Post Process Cynet Runs

ishanu chattopadhyay

1 Q1 & Q2: PRED. VISITORS & MEETINGS

- 2 scenarios (A. No change, B. Closed sites)
- Need confidence bounds for Q1 (prediction of visits to rec sites)

1a. Scenario A: No Change

Let $\{s^0\}$ be the set of cynet-run sites, and let $s^\star$ be a target site.

Let $\mathcal{N}_k(s^\star) \subset \{s^0\}$ be the set of k closest neighboring sites of $s^\star$ in $\{s^0\}$. Note, "closest" implies we are using a specific distance metric, the choice of which is described later.

Step 1. Compute predictions for all sites $\{s^0\}$, by computing the regressors on FTX files as done before. The predicted time series for site s is denoted as s_t.

Step 2. We then use: (Denoting a neighboring site as s^i)

$$s_t^\star = \sum_{s^i \in \mathcal{N}_k(s^\star)} \alpha_i s_t^i \tag{1}$$

where

$$\alpha_i = \frac{1/\theta(s^i, s^\star)}{\sum_{s^i \in \mathcal{N}_k(s^\star)} 1/\theta(s^i, s^\star)} \tag{2}$$

Note: we still need to choose k. We set it to some small value in the range $[10, 35]$.

1b. Distance Metric

Three different distance metrics may be used. We will try all of them:

1) Euclidean
2) Travel distance
3) Learned from prediction behavior (discussed later)

1c. Scenario B: Site Closing

Let $\mathscr{S}$ be the set of all rec sites.

Let s^c be a closed site, and let $\mathcal{N}_k^{\mathscr{S}}(s^c) \subset \mathscr{S}$ be the set of k neighbors of s^c. Note this is the set of neighbors among **all** rec sites, which is different from $\mathcal{N}_k(s^\star)$.

Step 1. Predict s_t for all $s \in \mathcal{N}_k^{\mathscr{S}}(s^c)$ using the approach in the previous section.

Step 2. Compute:

$$\forall s^i \in \mathcal{N}_k^{\mathscr{S}}(s^c),\ \beta_i = \frac{1/\theta(s^i, s^c)}{\sum_{s^i \in \mathcal{N}_k^{\mathscr{S}}(s^c)} 1/\theta(s^i, s^c)} \tag{3}$$

Step 3. Predict s_t^c using the approach in the previous section

Step 4. Change for sites $s^i \in \mathcal{N}_k(s^c)$ as:

$$\delta s_t^i = \beta_i s_t^c \tag{4}$$

Step 5. Do this for all closed sites (which updates a bunch of neighboring rec sites). Let the set of sites updated be denoted as $U \subset \mathscr{S}$. We know, the total change for each site in U, which is denoted as δs_t^u for $s^u \in U$.

Step 6. Update sites in $\{s^0\}$ as follows: For each $s^u \in U$, compute the α-coefficients as in previous section. In other words,

$$\alpha_i^u = \frac{1/\theta(s^i, s^u)}{\sum_{s^i \in \mathcal{N}_k(s^u)} 1/\theta(s^i, s^u)} \tag{5}$$

Then, for each $s^u \in U$:

$$s_t^i \leftarrow s_t^i + \sum_{s^u \in U} \alpha_i^u \delta s_t^u \tag{6}$$

Step 6. Update target sites using approach in previous section.

The last two sections are sufficient to answer both Q1 and Q2 for both scenarios. **Confidence Bounds Needs for Q1**

2 Q3 : Individual List of REC Sites Visited

2a. Scenario A

2a.i. Individuals Already in Cynet Run

Step 1. Compute Regressor to predict 3 variables for all uid that were already in cynet run1:

- distHome
- distWork
- RecSiteVisit

Step 2. Identify which recsites were actually visited (when RecSiteVisit predicts a visit), by looking up the nearest RecSite that satisfies the distHome and distWork distances at that time.

Step 3. Compute list of RecSites visited (in prediction period) by each individual

2a.ii. Individuals Not in Cynet Run

Please message me if there are indeed any individuals who are not included in the cynet run?

Let H^0 be the set of individuals in the cynet run.

Let $h^\star$ be an individual not in the run for whom we have to make predictions Let $\omega(h)$ be the coordinate of the workplace for individual h, and $\sigma(h)$ be the coordiante of the home.

Additionaly, let h_t^ω, h_t^σ be the predictions of distWork and distHome for indivdual h.

Assuming Step 1. above is done.

Step 2. For each day t in prediction period, set up and solve the following 2 regression problems:

$$\text{Given } h_t^\omega, h_t^\sigma \xrightarrow{predict} distHome \text{ (when at Rec site on day } t \text{)} \tag{7}$$

$$\text{Given } h_t^\omega, h_t^\sigma \xrightarrow{predict} distWork \text{ (when at Rec site on day } t) \tag{8}$$

Step 3. Use regressors computed in Step 2. to estimate distHome, distWork for $h^\star$.

Step 4. Compute list of sites

2b. Scenario B: Sites Close

We simply select the next closest site when computing the REC Site Ids

3 Q4: Average Number of Rec Site Visits Per Day

Step 1. Compute from RecSiteVisit Predictions directly on individuals in cynet run.

Confidence Bounds Needs for Q4

4 Confidence Bounds Calculation

Step 1. Vary the model slection at Cynet prediction stage for all calculations, to generate a set of different predictions (time series levels)

Step 2. Use this bank of time series to compute confidence bounds.