



Universidade de Vigo

Trabajo Fin de Máster

---

# Desagregación del consumo energético en un hogar inteligente: estado del arte y aplicación a datos reales.

---

Diego Alonso González

Máster en Técnicas Estadísticas

Curso 2017-2018



# Propuesta de Trabajo Fin de Máster

<b>Título en galego:</b> Desagregación do consumo enerxético nun fogar inteligente: estado do arte e aplicación a datos reais.
<b>Título en español:</b> Desagregación del consumo energético en un hogar inteligente: estado del arte y aplicación a datos reales.
<b>English title:</b> Disaggregation of energy consumption in a smart home: state of art and application to real data.
<b>Modalidad:</b> Modalidad B
<b>Autor:</b> Diego Alonso González, Universidad de Vigo
<b>Directores:</b> Javier Roca Pardiñas, Universidad de Vigo; Xose Ramón Sousa Vázquez, Optare Solutions
<b>Tutores:</b> Javier Roca Pardiñas, Universidad de Vigo; Xose Ramón Sousa Vázquez, Optare Solutions
<b>Breve resumen del trabajo:</b> <p>La monitorización y desagregación del consumo energético (NILM) juega un papel crucial en un mundo donde el consumo de los electrodomésticos representa el 30 % de la factura energética. En el presente trabajo, revisaremos el estado del arte del NILM y se propondrá la implementación de redes neuronales profundas para obtener la clasificación de las curvas de consumo de los electrodomésticos y la desagregación de la energía.</p>



Don Javier Roca Pardiñas, Profesor titular de la Universidad de Vigo, don Xose Ramón Sousa Vázquez, Director R&D and OSS Technologies de Optare Solutions, informan que el Trabajo Fin de Máster titulado

**Desagregación del consumo energético en un hogar inteligente: estado del arte y aplicación a datos reales.**

fue realizado bajo su dirección por don Diego Alonso González para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En Vigo, a 03 de julio de 2018.

El director:

Don Javier Roca Pardiñas

El tutor:

Don Xose Ramón Sousa Vázquez

El autor:

Don Diego Alonso González



# Agradecimientos

En primer lugar, me gustaría agradecer a Optare Solutions por brindarme la oportunidad de desarrollar mis prácticas en su empresa y, especialmente, a David e Irene, quienes me ayudaron en todo momento durante mi estancia en la empresa y por su amabilidad, predisposición y colaboración.

En segundo lugar, agradecer a Javier Roca su apoyo en la elaboración y revisión del presente trabajo y por aportar valor en las reuniones que mantuvimos. Por último, y no menos importante, agradecer a Javier Tarrío, profesor de la Universidad de Vigo, quien, de una manera totalmente desinteresada, se involucró en el proyecto desde el primer momento.

# Índice general

<b>Lista de figuras</b>	<b>xI</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Beneficios de la desagregación del consumo energético . . . . .	3
<b>2. NILM: Non Intrusive Load Monitoring</b>	<b>4</b>
2.1. Aproximación al problema . . . . .	4
2.2. Hardware necesario . . . . .	5
2.3. Conceptos básicos . . . . .	6
2.3.1. Métodos basados en evento y no evento. . . . .	6
2.3.2. Aprendizaje supervisado, semi-supervisado o no supervisado.	6
2.4. Reference Energy Disaggregation Dataset . . . . .	7
<b>3. Estado del arte</b>	<b>8</b>
3.1. La desagregación como un problema de optimización . . . . .	8
3.2. Blind source separation . . . . .	9
3.3. Extracción de estados estables de la señal de consumo . . . . .	10
3.4. El algoritmo de Hart . . . . .	12
3.4.1. Resultados del algoritmo de Hart . . . . .	13
3.5. Sparse coding . . . . .	13
3.6. Modelos ocultos de Markov . . . . .	14

3.6.1.	Otros modelos basados en HMM . . . . .	14
3.6.2.	Resultados de los modelos . . . . .	15
<b>4.</b>	<b>Estudios previos</b>	<b>16</b>
4.1.	Preprocesado . . . . .	16
4.1.1.	Detección de ciclos . . . . .	17
4.2.	Identificación de ciclos mediante clúster . . . . .	18
4.2.1.	Resultados obtenidos al aplicar técnicas de clusterización . . .	20
4.3.	Otras pruebas . . . . .	21
<b>5.</b>	<b>Un enfoque diferente</b>	<b>22</b>
5.1.	Nuevo preprocesado . . . . .	23
5.1.1.	Muestra sintética . . . . .	25
5.2.	Redes neuronales. Introducción . . . . .	25
5.2.1.	La función de activación . . . . .	27
5.3.	Aprendizaje . . . . .	29
5.4.	El perceptrón multicapa (MLP) . . . . .	30
5.4.1.	Métodos de optimización . . . . .	31
5.5.	Redes neuronales profundas . . . . .	32
5.5.1.	Las redes convolucionales . . . . .	33
5.6.	Clasificación por imagen . . . . .	35
5.6.1.	LeNet 5 . . . . .	35
5.6.2.	Pattern Recognition App . . . . .	37
5.7.	Modelos de redes neuronales implementados . . . . .	38
5.7.1.	Redes neuronales convolucionales . . . . .	38
5.7.2.	Red neuronal convolucional recurrente . . . . .	39
5.8.	Los datos de entrada . . . . .	40
5.9.	Datos numéricos . . . . .	46

5.9.1. Discretización de consumo desagregado . . . . .	47
5.9.2. Resultados de redes neuronales en el estudio de NILM . . . . .	49
<b>6. Conclusiones y líneas futuras</b>	<b>54</b>

# Índice de figuras

1.1. Número de artículos sobre NILM por año . . . . .	2
3.1. Ciclo de consumo de un frigorífico . . . . .	10
3.2. Ciclo de consumo de una lavadora . . . . .	11
3.3. El calentador únicamente utiliza resistencias por lo que no genera potencia reactiva. Sin embargo, el frigorífico sí genera potencia reactiva y activa. Diagrama cogido de la patente de los EEUU 4858141 [11] . .	13
4.1. Consumo de potencia agregada. . . . .	17
4.2. Ciclo de un frigorífico en el consumo agregado. . . . .	18
4.3. Clústers de los ciclos del frigorífico. . . . .	19
4.4. Ciclo de un frigorífico (negro) vs ciclo de potencia en el consumo agregado (rojo). . . . .	20
5.1. La importancia del ancho de ventana. . . . .	26
5.2. Tipos de modelos neuronales en función de su complejidad. Figura obtenida de Larrañaga et al [17] . . . . .	27
5.3. Estructura de la neurona $i$ -ésima. Figura obtenida de Larrañaga et al [17] . . . . .	28
5.4. Estructura de un perceptrón multicapa feedforward. Figura obtenida de Larrañaga et al [17] . . . . .	30

5.5. Interfaz de Pattern Recognition App . . . . .	37
5.6. Gráfica de consumo agregado de una casa. . . . .	41
5.7. Gráfica de consumo agregado de una casa. . . . .	42
5.8. Interfaz de visualización de datos. . . . .	43
5.9. Ventanas de consumo agregado capturadas cada 20 segundos. . . . .	43
5.10. Imágenes como <i>input</i> de la red. . . . .	44
5.11. Ciclo de un frigorífico con área rellena. . . . .	46
5.12. Con una altura adaptada al frigorífico se pierde información sobre el consumo del lavavajillas. . . . .	47
5.13. Resultados de la desagregación de consumo. . . . .	50
5.14. Resultados extraídos de Kelly [14] con redes neuronales (Autoencoders, rectangles o Long short-term memory (LSTM)) y métodos convencionales. . . . .	52
5.15. Algunos resultados extraídos de Nascimento [7] con diferentes modelos de redes neuronales. . . . .	53

# Resumen

## Resumen en español

Monitorizar y controlar el consumo energético juega un papel crucial tanto en el bienestar de las personas como en la conservación del medio ambiente en un mundo donde los electrodomésticos representan el 30 % del consumo energético de un hogar. En este contexto, la desagregación del consumo energético (NILM) pretende obtener el consumo detallado de cada electrodoméstico a partir de la señal de consumo agregado. En el presente trabajo se hará un repaso de lo acontecido en el periodo de prácticas en Optare Solutions comenzando por la presentación del conjunto de datos de referencia en NILM: RED dataset. Además, se estudiará y analizará desde un punto de vista crítico el estado del arte del NILM, repasando los algoritmos más convencionales y analizando sus resultados. Por último, se propondrá la implementación de redes neuronales profundas para obtener, en primer lugar, la clasificación de las curvas de consumo de los electrodomésticos y continuar con la desagregación del consumo. Se hará una breve introducción a las redes neuronales más elementales para seguir con las redes neuronales profundas. Finalmente, se analizarán los resultados obtenidos.

## **English abstract**

Controlling and controlling energy consumption plays a crucial role in the well-being of people and in the conservation of the environment in a world where household appliances represent 30 % of a household's energy consumption. In this context, the disaggregation of energy consumption (NILM) aims to obtain the detailed consumption of each household appliance based on the aggregate consumption signal. In the present work, what happened in the internship period in Optare Solutions will be repaired, starting with the presentation of the reference data set in NILM: RED dataset. In addition, the state of art of NILM is studied and analyzed from a critical point of view, reviewing the most conventional algorithms and analyzing their results. Finally, the implementation of neural networks will be proposed to obtain, in the first place, the classification of consumption curves of household appliances and will continue with the disaggregation of consumption. A brief introduction will be made to the most basic neural networks to continue with deep neural networks. Finally, the results obtained will be analyzed.

# Capítulo 1

## Introducción

A día de hoy, los electrodomésticos suponen el 30% del consumo energético en un hogar lo que representa una cuantía importante de la factura energética en edificios residenciales. El problema de la desagregación del consumo energético, que consiste en determinar el porcentaje de consumo perteneciente a cada aparato conectado a una fuente de energía, se ha convertido en un asunto con relativa importancia dado el aumento de aparatos electrónicos y las tendencias ecológicas.

Actualmente, nos encontramos con dos grandes tendencias en el estudio de la desagregación energética:

- **Non-Intrusive Load Monitoring: NILM.** La monitorización del consumo energético no intrusiva (NILM por sus siglas en inglés) consiste en medir el consumo de electricidad usando únicamente un medidor inteligente o smart meter que, normalmente, está ubicado en el panel eléctrico principal. Esta técnica consiste en obtener una desagregación del consumo de cada electrodoméstico a partir del consumo global de la casa.
- **Intrusive Load Monitoring - ILM** Esta monitorización del consumo energético se lleva a cabo midiendo el consumo de cada electrodoméstico de manera

individual, es decir, cada electrodoméstico lleva asociado un medidor.

Desde el año 2010 ha habido un drástico aumento del número de artículos publicados sobre la desagregación del consumo energético (ver Figura 1.1) y, desde 2013 se publican más de 100 artículos anuales. Tal es el interés por este tema que algunos autores afirman que la desagregación del consumo energético es el Santo Grial de la eficiencia energética. Por el lado económico, las pruebas del interés sobre este asunto son evidentes: en noviembre de 2015, la compañía Bidgely obtuvo una financiación de 16.6\$ millones de dólares y, actualmente, existen más de 30 empresas que ofrecen productos y servicios de desagregación.

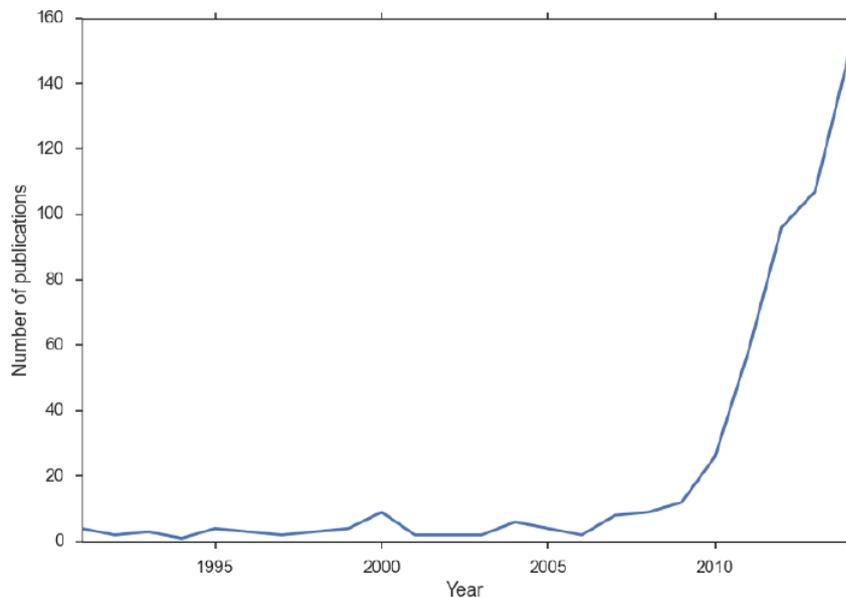


Figura 1.1: Número de artículos sobre NILM por año

## 1.1. Beneficios de la desagregación del consumo energético

Existen varios estudios acerca del comportamiento de los consumidores una vez que conocen con detalle la potencia consumida por sus electrodomésticos y sus hábitos de consumo (el lector interesado puede consultar [9] y [5]) de los que, como conclusión general, podemos extraer que el conocimiento de esta información por parte del consumidor deriva en un cambio en el patrón de comportamiento haciendo que ahorre más energía.

Existen diferentes maneras de hacer llegar esta información al consumidor pero podemos destacar dos por encima de todas: feedback directo e indirecto. La obtención de feedback de manera directa consiste en que el consumidor obtiene información la información directamente en su teléfono móvil o en el medidor inteligente. Por otro lado, el consumidor puede obtener información de manera indirecta, siendo la empresa proveedora quien le entregue esta información a través de las facturas de electricidad con un consumo detallado y desagregado de cada electrodoméstico.

Además, la monitorización del consumo de cada electrodoméstico puede servir para detectar posibles anomalías que puedan derivar en una avería de manera que el mantenimiento predictivo podría ser aplicado a cada uno de los aparatos eléctricos de los hogares.

Carece de sentido discusión entre las ventajas de la desagregación energética intrusiva o no intrusiva desde el momento en el que los resultados de la última no difieren excesivamente de los de la primera. Además de ofrecer un mejor servicio al cliente al ser menos intrusivo, el ahorro en medidores inteligentes y la reparación de sus posibles averías así como el interés asociado a la aplicación de técnicas de machine learning, hace del NILM el objeto principal de estudio del presente trabajo.

# Capítulo 2

## NILM: Non Intrusive Load Monitoring

### 2.1. Aproximación al problema

Los primeros estudios sobre desagregación de la energía datan de 1992 con los trabajos de Hart [10]. Hart pretendía, a partir de los cambios de potencia, agrupar en clúster aquellos que pareciesen ser ocasionados por el mismo aparato.

Sus trabajos sentaron los cimientos sobre los que se construyen los estudios actuales en base a una serie de suposiciones que se detallan a continuación.

Una de esas suposiciones es la conocida como *one – at – a – time* que asume que los aparatos eléctricos se encienden y se apagan de uno en uno, es decir, cada variación en el consumo eléctrico es debido a un único aparato. Esta premisa está supeditada a la frecuencia con la que se obtienen los datos ya que, si obtenemos una medición por minuto, habrá variaciones de potencia debidas al consumo de varios aparatos.

La segunda de sus suposiciones es conocida como *steady state load* que asume que la mayoría de los aparatos eléctricos consumen una cantidad constante de energía

si se mantienen en el mismo estado. En realidad, esta suposición no es correcta ya que el consumo de los aparatos sufren ligeras variaciones.

Años más tarde, en 1998, Cole y Albicki [3] sacaron a la luz otra solución al problema de la desagregación teniendo en cuenta los picos de potencia que, normalmente, se dan en el momento de encendido o apagado.

Con el paso del tiempo y el desarrollo de los medidores inteligentes, se han podido aplicar diferentes técnicas para obtener la desagregación del consumo tales como la transformada de Fourier o técnicas de machine learning.

## **2.2. Hardware necesario**

A pesar de que el presente trabajo se centrará exclusivamente en las técnicas y algoritmos, puede resultar interesante para el lector una breve explicación del hardware necesario en la obtención de los datos.

Existe una amplia diversidad de maneras de obtener los datos de consumo de un hogar. La forma más sencilla y directa de obtener esta información es a través de adaptadores especiales a los que se conectan directamente los aparatos. Esto lleva asociado un coste económico considerable ya que se requiere un adaptador por aparato y no es la solución más óptima de cara al usuario.

Los medidores inteligentes se presentan como la solución a este problema ya que las compañías ya han comenzado a sustituir los viejos medidores por éstos, que aportan una información sumamente valiosa para la desagregación del consumo.

Finalmente, cabe destacar que la propuesta de este trabajo por parte de la empresa Optare Solutions surge de la oportunidad de analizar los datos recogidos a través de la implantación del servicio de domótica para hogares inteligentes de la compañía R.

## 2.3. Conceptos básicos

En el contexto del NILM, es conveniente conocer la terminología propia del campo que se detalla en esta sección.

### 2.3.1. Métodos basados en evento y no evento.

Los métodos basados en evento tienen como objetivo clasificar los cambios de consumo que tienen lugar cuando un aparato cambia de estado. Para ello, estos algoritmos extraen un conjunto de características del consumo agregado: diferencial de consumo, duración del cambio, etc. y permiten identificar y clasificar un evento cuando este ocurre.

Los métodos basados en no eventos no ponen el foco sobre la detección de un evento específico en el consumo agregado sino que aprenden estos eventos automáticamente sin la necesidad de un bloque separado para detectar el evento.

### 2.3.2. Aprendizaje supervisado, semi-supervisado o no supervisado.

En el aprendizaje supervisado el modelo se entrena combinando la señal de consumo agregado con el consumo de cada aparato conectado. En este caso, se hace evidente la necesidad de aplicar un monitoreo intrusivo para obtener la información del consumo de cada aparato.

En el aprendizaje no supervisado, el modelo se entrena empleando para ello únicamente la señal del consumo global. Este es el método por excelencia en términos de optimización y calidad de servicio ya que se trata de un monitoreo no intrusivo. No obstante, es lógico pensar que para obtener unos resultados parecidos al aprendizaje supervisado se requiera un esfuerzo mucho mayor.

El aprendizaje semi-supervisado usa una combinación de datos etiquetados y no etiquetados para entrenar el modelo. En el caso de NILM, es el método más empleado y que mejor encaja con las características del problema.

Cabe destacar que en el estudio del NILM se usa una terminología que puede resultar confusa. En las técnicas de machine learning, al aplicar aprendizaje supervisado o no supervisado, los datos utilizados tanto para el entrenamiento como para el test provienen de la misma fuente. No obstante, en NILM se pueden utilizar para el entrenamiento datos etiquetados o no etiquetados pero, sin embargo, para el test únicamente se pueden emplear datos etiquetados que normalmente provienen de una casa no presente en el conjunto de datos empleado para entrenar el modelo.

## 2.4. Reference Energy Disaggregation Dataset

Reference Energy Disaggregation Dataset (REDD) es el conjunto de datos de referencia en el estudio de la desagregación del consumo energético. Con los datos recogidos por el Massachusetts Institute of Technology (MIT), REDD fue diseñado para convertirse en un conjunto de datos estandarizados sobre el que testear los algoritmos del NILM. Previo a su lanzamiento, se usaron diferentes conjuntos de datos que, normalmente, eran recogidos por el autor del estudio haciendo difícil comparar los resultados obtenidos con diferentes algoritmos.

El MIT recogió información de 6 casas de la región de Massachusetts, en EEUU. Uno de los mayores inconvenientes de este conjunto de datos es que no en todas las casas se recoge información de los mismos electrodomésticos sino que en cada casa, el muestreo difiere ligeramente del resto. Otro de los grandes problemas del REDD es la falta de datos en algunos intervalos de tiempo debido a caídas del sistema eléctrico o mal funcionamiento de los equipos, y es un aspecto a tener en cuenta a la hora de trabajar con estos datos.

# Capítulo 3

## Estado del arte

En este capítulo, se presentan y discuten algunos de los algoritmos que están siendo aplicados en el ámbito de la desagregación del consumo energético.

### 3.1. La desagregación como un problema de optimización

El problema de optimización se plantea como sigue: la señal del medidor inteligente es considerada como una serie temporal tal que  $Y = \{y_1, y_2, \dots, y_k\}$  donde  $k$  es el número de muestras e  $y_t$  es la medición en el instante  $t$ . Se describe el estado de cada electrodoméstico como un vector booleano  $A = \{a_1, a_2, \dots, a_n\}$  para  $n$  electrodomésticos.

Si el electrodoméstico  $i$  –ésimo está encendido, entonces  $a_i = 1$  y  $p_i = 100$ . El consumo total  $y_t$  en el instante  $t$  es la suma de la potencia consumida por todos los electrodomésticos:

$$y_t = \sum_{i=1}^n a_i p_i + e_t$$

Donde  $e_t$  es un término de error.

Si conocemos el consumo de cada electrodoméstico, podemos representar la desagregación como un problema de optimización combinatoria donde, para cada instante  $t$  se intenta encontrar un vector  $A_t^*$  tal que:

$$A_t^* = \operatorname{argmin}_A \left| y_t - \sum_{i=1}^n a_i p_i \right|$$

Este enfoque fue estudiado por George Hart quien concluyó que es computacionalmente impracticable. Argumenta que únicamente se podrá obtener una solución precisa si se enumera cada posible estado. Así, si contamos con datos de  $n$  electrodomésticos y cada uno de ellos puede estar en  $s$  estados, la complejidad computacional es de orden  $\mathcal{O}(s^n)$ . Veámoslo con un ejemplo: si tenemos 3 electrodomésticos que pueden ocupar 4 estados cada uno y el consumo se recoge cada 5 segundos durante un mes, tendremos que hacer  $10^{24}$  operaciones lo que supone aproximadamente unos 1700 años de cómputo en un ordenador relativamente potente.

Además, es poco probable que podamos conocer con exactitud el consumo de cada electrodoméstico o el número total de electrodomésticos conectados. Otro de los problemas que presenta este enfoque es que diferentes electrodomésticos del mismo tipo producen diferentes curvas de consumo.

## 3.2. Blind source separation

Blind Source Separation (BSS) hace referencia a un grupo de métodos que incluyen análisis de componentes principales, descomposición en valores singulares, análisis de componentes independientes y factorización de matrices no negativas.

En términos generales, BSS plantea como objetivo recuperar un conjunto de fuentes de señales individuales  $s(t) = (s_1(t), \dots, s_n(t))^T$  (donde  $t$  es el instante de tiempo desde 1 hasta  $T$  y  $n$  es el número de fuentes de señal) de un conjunto de señales mezcladas  $x(t) = (x_1(t), \dots, x_m(t))^T$  donde  $m$  es el número de señales mezcladas.

Un ejemplo de aplicación de esta técnica es el problema de la fiesta de cóctel donde se tiene un número de personas hablando y el mismo número de micrófonos instalados en la habitación. El objetivo: recuperar la señal de audio de cada persona a partir de la señal global. Sin embargo, en NILM normalmente se tiene menor número de señales mezcladas que fuentes de señal individual, es decir,  $m < n$ .

### 3.3. Extracción de estados estables de la señal de consumo

La mayor parte de los algoritmos aplicados en NILM comparten un primer paso: identificar estados estables de consumo. El objetivo es eliminar transiciones rápidas en la señal como por ejemplo el encendido de un motor como se puede observar en la Figura 3.1.

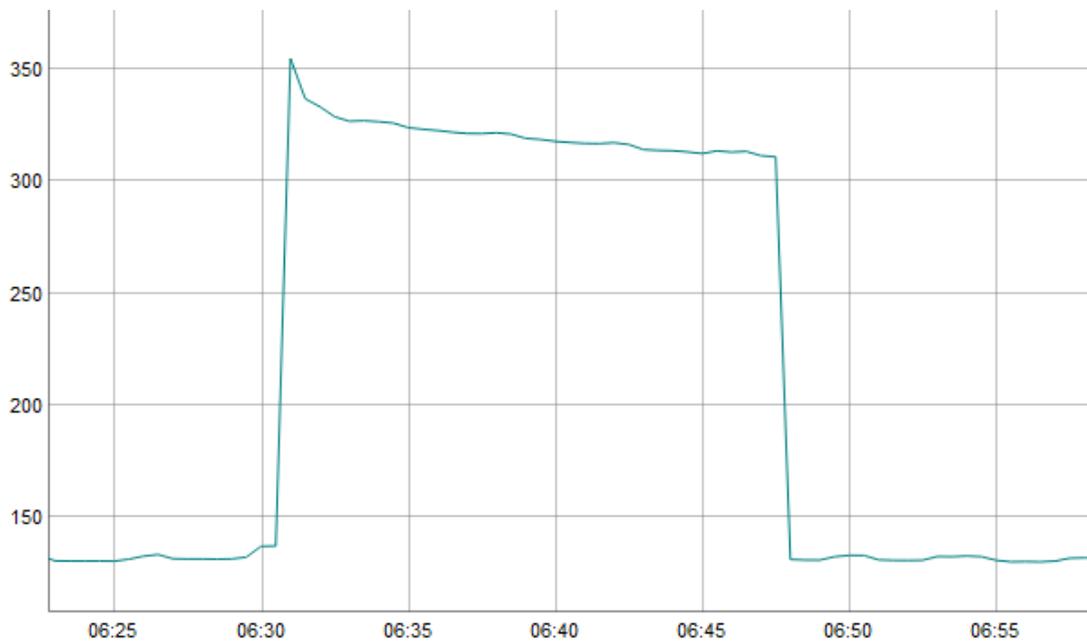


Figura 3.1: Ciclo de consumo de un frigorífico

### 3.3. EXTRACCIÓN DE ESTADOS ESTABLES DE LA SEÑAL DE CONSUMO11

Pero, ¿qué se entiende por estado estable? Hart estableció su propio criterio: en primer lugar, se divide la señal de consumo en secciones estables y de cambio. Una sección estable, es aquella que tiene una mínima longitud, que varía en función de cada aparato, y el consumo no varía más de 15 vatios. Toda sección que no sea clasificada como estable, se clasificará como en cambio. La muestra de una sección estable es promediada y con esto, quedan identificadas las secciones estables.

Para Kelly [14], no tiene sentido hacer esta división entre secciones estables y de cambio ya que ‘convertir una señal en una secuencia de estados estables hace que se pierda información y sería mejor aprovechar las transiciones y características de la curva de consumo para mejorar la desagregación’. Esta opinión cobra más sentido aún al observar la Figura 3.2.

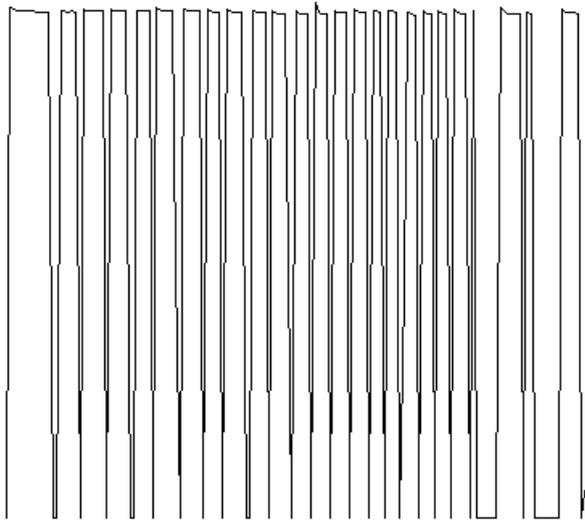


Figura 3.2: Ciclo de consumo de una lavadora

### 3.4. El algoritmo de Hart

El algoritmo de Hart toma como entrada las lecturas de la potencia real, la potencia reactiva <sup>1</sup> y el voltaje tomadas cada segundo. Una vez obtenida la muestra, las variaciones en el voltaje se corrigen normalizando la muestra para, posteriormente, identificar los estados estables y localizar los estados en cambio. En resumen, el algoritmo de Hart consta de 5 pasos:

1. Un detector de cambios identifica cambios en los estados estables de la señal de consumo agregado.
2. Se realiza un análisis clúster para situar estos cambios en un espacio bidimensional de potencia real y reactiva como se representa en la Figura 3.3.
3. Los clústers de magnitud similar y signo opuesto se emparejan (un calentador que consume 500W al encenderse, producirá un cambio de misma magnitud y signo opuesto cuando se apague). Esto es válido para aquellos aparatos que tengan un único estado (on u off) pero no captará aquellos aparatos con curvas de consumo más irregulares como las lavadoras.
4. Los clúster no emparejados se asocian con otros nuevos o ya existentes de acuerdo con un algoritmo de máxima verosimilitud.
5. Los eventos se asocian con etiquetas legibles (*horno* en vez de *load#125*) mediante una base de datos de potencias conocidas.

---

<sup>1</sup>La potencia reactiva no es una potencia (energía) realmente consumida en la instalación sino que aparece en una instalación eléctrica en la que existen bobinas o condensadores, y es necesaria para crear campos magnéticos y eléctricos en dichos componentes.

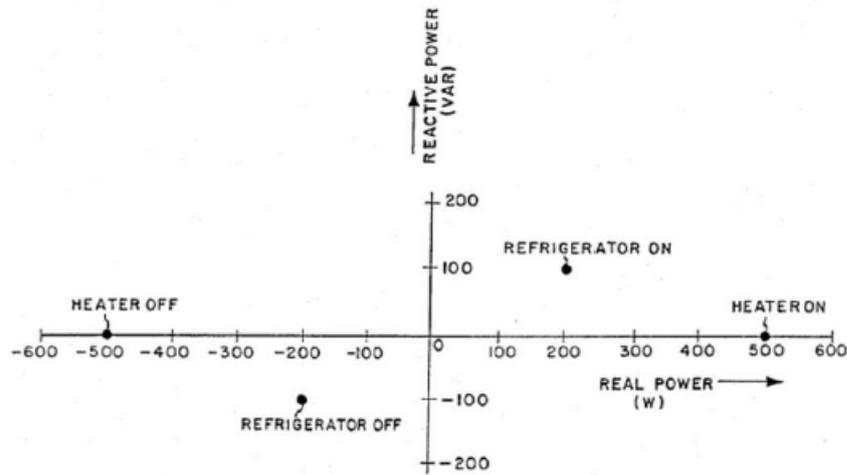


Figura 3.3: El calentador únicamente utiliza resistencias por lo que no genera potencia reactiva. Sin embargo, el frigorífico sí genera potencia reactiva y activa. Diagrama cogido de la patente de los EEUU 4858141 [11]

### 3.4.1. Resultados del algoritmo de Hart

El algoritmo de Hart fue probado en 26 lugares y 128 electrodomésticos siendo capaz de identificar aquellos con un único estado (on u off) alcanzando una precisión del 90 %. Los frigoríficos fueron identificados con una precisión del 80 % mientras otros aparatos con curvas de consumo más irregulares obtuvieron resultados muy pobres.

En los años sucesivos se fueron realizando diferentes modificaciones del algoritmo que mejoraron los resultados pero sin llegar a identificar de manera precisa los electrodomésticos con múltiples estados (lavadoras, lavavajillas, etc.)

## 3.5. Sparse coding

En el año 2010, Zico Colter desarrolló una nueva técnica de *machine learning* denominada *sparse coding* para desagregar el consumo de energía de una casa en el

plazo de una hora.

El algoritmo *sparse coding* se entrena sobre un modelo para cada curva de consumo de cada aparato tomando como muestra los datos recogidos durante una semana. Estos modelos entrenados se combinan para predecir el consumo de los aparatos de otras casas que no pertenecían al conjunto de entrenamiento. Esta técnica es capaz de conseguir una precisión del 55 % que, teniendo en cuenta el escaso tamaño muestral, se puede considerar un resultado satisfactorio.

## 3.6. Modelos ocultos de Markov

Los modelos ocultos de Markov han sido aplicados ampliamente en el estudio del NILM. Un modelo oculto de Markov (HMM por sus siglas en inglés) es un modelo estadístico en el que se asume que el sistema a modelar es un proceso de Markov de parámetros desconocidos. El objetivo es determinar los parámetros ocultos a partir de los parámetros observables.

La idea que subyace de la aplicación de estos modelos al NILM consiste en usar el estado oculto como el estado del aparato en cuestión y usar la potencia como dato observable. Una extensión de los modelos ocultos de Markov es los modelos ocultos factoriales (FHMM por sus siglas en inglés) que usan múltiples cadenas ocultas de Markov. En cada paso, la observación es una suma de observaciones de cada cadena de Markov. En NILM, se considera que la observación es la suma de las salidas de cada cadena de Markov por lo que cada cadena representa un aparato de la casa y la observación es el consumo agregado.

### 3.6.1. Otros modelos basados en HMM

Otros autores han trabajado sobre los modelos ocultos de Markov introduciendo ligeras modificaciones y obtuvieron como resultado 4 modelos que posteriormente

compararon. Los 4 modelos son los siguientes:

1. Un **HMM factorial** donde la variable de estado oculto simple de un HMM convencional se sustituye por variables de múltiples estados ocultos de manera que todos influyen sobre la variable observacional. En este modelo, cada estado oculto representa un aparato.
2. Un **FHMM condicional** que extiende el FHMM y le incorpora otras características como el tiempo, dependencia entre aparatos y otras medidas.
3. Un **Semi-modelo factorial oculto de Markov (FHSMM)** que incorpora funciones de probabilidad alternativas para la duración en un estado de un electrodoméstico.
4. Una mezcla de FHSMM y CFHMM que se denominó **Semi-modelo condicional factorial oculto de Markov (CFHSMM)**.

### 3.6.2. Resultados de los modelos

Como conclusión, el modelo que mejor resultado tuvo fue el CFHSMM. De nuevo, el algoritmo fue capaz de identificar aquellos aparatos con curvas de consumo sencillas pero no tuvo éxito con otros aparatos más complejos. Por otro lado, para que los modelos funcionasen correctamente, el usuario debería proporcionar el número de aparatos están activos en la señal de consumo agregado, cosa totalmente infactible.

# Capítulo 4

## Estudios previos

En este capítulo se expondrá el trabajo realizado por David e Irene, trabajadores de Optare Solutions, sobre la desagregación del consumo energético en un hogar inteligente.

### 4.1. Preprocesado

Los datos que se emplearon en las diferentes pruebas que se expondrán en las secciones siguientes fueron obtenidas del conjunto de datos expuesto en la Sección 2.4: REDD.

Contamos, por tanto, con un conjunto de datos que recoge la información de la potencia energética agregada y consumida por cada electrodoméstico de 6 casas. Una representación de la potencia agregada se puede observar en la Figura 4.1 mientras que en la Figura 4.2 podemos ver un ciclo de consumo de un frigorífico dentro de la potencia agregada.

Como se comentó, uno de los mayores inconvenientes del conjunto REDD es la existencia de datos faltantes o *gaps* que dificultan el tratamiento de los datos de manera directa. Con el objetivo de solventar este problema y suavizar las fluctua-

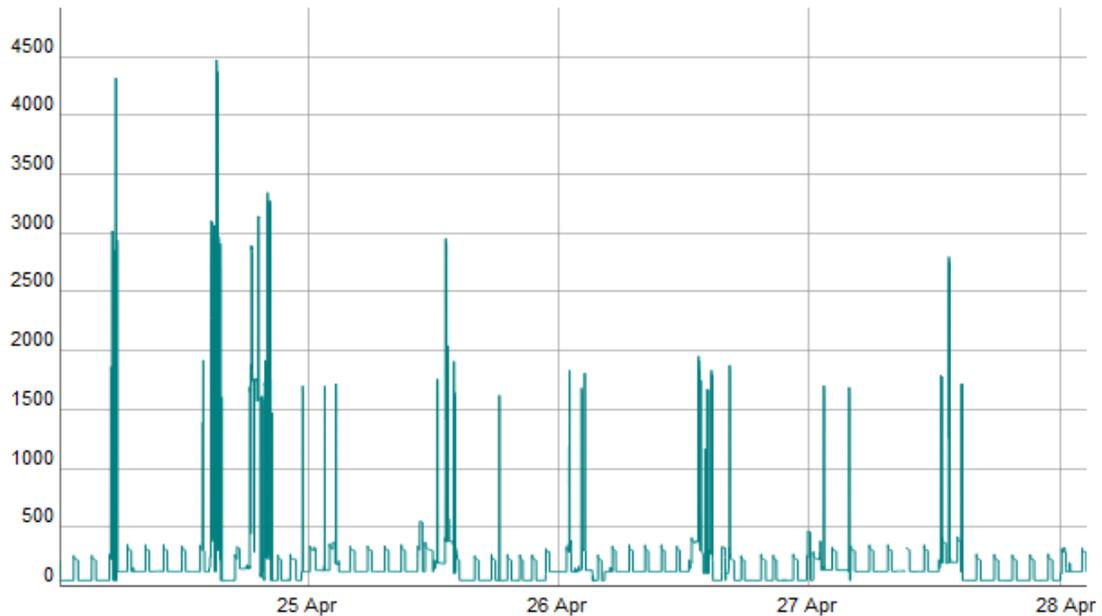


Figura 4.1: Consumo de potencia agregada.

ciones que se puedan cometer en la medición de instantes cercanos en el tiempo, se ha optado por hacer un remuestreo de los datos (tanto del consumo agregado como de cada electrodoméstico) cada 30 segundos tomando como valor la mediana de la potencia.

#### 4.1.1. Detección de ciclos

La detección de los ciclos se realiza sobre los datos de consumo de cada electrodoméstico. Para identificar un ciclo como tal, la potencia consumida ha de pasar de un estado latente a un estado de activación y dicho ciclo finalizará cuando se produzca el proceso inverso. Estos cambios requieren la definición de un umbral para cada electrodoméstico que se hará en base a su consumo ya que no todos los aparatos consumen la misma potencia. Así, un ciclo estará determinado por aquellas mediciones que superen el umbral establecido y las dos mediciones latentes en las que se encuentra.

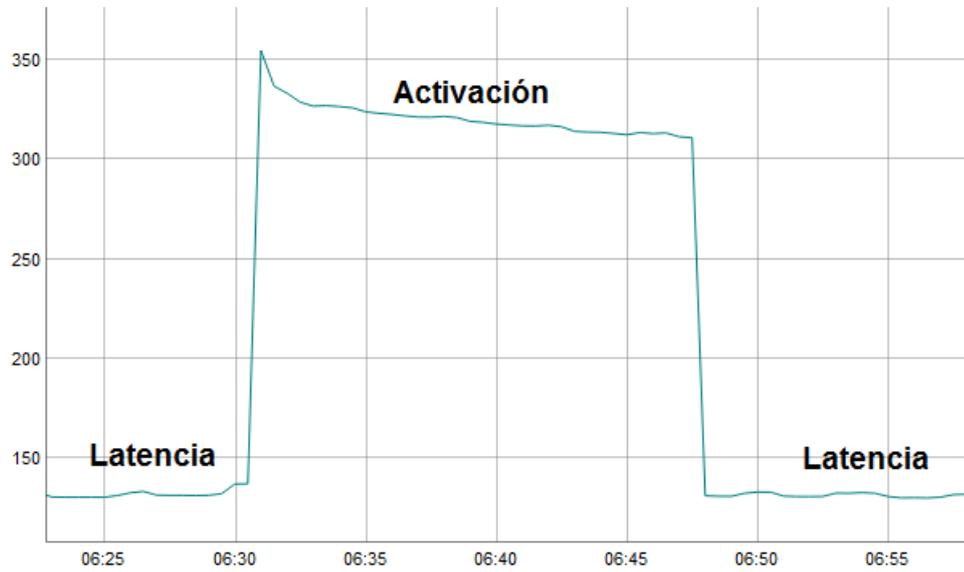


Figura 4.2: Ciclo de un frigorífico en el consumo agregado.

Finalmente, se crea un *data.frame* con las variables:

- Duración del ciclo.
- Potencia máxima.
- Potencia media.
- Potencia mediana.
- Tamaño de muestra del ciclo.

Este proceso se repite para cada casa y cada electrodoméstico.

## 4.2. Identificación de ciclos mediante clúster

Una vez obtenida la muestra, se comienza a trabajar la desagregación del consumo energético. Así, el siguiente paso fue la agrupación mediante clúster de los

distintos ciclos de un electrodoméstico en concreto. En nuestro caso, trabajaremos con los ciclos del frigorífico obteniendo la representación de la Figura 4.3.

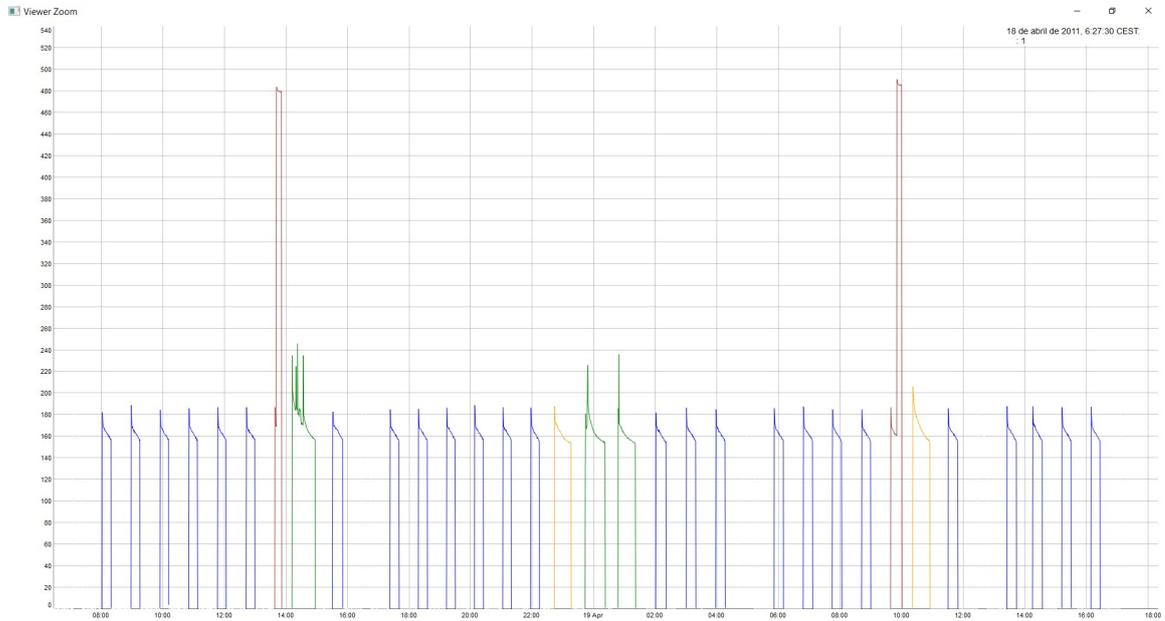


Figura 4.3: Clústers de los ciclos del frigorífico.

A continuación se pueden observar los clúster obtenidos por  $k - means$  y sus centroides:

K-means clustering with 4 clusters of sizes 45, 22, 9, 4

Cluster means:

	Duracion_min	Máximo	Pmax	Media	Mediana	nci	points	Bajadafinal
1	19.83333	186.4111	2.666667	155.4322	162.6944	1	40.64444	155.7111
2	31.02273	189.5455	3.545455	158.7766	162.5455	1	63.04545	155.0455
3	37.83333	230.2778	13.444444	163.4004	164.2778	1	76.66667	155.1667
4	21.37500	483.5000	23.500000	311.4517	248.2500	1	43.75000	417.7500

donde el clúster 1 se corresponde con el color azul, el 2 con el naranja, el 3 con el verde y el 4 con el marrón.

Estos clústers obtenidos a partir de la muestra del frigorífico fueron comparados con el consumo de potencia agregado en busca de patrones comunes como se puede observar en la Figura 4.4..

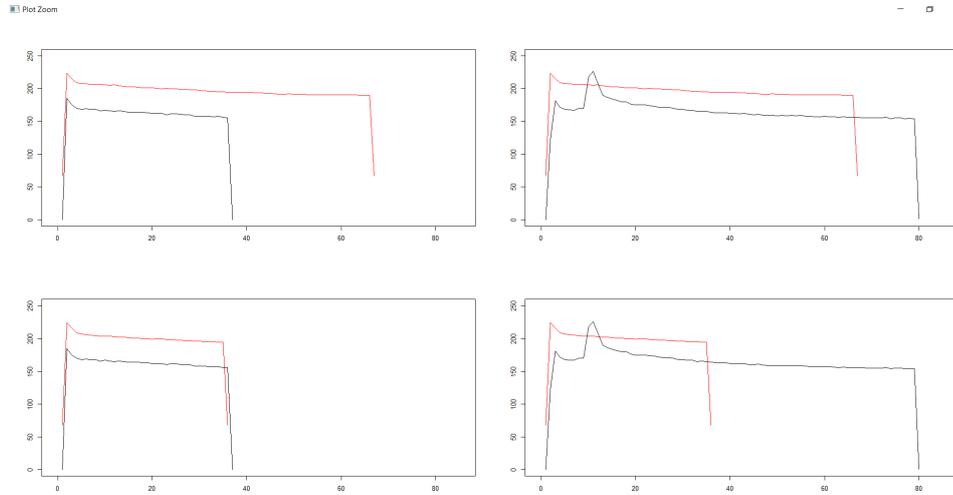


Figura 4.4: Ciclo de un frigorífico (negro) vs ciclo de potencia en el consumo agregado (rojo).

#### 4.2.1. Resultados obtenidos al aplicar técnicas de clusterización

Hemos visto una forma de identificar ciclos de un determinado electrodoméstico en el consumo agregado de una forma visual pero, para poder cuantificar la precisión en la identificación se requiere un valor numérico.

En primera instancia se cuantificó la correlación de los datos representados en la Figura 4.4 (dos ciclos del consumo agregado y dos del consumo de un frigorífico). Dado que, para poder obtener la correlación entre dos muestras, el tamaño muestral ha de ser el mismo, se optó por emplear los indicadores calculados anteriormente: duración del ciclo y potencia media y mediana. Los resultados no fueron los esperados ya que se obtuvo una mayor correlación entre los ciclos de la ventana superior izquierda en contraposición con lo observado.

Dado que no se obtuvieron buenos resultados, se optó por emplear el error cuadrático medio bajo la suposición de que dos ciclos que sean similares tendrán un

error cuadrático medio bajo. Los resultados corroboraron lo observado pero con un error muy alto por lo que se desecha el método.

### **4.3. Otras pruebas**

Se realizaron más pruebas con otros electrodomésticos siguiendo el método explicado en la sección anterior sin obtener resultados satisfactorios. Además, se implementaron otros métodos como las Cadenas Ocultas de Markov o mediante series temporales. Estas pruebas no están documentadas por lo que no se puede profundizar más en el tema. No obstante, los resultados obtenidos no fueron buenos por lo que, por necesidades de la empresa, se pospuso el estudio.

# Capítulo 5

## Un enfoque diferente

Dado que los resultados obtenidos con los métodos aplicados hasta el momento no fueron satisfactorios, mi propuesta, tras una amplia revisión de la literatura disponible sobre NILM, fue abordar el problema a través de un método hasta ahora poco aplicado en el estudio de la desagregación del consumo energético pero con muy buenos resultados [14] [7]: las redes neuronales profundas.

La capacidad de cómputo disponible a día de hoy hace de tareas relativamente complejas como el procesamiento lógico o el cálculo un asunto de dificultad irrisoria pero, sin embargo, asuntos ‘del mundo real’ como la identificación de un rostro en una imagen o la creatividad son aún un reto para la computación.

Para entender estas diferencias hay que poner el foco sobre la manera en la que se procesa la información en unos y otros. Mientras los ordenadores utilizan uno o varios microprocesadores increíblemente rápidos y eficaces, el cerebro humano está formado por millones de procesadores elementales, llamados neuronas, que están ampliamente conectados entre sí formando las redes neuronales.

A lo largo de la segunda mitad del pasado siglo, se hicieron grandes esfuerzos en acercar la computación a los modelos que rigen el funcionamiento del cerebro humano implementando modelos que conseguían acercarse cada vez más al procesamiento

biológico. Sin embargo, la falta de capacidad computacional de aquella época hizo relegar las redes neuronales a un segundo plano hasta que en los últimos años, con el avance de la tecnología, tuvieron un nuevo resurgir convirtiéndose en una de las herramientas matemáticas más interesantes y con mayor proyección.

En el presente trabajo abordaremos, en un primer lugar, la definición y evolución de las redes neuronales desde los modelos más sencillos, con la modelización matemática de una neurona, hasta modelos mucho más complejos como las redes neuronales convolucionales que aplicaremos a los datos del conjunto REDD.

Para el desarrollo e implementación de los diferentes algoritmos se emplearán principalmente dos lenguajes de programación: R para el preprocesado y visualización de datos y Matlab para el entrenamiento y validación de los modelos de redes neuronales.

## 5.1. Nuevo preprocesado

Como ya se ha comentado anteriormente, el conjunto REDD tiene un gran inconveniente: los datos faltantes (gaps). Esto hace que se necesite un preprocesado de datos para homogeneizar la muestra que alimentará a los diferentes modelos que se desarrollarán más adelante.

Dado el escaso tiempo disponible de estancia en la empresa, se utiliza como base el trabajo realizado por David e Irene y se lleva a cabo un remuestreo siguiendo las indicaciones de Nascimento [7]. Nascimento, propone rellenar los datos faltantes tomando como referencia el primer dato disponible tras la secuencia de faltantes en aquellos gap cuya duración no supere los 20 segundos. En caso de superar este umbral, el intervalo se mantiene sin rellenar.

El siguiente paso en la propuesta de Nascimento consiste en hacer un remuestreo a intervalos de 4 segundos tomando como valor del intervalo la media de los

puntos comprendidos en él.

Para finalizar, y es un paso que no se ha llevado a cabo en nuestro estudio, Nascimento propone reescalar los datos dividiéndolos por 500. Este paso se propone ya que según el autor, favorece la inicialización de la red.

Cabe destacar que este preprocesado de los datos de REDD, a pesar de parecer sencillo, es relativamente complejo y conlleva un tiempo de cómputo de unos 25 minutos por electrodoméstico y casa.

Además, el preprocesado no termina aquí sino que en función de las diferentes probaturas que se desarrollarán en lo que sigue, se requerirá un nuevo tratamiento de los datos para extraer los ciclos de cada electrodoméstico y sus respectivas representaciones gráficas. Notar que los scripts han de ser además adaptados para cada electrodoméstico dadas las características de las curvas de consumo de cada uno de ellos.

Un ejemplo de esto es el ancho de ventana de la entrada que proporcionaremos a las redes. El ancho de ventana viene determinado por el número de muestras, teniendo en cuenta que se toma una muestra cada 4 segundos. Dado un electrodoméstico, un ancho de ventana demasiado amplio puede perjudicar la desagregación pero ha de incluir el ciclo completo, por lo que no puede ser demasiado pequeña. Por tanto, el ancho de ventana ha de ser mayor que la duración máxima de los ciclos de un electrodoméstico. En la Tabla 5.1 podemos observar el número de muestras tomadas para el ancho de ventana de los lavavajillas, lavadora y frigorífico.

Electrodoméstico	Número de muestras
Lavavajillas	588
Lavadora	641
Frigorífico	2401

Tabla 5.1: Número de muestras seleccionadas para el ancho de ventana.

La problemática de la selección del ancho de ventana se puede observar mejor en la Figura 5.1.

### 5.1.1. Muestra sintética

Para el buen funcionamiento de las redes neuronales es necesario un tamaño muestral muy grande. En NILM, obtener este tamaño muestral, a veces, es relativamente complejo. Para solventar este problema, se propone la creación de datos sintéticos generados a raíz de los ya disponibles.

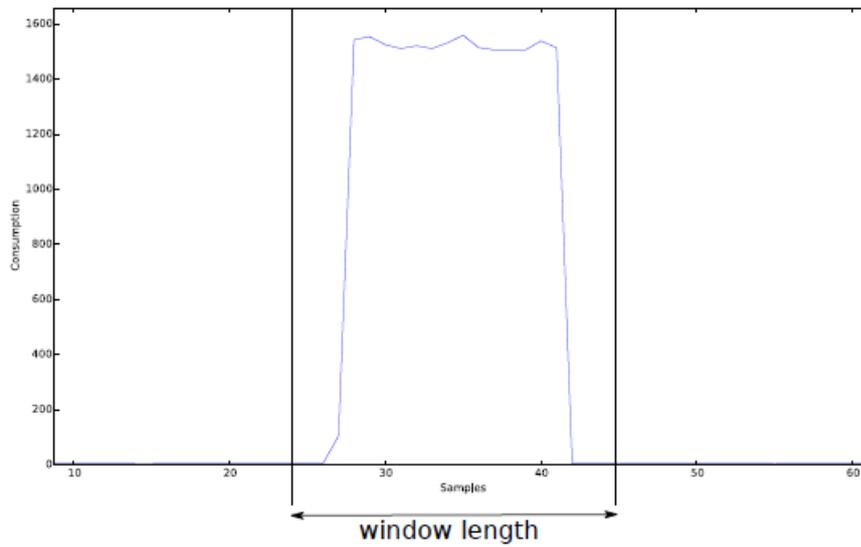
De esta manera, podemos obtener nuevos datos simplemente sumando los consumos de un microondas, un lavavajillas y un frigorífico obteniendo como resultado una casa con 3 electrodomésticos. Dados los datos disponibles en el conjunto REDD, el número de alternativas a la hora de crear nuevos datos es considerablemente alto lo que mejorará la capacidad de generalización de los modelos.

Cabe destacar que con esta generación de datos sintéticos se pierden algunos patrones de consumo como por ejemplo, encender una consola y la televisión al mismo tiempo.

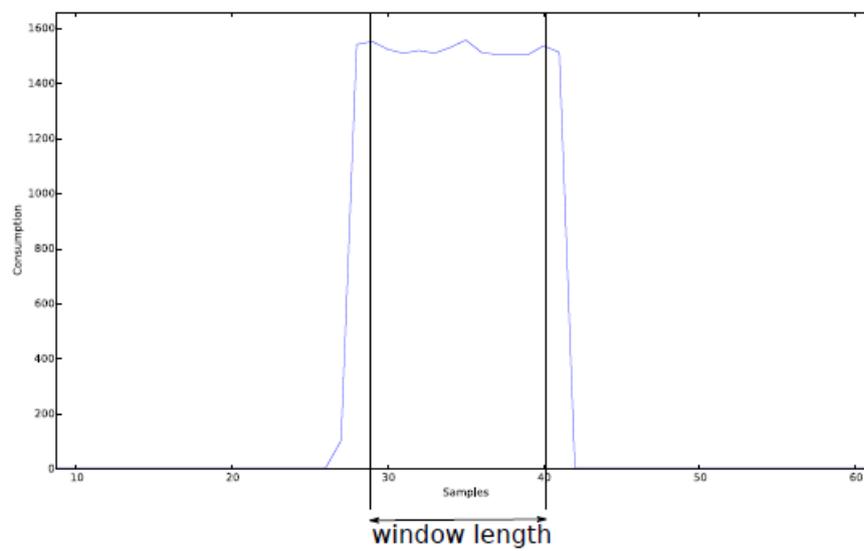
## 5.2. Redes neuronales. Introducción

Las redes neuronales (también conocidas como sistemas conexionistas) son un modelo computacional basado en un gran conjunto de unidades neuronales simples (neuronas artificiales), de forma aproximadamente análoga, al comportamiento observado en los axones de las neuronas en los cerebros biológicos.

Cada unidad neuronal está conectada con muchas otras y los enlaces entre ellas pueden incrementar o inhibir el estado de activación de las neuronas adyacentes. Cada unidad neuronal, de forma individual, opera empleando funciones de suma. Puede existir una función limitadora o umbral en cada conexión y en la propia



(a) Con un ancho de ventana demasiado amplio se perjudica el entrenamiento ya que aprende estados insignificantes



(b) Con un ancho de ventana demasiado corto, la red no es capaz de apreciar los cambios de estado.

Figura 5.1: La importancia del ancho de ventana.

unidad, de tal modo que la señal debe sobrepasar un límite antes de propagarse a otra neurona. Estos sistemas aprenden y se forman a sí mismos, en lugar de ser programados de forma explícita, y sobresalen en áreas donde la detección de soluciones o características es difícil de expresar con la programación convencional.

Según la complejidad del modelo, podemos distinguir varios tipos: neurona, capa, red y sistema neuronal como se puede observar en la Figura 5.1.

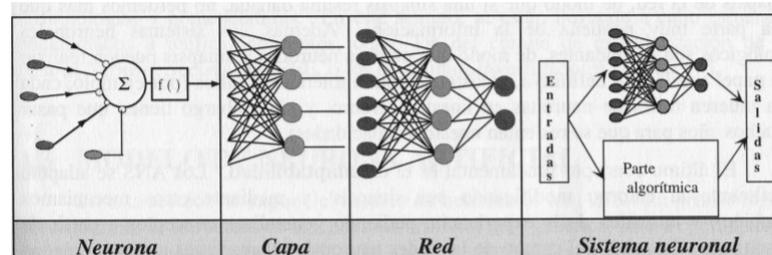


Figura 5.2: Tipos de modelos neuronales en función de su complejidad. Figura obtenida de Larrañaga et al [17]

Como se comentó anteriormente, cada neurona está conectada con muchas otras y opera empleando funciones de suma. Consta de 3 partes:

- Un conjunto de entradas  $x_j$  con sus respectivos pesos  $w_{i,j}$  que representan la intensidad de interacción entre una neurona y la siguiente.
- Una regla de propagación  $h_i = \sigma(w_{i,j}, x_j(t))$  definida a partir de  $x_j$  y  $w_{i,j}$ .
- Una función de activación  $y_i$  que representa la salida de la neurona y su estado de activación.

### 5.2.1. La función de activación

Al igual que ocurre con las neuronas biológicas, una neurona artificial puede estar activa (excitada) o inactiva (no excitada). La función de activación calcula

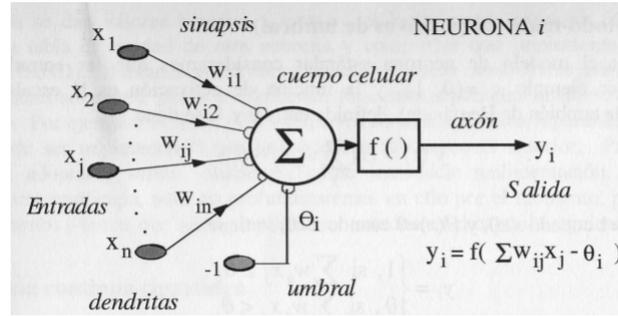


Figura 5.3: Estructura de la neurona  $i$ -ésima. Figura obtenida de Larrañaga et al [17]

el estado de actividad de una neurona transformando la función de entrada global ( $gin_i$ ) menos un umbral ( $\Theta_i$ ) en un valor (estado) de activación cuyo rango, normalmente, está comprendido entre 0 y 1 o -1 y 1.

Las funciones de activación más utilizadas son las siguientes:

### 1. Función lineal

$$f(x) = \begin{cases} -1 & \text{si } x \leq -1/a \\ a * x & \text{si } -1/a < x < 1/a \\ 1 & \text{si } x \geq 1/a \end{cases}$$

con  $x = gin_i - \Theta_i$  y el parámetro  $a > 0$ .

Los valores obtenidos por esta función de activación serán:

- $a(gin_i - \Theta_i)$  cuando el argumento esté comprendido entre  $(-1/a, 1/a)$
- 1 o -1 fuera del intervalo  $(-1/a, 1/a)$

Cuando  $a=1$  la salida es igual a la entrada.

### 2. Función sigmoidea

$$f(x) = \frac{1}{1 + e^{-gx}}$$

con  $x = gin_i - \Theta_i$ .

Los valores de salida que proporciona esta función están comprendidos dentro de un rango que va de 0 a 1. Al modificar el valor del parámetro  $g$  se ve afectada la pendiente de la función de activación.

### 3. Función tangente hiperbólica

$$f(x) = \frac{e^{gx} - e^{-gx}}{e^{gx} + e^{-gx}}$$

con  $x = gin_i - \Theta_i$ .

Los valores de salida de la función tangente hiperbólica están comprendidos dentro de un rango que va de -1 a 1. De nuevo, al modificar el valor del parámetro  $g$  se ve afectada la pendiente de la función de activación.

## 5.3. Aprendizaje

Puede definirse el aprendizaje como el proceso por el que se produce el ajuste de los parámetros libres de la red a partir de un proceso de estimulación por el entorno que rodea a la misma. En la mayoría de los casos, el aprendizaje consiste simplemente en determinar un conjunto de pesos sinápticos que permita a la red realizar correctamente el tipo de procesamiento deseado.

Al construir un sistema neuronal, se parte de un cierto modelo de neurona y de una determinada arquitectura de red estableciéndose los pesos sinápticos iniciales como nulos o aleatorios. Para que la red resulte operativa es necesario entrenarla. El entrenamiento o aprendizaje se puede llevar a cabo a dos niveles:

1. **Modificar los pesos sinápticos** siguiendo una cierta regla de aprendizaje que, normalmente, consiste en optimizar la función de error mediante un proceso iterativo. Un elemento importante de estas iteraciones es el ritmo de aprendizaje que regula la magnitud de la modificación en cada iteración.

2. **Modificando la estructura de la red** añadiendo o eliminando neuronas.  
Este tipo de ajuste es menos habitual.

## 5.4. El perceptrón multicapa (MLP)

La estructura más simple de las redes neuronales es el perceptrón multicapa o MLP, por sus siglas en inglés, que está formado por una capa de entrada que recibe la información, otra capa de salida que emite una decisión o predicción y, entre ellas, una o más capas ocultas que se ocupan de procesar los datos recibidos.

Normalmente, todas las neuronas de una capa reciben señales de una capa anterior y envían señales a una capa posterior, es decir, el flujo de información va en un único sentido. Este tipo de redes neuronales son conocidas como *feedforward*.

Sin embargo, también existe la posibilidad de conectar la salida de las neuronas a capas anteriores creando las redes neuronales conocidas como *feedback*.

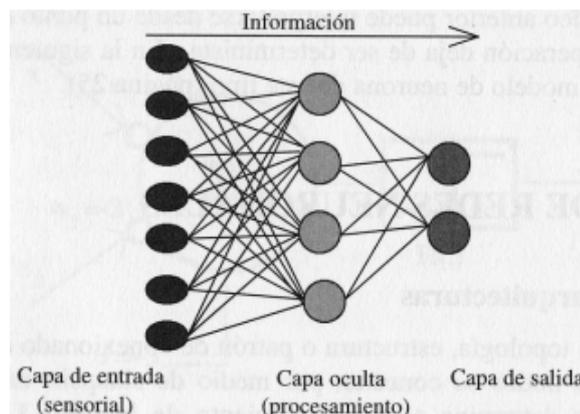


Figura 5.4: Estructura de un perceptrón multicapa feedforward. Figura obtenida de Larrañaga et al [17]

### 5.4.1. Métodos de optimización

La optimización del modelo y, por tanto, su aprendizaje, consiste en la corrección de los pesos sinápticos que rigen las interacciones entre las neuronas en función del error que se ha obtenido a la salida de la red.

En esencia, se trata de un proceso de optimización y, como tal, se podrán utilizar los algoritmos existentes en este campo para llevarlo a cabo. Entre todos los métodos disponibles, destaca por encima de todos el gradiente descendente popularmente impuesto por su facilidad y extensión de uso.

El gradiente del descenso se basa en construir el siguiente punto iterativo a partir de la dirección del gradiente del error ponderándolo por el ritmo de aprendizaje. El ritmo de aprendizaje está relacionado con la magnitud de la modificación que se implementa en cada iteración y que permite optimizar la eficiencia del entrenamiento. Matemáticamente, se corresponde con el factor de ponderación por el que se multiplica la corrección obtenida para cada uno de los parámetros de la red. Este ritmo de aprendizaje, no debe ser ni demasiado pequeño, ya que supondría tiempos de entrenamiento muy largos y el proceso se podría quedar atrapado en mínimos locales, ni demasiado grande ya que los valores oscilarán en torno al mínimo absoluto sin llegar a alcanzarlo.

Denotaremos como  $\alpha_i$  el vector de parámetros que se desea optimizar, donde  $i$  es el subíndice de la iteración  $i$ ésima y  $C_i$  será el error de la iteración. El ritmo de aprendizaje se denotará como  $r_i$ . Así, la ecuación del gradiente del descenso queda como sigue:

$$\alpha_{i+1} = \alpha_i - r_i \nabla C_i$$

A pesar de que no se profundizará en ello ni serán implementados en el presente trabajo, existen otros métodos de optimización como el de Newton. Este método hace uso de la matriz Hessiana (segundo orden) frente al gradiente del método anterior

(primer orden). Así, si trabajamos con  $n$  parámetros, la Hessiana constaría de  $n^2$  parámetros por lo que el entrenamiento de redes neuronales se volvería muy poco práctico.

## 5.5. Redes neuronales profundas

Las redes neuronales profundas han arrojado increíbles resultados en tareas de clasificación de imágenes o reconocimiento de voz. La mayor ventaja de este tipo de redes neuronales es que son capaces de aprender automáticamente una jerarquía de extractores de características a partir de los datos en bruto. Cada capa de la red aporta una serie de características a la siguiente, partiendo de detalles muy simples como los cambios de color en una imagen o el relieve de una figura hasta detalles tan minuciosos como el iris del ojo de una persona.

Este tipo de redes están presentes en muchos ámbitos de nuestras vidas:

- En 2015, una red neuronal profunda fue capaz de batir por 5 a 0 al campeón de Europa de Go <sup>1</sup> y, en marzo de 2016, venció al varias veces campeón mundial de Go, Lee Se-dol, por 4-1 (el lector interesado puede visionar el documental AlphaGo disponible en Netflix).
- Otra de las aplicaciones de las redes neuronales profundas es el procesamiento del lenguaje natural e incluso la identificación de una persona por su voz, como podemos comprobar en los asistentes personales como Siri o Cortana.
- El reconocimiento y clasificación de objetos es otro de los campos de aplicación de las redes neuronales profundas como el caso de AlexNet, una red neuronal convolucional que es capaz de clasificar más de 1000 objetos.

---

<sup>1</sup>El Go es un juego tradicional chino muy complejo en el que el número de movimientos posibles en una partida es de  $10^{170}$ , más que el número de átomos existentes en todo el universo.

De forma paralela al desarrollo del presente trabajo y, junto con otro compañero del Máster, hemos establecido colaboraciones con la Universidad de Oviedo y profesores del Grado de Telecomunicaciones de la Universidad de Vigo con el objetivo de desarrollar un sistema de identificación por imagen para perros perdidos de manera que, con una simple fotografía, el sistema sea capaz de identificar el perro y comunicar a sus dueños que ha sido encontrado.

Como acabamos de comprobar, las redes neuronales profundas parecen no tener límite en problemas de clasificación o reconocimiento de patrones. Tanto es así, que las compañías de telefonía móvil como Apple o Samsung están incorporando en sus dispositivos servicios de identificación de rostro para desbloquear el móvil o realizar pagos.

En esta sección, describiremos las estructuras de las redes neuronales adaptadas al estudio de la desagregación del consumo energético y las pruebas realizadas. Posteriormente, describiremos los trabajos de Jack Kelly [14] y Pedro Paulo [7] sobre los que me he basado para realizar las pruebas con datos numéricos en la última etapa de mi estancia en la empresa.

Como último apunte previo al desarrollo de los diferentes modelos, cabe destacar que se ha implementado una red por electrodoméstico de manera que cada una ha sido entrenada sobre los datos de un determinado conjunto de casas y, posteriormente, se ha llevado a cabo la fase de test sobre otro conjunto de casas no incluidas en el conjunto de entrenamiento. Las redes, toman como entrada los datos del consumo de la potencia agregada y su objetivo es el de obtener el consumo de un determinado electrodoméstico.

### **5.5.1. Las redes convolucionales**

Las convolucionales se presentan como un tipo de redes neuronales capaces de extraer información de una imagen a través de la aplicación de filtros basadas en el

sistema visual de los mamíferos siendo capaces de reconocer complejos patrones con gran precisión.

La función de estos filtros es extraer información acerca de la relación que tienen los píxeles de una imagen entre sí de manera local. Mientras que, con los modelos anteriores, todas las neuronas de cada capa se conectan entre sí, en las convolucionales un filtro extrae la interacción local de un grupo de neuronas y, mediante esto, se alimenta a una neurona de la siguiente capa.

Una red neuronal convolucional puede estar formada por infinidad de capas de distintos tipos. Actualmente, las capas más utilizadas son:

- **Capa convolucional.** Esta capa está compuesta por núcleos convolucionales<sup>2</sup> que se reparten sobre la imagen completa con el objetivo de extraer características de ella. A medida que se aumenta el número de capas se extraerán características más complejas.
- **Capa ReLU** Esta capa aplica una función de activación  $f(x) = \max(0, x)$  y aporta algunas ventajas sobre las funciones de activación tradicionales que vimos anteriormente.
- **Capa pooling** La capa de pooling reduce la muestra calculando el máximo o la media de una pequeña parte de los datos de entrada. Con esto, la red se vuelve más invariante a traslaciones y ganará en velocidad.
- **Capa dropout** Dado el elevado número de parámetros de una red neuronal profunda, son propensas a sobre ajustarse. Con esta capa, se reduce el overfitting. La idea que subyace de este método es eliminar aleatoriamente algunas

---

<sup>2</sup>Una convolución es un operador matemático que transforma dos funciones  $f$  y  $g$  en una tercera función que, en cierto sentido, representa la magnitud en la que se superponen  $f$  y una versión trasladada e invertida de  $g$ .

neuronas y sus conexiones durante el entrenamiento de manera que se reduce la dependencia de algunos pesos específicos en la propia capa.

Cabe destacar que esta clasificación de capas es lo más general y superficial posible y que toda capa puede ser modificada y adaptada a las necesidades del problema.

## 5.6. Clasificación por imagen

Una gran parte de las aplicaciones de las redes neuronales y, donde se encuentran los algoritmos más eficientes y eficaces, tiene que ver con la clasificación de imágenes. Como ya se comentó en la sección anterior, una de las redes neuronales más utilizadas y con más bagaje en la comunidad científica es la red AlexNet.

Es por esto que mi propuesta a la hora de aproximarnos al problema de la desagregación del consumo energético a través de las redes neuronales sea mediante la clasificación por imagen. Para ello, se programaron una serie de scripts en R que extraen en formato *.jpg* las representaciones gráficas de las curvas de consumo de cada electrodoméstico que serán pasadas a las redes neuronales como datos de entrada.

En esta sección describiremos en primer lugar las diferentes redes neuronales utilizadas y, posteriormente, detallaremos las variantes introducidas en los datos de entrada que, para agilizar las pruebas, se decidió trabajar únicamente con el 80 % de los datos de lavavajillas, lavadora y frigorífico de las casas 1,2 y 3. Para el conjunto de test, se utilizó el 20 % restante.

### 5.6.1. LeNet 5

Una de las primeras redes convolucionales fue la creada por Lecun con el objetivo de reconocer dígitos manuscritos obteniendo una tasa de error del 0.8 %. El

conjunto de datos utilizado para la clasificación de números manuscritos es conocido como MNIST y cuenta con 70000 imágenes de dígitos manuscritos comprendidos entre 0 y 9.

La red convolucional que hemos utilizado para realizar las primeras pruebas con las imágenes de los electrodomésticos tiene la siguiente estructura (se utilizará la nomenclatura de Matlab para la definición de las capas):

- **imageInputLayer([480 480 1])** Capa de entrada con 3 posibles parámetros  $[h, w, c]$  donde  $h$  es la altura de la imagen,  $w$  es el ancho y  $c$  es el parámetro que identifica el número de colores en la imagen siendo 1 para imágenes en blanco y negro o 3 para imágenes en color.
- **convolution2dLayer(5,20)** Capa convolucional con dos parámetros que representan el tamaño y el número de filtros respectivamente.
- **reluLayer**
- **maxPooling2dLayer(2, 'Stride', 2)** Capa pooling de tamaño  $[2, 2]$  con paso  $[2, 2]$
- **fullyConnectedLayer(10)** En esta capa se multiplica la entrada por una matriz de pesos y se suma el umbral  $\Theta$ . El parámetro representa el número de salidas.
- **softmaxLayer** Aplica la función SoftMax.
- **classificationLayer** Capa de clasificación.

Además de modificar los datos de entrada, se modificaron los parámetros de las capas en base a lo recogido en la literatura sobre NILM. Estos cambios no se detallarán ya que no aportaron resultados de relevancia significativa.

### 5.6.2. Pattern Recognition App

Una de las mayores ventajas de usar Matlab a la hora de introducirse en nuevas técnicas es la gran documentación y soporte de la comunidad. Esta ventaja se ve reflejada en la implementación de lo que en Matlab se conoce como *toolbox*, que consiste en herramientas paquetizadas que incluyen diversos modelos y destacan por su fácil implementación.

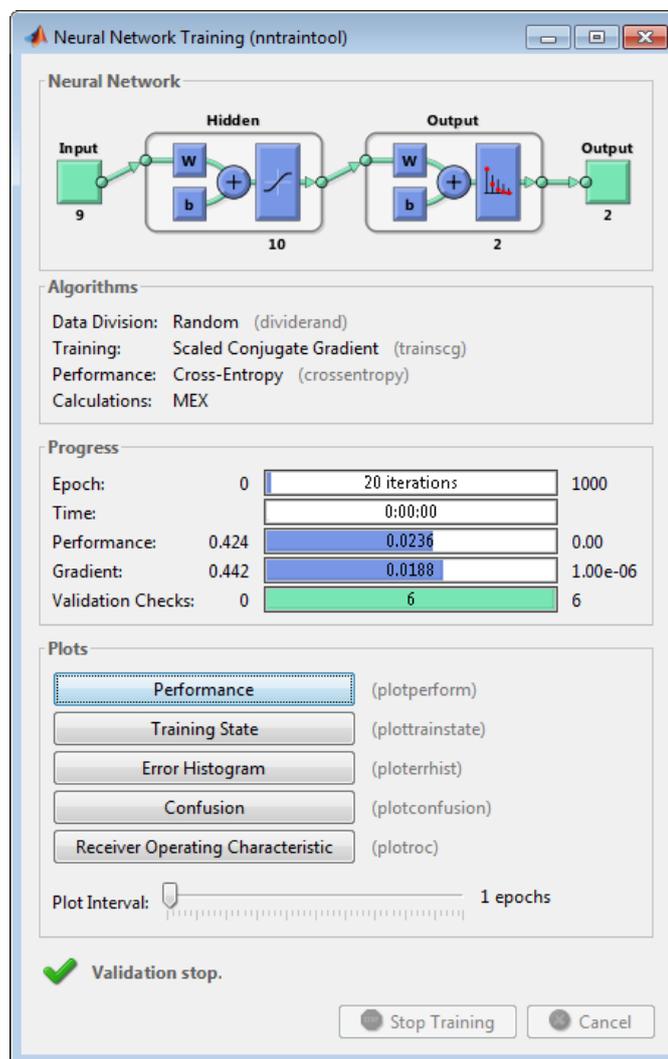


Figura 5.5: Interfaz de Pattern Recognition App

Una de estas *toolbox* con aplicación en el ámbito de las redes neuronales es la denominada *Neural Network Pattern Recognition App* que incluye diversos modelos basados en redes neuronales como reconocimiento de patrones, clustering o series de tiempo con modelos dinámicos.

En nuestro caso, hemos trabajado con el modelo de reconocimiento de patrones que permite ajustar, además de los datos de entrada, el número de capas ocultas y permite obtener una serie de informes con los resultados del modelo entrenado.

A pesar de que las posibilidades de adaptación de parámetros a las necesidades del problema son limitadas, esta *toolbox* es útil para adentrarse en el mundo de las redes neuronales y conocer un poco más su funcionamiento.

## 5.7. Modelos de redes neuronales implementados

A continuación se detallan algunos de los modelos de redes neuronales recogidos en la tesis de Nascimento [7] y que han sido aplicados en el presente estudio.

### 5.7.1. Redes neuronales convolucionales

La arquitectura de la red neuronal convolucional (CNN) está, básicamente, compuesta por capas convolucionales con un número creciente de capas. Se ha aplicado una capa de pooling con el objetivo de hacer el modelo invariante a traslaciones y reducir el número de parámetros que han de ser procesados.

Tipo de capa	Tamaño
Convolutional	8 filtros de tamaño 5
Convolutional	8 filtros de tamaño 3
Convolutional	8 filtros de tamaño 3
Max Pooling	Tamaño 4 paso 2
Convolutional	16 filtros de tamaño 3
Convolutional	16 filtros de tamaño 3
Convolutional	16 filtros de tamaño 3
Max Pooling	Tamaño 4 paso 2
Convolutional	32 filtros de tamaño 3
Convolutional	32 filtros de tamaño 3
Convolutional	32 filtros de tamaño 3
Max Pooling	Tamaño 4 paso 2
Convolutional	64 filtros de tamaño 3
Convolutional	64 filtros de tamaño 3
Convolutional	64 filtros de tamaño 3
Max Pooling	Tamaño 4 paso 2
Fully Connected	Número de clústers

Tabla 5.2: Arquitectura de CNN

### 5.7.2. Red neuronal convolutional recurrente

Las redes neuronales convolucionales recurrentes (RCNN) están compuestas por capas convolucionales recurrentes<sup>3</sup> (RCL) con un número constante de filtros.

---

<sup>3</sup>Se definen las capas convolucionales recurrentes con arquitecturas con bucles de realimentación.

Tipo de capa	Tamaño
Convolutacional	8 filtros de tamaño 5
RCL	8 32 filtros de tamaño 3. 3 iteraciones
Max Pooling	Tamaño 4 paso 2
RCL	8 32 filtros de tamaño 3. 3 iteraciones
Max Pooling	Tamaño 4 paso 2
RCL	8 32 filtros de tamaño 3. 3 iteraciones
Max Pooling	Tamaño 4 paso 2
Fully Connected	Número de clústers

Tabla 5.3: Arquitectura de RCNN

## 5.8. Los datos de entrada

Mi propuesta de abordar el problema de la desagregación energética por redes neuronales es principalmente debido a su capacidad de generalización y los buenos resultados en problemas de clasificación de imágenes. Así, los primeros conjuntos de pruebas contienen las representaciones gráficas de los ciclos de los electrónicos. Además, y es otra de las ventajas que hicieron decantarme por la utilización de imágenes, el tiempo de implementación y entrenamiento del modelo de clasificación de imágenes es mucho menor que si trabajásemos con datos numéricos.

A pesar de que la gráfica de consumo agregado (ver Figura 5.6) puede resultar bastante hostil a primera vista, una vez familiarizado con los patrones de cada electrodoméstico el ojo humano es capaz de identificar los ciclos de estos aparatos en la gráfica de consumo agregado.

La representación de estos datos fue implementada mediante el paquete de R *dygraph* que permite realizar un plot de los datos con la posibilidad de hacer

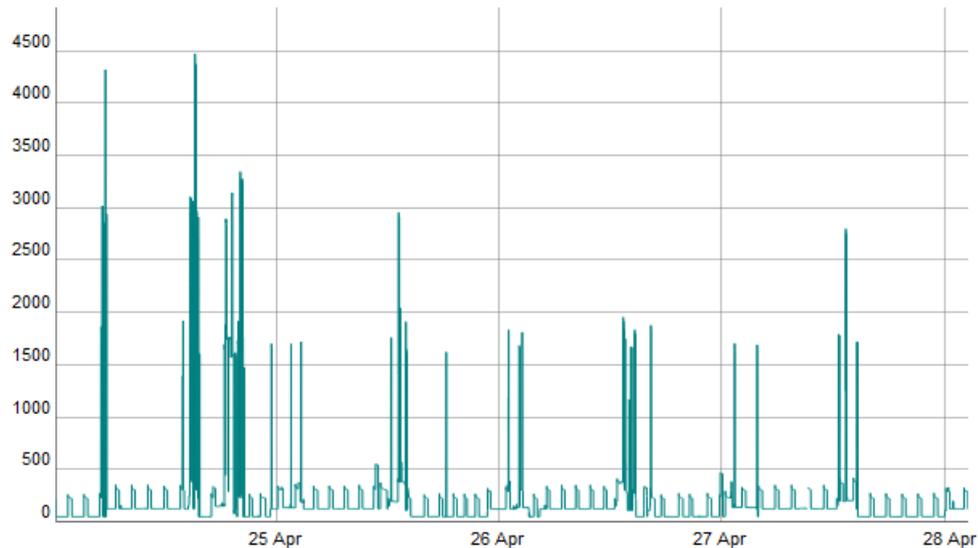


Figura 5.6: Gráfica de consumo agregado de una casa.

zoom sobre ellos. Así, si nos centramos en una parte de la gráfica anterior podemos observar algunos ciclos de electrodomésticos (ver Figura 5.7.)

Además, mediante el paquete *shiny* se ha creado la interfaz de visualización de datos que se puede observar en la Figura 5.8 donde se puede seleccionar la casa, el periodo de tiempo, el umbral de potencia observado e incluso se puede obtener la desagregación gracias a que el conjunto REDD contiene los datos del consumo agregado y desagregado.

Acabamos de comprobar que es posible afirmar que un ser humano que haya sido capaz de aprender los patrones de consumo de determinados electrodomésticos podría detectarlos en una señal de consumo agregada en un alto porcentaje. Por tanto, el reto será transformar esta idea en un modelo computacional de manera que la desagregación se pueda realizar de manera automática.

Con los modelos de redes neuronales expuestos anteriormente, se llevaron a cabo una gran multitud de pruebas. A continuación, se detallan algunas de las pruebas realizadas:

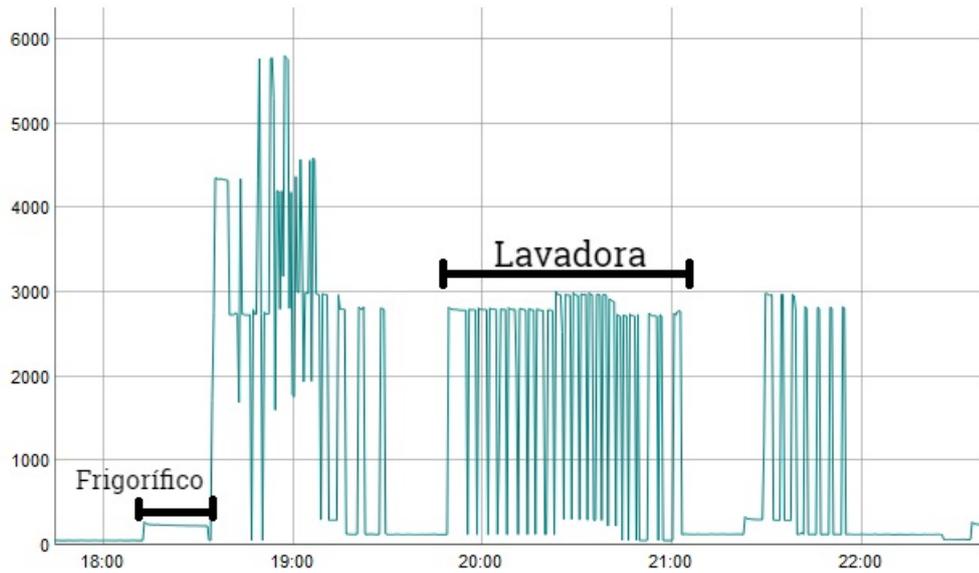


Figura 5.7: Gráfica de consumo agregado de una casa.

- Prueba 1.** En una primera aproximación al problema, se quiso comprobar si los modelos eran capaces de clasificar los electrodomésticos. Para ello, se usaron como datos de entrada las imágenes de los ciclos de los electrodomésticos y como target<sup>4</sup> las etiquetas de los electrodomésticos correspondientes a cada imagen.

Los datos de consumo agregado, que serán usadas para el test en una segunda fase, se representan en ventanas independientes cada 20 segundos con una altura fija correspondiente a la altura máxima de los ciclos del electrodoméstico que queremos identificar en el consumo agregado (en nuestro caso el frigorífico). Esto se puede observar mejor en la Figura 5.9. Uno de los problemas que surgen en el test con datos de consumo agregados es la falta de etiquetado lo que dificulta considerablemente la evaluación de los resultados. La eficacia del modelo se comprueba manualmente mediante la representación de las imágenes

<sup>4</sup>En redes neuronales, se denomina target al vector que contiene la salida esperada.



Figura 5.8: Interfaz de visualización de datos.

junto con las predicciones y mediante la herramienta de visualización de datos creada con *shiny*.

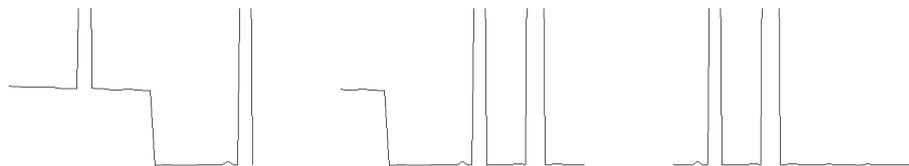
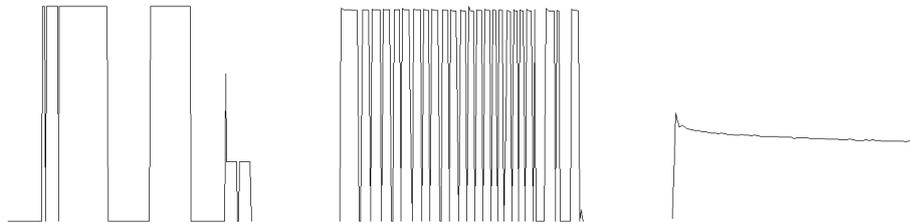


Figura 5.9: Ventanas de consumo agregado capturadas cada 20 segundos.

El resultado arrojó un  $\text{accuracy}^5$  del 97% en clasificación. Este resultado resul-

<sup>5</sup>La métrica utilizada para comprobar el nivel de desagregación es el  $\text{accuracy}$  que se define como la proporción de resultados correctos entre todos los posibles:  $\frac{TP+TN}{P+N}$  donde TP es el número de verdaderos positivos, TN verdaderos negativos, P número total de positivos, N número total de negativos

ta ligeramente engañoso ya que las clases estaban desbalanceadas. No obstante, es un excelente resultado para una primera prueba y el modelo ha sido capaz de clasificar correctamente lavavajillas, lavadora y frigorífico en un alto porcentaje. El test con el mismo modelo sobre los datos de consumo agregado no arrojó buenos resultados al clasificar casi todos los ciclos como frigorífico. Esto es debido a que las gráficas de consumo agregado tienen una altura fija y no se representan consumos por encima de ese máximo. Por tanto, los consumos de lavadora y lavavajillas solo aparecen representados hasta la altura fijada perdiendo así la posibilidad de identificar sus patrones característicos.



(a) Ciclo de un lavavajillas. (b) Ciclo de una lavadora. (c) Ciclo de un frigorífico.

Figura 5.10: Imágenes como *input* de la red.

- **Prueba 2.** Tras muchas modificaciones sin conseguir buenos resultados, se decide modificar el enfoque y, de ahora en adelante, se implementa una red por electrodoméstico para conseguir una clasificación binaria que identifique si un ciclo se corresponde con el electrodoméstico objetivo o no. Se fija el ancho y el alto de ventana para cada red como el máximo ancho y alto de todos los ciclos del electrodoméstico objetivo. Los resultados en clasificación siguen siendo buenos. Sin embargo, el test con el conjunto de datos agregados arroja un resultado bastante pobre.
- **Prueba 3.** Como se observa en la Figura 5.11, uno de los problemas que obser-

vamos es que la representación gráfica del consumo agregado incluye extremos laterales mientras que el desagregado no lo hace que las proporciones no sean las mismas. Esto implica que la red no es capaz de identificar un ciclo de un frigorífico en el consumo agregado ya que su representación es muy diferente a la del consumo desagregado. Se solventa este problema añadiendo extremos laterales a las imágenes del consumo desagregado y reescalando la imagen. Los resultados mejoran en clasificación pero no se identifica correctamente el electrodoméstico en el consumo agregado.

- **Prueba 4.** Se fija un ancho y alto de ventana correspondiente con los respectivos máximos de los datos del consumo desagregado del electrodoméstico objetivo. Para el consumo agregado, se fija la altura como el mínimo de la potencia de la ventana más el máximo de potencia del electrodoméstico objetivo. El mínimo de la curva de consumo agregado fluctúa en función de otros aparatos que estén actuando en segundo plano (iluminación, por ejemplo) por lo que, con esta medida, podemos obtener una ventana que capture el ciclo completo del electrodoméstico objetivo (si fijásemos la altura perderíamos información por ser demasiado alta o demasiado baja, en función de dónde se sitúe el mínimo del consumo agregado). Se obtiene un accuracy del 100% en clasificación. Los resultados del test con datos de consumo agregado mejoran pero no se consiguen grandes avances.

Tras llevar a cabo infinidad de modificaciones tanto en la representación de las imágenes: variaciones en las dimensiones, representación con área rellena (ver Figura 5.12.), etc. como en los parámetros de la red: número y tipo de capas, filtros, parámetros de inicialización, etc. se desiste en la tarea de la desagregación del consumo mediante imágenes.

Como conclusión, mediante imágenes hemos sido capaces de obtener resultados



Figura 5.11: Ciclo de un frigorífico con área rellena.

excelentes en la clasificación de los electrodomésticos pero no se ha cumplido con el objetivo de obtener una desagregación del consumo energético a partir de los datos de consumo agregados.

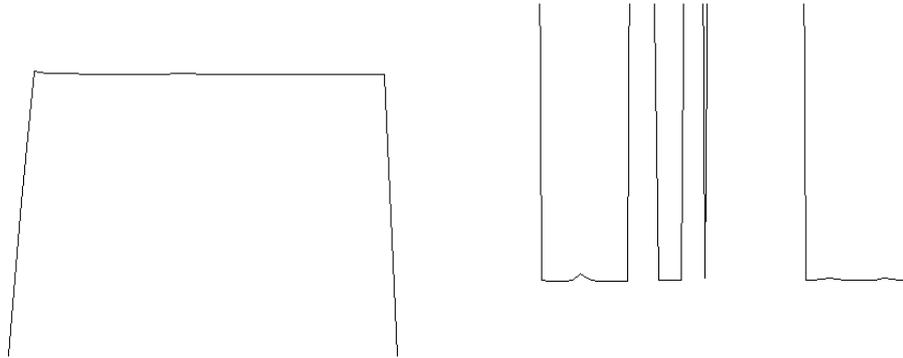
## 5.9. Datos numéricos

En la última sección de este trabajo, se detallará brevemente el trabajo realizado con datos numéricos en las últimas fechas de mi estancia en la empresa. A pesar de que no se han podido concluir las pruebas con datos numéricos por falta de tiempo, se comentarán brevemente los resultados obtenidos por Nascimento en la aplicación de redes neuronales a la desagregación del consumo energético.

Uno de los principales defectos del tratamiento de la desagregación con imágenes es la pérdida de información. Podemos abordar la representación de las curvas de consumo de dos formas principalmente:

- Adaptando el alto de la imagen al máximo de los datos de cada electrodoméstico. De esta manera, podemos obtener buenos resultados en clasificación pero no se podrá realizar la desagregación de una manera eficaz.

- La otra opción es fijar la altura como el máximo del consumo del electrodoméstico objetivo pero esto haría que no fuera posible capturar los patrones de potencia de otros electrodomésticos que tengan un consumo mayor.



(a) Frigorífico (electrodoméstico objetivo).

(b) Lavavajillas (imagen cortada).

Figura 5.12: Con una altura adaptada al frigorífico se pierde información sobre el consumo del lavavajillas.

Esta cuestión no supone un problema al trabajar con datos numéricos ya que se conocerá en todo momento el consumo de cada electrodoméstico.

### 5.9.1. Discretización de consumo desagregado

La discretización es un asunto que fue planteado desde las primeras pruebas realizadas en el presente estudio pero que, sin embargo, no se tuvo en cuenta hasta empezar a trabajar con datos numéricos. Esta discretización de la potencia consumida por un electrodoméstico es un asunto tratado también en la tesis de Nascimiento en lo que denominó *space quantization*.

La discretización de las potencias desagregadas se lleva a cabo con el objetivo de eliminar posibles *outliers* que afecten al entrenamiento de los diferentes modelos.

Además, las redes neuronales trabajan mejor con patrones regulares y las pequeñas variaciones de consumo son, en su mayoría, insignificantes.

La discretización se llevó a cabo de la siguiente manera (el algoritmo se aplica a cada electrodoméstico por separado):

1. Se realiza un clúster con el método de k-means y se fija el número de clústers siguiendo la propuesta de Nascimento (ver Tabla 5.4.)

Electrodoméstico	Número de clústers
Lavadora	5
Lavavajillas	5
Frigorífico	4

Tabla 5.4: Número de clúster para cada electrodoméstico

2. Una vez obtenidos los valores de los centroides, se asocia cada valor de la potencia con su centroide obteniendo la discretización de la potencia desagregada.

Para poder usar como datos de entrada las potencias de consumo de los electrodomésticos, éstos han de ser agrupados en matrices. Para ello, se fija el ancho de ventana (el máximo de la duración de los ciclos de cada electrodoméstico) y se introducen por filas todos los ciclos de ese electrodoméstico durante un periodo de tiempo. Como habrá ciclos en los que la duración sea menor que el ancho fijado, se rellena con ceros.

A continuación, se establece una correspondencia temporal entre los ciclos de consumo desagregado y el consumo agregado y se extraen los datos comprendidos entre el inicio y el fin de cada ciclo.

Los datos de consumo desagregado y agregado serán utilizados como datos de

entrada y test para los diferentes modelos de redes neuronales que serán implementados en Matlab.

```

layers = [
    imageInputLayer([1 2401 1])
    convolution2dLayer([1 5],8)
    convolution2dLayer([1 3],8)
    convolution2dLayer([1 3],8)
    maxPooling2dLayer([1 4], 'stride', [2 2])
    convolution2dLayer([1 3],16)
    convolution2dLayer([1 3],16)
    convolution2dLayer([1 3],16)
    maxPooling2dLayer([1 4], 'stride', [2 2])
    convolution2dLayer([1 3],32)
    convolution2dLayer([1 3],32)
    convolution2dLayer([1 3],32)
    maxPooling2dLayer([1 4], 'stride', [2 2])
    convolution2dLayer([1 3],64)
    convolution2dLayer([1 3],64)
    convolution2dLayer([1 3],64)
    maxPooling2dLayer([1 4], 'stride', [2 2])
    fullyConnectedLayer(5)
    softmaxLayer()
    classificationLayer()
];

```

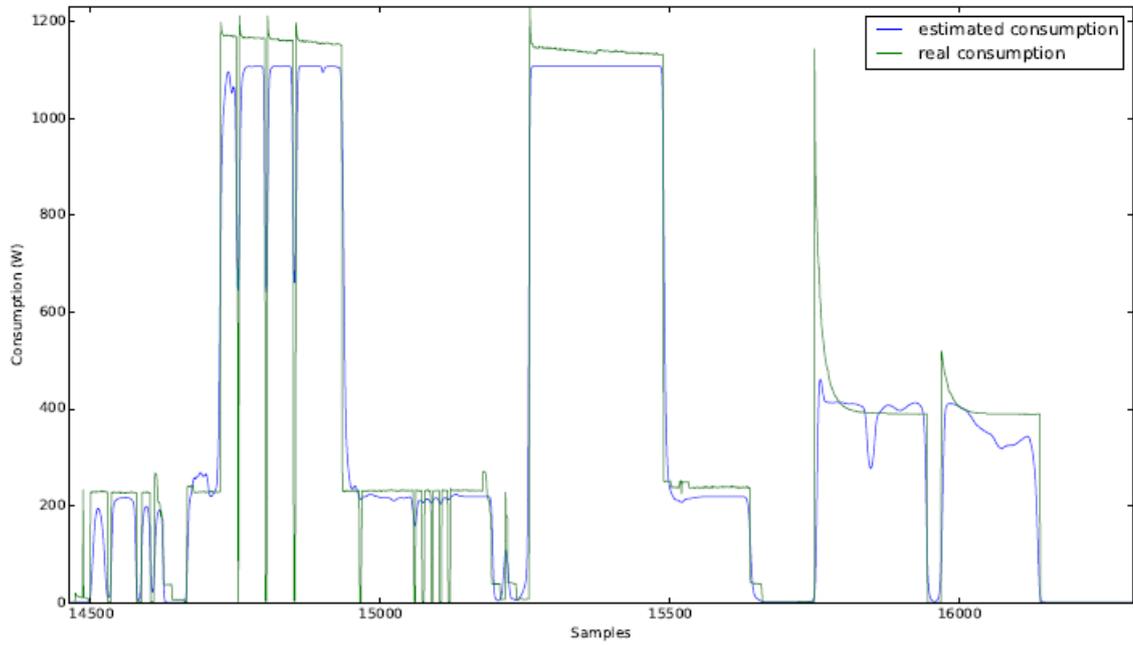
### 5.9.2. Resultados de redes neuronales en el estudio de NILM

A falta de resultados propios en el tratamiento de datos numéricos, comentaremos brevemente los resultados de la aplicación de redes neuronales en base a la literatura existente.

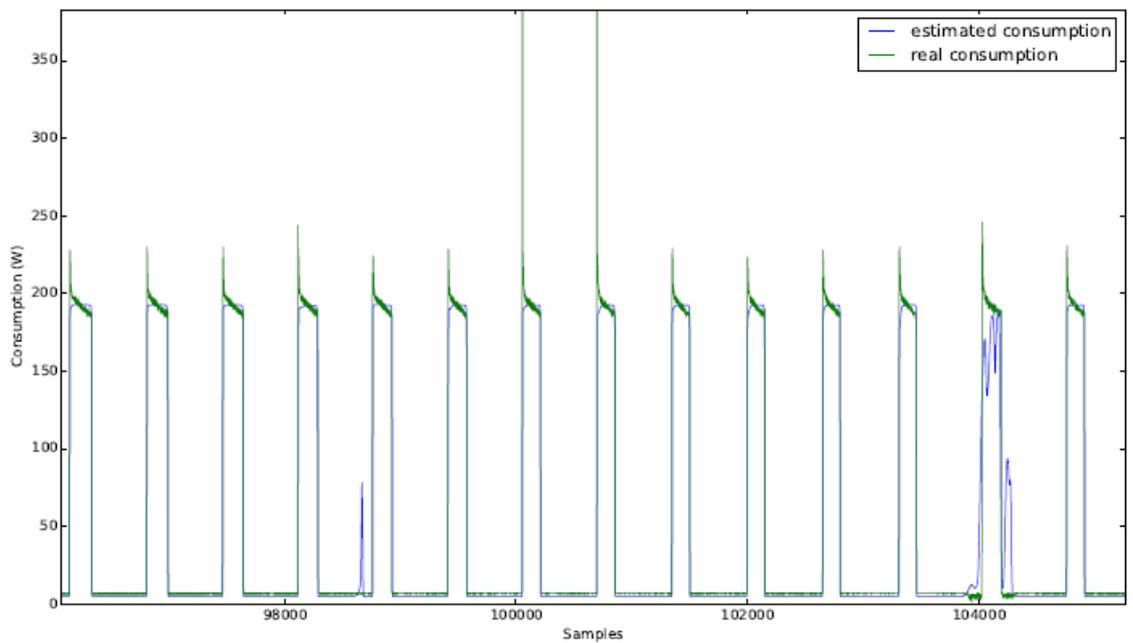
En líneas generales, los resultados obtenidos al aplicar redes neuronales son excelentes siendo capaces incluso de desagregar el consumo en casas no entrenadas previamente.

En el estudio de la desagregación del consumo existen algunas métricas que predominan por encima del resto y que permiten analizar el grado de desarrollo de un modelo. Distinguiremos entre métricas para calificar la clasificación y métricas para la desagregación. Así, para la clasificación tenemos:

$$\mathbf{Recall} = \frac{TP}{TP + FN}$$



(a) Resultados con el lavavajillas.



(b) Resultados con el frigorífico.

Figura 5.13: Resultados de la desagregación de consumo.

$$\mathbf{Precisión} = \frac{TP}{TP + FP}$$

$$\mathbf{F1} = 2 \frac{\text{precisión} \times \text{recall}}{\text{precisión} + \text{recall}}$$

$$\mathbf{Accuracy} = \frac{TP + TN}{P + N}$$

donde TP es el número de verdaderos positivos, TN el número de verdaderos negativos, FP el número de falsos positivos, FN el número de falsos negativos, P el número de positivos y N el número de negativos.

Para la desagregación, se definen las siguientes métricas:

$$\mathbf{Error\ relativo\ en\ la\ energía\ total} = \frac{\hat{E} - E}{\max(E, \hat{E})}$$

$$\mathbf{Error\ medio\ absoluto} = \frac{1}{T} \sum_{t=1}^T |\hat{y}_t - y_t|$$

donde E es la energía total actual,  $\hat{E}$  la energía total predicha,  $y_t^{(i)}$  la potencia actual del electrodoméstico  $i$  en el instante  $t$ ,  $\hat{y}_t^{(i)}$  la potencia estimada del electrodoméstico  $i$  en el instante  $t$  y  $\bar{y}_t^{(i)}$  la potencia agregada actual en el instante  $t$ .

Una vez definidas las métricas, podemos observar en la Figura 5.14 y Figura 5.15 como en ambos estudios las redes neuronales son capaces de proporcionar excelentes resultados en la desagregación del consumo energético y, dentro de las diferentes tipologías de redes, destaca el modelo convolucional de Nacimiento alcanzando un accuracy del 97.67% y una precisión del 98.75%.

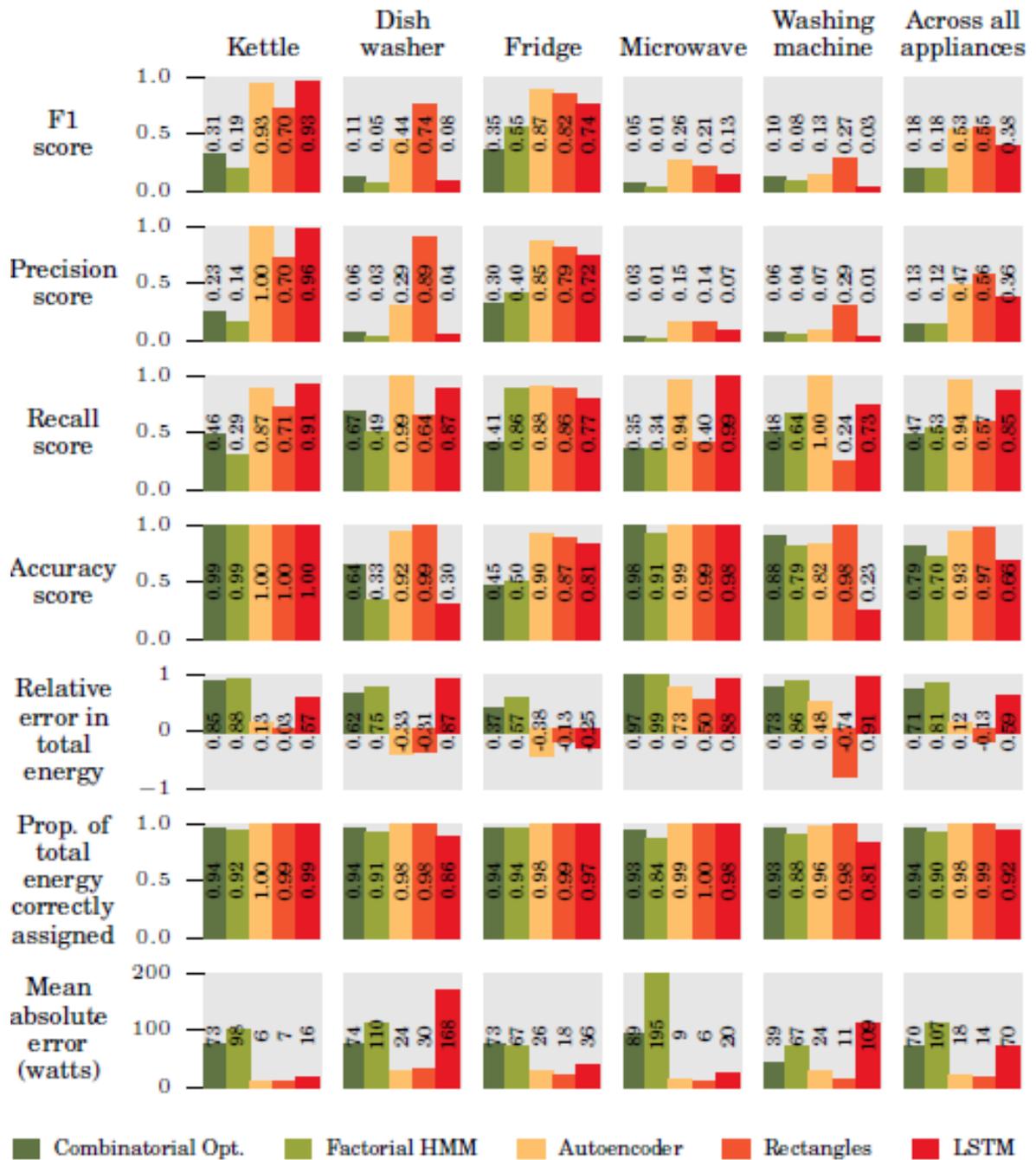


Figura 5.14: Resultados extraídos de Kelly [14] con redes neuronales (Autoencoders, rectangles o Long short-term memory (LSTM)) y métodos convencionales.

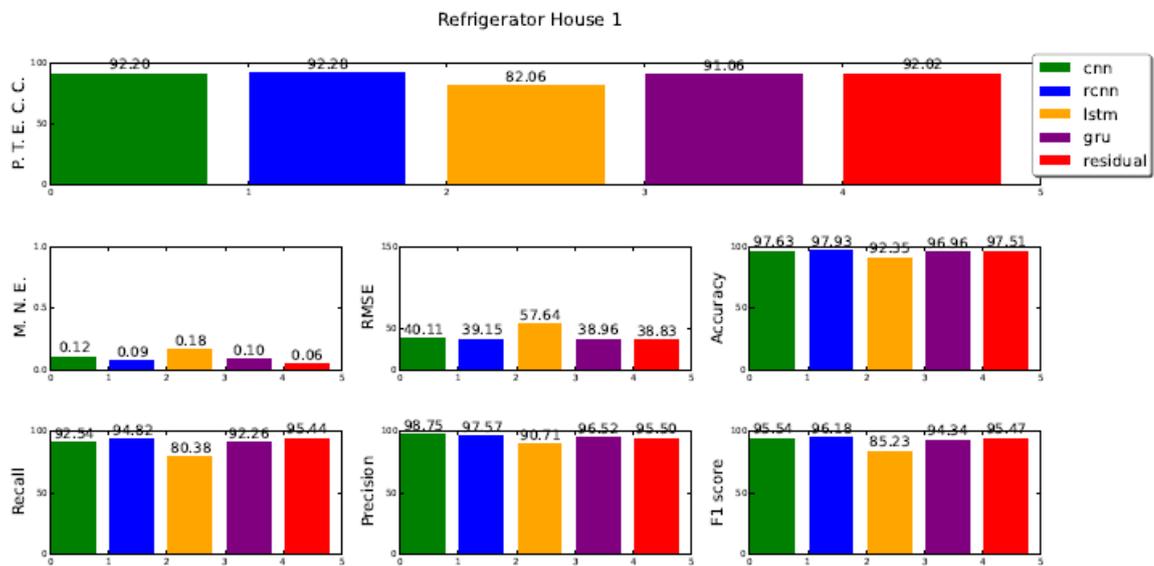


Figura 5.15: Algunos resultados extraídos de Nascimento [7] con diferentes modelos de redes neuronales.

# Capítulo 6

## Conclusiones y líneas futuras

La desagregación del consumo energético no intrusivo es, sin duda, un amplio campo de estudio y aplicación de los nuevos métodos de machine learning.

El presente trabajo hace un repaso de mi estancia en Optare Solutions comenzando por el objetivo principal del proyecto: una revisión crítica de los métodos de desagregación más tradicionales de aplicación en el estudio del NILM y la propuesta de nuevas metodologías. Algunos de estos métodos ya habían sido implementados en la empresa sin obtener buenos resultados por lo que, una vez revisada la literatura existente, mi propuesta fue la de abordar el problema mediante redes neuronales.

Dados los excelentes resultados de las redes neuronales en tareas de clasificación por imagen y la gran documentación y librerías disponibles sobre este asunto, se ha optado por comenzar una batería de pruebas mediante la clasificación por imagen de los ciclos de consumo de los diferentes electrodomésticos sin obtener resultados satisfactorios en el objetivo de la desagregación. No obstante, cabe destacar que el tratamiento por imagen de este asunto no había sido estudiada anteriormente por lo que, a pesar de que el resultado no fue satisfactorio en términos de desagregación, sí ha servido para descartar esta vía de estudio.

Finalmente, se propone el tratamiento del problema con datos numéricos. Para

ello, me he basado en los trabajos de Kelly y Nascimento con quien he llegado incluso a intercambiar mensajes y a quien agradezco su colaboración. Por falta de tiempo, el estudio con datos numéricos no pudo concluirse durante la estancia en la empresa pero es una tarea abierta que se llevará a cabo en el futuro y que ya ha sido trabajada fuera del entorno laboral en colaboración con Javier Tarrío, profesor de la Universidad de Vigo, a quien agradezco enormemente su colaboración.

En conclusión, las redes neuronales parecen ser una potente herramienta en la desagregación del consumo energético y han arrojado excelentes resultados en los estudios realizados a pesar de que el entrenamiento de un modelo complejo puede suponer hasta 10 días de cómputo en 3 GPUs.

A pesar de que las soluciones existentes han sentado una base, el desarrollo de los modelos tiene un gran campo de mejora aumentando la complejidad de los modelos y el tamaño muestral disponible mejorando también la recogida de datos con el desarrollo de nuevas herramientas de hardware. Además, otras técnicas de preprocesado, clustering y regularización pueden dar lugar a un menor error de generalización.

# Bibliografía

- [1] BHATTACHARJEE, S., KUMAR, A., AND ROYCHOWDHURY, J. Appliance classification using energy disaggregation in smart homes.
- [2] CASTRO, J. Algorithms for energy disaggregation. Master's thesis, Universidad politécnica de Cataluña, 2016.
- [3] COLE, A., AND ALBICKI, A. Algorithm for nonintrusive identification of residential appliances. *Circuits and Systems* (1998).
- [4] CUGLIARI, J., GOUDE, Y., AND POGGI, J. M. Disaggregated electricity forecasting using wavelet-based clustering of individual consumers. *IEEE International Energy Conference* (2016).
- [5] DARBY, S. The effectiveness of feedback on energy consumption. *A review for DEFRA of the Literature on Metering, Billing and direct Displays* (2016).
- [6] DEL BRÍO, B. M., AND MOLINA, A. S. *Redes neuronales y sistemas borrosos*. RA-MA, 2006.
- [7] DO NASCIMENTO, P. P. M. *Applications of deep learning techniques on NILM*. PhD thesis, Universidad Federal de Rio de Janeiro, 2016.

- [8] FARINACCIO, L., AND ZMEUREANU, R. Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses. *Elsevier* (1999).
- [9] FIOL, A., CARMONA, J., AND CASTRO, J. *Algorithms for Energy Disaggregation*. PhD thesis, Facutlad de Informática de Barcelona, 2016.
- [10] HART, G. W. Nonintrusive appliance load monitoring. *Proceedings of the IEEE* (1992).
- [11] HART, W., KERN, E., AND SCHWEPPE, F. Non-intrusive appliance monitor apparatus, 1989.
- [12] HE, W., AND CHAI, Y. An empirical study on energy disaggregation via deep learning. *Advances in Intelligent Systems Research* (2016).
- [13] KELLY, D. A. Disaggregating smart meter readings using device signatures. Master's thesis, Imperial College London, 2011.
- [14] KELLY, J. *Disaggregation of Domestic Smart Meter Energy Data*. PhD thesis, University of London, 2017.
- [15] KELLY, J., AND KNOTTENBELT, W. Neural nilm: Deep neural networks applied to energy disaggregation. *ACM BuildSys* (2015).
- [16] KOLTER, Z., AND JOHNSON, M. Redd: A public data set for energy disaggregation research. *SustKDD* (2011).
- [17] LARRAÑAGA, P., INZA, I., AND MOUJAHID, A. Departamento de ciencias de la computación e inteligencia artificial universidad del país vasco.
- [18] ORZÁEZ, J. M. A. *Non-Intrusive Load Monitoring techniques for Activity of Daily Living recognition*. PhD thesis, Universidad de Alcalá, 2016.

- [19] PARSON, O., GHOSH, S., WEAL, M., AND ROGERS, A. An unsupervised training method for non-intrusive appliance load monitoring. *Electronics and Computer Science* (2014).
- [20] RIDI, A., GISLER, C., AND HENNEBERT, J. A survey on intrusive load monitoring for appliance recognition. *2014 22nd International Conference on Pattern Recognition* (2014).
- [21] RIDI, A., AND HENNEBERT, J. Hidden markov models for ilm appliance identification. *Elsevier* (2014).
- [22] ROY, N., PATHAK, N., AND MISRA, A. Fine-grained appliance usage and energy monitoring through mobile and power-line sensing. *Pervasive and Mobile Computing* (2016).