

Proyectos GIM

Repositorio de casos prácticos



Alfredo Sánchez Alberca
asalber@ceu.es
<https://aprendeconalf.es>

Índice de contenidos

Prefacio	3
Licencia	3
1 Demanda eléctrica	4
1.1 Objetivos	4
1.2 Tareas	5
1.3 Datos	5
2 Envío de mensajes en una red de telefonía móvil	6
2.1 Objetivos	7
2.2 Tareas	7
2.3 Datos	8
3 CONVIVE: Big Data para la ciudad inteligente	9
3.1 Objetivos	10
3.2 Descripción técnica	10
3.2.1 Subsistema de adquisición de datos	11
3.2.2 Subsistema de gestión de datos	12
3.2.3 Subsistema de análisis de datos	12
3.3 Tareas	13
3.4 Datos	14
4 Sistema de cálculo simbólico de derivadas	15
4.1 Objetivos	15
4.2 Tareas	16

Prefacio

¡Bienvenida/os al repositorio de proyectos GIM!

Este repositorio contiene casos prácticos para desarrollar proyectos en el grado de Ingeniería Matemática.

Licencia

Esta obra está bajo una licencia Reconocimiento – No comercial – Compartir bajo la misma licencia 3.0 España de Creative Commons. Para ver una copia de esta licencia, visite <https://creativecommons.org/licenses/by-nc-sa/3.0/es/>.

Con esta licencia eres libre de:

- Copiar, distribuir y mostrar este trabajo.
- Realizar modificaciones de este trabajo.

Bajo las siguientes condiciones:

- **Reconocimiento** Debe reconocer los créditos de la obra de la manera especificada por el autor o el licenciador (pero no de una manera que sugiera que tiene su apoyo o apoyan el uso que hace de su obra).
- **No comercial** No puede utilizar esta obra para fines comerciales.
- **Compartir bajo la misma licencia** Si altera o transforma esta obra, o genera una obra derivada, sólo puede distribuir la obra generada bajo una licencia idéntica a ésta.

Al reutilizar o distribuir la obra, tiene que dejar bien claro los términos de la licencia de esta obra.

Estas condiciones pueden no aplicarse si se obtiene el permiso del titular de los derechos de autor.

Nada en esta licencia menoscaba o restringe los derechos morales del autor.

1 Demanda eléctrica



Las empresas energéticas necesitan predecir la demanda eléctrica de sus clientes para poder dimensionar la producción y hacer una compra de energía en el mercado diario. En este mercado se subasta la oferta de energía para el día siguiente y las compañías comercializadoras pujan para adquirir la potencia que estiman que sus clientes van a demandar en las próximas 24 horas.

Para las compañías eléctricas comercializadoras es fundamental predecir la curva de demanda de sus clientes, para ajustar la compra de energía a las demandas reales, y para las compañías productoras es igualmente importante predecir la demanda total para dimensionar la producción de energía del día siguiente. Estas predicciones se realizan habitualmente mediante complejos modelos matemáticos que combinan series temporales, procesos estocásticos y aprendizaje automático.

1.1 Objetivos

En este proyecto no se pretende construir modelos predictivos, pero si construir la curva de demanda eléctrica a partir de los consumos reales en momentos puntuales. Es decir, se trata de desarrollar distintos algoritmos de interpolación para ajustar la curva de demanda a una muestra de consumos puntuales.

Para ello se utilizarán, al menos, los siguientes métodos:

- Interpolación polinómica:
 - Algebraica

- Método de Newton
 - Método de Lagrange
- Interpolación mediante splines:
 - Splines lineales.
 - Splines cuadráticos.

1.2 Tareas

1. Investigar los fundamentos matemáticos de los distintos métodos de interpolación.
2. Programar algoritmos para cada uno de los métodos de interpolación en Python o Julia.
3. Programar una interfaz para acceder a los datos de demanda eléctrica de mediante la API de Red Eléctrica.
4. Desarrollar una aplicación en la que el usuario final elija un día y un método de interpolación, y la aplicación le muestre las curva de demanda para ese día interpolada mediante el método seleccionado.

1.3 Datos

Para probar los algoritmos de interpolación y la aplicación, se utilizarán datos de la web [Red Eléctrica](#). Esta web proporciona una [API](#) que permite acceder a base de datos de Red Eléctrica y que contiene, entre otra mucha información, el histórico de demandas reales.

2 Envío de mensajes en una red de telefonía móvil



La telefonía móvil utiliza una red de estaciones y subestaciones de telecomunicación 4G para la transmisión de datos entre dispositivos móviles. Muchas de estas estaciones están interconectadas por fibra y otras se comunican mediante conexiones inalámbricas por radio frecuencias.

Para que un mensaje llegue de un dispositivo emisor a otro receptor, el mensaje debe recorrer el camino que va de la subestación más próxima al emisor a la más próxima al receptor, pasando, a menudo, por varias estaciones de la red que interconectan las subestaciones de origen y destino. La distancia entre las estaciones y su capacidad de transmisión de datos son claves para que los mensajes lleguen lo más rápidamente posible del dispositivo emisor al receptor.

2.1 Objetivos

En este proyecto se debe desarrollar un algoritmo y una aplicación para buscar el camino óptimo para enviar un mensaje desde un punto geográfico a otro de una ciudad a través del grafo que representa la red de estaciones y subestaciones de telecomunicación de la ciudad.

La aplicación recibirá como entrada las coordenadas de los puntos de origen y destino, y el tamaño del mensaje, y debe devolver el camino más rápido para enviar el mensaje del origen al destino, es decir, el orden de las estaciones por las que el mensaje debe pasar, así como el tiempo que el mensaje tardaría en recorrer ese camino. En la búsqueda del camino óptimo debe tenerse en cuenta la distancia entre las subestaciones, así como la capacidad de transmisión de cada una de ellas.

Para la búsqueda del camino óptimo en el grafo conviene utilizar el famoso [algoritmo de Dijkstra](#) o alguna de sus variantes.

2.2 Tareas

1. Obtener las coordenadas geográficas de la ubicación de las estaciones de telefonía con una frecuencia específica y de un único operador en una determinada ciudad y representar la red en un plano cartesiano. Se puede asumir que las estaciones a menos de x kilómetros de separación están directamente conectadas por fibra.
2. Construir un grafo que represente la red de estaciones de telefonía. Los pesos de las aristas del grafo podrían ser las distancias en línea recta entre las estaciones, pero es mucho más realista es utilizar la velocidad de conexión entre ellas en Mbps, ya que las estaciones base suelen conectarse entre ellas con conexiones inalámbricas. Para ello se puede utilizar la capacidad de Shannon como se explica en este [artículo](#).
3. Identificar en el grafo las estaciones base más cercanas a las coordenadas de los puntos de origen y destino del mensaje.
4. Implementar en Python o Julia el algoritmo de Dijkstra o alguna de sus variantes para la búsqueda del camino óptimo.
5. Determinar cuál será la estación base que requiera más atención de mantenimiento, es decir, por la que pasen más mensajes. Para ello, usar la métrica de centralidad correspondiente. O lo que es lo mismo, si hubiera un agente malicioso (a.k.a. man in the middle, o Charlie), ¿dónde se ubicaría para interceptar la mayoría de mensajes.
6. Determinar cuánto tardará como máximo un mensaje en llegar de emisor a receptor. Usar la matriz de distancias tras ejecutar el algoritmo de Dijkstra y calcular el diámetro del grafo.

2.3 Datos

Para probar obtener los datos de la ubicación y frecuencia de transmisión de las estaciones de telefonía de una determinada ciudad puede consultarse [web del Ministerio de Asuntos Económicos y Transformación Digital](#) o bien esta otra web con el [mapa de estaciones de telefonía del Estado español](#).

3 CONVIVE: Big Data para la ciudad inteligente



Los datos abiertos (open data) es una iniciativa que muchas administraciones públicas están adoptando con el objetivo de mejorar la transparencia en lo que respecta a sus funciones de cara a la ciudadanía.

El ayuntamiento de Madrid es uno de los ayuntamientos que ha empezado a ofrecer datos en abierto, y en su portal de datos abiertos puede accederse a multitud de bases de datos con información de todo tipo, desde los alojamientos en la ciudad hasta las mediciones de contaminación de las estaciones de registro de contaminantes.

La puesta a disposición de la ciudadanía de toda esta información, no solo mejora la transparencia sobre la gestión municipal, sino que permite a ciudadanos o a empresas particulares poder analizar estos datos para mejorar sus decisiones.

3.1 Objetivos

El objetivo principal del proyecto CONVIVE es mejorar el conocimiento del estado real de la ciudad a través de la medición de datos de su entorno y de la información aportada por los ciudadanos. Para ello se desplegará toda la infraestructura necesaria para adquirir esa información y tratarla adecuadamente.

Un objetivo específico del proyecto es que el tratamiento de datos debe realizarse con tecnologías relacionadas con el Big Data. Así nos aseguraremos de que el sistema pueda integrar fuentes de información variadas incluso después de que el sistema haya sido creado.

3.2 Descripción técnica

El sistema a implementar consiste en una plataforma de análisis de varias fuentes de información existentes en el Ayuntamiento de Madrid. Para ello debe solucionar tres aspectos fundamentales:

1. La recogida de información de la ciudad. Este sistema incluye datos de dispositivos, de ciudadanos y de entidades.
2. La gestión de los datos recogidos. Debe ser lo suficientemente genérica para gestionar posibles incertidumbres en los datos propios de una ciudad: información incompleta, nuevas fuentes de información, nuevos campos en la misma información, etc.
3. El análisis de los datos almacenados. Los análisis pueden basarse en la combinación de varios conjuntos de datos integrados en el sistema. Es necesario tener en cuenta que estos datos pueden estar en ubicaciones diferentes y pueden tener un tamaño que no permita su adecuado procesamiento en una sola máquina.

En una ciudad el número de fuentes de información no está fijado ya que en cualquier momento pueden aparecer nuevas fuentes de información a medida que se implanten tecnologías nuevas o formas distintas de participar por parte de los ciudadanos. En el proyecto CONVIVE se definen algunas fuentes de información que deben ser tratadas y se emplean tecnologías que faciliten la integración de nuevas fuentes de información en el futuro. Las fuentes de información que debe gestionar CONVIVE son:

- Avisos de incidencias en la vía pública por parte de los ciudadanos. El sistema integrará estos avisos tanto de forma interactiva (para gestionar la incidencia) como de forma procesada (registros de avisos durante un periodo de tiempo).
 - Los avisos interactivos se almacenarán en un sistema que permita la actualización del estado de los mismos.

- Los avisos procesados suelen corresponder a avisos interactivos que han sido transformados a un formato que facilita su procesamiento estadístico. De esta forma se podrá relacionar estos avisos con otras fuentes estadísticas disponibles en la ciudad.
- Información acústica de la ciudad. El sistema integrará los datos del nivel de ruido existente en distintas partes de la ciudad. Para ello obtendrá la información de ruido en determinadas zonas y los registrará para su posterior análisis.

Para llevar a cabo el proyecto CONVIVE se divide su desarrollo en tres subsistemas distintos que coinciden con los tres aspectos a solucionar previamente comentados.

3.2.1 Subsistema de adquisición de datos

En este subsistema se deben solucionar aspectos relacionados con la recogida de la información. Para ello debe desplegar una red de sensores acústicos en la ciudad (encargados del proceso de recogida de información de ruido) y habilitar un sistema de comunicación para los ciudadanos (con el fin de generar avisos de incidencias en la vía pública).

En el proyecto CONVIVE tendremos dos tipos de información al finalizar el trabajo en este subsistema:

1. Una base de datos con información sobre las incidencias reportadas por los ciudadanos. En esta base de datos se almacenan las incidencias activas y cuyo estado se puede actualizar. Como criterio de diseño se almacenarán en una base de datos (no en un fichero) ya que tiene ciertas ventajas para su gestión.
2. Información relacionada con los niveles de ruido. Se recogerá información de los sensores desplegados en la ciudad y se almacenarán en ficheros con formato CSV.

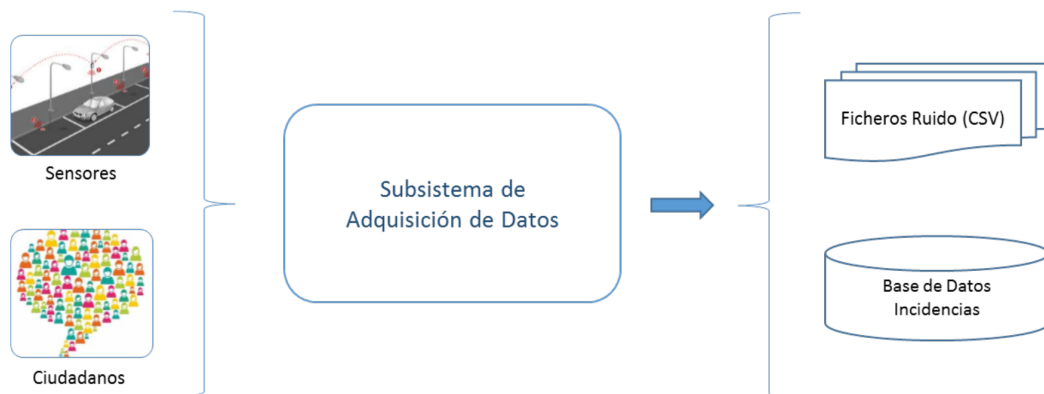


Figura 3.1: Subsistema de adquisición de datos

3.2.2 Subsistema de gestión de datos

En este subsistema se deben solucionar aspectos relacionados con el almacenamiento y sincronización de la información a través de distintas máquinas.

Dado que no se implementará realmente el subsistema de adquisición de datos (aunque sí se llevará a cabo un diseño detallado), se obtendrán datos del sistema AVISA que contiene las incidencias reportadas por los ciudadanos, y del nivel de ruidos del Ayuntamiento de Madrid.

Una vez que se disponga de todos los datos como ficheros en formato CSV (los datos de AVISA y los de ruido) utilizaremos un sistema que permita distribuirlos a través de varias máquinas y sincronizarlos para facilitar su acceso.

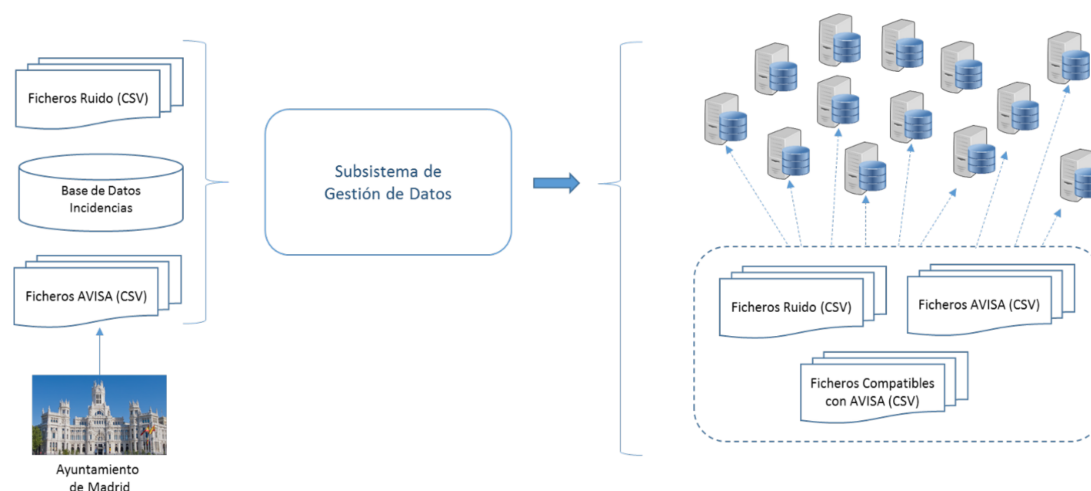


Figura 3.2: Subsistema de gestión de datos

3.2.3 Subsistema de análisis de datos

En este subsistema se obtiene información útil a partir de todos los datos disponibles. En un sistema real se cruzarían los datos del sistema AVISA, con los de niveles de ruido y con varios datos más para obtener información sintetizada y en muchos casos inferir relaciones entre distintos hechos que estén ocurriendo. En el caso del proyecto CONVIVE, por motivos de claridad a la hora de aprender conceptos, se utilizará únicamente un fichero con los datos del sistema AVISA en formato CSV. Se descartarán tanto los datos de ruido como cualquier otro tipo de fuente de datos que puede obtenerse del Ayuntamiento de Madrid.

En el proyecto CONVIVE se realizarán dos acciones con los datos:

1. Se analizarán los datos existentes del sistema AVISA para obtener información sintetizada de los incidentes reportados. Para ello debe tenerse en cuenta que el fichero con los datos puede estar repartido entre varias máquinas y que se debe utilizar un modelo de programación que facilite el acceso a todos esos datos sin cargar la máquina en la que se ejecute.
2. Se visualizarán los datos del sistema AVISA para que sean más fácilmente entendibles por los destinatarios finales. Las decisiones finales sobre las acciones a tomar las llevarán a cabo por personas con distintas capacidades de abstracción y síntesis. Para facilitar esas decisiones se mostrarán los datos mediante modelos visuales que permitan una mejor comprensión de éstos.

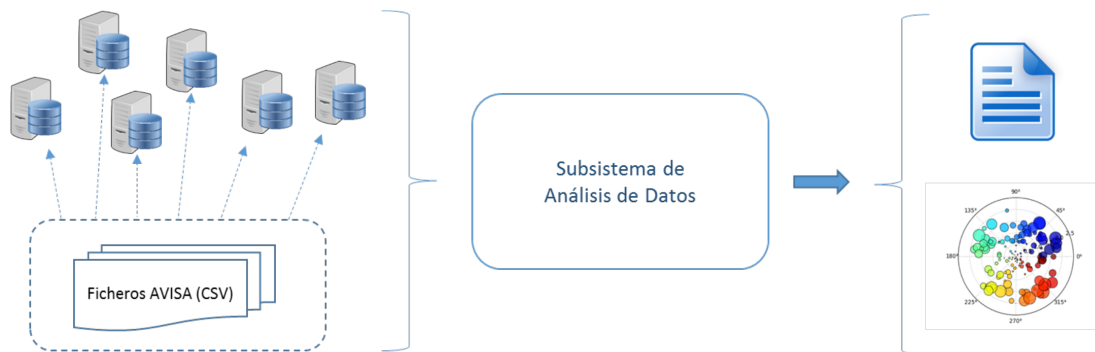


Figura 3.3: Subsistema de análisis de datos

3.3 Tareas

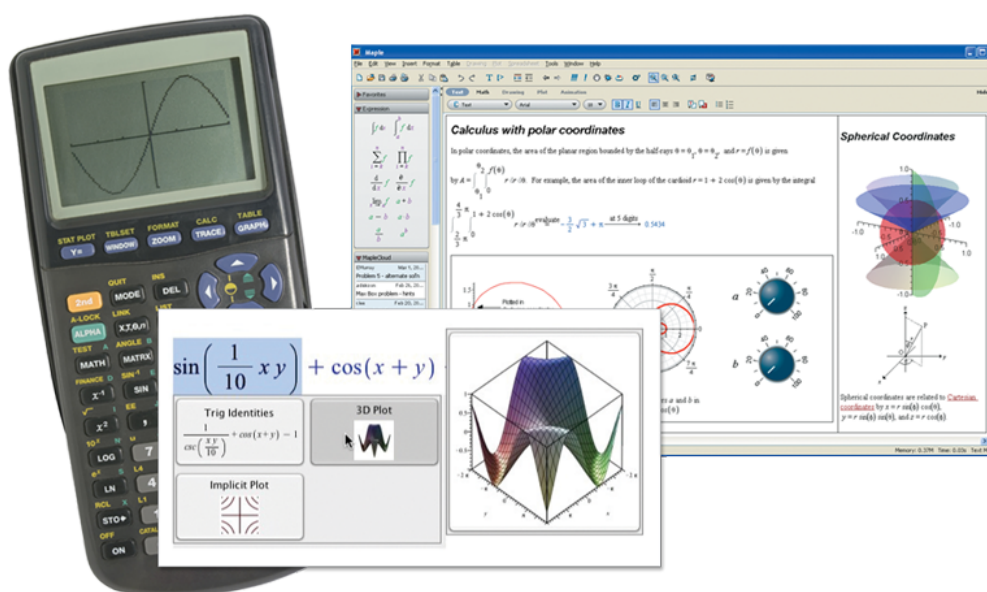
1. Obtener los ficheros de datos del sistema AVISA del Ayuntamiento de Madrid.
2. Obtener los ficheros de ruido del Ayuntamiento de Madrid.
3. Crear el procedimiento de transformación de entradas de la base de datos en ficheros en formato compatible con los del sistema AVISA.
4. Incluir todos los ficheros, los del sistema AVISA, los generados a partir de la base de datos y los de ruido, en un sistema de ficheros distribuido.
5. Preprocesar el conjunto de datos para prepararlo para el análisis estadístico.
6. Realizar un análisis estadístico de las principales variables incluidas en el conjunto de datos.
7. Crear gráficos de visualización de los datos correspondientes a las principales variables del conjunto de datos.

3.4 Datos

Se obtendrán ficheros con información del [portal de datos abiertos del Ayuntamiento de Madrid](#). Este portal tiene un catálogo de datos disponibles bajo los principios de la iniciativa Datos Abiertos (Open Data), que impulsa la publicación abierta, regular, reutilizable y autorizada de los datos de carácter público. En concreto se utilizarán los conjuntos de datos siguientes:

- Datos del Sistema AVISA. Tiene información de avisos de ciudadanos sobre incidencias en la vía pública. Es un fichero en formato CSV.
- Información de la contaminación acústica. Es otro fichero en formato CSV.

4 Sistema de cálculo simbólico de derivadas



Un *sistema de álgebra computacional* (CAS, del inglés computer algebra system) es un programa de ordenador o calculadora avanzada que facilita el cálculo simbólico. La principal diferencia entre un CAS y una calculadora tradicional es la habilidad del primero para trabajar con ecuaciones y fórmulas simbólicamente, en lugar de numéricamente. Esto permite trabajar con expresiones simbólicas (no numéricas) y realizar operaciones como la resolución de ecuaciones, el cálculo de límites, el cálculo de derivadas o el cálculo de integrales.

4.1 Objetivos

El objetivo de este proyecto es desarrollar un sistema de álgebra computacional para el cálculo simbólico de derivadas.

4.2 Tareas

1. Definir simbólicamente las derivadas de las funciones elementales.