

Análisis y visualización de datos de un medidor inteligente para estimar el consumo de electricidad

José Luis Hernández Camacho¹, Dora Luz Almanza Ojeda²

¹Licenciatura en Ingeniería en Mecatrónica

²Departamento de Ingeniería Electrónica

División de Ingenierías, Campus Irapuato-Salamanca, Universidad de Guanajuato

jl.hernandezcamacho@ugto.mx¹, dora.almanza@ugto.mx²

Resumen

La estimación del consumo de energía eléctrica registrado por los medidores caseros se ha convertido en un tema relevante que contribuye a la planeación en la generación de energía. Recientemente, es posible encontrar diferentes bases de datos que proveen las mediciones de consumo de energía eléctrica en casa habitación a lo largo de un año, lo cual permite visualizar, analizar y generar modelos precisos sobre la estimación de consumo en los hogares. En este contexto, este artículo presenta un módulo de lectura de datos de consumo de medidores eléctricos inteligentes domésticos desde archivos csv, que permite visualizar numéricamente y gráficamente el comportamiento del consumo desde diferentes variantes como consumo por horario en el día, por semana o por mes. Otras opciones de visualización se basan en el número de personas que habitan en la casa, pues los consumos también varían de acuerdo al número de habitantes del hogar. Sin embargo, estos datos son utilizados para plantear los posibles modelos de estimación que pueden aplicarse para obtener un patrón de consumo. El módulo de software no requiere instalación de librerías adicionales para su funcionamiento y solo se requiere elegir opciones de menú para visualizar la información. Algunas de las gráficas obtenidas son descritas y mostradas en la sección de resultados. Este módulo proporciona las estadísticas sobre los consumos de energía permitiendo la formulación de modelos de estimación.

Palabras clave: consumo de electricidad; visualización; archivos CSV, estadísticas.

Introducción

Estimar el consumo de electricidad a corto, mediano y largo plazo es una tarea importante para las compañías de electricidad de todo el mundo. Bajar las emisiones de carbono de una ciudad es posible bajo un esquema de predicción, estimación y planeación de consumo de energía eléctrica y las posibilidades de introducción fuentes de energía distribuida para lograr ciudades autosostenibles. Estudios recientes demuestran que el análisis de consumo de energía local provee información más precisa para estimar la distribución espacio temporal de la demanda de potencia eléctrica [1].

El uso de medidores de electricidad inteligentes, conocidos como Smart Meters (SM), permite monitorear los hábitos de consumo de potencia en los hogares. Esta información es almacenada de forma regular siendo diferente de acuerdo con la infraestructura con la que se cuenta en las regiones geográficas [2,3]. Estos intervalos de lectura permiten almacenar una gran cantidad de información sobre la carga consumida de electricidad medida en (kWh). Un análisis específico de estos datos muestra las horas de mayor demanda, daños en la línea de transmisión, aparatos de consumo excesivo, cortes en el servicio e incluso fraudes por parte de los usuarios. La gran cantidad de datos extraída por los medidores inteligentes requiere técnicas avanzadas de organización y separación de información, lo que implica utilizar técnicas rápidas de procesamiento y poderosos recursos de cómputo [3, 4].

Un factor importante del almacenamiento de los datos proporcionados por un SM es la frecuencia a la que debe registrarse la información de consumo. En [5] se plantea que puede realizarse cada hora para zonas residenciales y cada 15 minutos en comercios. Otras propuestas agregan que se realice por día o cada 30 minutos para registro de consumos. Sin embargo, se considera que los retos en el registro de estos datos dependen de, la frecuencia de muestreo de los datos, la resolución del sistema de digitalización y la capacidad para almacenar y transmitir la información a través del ancho de banda disponible [6]. La instalación masiva

de medidores inteligentes complicará las funciones de almacenamiento y análisis de los datos, por lo que se busca proveer de una estrategia logística o de una infraestructura que pueda garantizar los registros [7].

En [8], los autores se enfocan en el análisis de carga de los electrodomesticos, considerando la hora del día, el tiempo de uso, el clima para estimar el consumo en los hogares. Se consideran 3 tipos de uso de los electrodomésticos de dos estados, multi-estado y variable en el tiempo [9]. En el caso de los dos estados, se refiere a los aparatos que solo funcionan como encendido y apagado, por lo que su consumo de potencia se puede considerar constante, un ejemplo de ellos es el refrigerador. Un electrodoméstico multi-estado tiene un estado apagado y dos o mas encendidos, donde cada uno ellos consumen diferente potencia, como ejemplo de estos se tiene la secadora con diferentes ciclos de uso. En el caso de los aparatos variables en el tiempo no muestran un consumo de potencia estable debido a que utiliza diferentes velocidades de un motor de acuerdo con el uso. Esta clasificación de los electrodomésticos permite monitorear el consumo en tiempo real y considerar estimaciones de consumo de acuerdo con el tipo de electrodoméstico que se tiene en la casa. Una estrategia interesante busca detectar el encendido de los electrodomésticos y con ello considerar o estimar su consumo de acuerdo con el horario y tipo de uso estimado para el aparato, de dos estados, multi-estado o variable [10].

Los datos de consumo de energía eléctrica necesitan de técnicas de preprocesamiento y filtrado de datos para comenzar el diseño de modelos de estimación. Los retos en el preprocesamiento son proponer técnicas que permitan distinguir información incompleta, dañada, repetida o que no es útil para su análisis y propuesta en el modelo a estimar [11]. Después del preprocesamiento, se realiza una etapa de visualización para el análisis del comportamiento en el consumo de electrodomésticos en cada caso. Con ello, es posible proponer la mejor técnica de agrupamiento de usuarios con comportamiento similar y en seguida identificar el modelo para la predicción del consumo de potencia en las casas con n-horas de anticipación [1].

Los algoritmos de aprendizaje máquina o machine learning (ML) son de las más utilizadas para la estimación de consumo de potencia, entre ellos los mas comunes son: redes neuronales artificiales (RNA) [6], Support Vector Regression (SVR), Sistemas de inferencia adaptiva neuro difuso [11], herramientas de optimización [12], técnicas de agrupamiento [9, 13], Mapas de Auto-Organizados (Self-Organizing Map), redes bayesianas [14], entre otros. Todos estos algoritmos se utilizan en diferentes bases de datos con informacion sobre mediciones del consumo de potencia en diferentes países, obteniendo modelos interesantes para la estimación de consumo, similaridad de consumo entre usuarios, perfiles de consumo por temporadas, entre otras predicciones importantes para la industria eléctrica.

Las bases de datos entregados por los medidores inteligentes son accesibles a través de un registro y firma de un acuerdo de confidencialidad. Gracias a este acceso a la información, consideramos que la visualización de los perfiles de los usuarios permite dar un paso adelante a modelos de estimación sobre la cantidad de electricidad a generarse, las horas de mayor demanda, las zonas y las temporadas. Una mejor logística de producción permite reducir las emisiones de carbono al planeta, ya que se puede solo producir lo que se requiere y proponer esquemas de reducción de consumo de acuerdo con los hábitos de la region. Este artículo se focaliza en esta visualización y análisis de los datos mediante una plataforma desarrollada en python que nos permita seleccionar diferentes factores de búsqueda y filtrar solo la información importante para su uso posterior en modelos de predicción. La interfaz programada pretende ser un módulo genérico capaz de leer diferentes datos de los medidores inteligentes, apartir de archivos .csv y que de forma automática filtre y grafique los patrones de consumo correspondientes seleccionando entre diferentes periodos de tiempo. La intención es de una herramienta en línea personalizada para cada usuario y que este pueda consultar sus hábitos de consumo.

En la sección siguiente se describe la metodología general para la lectura y acomodo de los datos, las consideraciones que se llevan a cabo y los cálculos estadísticos obtenidos de los datos como parte del comportamiento de consumo. La sección de resultados muestra numérica y gráficamente los resultados obtenidos desde la interfaz programada. Las conclusiones y referencias de este proyecto se muestran al final del documento.

Metodología

Descripción de la base de datos

Las mediciones del consumo de electricidad que nos proporcionan las diferentes bases de datos son registradas a diferentes tiempos de muestreo. Mientras que algunas bases proporcionan mediciones cada hora, cada 30 minutos o hasta 15 minutos, lo cierto es que, las infraestructuras que se requieren para registrar estos datos permiten hacer registros incluso cada minuto. Sin embargo, este lapso de tiempo no es tan conveniente ya que implicaría una gran cantidad de información repetida, siendo lapsos de 30 minutos y una hora los más comunes.

Los datos proporcionados por el medidor inteligente son la fecha, el tiempo de captura y la potencia de consumo. Además, se registra el identificador del medidor para hacer referencia a los diferentes consumidores monitoreados. La mayoría de los datasets proporcionan datos del clima en las fechas de registro de las mediciones, estos datos son proporcionados por al menos una estación meteorológica de la ciudad, y resulta un dato esencial para la generación de modelos de predicción de consumo. Dependiendo de las bases de datos, algunas proporcionan las medidas por 4 años, 2 años, 1 año o hasta solo algunos pocos meses, siendo lo más importante el número de medidores que se reportan y si se trata de una ciudad con climas extremos de frío-calor o solo alguno de éstos. Como información adicional, algunas propuestas sugieren agregar datos de geocalización de los usuarios para involucrar eventos como, el tiempo que pasa fuera de casa, las horas de entrada salida y de forma particular la localización de varios usuarios en un mismo lugar. Lo anterior pudiera representar el que se encuentran en su lugar de trabajo, indicando un exceso de consumo en un mismo punto geográfico y un mínimo de consumo en zonas residenciales. Esta información al ser más de carácter privado es solo una propuesta para lograr modelos mucho más precisos en la estimación del consumo, sin llegar a ser proporcionada hasta ahora por alguna de las bases de datos disponibles en los diferentes sitios de internet.

En nuestro caso, elegimos la base de datos proporcionada en IEEEDataPort [15], la cual proporciona información de 3248 hogares en un lapso de enero a diciembre del año 2017, esta base de datos proporciona mediciones cada 30 minutos del consumo energético de diferentes hogares y además se incluyen datos adicionales de la provincia como el número de habitantes, aparatos eléctricos en el hogar y su estado o condición y datos del clima (temperatura) diaria promedio, mínima y máxima.

Diagrama general

La figura 1 muestra el diagrama a bloques de la metodología llevada a cabo para leer, visualizar, gráficar y analizar estadísticamente las mediciones.

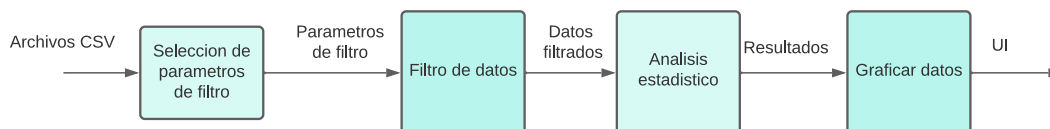


Figura 1. Diagrama general de la metodología propuesta.

Lectura y análisis de los datos

Los datos de los medidores son proporcionados en un archivo llamado consumption.csv, el cual contiene el identificador del medidor, la fecha, la hora y la potencia consumida por el usuario, mientras que en el archivo addInfo.csv se proporciona el mismo identificador del medidor, pero se reportan los datos acerca del hogar, como el número de habitantes. La tabla 1 muestra algunos de los registros proporcionados en la base de datos para el archivo consumption.csv.

Tabla 1. Mediciones de consumo de electricidad para 3 usuarios de la base de datos en la ciudad de Londres.

Identificador del medidor	Fecha	y hora de	consumo
	2017-12-31	2017-12-31	2017-12-31
	13:30:00	14:00:00	14:30:00
0xa62b9f23553ff183f61e2bf943aab3d5983d02d7	0.075	0.063	0.0780
0x459c834d1f6cfb5b734b82aa9f5410fa97fb70da	0.102	0.253	0.3830
0x4a1ed36825360a058cec2bdd409fc2459e1ce54f	0.069	0.067	0.0250

La lectura de los archivos se realizó mediante el lenguaje Python a través de las librerías pandas, matplotlib, numpy, PyQt5, entre otras, que nos facilitan la búsqueda, extracción y visualización de los datos. Para el diseño de la interfaz decidimos agregar datos numéricos y gráficas para la visualización cuantitativa y cualitativa de los datos. Es posible realizar las visualizaciones de datos para diferentes intervalos de tiempo, es decir, podemos mostrar los datos por día, por semana, por mes, para un grupo de usuarios de forma simultánea, la clave es encontrar los campos necesarios para filtrar esta información a como el usuario lo decida. La figura 2 muestra la interfaz diseñada con los campos de selección, las tablas de datos y las gráficas para visualización.

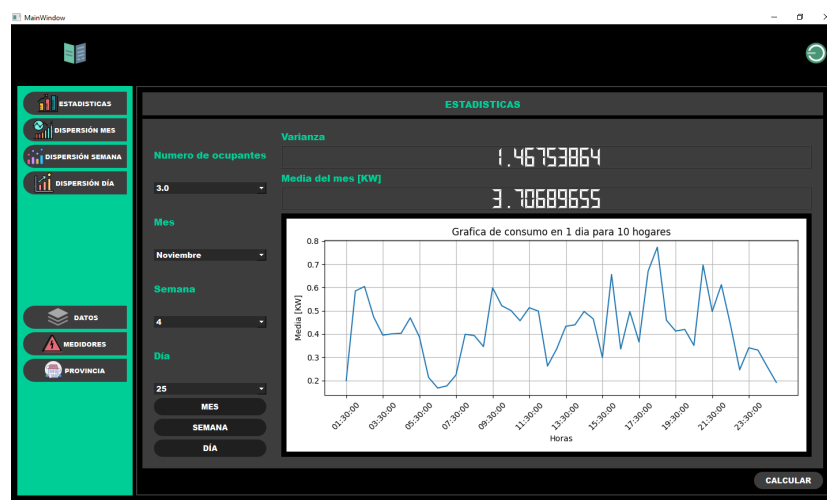


Figura 2. Interfaz de usuario para visualización de datos gráficos y numéricos.

Los datos son seleccionados de acuerdo con el filtro aplicado en la caja del número de ocupantes, con esto se calculan las medidas estadísticas de consumo para los hogares que coincidan con el filtro y la fecha seleccionada. Los resultados nos permiten hacer un análisis más detallado de lo que sucede con los hábitos de consumo en potencia, lo cual se presenta en la siguiente sección.

Resultados

En esta sección presentamos algunas de las gráficas que muestra la interfaz para diferentes períodos de tiempo o selecciones de búsqueda indicadas por el usuario. La figura 3 muestra el resultado para una búsqueda de 2017-11-25 para hogares con 3 habitantes. La figura 4 muestra otra vista para la misma fecha de búsqueda, pero para hogares con 4 habitantes, en la cual se muestra la dispersión de los datos y regresión lineal del mes para los mismos.

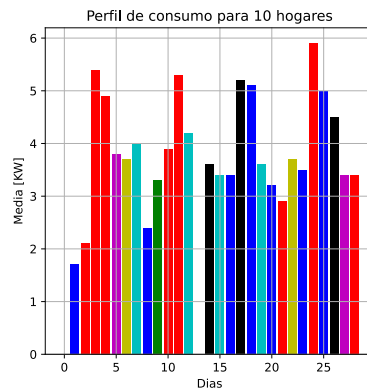


Figura 3. Visualización del perfil de consumo promedio para una búsqueda en 2017-11-25 para hogares con 3 habitantes.

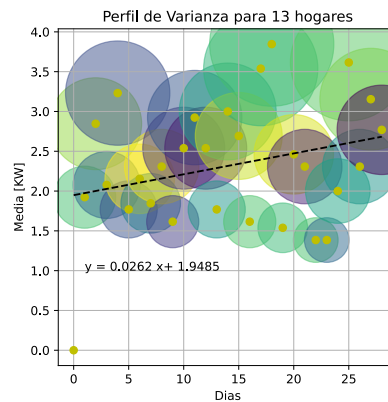


Figura 4. Visualización de la dispersión de los datos y regresión lineal para una búsqueda en 2017-11-25 para hogares con 4 habitantes.

Por otra parte, el análisis estadístico realizado permite obtener un promedio de consumo y su variación para diferentes fechas. Los datos de la media y varianza para las búsquedas indicadas en las figuras 3 y 4 además de un par extra, se muestran en la tabla 2.

Tabla 2. Estadísticas de consumo de electricidad para hogares con diferentes números de habitantes de la base de datos en la ciudad de Londres.

Filtro	Media del mes [KW]	Varianza del mes
2017-11-25 3 habitantes	3.70689655	1.46753864
2017-11-25 4 habitantes	2.31564987	0.063756165
2017-11-25 2 habitantes	2.06739812	0.7134327
2017-12-27 4 habitantes	2.43589744	2.7965812
2017-5-26 4 habitantes	0.64358974	0.17967784

Los datos que se muestran en la tabla 2 son resultado de diferentes búsquedas con cambios en la fecha y número de habitantes del hogar. Con estos datos se observa que un número más elevado de habitantes en

el hogar también se refleja en un mayor consumo promedio de energía eléctrica al mes. Del mismo modo, la fecha en que se realiza la búsqueda también refleja tiempos en que los consumos eléctricos son más elevados por cuestiones festivas o climáticas. Sin embargo, esto no siempre se cumple, pues comparando los resultados de la fila uno contra la fila dos, el único factor que cambia es el número de habitantes. Sin embargo, el consumo energético promedio de hogares con 3 habitantes es mayor al consumo de hogares con 4 habitantes, lo cual refleja que un hogar con más habitantes no significa que debe tener un consumo energético más elevado que uno con menos. Estos detalles son importantes y valdría la pena analizarlos más a fondo pues podría llegar a revelar hábitos en familias por el número de personas en el hogar.

Conclusión

Este artículo describe la importancia del análisis y visualización de datos en el contexto de la industria eléctrica. La generación de grandes cantidades de mediciones utilizando la infraestructura inteligente actual permite almacenar, visualizar y analizar los patrones de consumo, los aparatos de mayor consumo y un comportamiento de la demanda de energía eléctrica, en general para la estimación de necesidades. Existen muchas técnicas de aprendizaje máquina y de inteligencia artificial que facilitan el manejo de esta gran cantidad de datos y permiten la generación de modelos basados en datos reales de consumo. Se pudo notar la importancia de considerar el horario y el día de la semana que se toman las mediciones. Por ello concluimos que, es importante poner más atención a días y horarios laborales o a las llamadas “horas pico” de consumo como son, por la mañana, antes de salir de casa y al regresar por la tarde noche, pero de una manera más exacta, de acuerdo con la región y la cantidad de habitantes en la casa. Las herramientas de software facilitan mucho la lectura y visualización de los datos, pero se requieren de algoritmos rápidos y exactos, capaces de discriminar datos repetidos o incompletos para que la estimación se acerque mucho más a lo requerido. En perspectivas de este trabajo, se pretende mostrar la generalidad de la interfaz programada para incluir datos desde otros conjuntos y que sea posible probar diferentes modelos de aprendizaje profundo para la estimación del consumo, acorde a la base de datos.

Referencias

- Miyasawa, A., Akira, S., Fujimoto, Y. and Hayashi, Y. (2021), Spatial demand forecasting based on smart meter data for improving local energy self-sufficiency in smart cities. *IET Smart Cities*, 3: 107-120. <https://doi.org/10.1049/smc2.12011>
- Barai, G., Krishnan, S., & Venkatesh, B. (2015). Smart metering and functionalities of smart meters in smart grid - a review. 2015 IEEE Electrical Power and Energy Conference (EPEC), 138-145.
- Alahakoon, D., & Yu, X. (2016). Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey. *IEEE Transactions on Industrial Informatics*, 12, 425-436.
- Zhang, Y., Huang, T. & Bompard, E.F. (2018). Big data analytics in smart grids: a review. *Energy Informatics*, 1, 8. <https://doi.org/10.1186/s42162-018-0007-5>
- “Understanding energy use with Smart Meters”, PG&E company. [Online]. https://www.pge.com/en_US/residential/save-energy-money/analyze-your-usage/your-usage/view-and-share-your-data-with-smartmeter/smartmeter-benefits.page (accessed: July 2022).
- Sirojan, T., Lu, S., Phung, B.T., & Ambikairajah, E. (2019). Embedded Edge Computing for Real-time Smart Meter Data Analytics. 2019 International Conference on Smart Energy Systems and Technologies (SEST), 1-5.
- Dudek, G., Gawlak, A., Kornatka, M., & Szkutnik, J. (2018). Analysis of Smart Meter Data for Electricity Consumers. 2018 15th International Conference on the European Energy Market (EEM), 1-5.
- Pereira, L., Costa, D., Riveiro, M. (2022). A Residential Labeled Dataset for Smart Meter Data Analytics. <https://doi.org/10.17605/OSF.IO/JCN2Q>

- Dinesh, C., Makonin, S., & Bajić, I.V. (2020). Residential Power Forecasting Based on Affinity Aggregation Spectral Clustering. *IEEE Access*, 8, 99431-99444.
- Athanasiadis, C., Doukas, D., Papadopoulos, T., & Chrysopoulos, A. (2021). A Scalable Real-Time Non-Intrusive Load Monitoring System for the Estimation of Household Appliance Power Consumption. *Energies*, 14(3), 767. <https://doi.org/10.3390/en14030767>
- Sulaiman, S.M., Jeyanthi, P.A., & Devaraj, D. (2019). Smart Meter Data Analysis Issues: A Data Analytics Perspective. 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), pp. 1-5, doi: 10.1109/INCOS45849.2019.8951377
- Wang, C., Bäck, T., Hoos, H.H., Baratchi, M., Limmer, S., & Olhofer, M. (2019). Automated Machine Learning for Short-term Electric Load Forecasting. 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 314-321.
- Celis, S., Giraldo, L.F., Oliveira-De Jesus, P., & Rincon, C.A. (2018). A Clustering Approach for Domestic Smart Metering Data Preprocessing, 2018 IEEE ANDESCON, pp. 1-3, doi: 10.1109/ANDESCON.2018.8564597.
- Pawar, S., & Momin, B.F. (2017). Smart electricity meter data analytics: A brief review. 2017 IEEE Region 10 Symposium (TENSYP), 1-5.
- Triguero, I. (2020). IEEE-CIS Technical Challenge On Energy Prediction From Smart Meter Data. *IEEEDataPort*. IEEE-CIS Technical Challenge on Energy Prediction from Smart Meter Data | IEEE DataPort (iee-dataport.org)