3: Test Allocation Sub-Routine

Input: Posterior distribution estimates $\hat{\alpha}_k, \hat{\beta}_k$ for each type $k \in K_t$, arrivals $A_{ke}(t)$ and budgets $B_e(t)$ for each type k and point of entry e, number of pipeline tests Q_k for each type k, and tuning parameters $\{c, \gamma\}$

for $k = 1: K_t$

$\hat{\alpha}_k \leftarrow \mathrm{c}\,\hat{\alpha}_k,\ \hat{\beta}_k \leftarrow \mathrm{c}\,\hat{\beta}_k$ # *prior widening*

Compute λ_k from Eq. (6) using $\hat{\alpha}_k, \hat{\beta}_k$

$\lambda_k \leftarrow$pseudo-update index of type k (repeat Q_k times) # *account for pipeline tests*

$Y_k \leftarrow \sum_e A_{ke}(t)$ # *type k passengers not yet allocated a test*

for $e = 1: \mathcal{E}$

$Y_{ke} \leftarrow A_{ke}(t)$ # *type k passengers at point of entry e not yet allocated a test*

$C_e \leftarrow B_e(t)$ # *remaining (un-allocated) tests at point of entry e*

$N_{ke} \leftarrow 0$ # *initialize allocations*

end

end

while $\max_{\mathrm{e}} C_e > 0$ **and** $\max_{\mathrm{k},\, e' \in \{e \mid C_e > 0\}} Y_{ke'} > 0$

$k^* = \mathrm{argmax}_k\{\lambda_k : Y_k > 0\}$

$e^* = \mathrm{argmax}_e\{C_e : Y_{k^*e} > 0\}$

$N_{k^*e^*} \leftarrow N_{k^*e^*} + 1$ # *allocate a test to a type k passenger at point of entry e*

$C_{e^*} \leftarrow C_{e^*} - 1,\ Y_{k^*e^*} \leftarrow Y_{k^*e^*} - 1,\ Y_{k^*} \leftarrow Y_{k^*} - 1$

$\lambda_{k^*} \leftarrow$pseudo-update index of type k^*

end

return $\{N_{ke}\}_{k \in K_t, e \in \{1,\ldots,\mathcal{E}\}}$ # *number of tests allocated to each type at each port of entry*

Gittins Pseudo-Update for type $\boldsymbol{k}$

Input: $\hat{\alpha}_k, \hat{\beta}_k$ for type k

$\hat{r}_k^{EB} \leftarrow \dfrac{\hat{\alpha}_k}{\hat{\alpha}_k + \hat{\beta}_k}$

$\hat{\alpha}_k \leftarrow \hat{\alpha}_k + \hat{r}_k^{EB}$ and $\hat{\beta}_k \leftarrow \hat{\beta}_k + 1 - \hat{r}_k^{EB}$

Compute λ_k from Eq. (6) using $\hat{\alpha}_k, \hat{\beta}_k$

return λ_k