



# Modelado lingüístico y síntesis de series temporales heterogéneas de consumo energético

S. Martínez-Municio\*, L. Rodríguez-Benítez\*, E. Castillo-Herrera\*, J. Giralt-Muiña\*, L. Jiménez-Linares\*

\*Escuela Superior de Informática, Universidad de Castilla-La Mancha, Paseo de la Universidad s/n. 13071. Ciudad Real, España  
{sergio.martinez, luis.rodriquez, ester.castillo, juan.giralt, luis.jimenez}@uclm.es

**Resumen**—En la actualidad, gracias a la presencia de sensores o al auge de tecnologías propias del *internet de las cosas*, podemos monitorizar y registrar los consumos energéticos obtenidos en los edificios a lo largo del tiempo. Mediante un análisis efectivo de estos datos que capturen los patrones de consumo, se pueden conseguir reducciones significativas de los mismos, contribuyendo a su sostenibilidad. En este trabajo proponemos un marco de trabajo a partir del cual definir modelos que capturen esta casuística, haciendo acopio de un conjunto de series temporales de consumo eléctrico. El objetivo de estos modelos es obtener un resumen lingüístico que describa en lenguaje natural cual es la situación consumista de un edificio o conjunto de edificios en un periodo temporal concreto. La definición de estas descripciones se ha resuelto mediante resúmenes lingüísticos difusos, y como novedad en este campo, proponemos una extensión de los mismos que capturen situaciones donde la pertenencia a los conjuntos difusos no resulte muy marcada (i.e. que no supere un cierto valor umbral, como por ejemplo, 0.8), obteniendo una semántica enriquecida. Para la experimentación, se hará uso de datos de consumo energético de una organización educativa sobre los que se definirán los modelos propuestos y se obtendrán los resúmenes lingüísticos asociados a los mismos, para demostrar la capacidad que otorgan estas técnicas a la hora de obtener conclusiones sobre los diferentes escenarios de consumo en términos lingüísticos.

**Palabras Clave**—consumo energético, modelo difuso, resúmenes lingüísticos, series temporales

## I. INTRODUCCIÓN

La energía constituye uno de los pilares fundamentales para el desempeño adecuado de las actividades imperantes en la sociedad actual. De acuerdo con la Agencia de la Energía [1], los edificios representan el ámbito de mayor consumo energético, lo que conlleva a una responsabilidad directa en un tercio de las emisiones mundiales de dióxido de carbono. Del mismo modo, se ha reportado que los edificios operan de manera ineficiente [2] y por lo tanto, poder entender los patrones de consumo de los mismos resulta de especial interés para optimizar los recursos de las organizaciones, tanto económicos como de uso de la energía, y ser más sostenibles desde el punto de vista medioambiental [3]. En la actualidad, gracias a las tecnologías existentes así como al auge de otras nuevas, como los sensores inteligentes o redes complejas que configuran el *internet de las cosas*, junto con el interés cada vez más acuciante por las técnicas de aprendizaje máquina [4], se dispone de una cantidad de datos sin precedentes que permiten obtener un conocimiento detallado del uso que se está haciendo de la energía en un instante temporal dado en

términos cuantitativos; sin embargo, resulta fundamental incorporar modelos que caractericen estos patrones de consumo energético en términos cualitativos, de modo que permitan aumentar su expresividad mediante descripciones breves en lenguaje natural con el propósito de planificar políticas que contribuyan a una mayor sostenibilidad de los edificios en un marco de revisión y ejecución continua. Dichas descripciones vendrán dadas por medio de resúmenes lingüísticos, ya que han demostrado su utilidad en diferentes ámbitos, como a la hora de describir el tráfico [5] o evaluar la calidad de la forma de andar de las personas [6]. Estos resúmenes están basados en la propuesta de Yager [7], quien proporciona una definición formal basada en el concepto de protoformas introducido por Zadeh [8]:

$$Q[R]y \text{ are } P \quad (1)$$

donde  $y$  es un objeto caracterizado por un atributo con un dominio finito;  $Q$  y  $R$  (opcional) son cuantificadores, que llevan asociado un valor lingüístico definido sobre el dominio del atributo de  $y$ , y  $P$  un resumen que lleva asociado otro valor lingüístico. Además, se han realizado propuestas de generación automática de descripciones lingüísticas alineadas con la temática de este trabajo, donde la principal diferencia radica en que generan dichas descripciones partiendo de las series temporales de un día, sin definir un modelo previo que segmente dichos datos temporales en patrones de consumo [9].

La estructura del trabajo es la siguiente: en la Sección II se introducen los conceptos de modelo de edificio y organizacional, como principales abstracciones de caracterización de consumos, así como la definición de resumen lingüístico. En la Sección III se estudia la validez de los modelos definidos en términos lingüísticos. Una extensión de los resúmenes clásicos que aumenta las capacidades semánticas de los mismos es presentada en la Sección IV, mientras que en la Sección V se lleva a cabo la experimentación mediante la aplicación de los modelos definidos sobre datos de consumo real. Finalmente, la Sección VI presenta las conclusiones obtenidas y establece futuras líneas de trabajo.

## II. DEFINICIÓN DE MODELOS

Se propone la definición de modelos para caracterizar el consumo energético obtenido en grandes instituciones con el fin de obtener en términos lingüísticos las conclusiones

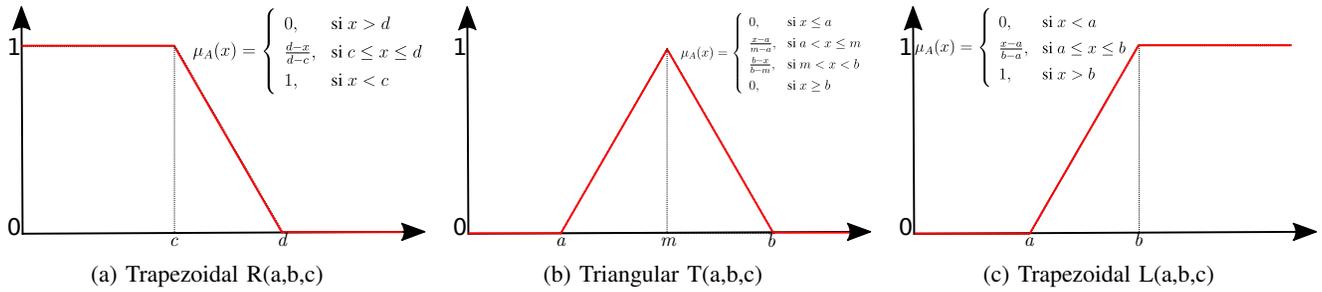


Figura 1: Funciones de pertenencia difusas empleadas

derivadas de los mismos. Para ello, en primer lugar, se define un modelo base (modelo de edificio) que permita categorizar a nivel de edificio cómo ha sido su consumo por medio de una segmentación de los datos que es resuelta con el algoritmo  $k$ -means, para posteriormente, definir un modelo general (modelo de organización) que permita establecer un marco comparativo entre los modelos básicos para saber cómo se han comportado con respecto a los demás y obtener esta conclusión en términos lingüísticos. Como caso de estudio, se emplearán datos de consumo provenientes de una institución educativa española geográficamente distribuida: la Universidad de Castilla-La Mancha (UCLM), que cuenta con un total de 97 edificios repartidos por toda la región. En concreto, se utilizarán datos de consumo por día del año 2016 para la construcción de los modelos, y del año 2017 para la validación de los mismos.

#### II-A. Modelo de Edificio

Un modelo  $M$  de un edificio  $i$  se define como un conjunto de  $k$  grupos que tienen una semántica asociada:

$$M^i = \{G_0, G_1, \dots, G_{k-1}\} \quad (2)$$

donde cada  $G_j$  viene definido por un prototipo,  $c_j$ :

$$G_j = \{c_j\} \quad (3)$$

Cada uno de los modelos de edificio  $M^i$  estará compuesto por dos grupos que caracterizarán los siguientes patrones de consumo: periodos con unos niveles bajos o de no actividad, que se corresponderá con  $G_0$ , y periodos donde existe una actividad relevante, que se corresponderá con  $G_1$ . Por tanto, los términos {no actividad, actividad} representan la semántica asociada a los grupos del modelo de edificio  $M^i$ . De este

Tabla I: Modelos de Edificio de la UCLM - Año 2016

		Modelos definidos	
		$G_0$	$G_1$
Edificios	0	500.44	1215.19
	1	570.23	1213.66
	...		
	95	276.31	1994.66
	96	211.07	406.82

modo, en la Tabla I aparecen recogidos un subconjunto de los diferentes modelos de edificio que han sido generados, donde

cada columna, representa la caracterización del consumo como no actividad ( $G_0$ ) y actividad ( $G_1$ ) respectivamente, y cada fila el edificio en cuestión. Así, por ejemplo, tenemos que el modelo para el edificio 95 quedaría formalizado como:  $M^{95} = \{276.31, 1994.66\}$ , de manera que, si se dispone del consumo real  $x$  para un día concreto, podemos saber cómo ha sido dicho consumo de acuerdo a su modelo, es decir, obtener su categoría semántica (caracterización).

#### II-B. Modelo de Organización

Un modelo organizacional es un modelo de modelos compuesto por cada uno de los sub-modelos de edificio  $M^i$  que queda definido como:

$$O = \left\{ \begin{array}{l} O_{G_0} = \{M_{G_0}^0, M_{G_0}^1, \dots, M_{G_0}^{|M|}\} \\ O_{G_1} = \{M_{G_1}^0, M_{G_1}^1, \dots, M_{G_1}^{|M|}\} \\ \vdots \\ O_{G_k} = \{M_{G_k}^0, M_{G_k}^1, \dots, M_{G_k}^{|M|}\} \end{array} \right\} \quad (4)$$

donde cada  $O_{G_j}$  se refiere al conjunto de modelos de edificio  $M_{G_j}^i$  que caracteriza cada patrón de consumo conforme a lo expuesto en el apartado anterior. Así pues, haciendo uso de (4), definimos el modelo organizacional para la UCLM como:

$$O_{UCLM} = \left\{ \begin{array}{l} O_{G_0} = \{M_{G_0}^0, M_{G_0}^1, \dots, M_{G_0}^{96}\} \\ O_{G_1} = \{M_{G_1}^0, M_{G_1}^1, \dots, M_{G_1}^{96}\} \end{array} \right\} \quad (5)$$

donde  $O_{G_0}$  vendrá dado por el dominio formado por todos los prototipos de los modelos de edificio  $M^i$  correspondientes a la semántica de no actividad:

$$Dom(O_{G_0}) = \{500.44, 570.23, \dots, 211.07\} \quad (6)$$

y  $O_{G_1}$  por los de la semántica de actividad:

$$Dom(O_{G_1}) = \{1215.29, 1213.66, \dots, 406.82\} \quad (7)$$

Partiendo de este modelo de organización  $O_{UCLM}$ , en el siguiente apartado se discutirá la obtención de los resúmenes lingüísticos basados en técnicas difusas que describirán el comportamiento consumista de la organización.



### II-C. Caracterización lingüística

La caracterización lingüística será realizada en base a (1), donde  $y$  será un sintagma significativo,  $R$  vendrá dado por la semántica asociada al modelo, y  $P$  será definido en términos difusos. Un caso especial ocurre con  $Q$ , pues será tratado de manera conjunta con  $y$  y no como un modificador independiente con el que evaluar la sentencia completa (i.e. obtener el grado de verdad de la frase cuantificada lingüísticamente). De esta manera, (1) puede ser simplificada a la siguiente forma:

$$[R]y \text{ are } P \quad (8)$$

La formalización del resumen  $P$  se lleva a cabo haciendo uso de una variable lingüística,  $\mathcal{L}_v$ , que está definida mediante  $n$  conjuntos difusos ( $S'$ ) que configuran una partición difusa del dominio cuyo universo de discurso viene dado por el conjunto de todos los prototipos que conforman cada dominio del modelo  $O$ :

$$\text{Dom}(\mathcal{L}_v) = \text{Dom}(O_{G_j}) \quad (9)$$

A su vez, cada conjunto difuso de  $\mathcal{L}_v$  queda definido mediante una función de pertenencia compuesta por tres parámetros  $P_r$ , que representan el valor del percentil  $r$ , pues permiten conocer el posicionamiento de cada edificio  $i$  en lo que a consumo se refiere, con respecto al total de la organización. En la Fig. 1 se muestran las funciones de pertenencia consideradas para la definición de los valores de  $\mathcal{L}_v$ .

$$\mathcal{L}_v = \{S'_0, S'_1, \dots, S'_m\} \quad (10)$$

Dado que el modelo  $O$  está compuesto de tantos dominios como grupos tengan los modelos de edificios que lo conforman, es necesario identificar el dominio de  $O$  al que pertenece un consumo dado  $x$  de un edificio específico  $i$  para poder aplicar la variable lingüística  $\mathcal{L}_v$  en las magnitudes adecuadas. Para ello, partiendo del modelo de edificio  $M^i$  que categoriza el consumo de un edificio  $i$ , obtenemos el grupo semántico al que pertenece dicho consumo  $x$  a través de la siguiente fórmula:

$$\arg \min \text{dist}(x, G_j) \quad (11)$$

donde  $\text{dist}(x, G_j)$  es una medida de distancia que cuantifica la similitud entre el consumo  $x$  y el prototipo del grupo  $G_j$ , típicamente la euclídea. Sabiendo el grupo semántico al que pertenece  $x$ , podemos obtener el dominio de  $O$  adecuado y aplicar la variable lingüística  $\mathcal{L}_v$  para obtener la etiqueta lingüística que definirá el resumen  $P$ , aplicando la operación de la t-conorma del máximo. Por lo tanto, si aplicamos estos conceptos al modelo de estudio,  $O_{\text{UCLM}}$ , la caracterización lingüística vendrá dada de acuerdo con (8), donde  $y$  será consumo,  $R$  será actividad o no actividad y  $P$  quedará definido mediante dos variables lingüísticas  $\mathcal{L}_v$ ,  $\mathcal{L}_v^0$  y  $\mathcal{L}_v^1$ , una para cada conjunto de modelos de edificio  $M^i$  que caracterizan los patrones de consumo de actividad y no actividad en  $O_{\text{UCLM}}$ , es decir, para  $O_{G_0}$  y  $O_{G_1}$  respectivamente. Dichas  $\mathcal{L}_v$  son definidas mediante cinco conjuntos difusos haciendo uso de (10), tal y como se aprecia en la Tabla II, donde la única diferencia entre  $\mathcal{L}_v^0$  y  $\mathcal{L}_v^1$  radica en el dominio empleado.

Tabla II: Definición de  $\mathcal{L}_v$  para  $O_{\text{UCLM}}$

Etiqueta	Función de Pertenencia ( $\mu$ )
Insignificante	$R\{P_0, P_{10}, P_{25}\}$
Leve	$T\{P_{10}, P_{25}, P_{50}\}$
Normal	$T\{P_{25}, P_{50}, P_{75}\}$
Grande	$T\{P_{50}, P_{75}, P_{90}\}$
Enorme	$L\{P_{75}, P_{90}, P_{100}\}$

La caracterización lingüística vista hasta ahora es aplicable a la hora de obtener conclusiones acerca de cómo ha sido el consumo obtenido durante el periodo de vigencia del modelo, pero no nos proporciona la información necesaria sobre el error cometido en la estimación del modelo de edificio  $M^i$  para poder concluir sobre la validez o bondad del mismo. Por ello, en la siguiente sección se propondrá un mecanismo que permita la validación de los modelos en términos lingüísticos.

### III. VALIDACIÓN DE MODELOS

En esta sección se propondrá un mecanismo para conocer la validez de cada uno de los modelos de edificio  $M^i$  definidos, así como del modelo organizacional  $O$ , y de este modo saber si los modelos definidos arrojan las conclusiones adecuadas.

#### III-A. Validación del Modelo de edificio

En primer lugar, se debe asociar el consumo real  $x$  obtenido en un instante de tiempo en cada edificio  $i$ , con el grupo semántico  $G_j$  del modelo que mejor lo define para obtener el error cometido en la estimación del modelo por día de consumo real. A este grupo semántico lo denominaremos  $\tilde{G}$ , y será el prototipo que mejor describe el consumo real  $x$ . Formalmente,  $\tilde{G}$  viene definido en términos generales mediante la siguiente función  $f$ , que combina los diferentes grados de pertenencia a los distintos grupos que conforman el modelo de edificio  $M^i$ :

$$\tilde{G} = f(\mu_{G_0}(x), \mu_{G_1}(x), \dots, \mu_{G_{k-1}}(x)) \quad (12)$$

donde  $\mu_{G_j} = \text{dist}(x, G_j)$  y el grupo semántico vendrá dado por  $j = \arg \max \{\mu_{G_0}(x), \mu_{G_1}(x), \dots, \mu_{G_{k-1}}(x)\}$ . De este modo, el error obtenido a la hora de catalogar el consumo real  $x$  en uno de los grupos semánticos vendrá dado por:

$$\epsilon = \tilde{G} - x \quad (13)$$

lo que nos proporciona un valor que es independiente de la escala escogida, permitiéndonos comparar el error obtenido para cada grupo semántico  $\tilde{G}$ . Así, para catalogar si dicho error cometido con respecto a la estimación de su modelo de edificio  $M^i$  es significativo, con un nivel de confianza del 95%, consideramos el intervalo definido por  $\pm 2\sigma$ , con  $\sigma$  siendo la desviación típica propia del grupo semántico  $G_j$  al que pertenece el consumo real  $x$ , de modo que:

- Si  $\epsilon > 2\sigma$ , entonces se ha predicho un mayor consumo que el real, y por lo tanto está siendo **sobreestimado**.
- Si  $-2\sigma \leq \epsilon \leq 2\sigma$ , entonces se ha predicho un consumo que se muestra acorde con el real, y por lo tanto está siendo **adecuado**.

- Si  $\epsilon < -2\sigma$ , entonces se ha predicho un menor consumo que el real, y por lo tanto está siendo **subestimado**.

Con el error  $\epsilon$  cometido por cada consumo real  $x$ , el interés ahora se centra en categorizar la significatividad de dicho error a nivel del modelo de edificio  $M^i$ , es decir, si éste se sitúa en los márgenes establecidos como adecuados o si por el contrario, se sitúa en el 5% de las observaciones restantes, con el fin de obtener una descripción lingüística que resuma la validez del modelo estimado. Para ello, emplearemos una variable lingüística  $\mathcal{L}_m$  (Tabla III) compuesta por cinco conjunto difusos, cuyo dominio vendrá dado por la cantidad de observaciones (errores) que encajan en cada categoría descrita anteriormente de acuerdo con la siguiente fórmula:

$$L_j = \frac{|\epsilon_j|}{|\epsilon|} \times 100 \quad (14)$$

donde  $j \in \{\text{sobreestimado}, \text{adecuado}, \text{subestimado}\}$ ,  $|\epsilon_j|$  será la cantidad de errores que encajan en una de las categorías definidas, y  $|\epsilon|$  será la cantidad total de error cometido. Dado

Tabla III: Definición de  $\mathcal{L}_m$  para validar el modelo

Etiqueta	Función de pertenencia ( $\mu$ )
Insignificante	R{0,10,25}
Leve	T{10,25,50}
Normal	T{25,50,75}
Grande	T{50,75,90}
Enorme	L{75,90,100}

que se tienen tres categorías: *sobreestimado*, *adecuado* y *subestimado*, la descripción lingüística que resuma de manera global la validez del modelo de edificio  $M^i$ , vendrá dada por la categoría cuyo  $L_j$  sea máximo. Así pues, aplicando (8), el resumen lingüístico vendrá dado por:  $y$ , que será la categoría donde  $L_j$  es máximo, y  $P$ , que será la etiqueta lingüística de  $\mathcal{L}_m$  asociada a la categoría  $j$ .

### III-B. Validación del Modelo de organización

Los conceptos vistos en los modelos de edificio pueden ser extendidos para determinar la bondad o validez del modelo organizacional  $O$ . Para ello, hay que realizar una agregación de cada una de las categorías: *subestimado*, *adecuado* y *sobreestimado* para cada modelo de edificio  $M^i$  que componen el metamodelo  $O$  y aplicar la misma variable lingüística  $\mathcal{L}_m$  que en los modelos de edificio  $M^i$ . Dicha agregación,  $L_j$ , viene definida por la siguiente ecuación:

$$L'_j = \frac{\sum L_j}{|L_j|} \quad (15)$$

y al igual que en el caso anterior, la descripción lingüística que resuma de manera global la validez del modelo organizacional  $O$ , vendrá dada por la categoría cuyo  $L'_j$  sea máximo. De este modo, aplicando nuevamente (8), el resumen lingüístico vendrá dado por:  $y$ , que será la categoría donde  $L'_j$  es máximo,  $R$  que será el sintagma *organizacional* para distinguirlo del caso anterior, y  $P$ , que será la etiqueta lingüística de  $\mathcal{L}_m$  asociada a la categoría  $j$ .

A lo largo de esta sección, se ha propuesto un método que permite evaluar cada modelo propuesto gracias al uso de resúmenes lingüísticos difusos. El empleo de conjuntos difusos para su definición, brinda la posibilidad de trabajar con límites poco definidos, de modo que la conclusión derivada de los mismos puede hacer uso con *cierto nivel de pertenencia* de varios de estos conjuntos. Cuando se produce esta casuística, los resúmenes lingüísticos basados en (8) no resultan lo suficientemente expresivos como para poner de manifiesto dicha situación. Por ello, en la siguiente sección introducimos un nuevo concepto de resumen extendido que permite expresar en términos lingüísticos conclusiones cuyos niveles de pertenencia difusa no resulten muy marcados.

## IV. RESÚMENES EXTENDIDOS

Los resúmenes lingüísticos deben poseer la suficiente capacidad expresiva como para no enmascarar u ocultar información que induzcan a conclusiones inexactas. Para tratar de capturar esta casuística, proponemos una modificación sobre la definición dada en (8), permitiendo la adición de un cuantificador absoluto (*cercano a, próximo a...*) [10],  $W$ , al resumen  $P$ , dando lugar a un resumen extendido  $P'$ :

$$P' = WP \quad (16)$$

de modo que es capaz de modelar la descripción lingüística en términos de dos etiquetas lingüísticas con un *nexo* que le proporciona la semántica adecuada. El criterio para discernir si resulta necesario la adición del cuantificador  $W$  al resumen  $P$  para obtener una descripción que capture este tipo de casuísticas en términos lingüísticos vendrá dado por un umbral de pertenencia  $\delta$  asociado a cada conjunto difuso, de modo que: si el valor de pertenencia  $\mu$  es inferior a dicho umbral, por ejemplo  $\delta = 67\%$ , entonces es necesario emplear un resumen extendido  $P'$ ; en caso contrario, un resumen  $P$  clásico resulta suficiente. De este modo, la definición dada en (8) quedaría como sigue:

$$[R]y \text{ are } P' \quad (17)$$

## V. RESULTADOS EXPERIMENTALES

En esta sección se prueba la capacidad expresiva que proporcionan los resúmenes lingüísticos a la hora de resumir el estado relativo al consumo energético en una organización, que como se mencionó en la Sección II, será la Universidad de Castilla-La Mancha. Para ello, se emplearán datos de consumo diario del año 2017 para el edificio 62, cuya área de desempeño es docente e investigadora, en el cual destacamos un primer periodo de no actividad, correspondiente al periodo vacacional de Semana Santa, y otro segundo de actividad, correspondiente a un periodo lectivo, seleccionando un día cualesquiera de cada periodo con los que formalizar los resúmenes lingüísticos:  $x_{na} = 1203.42$  y  $x_a = 2036.24$ .

### V-A. Caracterización del consumo respecto a $M^{62}$

De acuerdo a (2), el modelo de edificio viene dado por:

$$M^{62} = \{1070.84, 1563.12\}$$



mientras que el modelo de la organización  $O_{UCLM}$  es el mismo que (4). Para obtener la caracterización lingüística asociada a los consumos de no actividad ( $x_{na}$ ) y actividad ( $x_a$ ), en primer lugar definimos la variable lingüística  $\mathcal{L}_v$  con la que formalizar el resumen lingüístico, cuya definición aparece en la Tabla IV. Acto seguido, se identifica el dominio de  $O_{UCLM}$

Tabla IV: Definición de  $\mathcal{L}_v$  para  $O_{UCLM}$

Etiqueta	Función de Pertenencia ( $\mu$ )
Insignificante	$R\{P_0, P_{10}, P_{25}\}$
Leve	$T\{P_{10}, P_{25}, P_{50}\}$
Normal	$T\{P_{25}, P_{50}, P_{75}\}$
Grande	$T\{P_{50}, P_{75}, P_{90}\}$
Enorme	$L\{P_{75}, P_{90}, P_{100}\}$

sobre el que aplicar la variable lingüística  $\mathcal{L}_v$  para cada uno de los consumos haciendo uso de (11):

$$\arg \min\{dist(1203.42, 1070.84), dist(1203.42, 1563.12)\} = 0$$

$$\arg \min\{dist(2036.24, 1070.84), dist(2036.24, 1563.12)\} = 1$$

lo que nos indica que para  $x_{na}$ , su grupo semántico se corresponde con  $G_0$ , y que para  $x_a$  es  $G_1$ , algo lógico si atendemos a su modelo de edificio. Sabiendo los grupos semánticos de  $x_{na}$  y  $x_a$ , se concluye que el dominio de  $O_{UCLM}$  sobre el que aplicar  $\mathcal{L}_v$  para cada caso son los definidos en (6) y (7) respectivamente. De este modo, obtenemos la pertenencia de  $x_{na}$  y  $x_a$  a cada conjunto difuso definido en  $\mathcal{L}_v$ , y mediante la t-conorma del máximo, se obtiene el conjunto al que pertenece  $x_{na}$ :

$$\begin{aligned} \max\{\mu_{insignificante}(x) = 0, \mu_{leve}(x) = 0, \\ \mu_{normal}(x) = 0, \mu_{grande}(x) = 0, \\ \mu_{enorme}(x) = 1\} = \mu_{enorme}(x) \end{aligned}$$

y  $x_a$ :

$$\begin{aligned} \max\{\mu_{insignificante}(x) = 0, \mu_{leve}(x) = 0, \\ \mu_{normal}(x) = 0, \mu_{grande}(x) = 0, \\ \mu_{enorme}(x) = 1\} = \mu_{enorme}(x) \end{aligned}$$

Una vez se dispone del conjunto difuso al que pertenece el consumo, ya se puede formalizar el resumen lingüístico de acuerdo a (8). Para el caso del consumo de no actividad,  $x_{na}$ , se tiene que:

«el consumo está siendo *enorme*»

lo cual es debido a que el edificio seleccionado posee un centro de datos que está funcionando continuamente, y para el caso del consumo de actividad,  $x_a$ , se tiene que:

«el consumo está siendo *enorme*»

lo cual es debido a un pico de consumo, ya sea por ser un día veraniego, con el consiguiente consumo en los sistemas de aire acondicionado, o acumulación de horas lectivas en ese periodo.

## V-B. Comparación de $M^{62}$ respecto a $O_{UCLM}$

El experimento anterior proporciona información sobre el consumo obtenido por el edificio 62 comparado con la estimación resuelta por su modelo de edificio  $M^{62}$ . Sin embargo, también es posible obtener cómo se sitúa dicho modelo de edificio  $M^{62}$  con respecto al modelo de la organización  $O_{UCLM}$ . Para ello, debemos seleccionar el prototipo del modelo de edificio que se quiere comparar respecto al modelo organizacional. Si se selecciona  $G_0 = 1070.84$ , empleando la misma variable lingüística  $\mathcal{L}_v$  usando como dominio el definido en (6), tenemos que el conjunto difuso al que pertenece  $G_0$ , aplicando la t-conorma del máximo es:

$$\begin{aligned} \max\{\mu_{insignificante}(x) = 0, \mu_{leve}(x) = 0, \\ \mu_{normal}(x) = 0, \mu_{grande}(x) = 0, \\ \mu_{enorme}(x) = 1\} = \mu_{enorme}(x) \end{aligned}$$

mientras que si se selecciona  $G_1 = 1563.12$ , usando  $\mathcal{L}_v$  sobre el dominio definido en (7), tenemos que el conjunto difuso al que pertenece  $G_1$ , aplicando la t-conorma del máximo es:

$$\begin{aligned} \max\{\mu_{insignificante}(x) = 0, \mu_{leve}(x) = 0, \\ \mu_{normal}(x) = 0, \mu_{grande}(x) = 0.49, \\ \mu_{enorme}(x) = 0.51\} = \mu_{enorme}(x) \end{aligned}$$

de modo que los resúmenes asociados son, para el caso de  $G_0$ :

«el consumo de *no actividad* está siendo *enorme*»

y para el caso de  $G_1$ :

«el consumo de *actividad* está siendo *enorme*»

lo que nos sugiere que el edificio seleccionado es de los que más consumen de toda la Universidad, tanto en periodos donde no existe gran actividad docente o investigadora, como en periodos donde sí la hay. Destacar que para el caso de  $G_1$ , hubiese sido más adecuado emplear un resumen extendido de acuerdo a (17), dándonos como resultado:

«el consumo de *actividad* está siendo *grande con tendencia a enorme*»

## V-C. Validación de Modelos

En este apartado llevaremos a cabo la validación del modelo de edificio  $M^{62}$  y el de la organización  $O_{UCLM}$ . Para el primer caso es necesario calcular el error  $\epsilon$  cometido a la hora de estimar el modelo  $M^{62}$ . Por cada día de consumo del año 2017, se debe obtener el grupo semántico  $\tilde{G}$  de su modelo al que pertenece,  $M^{62}$ , para poder calcular dicho error de acuerdo a (12) y (13) respectivamente. Por ejemplo, utilizando los datos de  $x_{na}$  y  $x_a$  del experimento anterior, se tiene que,  $\tilde{G}_{na} = G_0 = 1070.84$  y  $\tilde{G}_a = G_1 = 1563.12$ , por lo que los errores obtenidos en esos días concretos son:  $\epsilon_{na} = 1070.84 - 1203.42 = -132.58$  y  $\epsilon_a = 1563.12 - 2036.24 = -473.12$ . Una vez se tienen todos los errores

$\epsilon$  cometidos en la estimación, el siguiente paso consiste en categorizar su significatividad a nivel de  $M^{62}$  aplicando la variable lingüística  $\mathcal{L}_m$  definida en la Tabla III, donde el  $L_j$  para cada categoría junto con la pertenencia asociada a cada conjunto difuso de  $\mathcal{L}_m$  identificada viene expresada en la Tabla V.

Tabla V: Definición de  $\mathcal{L}_m$  para validar el modelo

Sobreestimado		Adecuado		Subestimado		
$L_j$	$\mu$	$L_j$	$\mu$	$L_j$	$\mu$	
0 %	<b>1</b>	89 %	0	11 %	<b>0.9</b>	<b>Insignificante</b>
	0		0		0.1	<b>Leve</b>
	0		0		0	<b>Normal</b>
	0		0		0	<b>Grande</b>
	0		<b>1</b>		0	<b>Enorme</b>

Luego como el valor de  $L_{\text{adecuado}}$  resulta ser máximo, con una pertenencia total a la etiqueta enorme, podemos concluir que:

«el modelo resulta adecuado de manera enorme»

o lo que es lo mismo, que el modelo  $M^{62}$  captura correctamente los patrones de consumo subyacentes y nos arroja estimaciones que son correctas.

Por otro lado, para validar el modelo  $O_{\text{UCLM}}$ , se agrega cada una de las categorías de acuerdo a (15), cuyo resultado aparece reflejado en la Tabla VI. Aplicando la variable lingüística  $\mathcal{L}_m$

Tabla VI: Pertenencias de  $L_j$  sobre  $\mathcal{L}_m$

counter	$L_{\text{sobre}}$	$L_{\text{adecuado}}$	$L_{\text{sub}}$
0	9 %	91 %	0 %
1	4 %	96 %	0 %
...			
95	13 %	87 %	1 %
96	23 %	77 %	0 %
	<b>9.35 %</b>	<b>87.14 %</b>	<b>3.51 %</b>

sobre cada categoría  $L'_j$  agregada, obtenemos las pertenencias a cada conjunto difuso de  $\mathcal{L}_m$ , cuyos valores se pueden apreciar en la Tabla VII. Así pues, como el valor  $L'_{\text{adecuado}}$

Tabla VII: Pertenencias de  $L'_j$  sobre  $\mathcal{L}_m$

Sobreestimado		Adecuado		Subestimado		
$L'_j$	$\mu$	$L'_j$	$\mu$	$L'_j$	$\mu$	
9 %	<b>1</b>	87 %	0	4 %	<b>1</b>	<b>Insignificante</b>
	0		0		0	<b>Leve</b>
	0		0		0	<b>Normal</b>
	0		0.20		0	<b>Grande</b>
	0		<b>0.80</b>		0	<b>Enorme</b>

resulta ser máximo, con una pertenencia de 0.8 a la etiqueta enorme, podemos concluir que:

«el modelo *organizacional* resulta adecuado de manera enorme»

o lo que es lo mismo, que el modelo  $O_{\text{UCLM}}$  captura correctamente los patrones de consumo encapsulados por cada modelo de edificio que lo conforma, y que por tanto, arroja estimaciones adecuadas.

## VI. CONCLUSIONES Y TRABAJO FUTURO

Este trabajo propone un nuevo enfoque a la hora de analizar y extraer conclusiones a partir de un conjunto de series temporales de datos de consumo energético heterogéneas (i.e. provenientes de múltiples edificios enmarcados en una misma organización), mediante la definición de modelos que resuman en términos lingüísticos difusos la situación consumista de la organización, con el propósito de servir de apoyo a la toma de decisiones de la alta dirección a la hora de acometer políticas energéticas que contribuyan a la configuración de edificios sostenibles. Además, se ha propuesto una extensión de los resúmenes lingüísticos clásicos que permite tratar la casuística donde la conclusión está mejor definida en término de dos etiquetas si el umbral de pertenencia a la mismas no resulta muy marcado. Finalmente, futuras líneas de trabajo deberían estar encaminadas a obtener un aumento del rendimiento del modelo propuesto, ya sea realizando una segmentación más fina de los grupos, o empleando modelos basados en otro tipo de técnicas, como *deep learning*; definir una arquitectura *big data* basada en micros servicios que dé soporte a la definición y manipulación de los modelos o incorporar al modelo un sistema de alertas basado en resúmenes lingüísticos.

## AGRADECIMIENTOS

Los autores quieren agradecer al Ministerio de Economía, Industria y Competitividad de España por el apoyo ofrecido mediante el proyecto TIN2015-64776-C3-3-R, cofinanciado por el Fondo Europeo de Desarrollo Regional (FEDER).

## REFERENCIAS

- [1] O. for Economic Co-operation and Development, *Transition to Sustainable Buildings: Strategies and Opportunities to 2050*, ser. Energy technology perspectives. OECD, 2013.
- [2] M. A. Piette, S. K. Kinney, and P. Haves, "Analysis of an information monitoring and diagnostic system to improve building operations," *Energy and Buildings*, vol. 33, no. 8, pp. 783–791, 2001.
- [3] L. Hernández, C. Baladrón, J. M. Aguiar, B. Carro, and A. Sánchez-Esguevillas, "Classification and clustering of electricity demand patterns in industrial parks," *Energies*, vol. 5, no. 12, pp. 5215–5228, 2012.
- [4] J. Patterson and A. Gibson, *Deep Learning: A Practitioner's Approach*. O'Reilly Media, Inc., 2017.
- [5] A. Alvarez-Alvarez, D. Sanchez-Valdes, G. Trivino, Á. Sánchez, and P. D. Suárez, "Automatic linguistic report of traffic evolution in roads," *Expert Systems with Applications*, vol. 39, no. 12, pp. 11293–11302, 2012.
- [6] A. Alvarez-Alvarez and G. Trivino, "Linguistic description of the human gait quality," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 1, pp. 13–23, 2013.
- [7] R. R. Yager, "A new approach to the summarization of data," *Information Sciences*, vol. 28, no. 1, pp. 69–86, 1982.
- [8] L. A. Zadeh, "A computational approach to fuzzy quantifiers in natural languages," *Computers & Mathematics with applications*, vol. 9, no. 1, pp. 149–184, 1983.
- [9] A. van der Heide and G. Triviño, "Automatically generated linguistic summaries of energy consumption data," in *Intelligent Systems Design and Applications, 2009. ISDA'09. Ninth International Conference on*. IEEE, 2009, pp. 553–559.
- [10] L. A. Zadeh, "Fuzzy logic= computing with words," *IEEE transactions on fuzzy systems*, vol. 4, no. 2, pp. 103–111, 1996.