

Casos de uso: ENTROPY

Aurora González

Universidad de Murcia

aurora.gonzalez2@um.es

19 de Marzo, 2025

Overview

Introducción

Pilotos

IoT y comunicaciones

Modelado y fusión de datos

Diseño y despliegue de algoritmos

- Predicción de consumo energético

- Perfiles de comportamiento a partir del uso del aire acondicionado

- Calidad de los datos

Provisión de servicios y aplicaciones

Introducción

- ▶ Horizon 2020 es un programa de financiación para la innovación y la investigación en Europa (2014-2020) cuyo sucesor es Horizon Europe (2021-2027)
- ▶ Organismos públicos y privados, incluyendo PyMES.
- ▶ Su objetivo es fomentar la cooperación en investigaciones internacionales europeas e integra ciencias sociales y humanidades a los proyectos ICT.

Los proyectos los solicita un Consortium (grupo de organizaciones de distintos países de la UE) entre los que se encuentra el coordinador, que se considera el líder del proyecto.

En 2015 comenzó el proyecto ENTROPY, siendo coordinadora la Universidad de Murcia.

Consortium: España, Chipre, Grecia, Italia, Serbia, Austria y Suiza.

ENTROPY: DESIGN OF AN INNOVATIVE ENERGY-AWARE IT ECOSYSTEM FOR MOTIVATING BEHAVIOURAL CHANGES TOWARDS THE ADOPTION OF ENERGY EFFICIENT LIFESTYLES



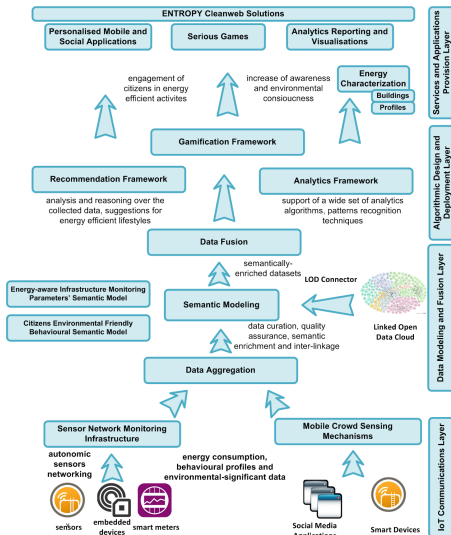
Área temática: Nuevas soluciones basadas en ICT para la eficiencia energética

Programa: Reduccción del consumo energético y la huella de carbón por medio de usos inteligentes y sostenibles

Ver el vídeo Introduction to ENTROPY:

<https://www.youtube.com/watch?v=PC5fQuXsQI4>

Objetivo: cambiar el comportamiento para beneficiar la eficiencia energética



Pilotos

- ▶ Parque Tecnológico Navacchio (cerca de Pisa, Italia)
- ▶ Universidad de Murcia (parque tecnológico de Fuente Álamo, facultad de Química y edificio Pleiades)
- ▶ Technopole en Sierre (Suiza)

IoT y comunicaciones

- ▶ Mecanismos de crowd-sensing
- ▶ Datos estáticos
- ▶ Smart devices

Mecanismos de crowd-sensing

ENTROPY utiliza crowdsensing híbrido:

- ▶ Participatorio: el feedback que nos dan los usuarios de la aplicación y el juego a través identificación de problemas en el uso de los equipos , afirmando haber realizado una tarea...
- ▶ Oportunista: a través de las redes y de los indicadores extraídos del uso de la aplicación y el juego. Por ejemplo, que lean las recomendaciones, que realicen las actividades, que cambien de habitación...

Mecanismos de crowd-sensing

Las multitudes se motivan de forma diferente dependiendo de la iniciativa. Hay 4 tipos de motivación:

- ▶ Contribuir a una causa mayor
- ▶ Posibilidad de ganar dinero
- ▶ Retos y ganar premios
- ▶ Diversión

A través de los datos extraídos del crowdsensing y de encuestas, ENTROPY es capaz de extraer un perfil de comportamiento que depende de estas motivaciones y que lleva a la creación de recomendaciones personalizadas que tendrán una mejor aceptación por parte del usuario.

- ▶ Geolocalización (posición de los usuarios)
- ▶ Preferencias recogidas a través de la ejecución o con cuestionarios
- ▶ Archivos multimedia asociados con un punto de interés (foto de un aparato activo que podría ser desactivado)
- ▶ Datos que el usuario provee sobre variables como: luminosidad, temperatura
- ▶ Percepción del comfort del usuario: temperatura, luminosidad, ruido, condiciones CO2...

Datos procedentes de crowd-sensing

- ▶ Datos y análisis procedentes de la app y el serious game sin la intervención del usuario.
 1. Ránking de eficiencia energética
 2. Categorización de estilo de vida
 3. Adopción de las recomendaciones
 4. Proceso del juego
 5. Compromiso (*engagement*) con la aplicación
- ▶ Feedback sobre el contenido personalizado (“consejo útil”, “objetivo alcanzado”...)
- ▶ Estadísticas sobre la influencia de un usuario en la motivación de otros a tomar acción.
- ▶ Presencia en las redes e influencia (likes, cuentas conectadas, compartidos..)

Datos, datos, datos

- ▶ Datos estáticos relacionados con los ocupantes: demográficos, información socioeconómica, educación y nivel tecnológico, actividades diarias, hábitos, patrones de actividades dentro de los edificios, etc.
- ▶ Datos estáticos relacionados con los edificios: diseño de planta, área total, nivel de certificación energética, aparatos instalados, consumo medio, aislamiento, etc.
- ▶ Datos relacionados con la energía: consumo por edificio, habitación y aparato, facturas, tipos de aparatos, horas de operación... y, en particular, de los aparatos de aire acondicionado (HVAC): on/off, temperatura deseada...
- ▶ Datos relacionados con el lugar: temperatura, humedad, zona climática, longitud y latitud...

Modelado y fusión de datos

- ▶ Mecanismos de agregación de datos
 1. Registro de edificio y sensores
 2. Monitoreo de streams de datos
- ▶ Mecanismos de fusión de datos: colecciones MongoDB
- ▶ Modelos semánticos

Se utiliza FIWARE para conectar los sensores. Para ello se basa en diferentes tecnologías a través del ORION Context Broker (OCB) y los agentes IoT relacionados.

El propósito de utilizar un Context Broker es homogeneizar las fuentes de datos heterogéneas en un modelo de datos común.

Monitoreo de streams de datos

Una característica clave del ORION context broker (OCB) es que se ha diseñado para guardar el estado actual de las entidades. Por ello, sólomente guarda el último valor de los atributos de una entidad y carece de capacidad de almacenamiento.

Es por eso que el catálogo de FIWARE incluye en su ecosistema el componente COMET. El componente COMET se encarga de manejar (guardar y recuperar) datos crudos históricos y agregar la información de las series temporales relacionada con su evolución en el tiempo.

Entre otras cosas COMET provee de un *Rate Change Estimation* que se ejecuta cada cierto tiempo para detectar cambios (suma / sampleo).

Monitoreo de streams de datos

COMET utiliza algunos métodos de agregación para los atributos numéricos: mínimos, máximos, medias... De esta forma, una petición a COMET para alguna de estas agregaciones presentaría el formato:

```
http : // < sth - host > : < sth - port >
/STH/v1/contextEntities/type/ < entityType > /id/ <
entityId > /attributes/ < attrName > ?aggrMethod = <
aggr_method > &aggrPeriod = < aggrPeriod > &dateFrom = <
initDate > &dateTo = < finalDate >
```

Colecciones MongoDB

Después se definen unas colecciones MongoDB.

Se ha seleccionado MongoDB para la implementación del repositorio de datos ENTROPY principalmente por su disponibilidad, horizontal scaling, características avanzadas de seguridad y por el formato de datos que acepta.

Específicamente el formato JSON-LD admite el mapeo de la información recogida a los modelos semánticos de ENTROPY (comportamiento y iot-energía) y proporciona almacenamiento de la información semántica de los datos de acuerdo a los modelos

Ejemplo colección MongoDB: Persona

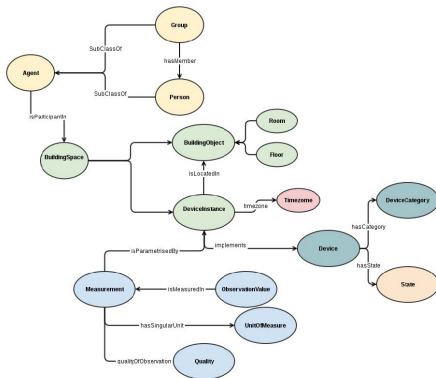
- ▶ `_class`: clase java que describe el objeto en la plataforma ENTROPY
- ▶ `@id`: URI que describe a una persona beneficiaria de la plataforma ENTROPY
- ▶ `@type`: tipo de persona (estudiante, empleado)
- ▶ `educationalLevel`
- ...
- ▶ `hasActivitiesAt`: array JSON de los espacios donde esta persona tiene actividades

Los hay para: espacio del edificio, sensor, valor observado, recomendación, plantilla para proceso analítico, proceso analítico, query de entrada analítica

UMU

Modelos semánticos

Energético



Más detalles en Şimşek, U., Fensel, A., Zafeiropoulos, A., Fotopoulou, E., Liapis, P., Bouras, T., ... & Skarmeta, A. (2016, September). A semantic approach towards implementing energy efficient lifestyles through behavioural change. In Proceedings of the 12th international conference on semantic systems (pp. 173-176).

Diseño y despliegue de algoritmos

Aunque las empresas dicen que utilizan analytics, en realidad lo más habitual es que utilicen herramientas de business intelligence (BI) que devuelven gráficas simples y análisis descriptivos en vez de análisis en tiempo real y predictivos que utilizan métodos más complejos o de aprendizaje automático.

Vamos a mostrar 4 de los varios análisis que se han llevado a cabo en ENTROPY:

- ▶ Predicción de consumo energético
- ▶ Agrupación por uso del aire acondicionado
- ▶ Calidad de los datos
- ▶ Cálculo de indicadores

Proceso de modelado

Para el modelado del **consumo energético** utilizamos el software **R** y varios paquetes siendo **CARET** el más importante entre todos.

El proceso incluye

- ▶ Limpieza y descripción de los datos
- ▶ Definición de los modelos
- ▶ Preprocesamiento
- ▶ Validación y entrenamiento
- ▶ Evaluación

Predicción de consumo energético

Predicción de consumo energético: motivación

Para medir la efectividad de las medidas de cambio de comportamiento se han construido modelos *baseline* o de referencia del consumo energético. De esta forma, se comparará el consumo que se ha tenido mientras las medidas estaban activas con el consumo que se debería haber tenido según los modelos de predicción.

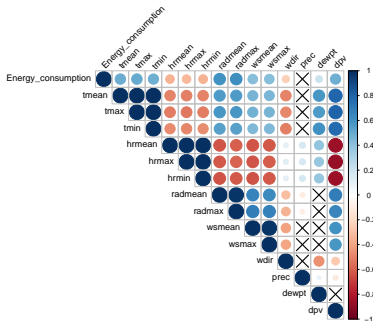
Predicción de consumo energético: variables de entrada

Variables de entrada:

- ▶ Consumo en días anteriores (?)
- ▶ Variables medioambientales: temperatura, humedad, radiación, lluvia,...
- ▶ Información sobre la ocupación: día de la semana, vacaciones

Predicción de consumo energético: Limpieza y descriptivos

- ▶ Eliminar ruido y valores atípicos, entender variables de entrada y salida, transformar variables
- ▶ Visualización



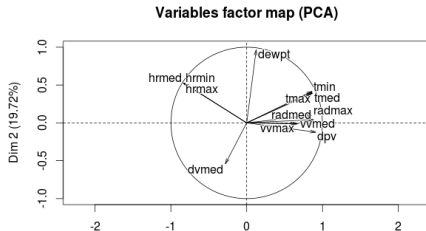
Predicción de consumo energético: Preprocesamiento

► Estandarización o Normalización

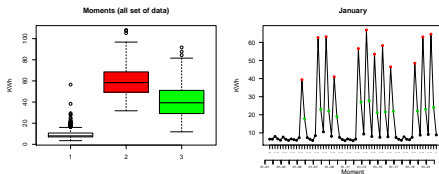
$$x_{est} = \frac{x - \mu}{\sigma}$$

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

► PCA



Predicción de consumo energético: definir modelos



- ▶ Momento 1: vacaciones, fines de semana y noches (22:00 PM - 06:00 AM). REC = [3.578, 14.1] KWh, MEC = 7.904 KWh.
- ▶ Momento 2: mañanas habituales (06:00 AM - 14:00 PM). REC = [26.01, 86.19] KWh, MEC = 54.27 KWh.
- ▶ Momento 3: tardes habituales (14:00 PM - 22:00 PM). REC = [6.357, 53.290] KWh, MEC = 31.48 KWh.

REC = Rango de consumo energético MEC = Media de consumo energético

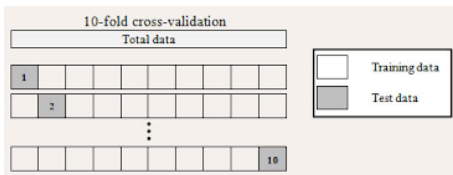
Predicción de consumo energético: Justificación

- ▶ La visualización no es normalmente suficiente
- ▶ Testeo de hipótesis: testar si existen diferencias estadísticas entre estos grupos

Kruskal-Wallis H test: $H(2) = 547.7$, $p\text{-value} < 0.01$ y post-hoc Wilcoxon test

Predicción de consumo energético: Validación y entrenamiento

- ▶ Dividimos los datos en train (75%) y test (25%) set
- ▶ Usamos 10-fold cross validation y 5 repeticiones en el conjunto de training para encontrar la mejor combinación de hiperparámetros.



Predicción de consumo energético: Búsqueda de hiperparámetros (grid search)

Aplicando la estrategia de cross validation al conjunto de entrenamiento se pueden seleccionar los mejores hiperparámetros.

Técnica	Función en R	Parámetros	Valores para tunear
Multi-Layer Perceptron (MLP)	mlp	size	{33, 34, 35, 36, 37, 38, 40}
Support Vector Machines with Radial Basis Function Kernel (SVM)	svmRadialCost	cost	{1, 2, 3, 4, 5, 8, 10}
Gaussian Process with Radial Basis Function Kernel (Gauss)	gaussprRadial	sigma	{0.01, 0.05, 0.1, 0.5}
Bayesian Regularized Neural Networks (BRNN)	brnn	neurons	{2, 3, 4, 5, 10}
Random forest (RF)	rf	mtry	{2, 3, 4, 5, 6, 7}

Predicción de consumo energético: Evaluación

Aplicamos cada modelo al conjunto de test y calculamos el root mean square error (RMSE) y su coeficiente de variación.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

$$CVRMSE = \frac{RMSE}{\bar{y}} \quad (2)$$

Inputs finales: día de la semana, mes, estación, radiación, humedad, temperatura media y temperatura media de los días anteriores.

- ▶ Momento 1 (noche): RF, mtry = 4, RMSE = 1 KWh, CVRMSE = 12.18 %
- ▶ Momento 2 (mañana): BRNN, neurons = 2, RMSE = 7.08 KWh, CVRMSE = 12.86 %
- ▶ Momento 3 (tarde): RF, mtry = 3, RMSE = 3.87 KWh, CVRMSE = 12.09 %

Perfiles de comportamiento a partir del uso del aire acondicionado

Comportamiento HVAC: variables de entrada

HVAC: Heating, Ventilation, and Air Conditioning

Variables de entrada:

- ▶ Temperatura dentro de la sala
- ▶ Estado on/off cada 10 minutos
- ▶ Temperatura deseada por el usuario

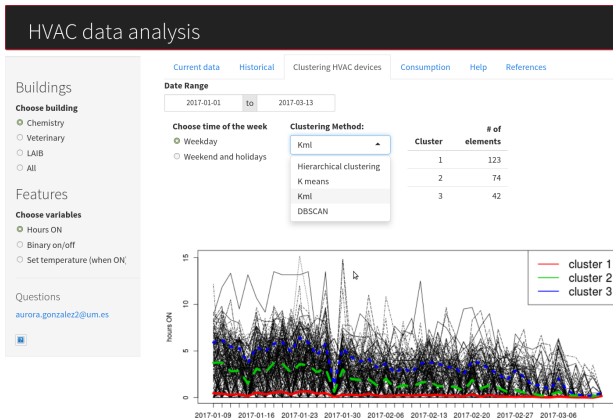
Comportamiento HVAC: modelado

Tras la agregación diaria de los datos (horas de encendido, temperatura de consigna media, interacciones con el mando....), se prueban los siguientes modelos

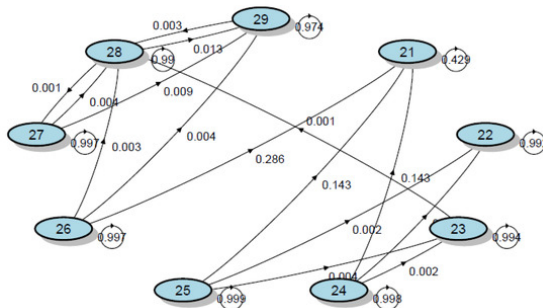
- ▶ Clustering jerárquico
- ▶ K-means normal y longitudinal
- ▶ DBSCAN
- ▶ Clustering espectral

Perfiles de comportamiento a partir del uso del aire acondicionado

Comportamiento HVAC: exploración



Comportamiento HVAC: exploración



González-Vidal, A., Ramallo-González, A. P., & Skarmeta, A. (2018, June). Empirical study of massive set-point behavioral data: Towards a cloud-based artificial intelligence that democratizes thermostats. In 2018 IEEE International Conference on Smart Computing (SMARTCOMP) (pp. 211-218). IEEE.

Comportamiento HVAC: utilidad

- ▶ Lanzar acciones específicas diferenciando entre grupos de usuarios con comportamiento similar (clusters)
- ▶ Mejorar la predicción del consumo energético a partir de las agregaciones por grupos (clusters)
- ▶ Detectar valores atípicos que sean fallos de los aparatos
- ▶ Reducir el número de aparatos de aire acondicionado que deben ser monitorizados para obtener los mismos resultado

Calidad de los datos

Valorar la calidad de los datos es crucial en procesos donde los datos recogidos se transmiten a un repositorio central.

Definir la calidad del sensor es importante porque tendrá un impacto en

- ▶ La selección del modelo
- ▶ La estimación de los parámetros

y, como consecuencia, en las predicciones y otros resultados de los análisis.

Para determinar la calidad se ha implementando un sistema de detección de valores faltantes en series temporales

$$\%Outliers = \frac{num(outliers)}{longitud(stream)}$$

$$Q = 1 - \%Outliers$$

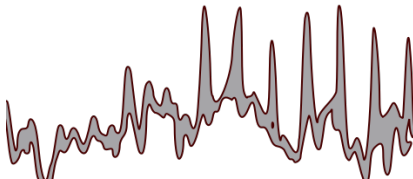
Tipos de outliers

- ▶ Additive outliers (AO): Un AO afecta el nivel de la serie observada solo en el punto donde ha ocurrido.
- ▶ Innovational outliers (IO): Un IO se caracteriza por un impacto extraordinario cuyo efecto persiste en algunas observaciones siguientes.
- ▶ Level shifts (LS): En un LS todas las observaciones que aparecen después del outlier se han movido a un nuevo nivel. Es permanente

Tipos de outliers

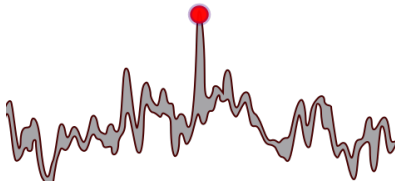
- ▶ Temporary change (TC): Los outliers TC son similares a los IO pero el efecto disminuye exponencialmente en las siguientes observaciones hasta volver a su comportamiento habitual
- ▶ Seasonal Level Shifts (SLS): Un SLS aparece como un valor sorprendentemente alto o bajo que se repite en intervalos regulares.

Identifica el tipo de outlier



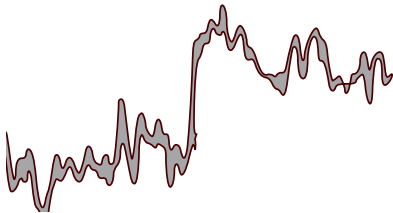
- ▶ AO: Additive outliers
- ▶ SLS: Seasonal Level Shifts
- ▶ IO: Innovational outliers
- ▶ LS: Level shifts

Identifica el tipo de outlier



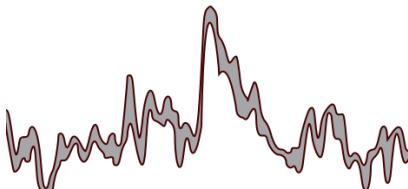
- ▶ AO: Additive outliers
- ▶ SLS: Seasonal Level Shifts
- ▶ IO: Innovational outliers
- ▶ LS: Level shifts

Identifica el tipo de outlier



- ▶ AO: Additive outliers
- ▶ SLS: Seasonal Level Shifts
- ▶ IO: Innovational outliers
- ▶ LS: Level shifts

Identifica el tipo de outlier



- ▶ AO: Additive outliers
- ▶ SLS: Seasonal Level Shifts
- ▶ IO: Innovational outliers
- ▶ LS: Level shifts

Cálculo de indicadores

Hay un procedimiento para calcular indicadores relacionados con la eficiencia energética y la aceptación y utilidad del proyecto. Estos indicadores son: **Compromiso**, **Conocimiento** y **Efectividad**. La evolución de estas métricas se monitorea. Éstas métricas están basadas en el feedback medido con crowd-sensing a través de los móviles.

Cálculo de indicadores

Métrica de comportamiento	Fórmula	
Compromiso	$Y=0.4(\text{logins últimos 30 días usuario / top usuario logins})+0.4(\text{interacciones contenido últimos 30 días usuario / top usuario interacciones})+0.2(\text{fallos registrados por el usuario/fallos del top})$	Interacción del usuario con la app y su contenido
Conocimiento	$Y=0.2(\text{consejos leídos / enviados})+0.3(\text{respuestas correctas / leídas})+0.3(\text{compromisos rellenados / compromiso})+0.2(\text{fallos resueltos por el usuario / fallos})$	Nivel de conocimiento de los usuarios adquirido por su interacción con las apps
Efectividad	$Y=0.4(\text{tiempo medio que tarda en leer las notificaciones / tiempo que están disponibles})+0.1(\text{tips leídas/enviadas})+0.1(\text{preguntas leídas/enviadas})+0.1(\text{respuestas correctas/leídas})+0.1(\text{compromisos llevados a cabo / compromiso})+0.2(\text{fallos registrados})$	Efectividad y rapidez en la interacción

Provisión de servicios y aplicaciones

Plataforma ENTROPY

- ▶ Acceso al ecosistema ENTROPY
- ▶ Utiliza: Internet de las cosas, fusión de la información, recomendaciones basadas en reglas y mecanismos de Big Data analytics.
- ▶ Provee: gestión avanzada de la energía y servicios de concienciación energética en el sector de los edificios

plataforma ENTROPY

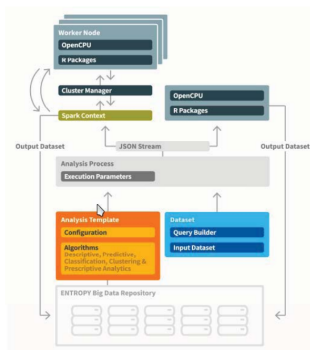
1. Gestión de áreas, sensores y *streams* de datos
 - ▶ Registrar y gestionar áreas y subáreas
 - ▶ Sincronizar los sensores y asignarlos por área
 - ▶ Activar y gestionar los streams de datos
2. Fusión de datos y análisis
 - ▶ Diseñar y hacer búsquedas (*queries*) en los datos disponibles
 - ▶ Diseñar y activar procesos de minería y análisis de datos
3. Recomendaciones personalizada
 - ▶ Diseñar las plantillas de recomendación (consejos, tareas, preguntas)
 - ▶ Componer reglas para mandar mensajes a los usuarios de acuerdo al contexto medido por los sensores

Ver la plataforma en acción:

<https://www.youtube.com/watch?v=qKJrcwUG88w>

plataforma ENTROPY - Arquitectura análisis de datos y OpenCPU

La arquitectura para realizar análisis de datos



plataforma ENTROPY - Arquitectura análisis de datos

OpenCPU¹ - Sistema para realizar computación embebida e investigación reproducible

- ▶ API HTTP
- ▶ Integra R y Javascript (equivalente para Python es Flask)

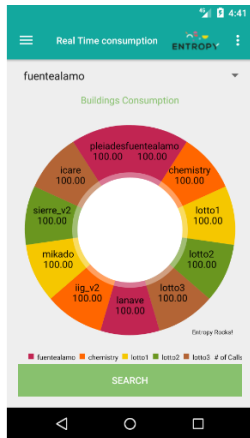
The screenshot shows the 'HTTP request options' dialog in OpenCPU. It includes a 'Method' dropdown set to 'POST' and an 'Endpoint' text field containing '.../library/ParkForecast/R/metrics'. Below this, a note states 'Method and Endpoint are required. Click below to add additional parameters.' A table lists parameters: 'n' with value '100', 'W' with value '600', and 'metric' with value '1'. Each parameter has a 'Param Name' column and a 'Param Value' column. The 'Param Value' column has red text labels: 'Number of retrieved data' for 'n', 'Time aggregation in seconds' for 'W', and 'Do you want to calculate metrics? yes = 1/ no = 0' for 'metric'. At the bottom, there are buttons for '+ Add parameter', '+ Add file', and a green 'Ajax request' button.

OpenCPU crea un servicio web, empujando la funcionalidad de R a la web con la misma facilidad con la que se escribe un paquete de R, es muy flexible y accesible.

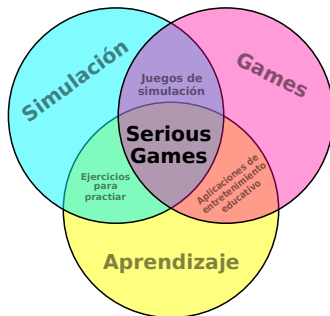
¹<https://www.opencpu.org/>

Aplicación personal

- ▶ Datos de sensores en tiempo real
- ▶ Notificaciones para ahorrar energía basadas en contexto
- ▶ Visualización de indicadores de energía y consumo
- ▶ Quizzes educativos
- ▶ Consejos educativos
- ▶ Personalización y adaptación al perfil del usuario
- ▶ Seguimiento de las KPIs de comportamiento



Serious game



Un *serious game* es un juego cuyo objetivo principal va dirigido al entrenamiento y no al entretenimiento.

Serious game

- ▶ Búsqueda del tesoro con realidad aumentada
- ▶ Orientado a objetivos y retos
- ▶ Cupones y premios
- ▶ Preguntas de múltiple respuesta
- ▶ Marcadores de realidad aumentada para descubrir pistas para las preguntas
- ▶ Android and iOS Versions

