0826~

논문 핵심 알고리즘

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 $k = 1: K_t$

$\hat{\alpha}_k \leftarrow c\,\hat{\alpha}_k,\ \hat{\beta}_k \leftarrow c\,\hat{\beta}_k$ # *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 $\max_e C_e > 0$ **and** $\max_{k,\, e' \in \{e \mid C_e > 0\}} Y_{ke'} > 0$

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

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

$N_{k^*e^*} \leftarrow N_{k^*e^*} + 1$ # *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*

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)

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

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

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

end

end

Specifically, consider countries of color $l \in \{\mathcal{W}_t, \mathcal{G}_t, \mathfrak{B}_t\}$, and the corresponding set of types $\mathcal{L}_l = \{k \in K_t : k \text{ is from a country in } l\}$. Let $T_k = P_k + N_k$ be the total number of tests (with results) allocated to type k over the last 16 days. Conditional on T_k, P_k is approximately binomially distributed with T_k

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

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

$$M_{1l} = \frac{1}{|\mathcal{L}_l|} \sum_{k \in \mathcal{L}_l} \frac{P_k}{T_k} \tag{2}$$

$$M_{2l} = \frac{1}{|\mathcal{L}_l|} \sum_{k \in \mathcal{L}_l} \frac{P_k(P_k-1)}{T_k(T_k-1)} \tag{3}$$

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

Gittins Pseudo-Update for type *k*

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

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

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

세팅

- L(W, G, B) → 20개의 그룹
  - 컬럼 40개 이용하여 KMeans
    - 40개 컬럼: app+기초 변수(ip, device, channel, os), 기초 변수 + app LDA topic 5개씩
    - Kmeans에 사용되지 않는 나머지 4개의 컬럼: ip, app, time_dh , is_attributed

  - 20개의 그룹에서 ip가 독립적으로 존재하지 않고 흩어져 있는 문제

    → 각각의 ip에 대해 평균적으로 20개의 그룹 중 8개의 그룹에 존재하고 있음

    ```
    0  120571
    1  101588
    2  87611
    3  88804
    4  77213
    5  74833
    6  101787
    7  21624
    8  99143
    9  100748
    10  91925
    11  89932
    12  56543
    13  57005
    14  99180
    15  104129
    16  67226
    17  78827
    18  76366
    19  58881
    8.01403236747747
    ```

    - 총 ip 유저 수 : `206,380`
- k(L그룹 안의 나라 목록) → ??
  - L그룹을 더 세분화해야 될까?

  → L: 대그룹, K:소그룹
- 개개인의 승객 → ip

규칙

[시간 단위로 반복 Ti]

[L그룹 별로 반복 Li]

Case1. Li 가 처음 접하는 그룹인 경우

- inspection rate 무작위로 ip 추첨해서 검사

Case2 Li 가 이미 접한 적 있는 그룹인 경우

- Optimistic Gittins Indices(람다식) 계산

Case2에 해당하는 그룹 중 argmax해서 k 소그룹에 있는 ip rate의 비율로 선택 후 검사

검사 대상 ip가 fraud_ip면 해당 ip 끝까지 블락

검사 결과를 바탕으로 Al, Bl, Ak, Bk 파라미터 업데이트

- 시간 단위 : 1시간
- 답지를 깔 때 걸리는 딜레이 시간: ? 시간
  - 딜레이 반영 알고리즘

  관찰 결과 P/N 확률을 먼저 정해두고 반영

Gittins Pseudo-Update for type *k*

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

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

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

- 무작위: inspection rate
- detect/undetect

→ 질문: 분모 분자가 지워져서 결국 양성 판별 횟수 아닌가, 즉, 정확도에 무관하게 검사를 많이 하면 할수록 좋은 값이 되는거 아닌가??

알고리즘 평가

Objective

Note that, conditional on $T_{xe}(t)$, the number of positive tests observed at entry e is binomially distributed with $T_{xe}(t)$ trials and (unknown) success probability $R_x(t)$. Our goal is to maximize the expected total number of infections caught at the border, i.e.,

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