SocioBee

I. ESTADO INICIAL

Tenemos una campaña que se compone de n hexágonos: $\{h_1, \cdots, h_n\}$. Por cada hexágono tenemos un numero diferente de datos. Supongamos que para el i-enésimo hexágono tenemos f_i mediciones. Todas las mediciones se componen de diferentes datos, todos los datos asociados al i-esimo hexágono son los siguientes:

- Momento temporal en el que se tomo cada dato. Como del hexágono i-esimo se han tomado f_1 datos, tenemos f_i tiempos diferentes que denotaremos de la siguiente manera: $\{t_{i,1}, \cdots, t_{i,f_i}\}$. El primer índice representa el hexágono al que pertenecen las marcas temporales y el segundo subíndice a la posición correspondiente para ese tiempo. Como se han ido captados de forma continua en el tiempo están ordenados, si el segundo índice es menor entonces el valor de esa variable es menor o igual.
- Al igual que con los momentos temporales tenemos f_i datos: $\{d_{i,1}, \cdots, d_{i,f_i}\}$. En cada medición se ha obtenido valores de a variables diferentes y cada medición se representa del siguiente modo:

$$d_{i,j} = \begin{pmatrix} CO2 \\ No2 \\ .. \\ Otros \end{pmatrix} = \begin{pmatrix} d_{i,j}^1 \\ d_{i,j}^2 \\ .. \\ d_{i,j}^a \end{pmatrix} \tag{1}$$

- Cada medición la hace un usuario, por lo que tenemos f_i usuarios, por lo que tenemos $\{u_{i,1}, \cdots, u_{i,f_1}\}$ usuarios para el hexágono i-esimo.

De este modo en el hexágono i-esimo, la j-esima medición tiene los datos $d_{i,j}$, que los ha captado el usuario $u_{i,j}$ en el momento $t_{i,j}$. Conjuntamente todos los datos del hexágono i-esimo:

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

II. NUEVA AGRUPACIÓN

Ahora lo que vamos a hacer es agrupar las mediciones tomadas según los slot de tiempo que nos interesan. De este modo realizaremos una estadística con ello para tomar por slot de tiempo un único valor representativo. Supongamos que nosotros queremos tener un valor cada x horas y hemos empezado en el tiempo t_0. Por tanto los primero l slot de tiempo son: $\{[t_0, t_0+x), [t_0+x, t_0+2x), \cdots, [t_0+(l-1)x, t_0+lx)\}$

Lo que vamos a hacer es agrupar los tiempos, los usuarios y los datos por cada slot de tiempo al que pertenecen. Por tanto que se va a realizar es agrupar los subíndices, por ejemplo $\{1, \cdots, f_i\}$ en el caso del hexágono i-esimo, en diferentes l subconjuntos cada uno de ellos correspondiente a un slot de tiempo concreto. De este modo en el conjunto J_i^m con $m \in \{1, \cdots, l\}$, tendremos los índices correspondientes a los datos tomados en el hexágono i eximo en el slot de tiempo $[t_0 + (m-1)x, t_0 + mx)$.

Para ello, definimos este conjunto siguiendo esta definición:

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

Por lo que $\forall j \in J_i^m, t_{i,j}$ esta en el slop de tiempo indiciado. Esto también nos permite saber que el subconjunto de tiempos captados en ese slot de tiempo es el siguiente:

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

y el conjunto de datos captados en este periodo y correspondientes a estas mediciones esta son los siguientes:

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

y los usuarios que los han captado son los siguientes:

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

II-A. *Un único valor por slot temporal*

Si queremos un único dato para el m-esimo slot de tiempo, tenemos que realizar una estadística con todos los datos de D_i^m. La estadsitica propuesta es calcular la media de los valores obtenidos durante todo el el tempo del slot indicado, de este modo cada uno de la columna de este nuevo dato (y único para este slot):

$$d'^p_{i,j} = \frac{1}{|J_i^m|} \sum_{j \in J_i^m} d_{i,j}^p \qquad \forall p \in 1, \cdots, a \tag{2}$$

III. Prioridad temporal

En las series temporales, predecir un valor que falta no es un problema sino que el problema es cuando hay varios valores seguimos sin nada. Por tanto, la medición temporal debe tener en cuenta si en el slot de tiempo anterior ha habido suficientes datos o no así como la cantidad de datos que hay este nuevo slot de tiempo. Basándonos en esto sugerimos que la prioridad temporal se formule del siguiente modo:

$$P_i^t = \log_{|J_i^m|+2}(max\{2, (\text{Area's m2/15m2}^1) - |J_i^{m-1}|\}) \quad (3)$$

Con el objetivo de que este peso se pueda potencial en función de cuantas instancias necesitamos para conseguir el numero que consideramos como el mínimo necesario para realizar una estadística buena:

$$P_i^t = max\{1, \log(max\{1, |J_i^m| - (\text{Area's m2/15m2})\})\} * \log_{|J_i^m|+2}(max\{2, (\text{Area's m2/15m2}) - |J_i^{m-1}|\}) \quad (4)$$

Los máximos es únicamente para que cuando el numero valla bien el logaritmo no de un negativo directamente o un infinito. en realidad lo que hacemos con los máximos es quitar esos casos y que en los casos en los que valla bien multiplicar por un 1.

IV. Prioridad de moda

Esta prioridad nos informa cuando una celda esta siendo muy polinizada en relación con las demás. Lo ideal seria que todas las celdas tuvieran un 1 es ente valor. De modo que si tienen una cifra menor implica que no es popular y una cifra mayor implica exceso de popularidad respecto a las demás. El número de arriba de la ecuación es cuantas mediciones ha tenido esa celda en todos los slot hasta el l-esimo y el de abajo es cuantas mediciones se han realizado en esa campaña en todas las celdas llegados al l-esimo slot.

$$P_i^l = \frac{\sum_{j=1}^{l}(|J_i^j|)}{\sum_i^n(\sum_{j=1}^{l}(|J_i^j|))} * n \quad (5)$$

V. Prioridad de distancia

NO ESTOY SEGURA!

Supongamos que d es la distancia a recorrer entre el la posición del usuario y la posición de la celda.

$$P_i^d = \frac{P_i^t + P_i^n}{d} \quad (6)$$

VI. Glosario

n = numero de hexágono en una campaña
a= numero de variable a medir en una campaña
i = hexágono que estamos mirando.

[1]Esto es de Babis y Alexandros