Intervalos semánticos difusos para la transferencia de conocimiento entre sistemas de bicicletas compartidas

Francisco J. Márquez-Saldaña[1], Francisco A. Márquez[2], Gonzalo A. Aranda-Corral[2], and Joaquín Borrego-Díaz[1]

[1] Dpto. de Ciencias de la Computación e Inteligencia Artificial, Universidad de Sevilla, Sevilla, España
{fmarquez,jborrego}@us.es
[2] Dpto. de Tecnologías de la Información, Universidad de Huelva, Huelva, España
{alfredo.marquez,gonzalo.aranda}@dti.uhu.es

Resumen La identificación interoperable de patrones diarios de uso en sistemas de bicicletas compartidas (BSS) se ha abordado recientemente mediante la semantización de la curva de demanda en cinco intervalos —*noche*, *trabajo*, *mañana*, *almuerzo* y *tarde*— con fronteras temporales nítidas ajustadas manualmente por sistema. En este trabajo en curso proponemos una reformulación basada en lógica difusa donde cada intervalo se modela como un conjunto difuso sobre $[0, 24)$ mediante funciones de pertenencia trapezoidales aprendidas de forma no supervisada a partir de la curva tipo de cada ciudad. La nueva representación sustituye a la etapa de semantización, habilita una agregación robusta de la demanda mediante la mediana ponderada difusa y permite cuantificar el solapamiento conceptual entre franjas como huella cultural del sistema.

Keywords: Lógica difusa · Variables lingüísticas · Series temporales · Sistemas de bicicletas compartidas · Transferencia de conocimiento.

1. Introducción

Los Sistemas de Bicicletas Compartidas (BSS, del inglés *Bike Sharing Systems*) son un área de notable interés dentro de la movilidad urbana sostenible [1]. La naturaleza estocástica de la demanda y la influencia de factores exógenos –meteorología, infraestructura, eventos– condicionan su operación, y la identificación de patrones de uso a nivel de estación constituye un paso previo a la predicción de demanda o el reequilibrado de la flota.

Los autores hemos propuesto recientemente *Pearclustering* [3], un algoritmo de clasificación no supervisada basado en la correlación de Pearson como distancia de disimilitud que, combinado con una representación *semantizada* de la curva diaria en cinco *intervalos* —*noche*, *trabajo*, *mañana*, *almuerzo* y *tarde*— con fronteras nítidas ajustadas manualmente por ciudad, transfiere el conocimiento aprendido a sistemas no contemplados en el entrenamiento. Esta estrategia adolece de tres limitaciones: (i) los conceptos manejados son intrínsecamente

graduales –las 13:45 son *almuerzo* "casi total" en Sevilla y solo "parcialmente" en Dublín–, por lo que la asignación nítida pierde información; (ii) la fijación manual de los cortes es subjetiva y no escala a la incorporación automática de nuevos sistemas; y (iii) la representación *crisp* no permite cuantificar el solapamiento conceptual entre franjas.

En esta investigación en curso reformulamos los intervalos semánticos como variables lingüísticas [4] mediante conjuntos difusos sobre $[0, 24)$, con funciones de pertenencia aprendidas a partir de la curva tipo de cada sistema.

2. Trabajo previo: intervalos semánticos crisp

Sea i un sistema BSS y $D^i(t)$ la demanda registrada (bicicletas tomadas o devueltas) en el instante $t \in [0, 24)$, agregada por hora. En [3] la curva diaria se semantiza dividiendo el dominio en $m = 5$ intervalos disjuntos $\{[t_n, t_{n+1})\}_{n=1}^{m}$ con $t_1 = 0$, $t_{m+1} = 24$, asociando cada uno a un concepto c_n del conjunto

$$\mathcal{C} = \{\text{NOCHE}, \text{TRABAJO}, \text{MAÑANA}, \text{ALMUERZO}, \text{TARDE}\}.$$

La demanda agregada al intervalo n es $\widehat{D}_n^{\,i} = \text{Med}\{D^i(t_k) : t_n \le t_k < t_{n+1}\}$, donde la mediana mitiga el efecto de los valores anómalos. Los puntos de corte $\{t_n\}$ se eligen manualmente por inspección visual de la curva tipo, garantizando que un mismo c_n refleje el mismo concepto operativo en sistemas distintos pese a las diferencias horarias entre ciudades. No obstante, los cortes nítidos producen discontinuidades artificiales y dificultan la captura de comportamientos que "viven" en la frontera entre dos conceptos.

3. Propuesta: intervalos semánticos difusos

3.1. Definición formal

Reformulamos cada concepto $c_n \in \mathcal{C}$ como un conjunto difuso $\tilde{c}_n$ sobre el dominio temporal con función de pertenencia $\mu_n^{\,i} : [0, 24) \to [0, 1]$, específica de cada sistema i. Adoptamos la condición de *partición difusa de Ruspini* [5]

$$\sum_{n=1}^{m} \mu_n^i(t) = 1 \quad \forall\, t \in [0, 24), \tag{1}$$

que preserva la interpretación de los grados de pertenencia como pesos normalizados y recupera la partición *crisp* como caso límite cuando cada $\mu_n^{\,i}$ degenera en la función indicadora del intervalo correspondiente. Para mantener la interpretabilidad, modelamos cada $\mu_n^{\,i}$ mediante una función trapezoidal definida por cuatro parámetros $(a_n^i, b_n^i, c_n^i, d_n^i)$ con $a_n^i \le b_n^i \le c_n^i \le d_n^i$:

$$\mu_n^{\,i}(t) = \text{máx}\left\{0,\ \text{mín}\left\{\frac{t - a_n^i}{b_n^i - a_n^i},\ 1,\ \frac{d_n^i - t}{d_n^i - c_n^i}\right\}\right\}, \tag{2}$$

imponiendo $d_n^i = b_{n+1}^i$ y $c_n^i = a_{n+1}^i$ entre conjuntos vecinos para satisfacer (1) por construcción.

3.2. Aprendizaje de las funciones de pertenencia

Frente al ajuste manual del enfoque *crisp*, los parámetros se estiman automáticamente a partir de la curva tipo $\bar{D}^{\,i}(t) = \mathrm{Med}_d\, D_d^i(t)$. El procedimiento, desatendido y reproducible, consta de cuatro pasos: (1) suavizado de $\bar{D}^{\,i}(t)$ por estimación de núcleos (KDE [6]) con ancho de banda por validación cruzada; (2) detección de máximos locales, cuyos entornos anclan los núcleos b_n^i y c_n^i (donde $\mu_n^i = 1$) correspondientes a los picos conceptuales de *trabajo* y *almuerzo*; (3) detección de mínimos locales, que delimitan las transiciones *noche* $\rightarrow$ *trabajo*, *trabajo* $\rightarrow$ *mañana*, *mañana* $\rightarrow$ *almuerzo* y *almuerzo* $\rightarrow$ *tarde*; y (4) cierre por Ruspini, fijando a_n^i y d_n^i con el solapamiento necesario entre conjuntos vecinos. El método permite incorporar nuevos sistemas sin reajuste manual.

3.3. Agregación difusa y compatibilidad con el pipeline existente

La generalización natural de la agregación *crisp* es la *mediana ponderada difusa*. Sean $\{D^{\,i}(t_k)\}_{k=1}^{K}$ los valores horarios de demanda en un día y $w_{n,k} = \mu_n^i(t_k)$ sus grados de pertenencia al intervalo n. Definimos la demanda difusa agregada como

$$\widetilde{D}_n^{\,i} = \mathrm{Med}_{w_{n,\cdot}}\big\{D^{\,i}(t_k)\big\}_{k=1}^{K}, \tag{3}$$

donde Med_w es la mediana ponderada (el valor que separa la masa total de pesos en dos mitades). Esta definición conserva la robustez frente a anomalías del enfoque *crisp* y degrada de forma suave los valores cercanos a las fronteras, eliminando los saltos artificiales introducidos por los cortes nítidos. Como el vector $\widetilde{\mathbf{D}}^{\,i,j} = (\widetilde{D}_1^{\,i,j}, \ldots, \widetilde{D}_m^{\,i,j})$ asociado a la estación j del sistema i tiene la misma dimensionalidad que su contraparte *crisp*, todo el pipeline posterior de *Pearclustering* —matriz de correlaciones, agregación de bicicletas tomadas y devueltas, e inferencia sobre sistemas no vistos— se mantiene sin modificación.

4. Diseño experimental en curso

Datos y comparativa. Replicamos el banco de pruebas de [3]: entrenamiento sobre los sistemas de Sevilla, Lyon y Dublín; validación sobre Valencia, Bruselas y Amiens. El histórico cubre un año completo de información extraída del proceso ETL multi-sistema descrito en [2]. Se contrastará el enfoque difuso frente al *crisp* a tres niveles: (i) estabilidad de las funciones de pertenencia aprendidas frente a la sensibilidad observada en la elección manual; (ii) calidad del clustering medida mediante el coeficiente de silueta sobre la matriz de correlaciones de Pearson; y (iii) coherencia geográfica de las etiquetas inferidas en los sistemas de validación, evaluada mediante contraste GIS con capas de oficinas, ocio y áreas residenciales.

Solapamiento conceptual como subproducto. La naturaleza difusa de la representación permite definir el grado de *solapamiento conceptual* entre los conjuntos difusos $\tilde{c}_n$ y $\tilde{c}_{n'}$ como

$$\Omega_{n,n'}^{i} = \int_0^{24} \mathrm{mín}\big\{\mu_n^i(t),\, \mu_{n'}^i(t)\big\}\, dt, \tag{4}$$

indicador novedoso de la borrosidad cultural del sistema. La hipótesis preliminar es que las ciudades con horarios sociolaborales rígidos (p. ej. Dublín) presentarán valores $\Omega^i_{n,n'}$ menores que las de horario flexible (p. ej. Sevilla), lo que proporcionaría una caracterización cuantitativa de la *huella horaria* de cada sistema.

5. Conclusiones y trabajo futuro

Hemos presentado una reformulación difusa de los intervalos semánticos para la identificación interoperable de patrones diarios en BSS. La propuesta sustituye las fronteras nítidas por funciones de pertenencia trapezoidales aprendidas automáticamente de la curva tipo, eliminando la subjetividad del ajuste manual, proporcionando una agregación de la demanda más fiel a su carácter gradual y aportando, sin coste adicional, un indicador cuantitativo de la borrosidad horaria del sistema.

Las líneas inmediatas de continuación incluyen: (i) la extensión natural a un *Fuzzy C-Means* [7] con distancia de Pearson, sustituyendo la asignación dura de clusters por una pertenencia gradual; (ii) la integración de las demandas difusas $\widetilde{D}^i_n$ como variables lingüísticas en un sistema Mamdani de predicción a medio plazo; y (iii) la formulación de la transferencia entre sistemas como morfismos entre retículos difusos en el marco del Análisis Formal de Conceptos Fuzzy [8], conectando este trabajo con la línea de *semántica emergente* desarrollada por los autores.

Agradecimientos. Este trabajo está financiado por el proyecto PID2023 - 147198NB-I00 del Ministerio de Ciencia, Innovación y Universidades (y fondos FEDER) y la Cátedra US - Google TSI-100930-2023-2, financiada por Google y Ministerio de Asuntos Económicos y Transformación Digital. .

Referencias

1. Miriam Ricci: Bike sharing: A review of evidence on impacts and processes of implementation and operation. Research in Transportation Business & Management, 15:28–38, 2015. Managing the Business of Cycling.
2. Márquez-Saldaña, F., Aranda-Corral, G.A., Borrego-Díaz, J.: Bike sharing systems data interoperability by a unified station status concept and big data solutions. J. Traffic Transp. Eng. **12**(2), 420–433 (2025)
3. Márquez-Saldaña, F., Aranda-Corral, G.A., Borrego-Díaz, J.: Pearclustering: a novel clustering algorithm with an application to bike mobility. Evol. Intell. **18**, art. 73 (2025)
4. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning—I. Inf. Sci. **8**(3), 199–249 (1975)
5. Ruspini, E.H.: A new approach to clustering. Inf. Control **15**(1), 22–32 (1969)
6. Silverman, B.W.: Density Estimation for Statistics and Data Analysis.
7. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
8. Bělohlávek, R.: Concept lattices and order in fuzzy logic. Ann. Pure Appl. Log. **128**(1–3), 277–298 (2004)