Priority of pollination

Maite Puerta
Universidad de Deusto

Thursday 6th October, 2022

Contents

Initial information

For each campaign we have n hexagons: $\{h_1, \cdots, h_n\}$, each of them with f_i samples of air quality, so in the i-th hexagon we have:

- f_i timestamps: $\{t_{i,1}, \cdots, t_{i,f_i}\}$
- f_i users (can be the same): $\{u_{i,1}, \cdots, u_{i,f_1}\}$.
- f_i samples: $\{d_{i,1}, \cdots, d_{i,f_i}\}$. Each of these samples is an array with the information of the values of each air quality variable (in total a variable):

$$d_{i,j} = \begin{pmatrix} CO2 \\ .. \\ Others \end{pmatrix} = \begin{pmatrix} d^1_{i,j} \\ .. \\ d^a_{i,j} \end{pmatrix} \qquad (1)$$

The i-th hexagon can be represented as a set (ordered by time) of f_i samples, each of which is composed of a timestamp of the moment that the data was captured, the air quality data captured, and the user who captured this information.

$$\{(t_{i,1}, d_{i,1}, u_{i,1}), \cdots, (t_{i,f_1}, d_{i,f_1}, u_{i,f_1})\}$$

Time Slots

Time Slots

Time slots since the campaign began to the present moment.
$\{[t_0, t_0 + x), [t_0 + x, t_0 + 2x), \cdots, [t_0 + (l - 1)x, t_0 + l * x)\}$

Consideration:

- t_0 = timestamp of the beginning of the campaign
- x = seconds during which samples will be grouped by campaign Example: 60 seconds (1 hour), 300 (5 hours).
- actual moment $< t_0 + l * x$
- if *campaign_duration* is the duration of the campaign in seconds then $campaign\_duration + t_0$ is the final moment of the campaign. THEN actual moment $< campaign\_duration + t_0$

Reorganize data by time slots

The idea is to reorganise the timestamps of the i-th hexagon for the defined time slots. This equation only means that I recollect the index of the f_i samples that are related to the m-th time slot $([t_0 + (m-1)x, t_0 + m * x))$.

$$J_i^m = \{l \in \{1, \cdots, f_i\} \ \text{ s.h } \ t_{i,l} \in [t_0 + (m-1)x, t_0 + m * x)\}$$

It identifies the timestamp, data, and user associated with the samplings recollected in the i-th hexagon in the time slot $([t_0 + (m-1) * x, \ t_0 + m * x))$. They are:

$$T_i^m = \{t_{i,j}, \ \text{ t.q } \ j \in J_i^m\}$$

$$D_i^m = \{d_{i,j} \ \text{ t.q } \ j \in J_i^m\}$$

$$U_i^m = \{U_{i,j} \ \text{ t.q } \ j \in J_i^m\}$$

Time priority Ideas

In a time series, missing one value is not a dramatic case, but if we have several missing consecutive values, it is. In determining the priority, we have to consider this information.

Mathematical base

If $base1 < base_2$ and a>=1, then:

$$\log_{base_1}(a) >= \log_{base_2}(a) \tag{2}$$

If a>=1 and b>=1 and a>=b then:

$$\log_{base}(a) >= \log_{base}(b) \tag{3}$$

In order to figure out the temporal priority, we want to look at the number of samples a cell had in the last time slot and the current time slot.

Idea of the temporal priority

Let $|J_i^m|$, $|J_i^{m-1}|$ the number of samples in the m-th and $m-1$-th timeslots of the i-th hexagon. We propose this equation:

$$\log_{|J_i^m|+2}(10 - |J_i^{m-1}|) \tag{4}$$

We also want to evaluate the number of samples in the actual time stamp so we added another factor to this equation to evaluate this priority.

Idea of the temporal priority

$$P_i^t = \log(10 - |J_i^m|) * \log_{|J_i^m|+2}(10 - |J_i^{m-1}|) \tag{5}$$

- i-th Hexagon
- $10 \approx$ (Area's m2 / 15m2)
- $|J_i^m| =$ number of samples of the i-th hexagon in the m-th time slot of the campaign
- $m =$ it's the actual time slot.

Problems

This equation has several problems that we can fix by complicating it, but the idea is the same. We only fix the next problems:

- if $|J_i^m| == 10$, then this log doesn't exist.
- The same but with the second factor of the equation.
- $|J_i^m|$ has to be bigger than 2.

Correcting these values, we have the next equation:

Temporal priority

$$PT_i^m = \log(max\{2, 10-|J_i^m|\}) * \log_{|J_i^m|+2}(max\{2, 10-|J_i^{m-1}|\}) \quad (6)$$

I put 10 as an approximation of (Area's m2/15m2). It can be another value.

Trendy priority

The idea of this priority is evaluate the trend of a cell.

Trendy priority

$$P_i^l = \frac{\sum_{j=1}^{l}(|J_i^j|)}{\sum_i^n(\sum_{j=1}^{l}(|J_i^j|))} * n =$$

$$(\#\ \text{hexagon in the campaign}) * \frac{\#\ \text{of sampling in the i-cell}}{\#\ \text{sampling in the campaign}}$$

If all cells had the same popularity, then all would have a 1.0 value. If the value is bigger, then it means that its popularity is greater than the median value expected. If the value is lower, then it means that its popularity is less than the median value expected.