

Clasificación de barrios

Reading 01

Introducción a la ciencia de datos

Autor

Franklin Enmanuel Magallanes Pinargote

CONTENIDO

1. Desarrollo.....3

2. Referencias.....3

1. Desarrollo

En base a lo expuesto en (Huina Mao et al., 2017), se tenía como propósito obtener patrones espacio-temporales de las actividades de los usuarios de telefonía. Para esto, la data raw, pasó por los siguientes tres pasos:

1. Agregar a cada antena el volumen total de llamadas por cada hora y normalizar cada serie temporal original de volumen de llamadas en *z-score*, denotado como $z_i^e(t) = \frac{v_i(t) - \mu_i}{\sigma_i}$, donde $v_i(t)$ representa el volumen total de llamadas por la antena i en la hora t , μ_i es el promedio de volumen de llamadas de la antena i . Esta escala de normalización permite el comparar diferentes antenas de telefonía.
2. Con el volumen de llamadas normalizado, calculamos el promedio de actividad de todas las áreas durante una hora, para restarlo al *z-score*. Obteniendo el *z-score residual espacial*. Denotado como: $z_i^e(t) = z_i(t) - \overline{z(t)}$, donde $\overline{z(t)}$ representa el promedio normalizado del volumen de llamadas en el tiempo t .
3. Calcular $z_i^e(t)$, de cada antena por hora del día (0 AM a 11 PM) y día de la semana (Lunes hasta Domingo). Obteniendo así $\overline{z_i^e(t)}$.

Luego de realizado los tres pasos obtenemos un vector de 168 dimensiones (7 días x 24 horas) para cada antena de telefonía. En el artículo también se menciona, que se descartan para el estudio antenas que no tengan todas las 168 dimensiones.

Estas dimensiones, nos reflejan la intensidad del volumen de llamadas por series de tiempo (cada hora del día). De esta manera podemos separar el análisis en dos franjas horarias, que empíricamente podemos detectar como comerciales o residenciales. Por ejemplo, en áreas con alta actividad durante el día (6:00 AM a 5 PM) y en días laborales (Lunes a Viernes) muy probablemente pertenecerían al tipo comercial. Por otro lado, áreas con alta actividad durante la noche (6:00 PM a 11:00 PM) y fines de semana (Sábados a Domingos) encajarían mejor en el tipo de áreas residenciales.

2. Referencias

Huina Mao, Yong-Yeol Ahn, Budhendra Bhaduri & Gautam Thakur (2017) *Improving land use inference by factorizing mobile phone call activity matrix*, Journal of Land Use Science, 12:2-3, 138-153, DOI: [10.1080/1747423X.2017.1303546](https://doi.org/10.1080/1747423X.2017.1303546)