

Modelos de aprendizaje profundo para la predicción de comportamiento sedentario futuro

Deep learning models for predicting future sedentary behavior

Martín Santillán Cooper, Marcelo Armentano

ISISTAN (CONICET-UNCPBA), Campus Universitario, Tandil, Argentina.
mscooper@exa.unicen.edu.ar
marcelo.armentano@isistan.unicen.edu.ar

Resumen. Es sabido que el comportamiento sedentario posee consecuencias negativas para la salud y, por lo tanto, alentar a los individuos a evitar este tipo de comportamiento puede colaborar en la reducción de diferentes indicadores de riesgo. En este trabajo, se evaluaron diferentes arquitecturas de aprendizaje profundo para predecir el comportamiento sedentario futuro de un individuo a partir de los registros capturados de diferentes sensores disponibles en los dispositivos móviles actuales. Se analizaron usuarios con diferentes niveles de gasto energético, y se obtuvieron resultados alentadores que demuestran la eficiencia de las arquitecturas propuestas.

Abstract. It is well known that sedentary behavior has negative consequences for health. Therefore, encouraging individuals to avoid this type of behavior can help to reduce different risk indicators. In this work, different deep learning architectures were evaluated to predict the future sedentary behavior of an individual from the captured records of different sensors available on mobile devices. Users with different levels of energy expenditure were analyzed, and encouraging results were obtained that demonstrate the efficiency of the proposed architectures.

Keywords: aprendizaje profundo; aprendizaje de máquina; comportamiento sedentario

1 Introducción

La definición de comportamiento sedentario ha ido evolucionando a lo largo de los años, al mismo tiempo que lo hizo la forma de medirlo. La Red de Investigación del Comportamiento Sedentario¹ define el comportamiento sedentario como cualquier comportamiento caracterizado por un gasto de energía menor o igual a 1.5 MET (*Metabolic Equivalent of Tasks*, Equivalente Metabólico de Tareas) mientras se está sentado o reclinado [1]. MET mide la intensidad de una actividad en múltiplos del gasto energético en reposo. Ejemplos de actividades sedentarias son mirar televisión (1.0 MET), comer sentado (1.5 MET), jugar videojuegos (1.0 MET) y conducir (1.3 MET).

La investigación realizada en el área demuestra asociaciones sólidas y consistentes entre el tiempo sedentario y la diabetes, las enfermedades cardiovasculares y la mortalidad por todas las causas [2, 3]. Sin embargo, las asociaciones informadas fueron en gran medida independientes de la actividad física. Por lo tanto, es importante tener en cuenta que el comportamiento sedentario no representa lo contrario de la actividad física y que es posible que un individuo tenga niveles altos de actividad física moderada a vigorosa (MVPA, por sus siglas en inglés) y, a su vez, comportamiento sedentario. En general, se ha demostrado que el tiempo sedentario está asociado de forma perjudicial con la salud y con marcadores de riesgo metabólico en diversos grupos de población. Además, se ha destacado la importancia de no solo estimular la MVPA sino también de reducir el tiempo de sedentarismo, ya que la conducta sedentaria es un factor de riesgo para la mortalidad independiente de la MVPA [3, 4]

Se han investigado también las asociaciones entre los descansos breves en el tiempo sedentario con los resultados metabólicos [5] y con la optimización de las operaciones cognitivas [5, 6, 7, 8]. Benatti y Ried-Larsen [9] afirmaron que existe evidencia suficiente para demostrar los efectos positivos de romper el tiempo prolongado de estar sentado en los resultados metabólicos.

Los métodos utilizados para medir el comportamiento sedentario se pueden clasificar en subjetivos (cuestionarios y diarios de autoevaluación) y objetivos (sensores disponibles en dispositivos ubicuos). Los métodos subjetivos están siendo superados por las nuevas tecnologías que pueden proporcionar, para todos los grupos de la población, información de segundo a segundo sobre la postura, el movimiento (o la falta de movimiento) y los patrones dentro y entre días [10].

Si bien los dispositivos móviles pueden considerarse como una de las causas del comportamiento sedentario [10, 11], también ofrecen nuevas oportunidades para prevenirlo. Hoy en día, los dispositivos móviles portátiles, como teléfonos inteligentes (*smartphones*), relojes inteligentes (*smartwatches*) y rastreadores de ejercicios están equipados con una amplia variedad de sensores que se pueden usar para la actividad humana y el análisis de comportamiento. El uso de métodos objetivos para evaluar el comportamiento sedentario está creciendo en popularidad a medida que los costos de los dispositivos móviles portátiles disminuyen y son más fáciles de usar. En 2018, el

¹ <https://www.sedentarybehaviour.org/>

91% de las personas entre 18 y 49 años en los EE. UU poseían un teléfono inteligente². En este contexto, estos dispositivos pueden verse como una oportunidad para desarrollar métodos objetivos complejos para medir el comportamiento sedentario.

Se han implementado muchas aplicaciones para teléfonos inteligentes con el objetivo de alertar al usuario cuando se reconoce un comportamiento sedentario [11, 12, 13]. Predecir conductas sedentarias futuras puede ayudar a habilitar intervenciones preventivas, como recordatorios y sugerencias para diferentes actividades basadas en la Teoría de la conducta planificada (TPB) [13,14]. TPB postula que es más probable que un sujeto participe en las intervenciones recomendadas para reducir las conductas sedentarias si tales actividades están incluidas en sus planes. Siguiendo esta idea, nuestra hipótesis es que si pudiéramos predecir en el momento t que un sujeto será sedentario en el tiempo $t + 1$, podríamos recomendar actividades para cambiar su rutina sedentaria a largo plazo. Este tipo de intervenciones puede dar lugar a mejores oportunidades para cambiar el comportamiento de los sujetos a resultados más saludables.

En este artículo, se presenta un enfoque novedoso para predecir el comportamiento sedentario futuro de un sujeto en términos de su nivel de MET a partir de diferentes algoritmos de aprendizaje profundo (*Deep Learning*), basado en la observación de valores obtenidos de múltiples sensores de dispositivos portátiles y móviles.

El artículo está organizado de la siguiente manera. En la Sección 2, se presentan los trabajos relacionados. En la sección 3, se describen las arquitecturas propuestas en este trabajo para predecir el comportamiento sedentario futuro. En la sección 4, se detalla el procedimiento experimental llevado a cabo, describiendo el *dataset* utilizado, las características extraídas del mismo y los resultados obtenidos para los casos de estudio seleccionados. Finalmente, en la Sección 5, se exponen las conclusiones a las que se arribaron.

2 Marco Conceptual

A continuación, se presentan diferentes modelos pertenecientes a lo que se conoce como *Aprendizaje Profundo* y se discute su uso en el marco de problemas donde la secuencialidad juega un rol importante.

2.1 Perceptrón Multicapa

El Perceptrón Multicapa (MLP) es la arquitectura más sencilla dentro del campo de Aprendizaje Profundo. Básicamente es una red de unidades computacionales (neuronas) unidas por conexiones directas. Cada neurona lleva a cabo cierto cálculo y produce como salida un valor que es luego emitido a hacia todas sus conexiones salientes. Cada conexión tiene un peso que corresponde a cuan fuertemente están conectadas dos neuronas. Típicamente, la computación que lleva a cabo está separada

² <http://www.pewresearch.org/fact-tank/2018/09/28/internet-social-media-use-and-device-ownership-in-u-s-have-plateaued-after-years-of-growth/>

en dos etapas: primero realiza la suma ponderada de todas sus conexiones entrantes y luego aplica una función de activación [15].

Una de las desventajas de utilizar este tipo de redes para el modelado de secuencias es que ignora completamente la topología de la entrada, es decir, timesteps y características se representan en la misma dimensión, por lo que no es posible “obligar” a la red a que tenga en cuenta estas representaciones. Dicho de otra manera, las variables de entrada podrían ser presentadas en cualquier orden sin afectar el resultado del entrenamiento [15].

En este trabajo se consideró dicha arquitectura con el fin de evaluar si utilizar arquitecturas que tengan en cuenta la estructura secuencial de los datos, representa o no una mejora en la precisión de los modelos .

2.2 Redes Neuronales Convolucionales

Las Redes Neuronales Convolucionales (CNN) son muy comunes en el procesamiento de imágenes y presentan varias diferencias con respecto a las MLP. Por un lado, las CNN toman como entrada datos claramente estructurados, en los que las variables cercanas están fuertemente correlacionadas (por ejemplo, imágenes). Por otro lado, este tipo de redes reduce en gran medida el número de parámetros a entrenar, ya que éstos son compartidos por todas las neuronas de una determinada capa. Estos parámetros son llamados filtros, y a partir de aplicar la convolución a cada sector de los datos de entrada (sean de la entrada de la red o de una capa intermedia), extraen características locales que luego son combinadas sucesivamente generando características globales [15]. Por lo tanto, dichos filtros pueden ser interpretados como características que crecen en complejidad con cada capa de la red. Comúnmente, entre cada capa de convolución se aplica un método llamado *pooling* que reduce la cantidad de neuronas y, por lo tanto, la resolución de los datos procesados, obteniendo invariancia ante la translación y mejorando la capacidad de generalización de la red.

Las CNN han sido utilizadas con éxito en tareas de predicción de series de tiempo en diversas investigaciones [16, 17, 18]. Más recientemente, este tipo de redes, han mostrado un rendimiento del estado del arte en tareas que eran dominadas por las Redes Neuronales Recurrentes (RNN), como las síntesis de audio, modelado de lenguaje a nivel de palabra, traducción de máquina [19, 20, 21]. A partir de estos trabajos, se han implementado diferentes tipos de CNN para el modelado de secuencia, y, en especial, para la predicción en series de tiempo.

2.3 Redes Temporales Convolucionales

En Bai et al. 2018 [36] los investigadores presentan un tipo de CNN especialmente diseñada para las tareas de predicción y modelado de secuencias, llamada Redes Temporales Convolucionales (TCN). En dicho trabajo, los investigadores realizan una evaluación sistemática de arquitecturas convolucionales y recurrentes. Específicamente, se ponen a prueba tareas que han sido utilizadas en repetidas ocasiones para comparar la efectividad de diferentes RNNs. Los resultados muestran

que la común asociación entre el modelado de secuencias y las RNN debe ser reconsiderada, ya que fueron las TCN las arquitecturas con mayor precisión.

Las TCN ya habían sido utilizada en otros trabajos [19, 22], pero en Bai et al. 2018 [36] se implementa su versión más simple, y es, además, la versión utilizada en el presente trabajo. En dicha versión, se busca que las TCN tengan todas las ventajas que presentan las CNN más simples –alta paralelización, no sufrir *vanishing gradient* o *exploding gradient*, baja utilización de memoria-, pero siendo adaptadas para el modelado de secuencias y la predicción de series de tiempo. Por esta razón, existen varias diferencias claves entre ambas. Las diferencias más importantes son que las TCN:

- No utilizan las capas de *pooling* para que la dimensión de entrada sea igual a la de salida,
- Siempre aplican la convolución de una dimensión (la del tiempo).
- Son causales, es decir, las convoluciones solo consideran información del pasado y,
- Utilizan dilataciones para aumentar el campo receptivo exponencialmente y así poder cubrir largas dependencias en los datos secuenciales.

2.4 Redes Neuronales Recurrentes

Las Redes Neuronales Recurrentes son modelos dedicados a secuencias que mantienen un estado interno que es propagado a través del tiempo. Dicho estado interno funciona como la memoria y puede actuar como una representación de todo aquello que la red haya visto en la secuencia de entrada hasta el momento. En síntesis, las RNN tienen como entrada un vector que representa una secuencia (por ejemplo: una oración) y las neuronas procesan cada elemento de dicho vector uno a la vez e iterativamente [23].

Las RNN han ganado mucha popularidad por su buen desempeño en tareas relacionadas al modelado del lenguaje [24] y traducción de máquina [25]. Desgraciadamente, las arquitecturas más simples de RNN son difíciles de entrenar [26], ya que presentan problemas como *vanishing gradient* y *exploding gradient*, por lo que en la práctica se utilizan arquitecturas más complejas que logran sortear en cierta medida dichos problemas. Estas variantes utilizadas son las LSTM [27] y las GRU [28] que añaden mayor cantidad de información a cada neurona en comparación con las RNN más simples, como las compuertas de olvido.

3 Trabajos Relacionados

Los trabajos que buscan resolver el problema de predecir el comportamiento sedentario a futuro, son escasos y hacen suposiciones diversas a la hora de tratar de resolverlo.

Q. He and E. Agu [15] son los únicos dos autores que han propuesto, analizado y comparado modelos para la predicción de comportamiento sedentario futuro. A diferencia del enfoque propuesto en el presente trabajo, Q. He and E. Agu determinan

el nivel de sedentarismo como el porcentaje de registros de actividad estacionarios que se tomaron de los datos de actividad física de cada usuario en cada hora, con respecto a los demás tipos de registros de actividad (corriendo, caminando, etc.).

Primero, propusieron un algoritmo de dominio de frecuencia para detectar patrones sedentarios recurrentes a partir de los datos de los usuarios. Este trabajo buscó ajustar funciones periódicas (seno, coseno) al porcentaje de registros de actividad estacionaria de cada hora y de cada usuario, tratando así de identificar los patrones subyacentes en su actividad física [29].

En un segundo estudio, los mismos autores exploraron la idea de si el contexto de un usuario puede ser usado para predecir su comportamiento sedentario futuro. Usando regresión logística, fueron capaces de clasificar las variables de contexto de cada usuario para predecir si iba a ser "muy sedentario" en la próxima hora, con una precisión de 73.1% y un *recall* de 87.7% [11]. Paralelamente, propusieron un enfoque para descubrir automáticamente los patrones de actividad sedentaria de los usuarios usando modelos auto-regresivos [30].

Finalmente, se enfocaron en detectar los ritmos predominantes de comportamiento sedentario modelando los ritmos cíclicos y ritmos lineales expuestos por la filosofía de Lefebvre [31].

Es importante remarcar que en ninguno de los trabajos anteriormente expuestos se toma en cuenta el MET de cada actividad, que es el estándar en el área de la salud para decidir si una actividad es sedentaria o no. Consideramos que es muy importante tener en cuenta el concepto de MET al momento de diseñar modelos para la predicción de comportamiento sedentario, ya que este problema es de carácter interdisciplinario, por lo que para que dichos modelos tengan mayor performance, es necesario tener en cuenta, además, los avances de las ciencias médicas. Adicionalmente, es importante resaltar que hasta el momento no se han utilizado modelos que usen algoritmos de *Deep Learning* para tratar el problema de la predicción del comportamiento sedentario.

4 Predicción de Comportamiento Sedentario Futuro

El problema que nos proponemos resolver es el de predecir el gasto de energía que tendrá una persona en el futuro en base a datos recolectados de sensores de su *smartphone*. El gasto energético es medido en MET, que es la medida estándar en la comunidad científica que estudia la salud en relación con la actividad física. En el ámbito de la salud, se ha llegado a un acuerdo entre los investigadores en determinar cómo actividad sedentaria a toda aquella actividad cuyo MET asociado sea menor o igual a 1,5.

En el contexto de este trabajo, se define la predicción del comportamiento sedentario futuro (FSBP) como la tarea de decidir si la actividad física de un usuario superará o no, en promedio, 1,5 MET en un futuro próximo.

En el marco del aprendizaje de máquina, al ser el MET un valor continuo, el problema a resolver es una tarea de regresión. Dentro de los algoritmos de aprendizaje profundo, las Redes Neuronales Recurrentes (RNN), las Redes Neuronales

Convolutivas (CNN) y las Redes Temporales Convolutivas (TCN) son especialmente aptas para el problema a tratar porque los datos registrados de la actividad del usuario están organizados en series de tiempo en las que la secuencialidad es importante.

En este trabajo, se buscará predecir el valor de MET para un usuario únicamente en la hora siguiente. Es decir, dados los datos de las horas $t - n, \dots, t - 2, t - 1, t$ se utilizarán Redes Neuronales para hallar el gasto energético en el tiempo $t + 1$. Nuestro objetivo principal es evaluar la factibilidad de utilizar redes neuronales capaces de aprender, a partir de una secuencia de datos de la actividad de un usuario, patrones de comportamiento que sean útiles para el problema de FSBP. Además, nos interesa evaluar la performance de las TCN en el contexto del problema a abordar, ya que, para tareas de modelado de secuencias, han demostrado superar a las RNN que han sido comúnmente utilizadas como punto de referencia en una gran variedad de dominios [36].

Las TCN son un tipo de CNN [32] que aplican la convolución de una dimensión, son causales (es decir, solo consideran información del pasado), utilizan conexiones residuales y dilataciones para aumentar el campo receptivo. Las TCN demostraron superar ampliamente a las clásicas RNN basadas en GRU o LSTM en los problemas típicos como el problema de sumar o de copiar [36]. Además, las TCN han demostrado tener más memoria que redes recurrentes de la misma capacidad. Por otro lado, y no menos importante, las TCN y las CNN en general, poseen la característica de ser altamente paralelizables, en contraste con las RNN, por lo que el proceso de la búsqueda óptima del valor de los hiper-parámetros se realiza de una forma más rápida.

Uno de los hiper-parámetros que varía en las distintas arquitecturas propuestas es la cantidad de *time-lags* utilizados. Los *time-lags* son la cantidad de información que se le da a la arquitectura sobre el pasado. Es decir, si se quiere predecir el gasto de energía de un usuario en un tiempo t , y la cantidad de *time-lags* es 3, la entrada de la red neuronal estará formada por las características (*features*) de los tiempos $t - 1, t - 2$ y $t - 3$.

En general, los hiper-parámetros a configurar son: tipo de capas utilizadas, cantidad de capas, cantidad de unidades por capa, técnica de regularización a utilizar, *time-lags* a utilizar, algoritmo de optimización y la función de pérdida. Además, las CNN poseen ciertos parámetros que no poseen las demás arquitecturas, como la cantidad de filtros por capa o el tamaño del núcleo (*kernel*). Más aún, las TCN poseen hiper-parámetros que no poseen las CNN: como la cantidad de bloques residuales o la lista de dilataciones. Estos últimos hiper-parámetros se tunean de una forma menos aleatoria que los demás, ya que lo que se intenta calcular en este caso es que el campo receptivo sea igual o mayor a la cantidad de *time-lags*.

En resumen, se evaluarán 4 arquitecturas de redes neuronales: la arquitectura 1 es una RNN, la arquitectura 2 es una CNN, la arquitectura 3 es una TCN y la arquitectura 4 es una NN. La arquitectura 4 utiliza únicamente información de la hora anterior para predecir el valor del MET de la hora siguiente y fue agregada para evaluar la utilidad de utilizar un enfoque de modelado de secuencias para FSBP.

A continuación, la Tabla 1 muestra una descripción de las arquitecturas propuestas. En todos los casos se utiliza el Error Cuadrático Medio (MSE) como función de

pérdida, ADAM como algoritmo de optimización, siendo 128 la cantidad de épocas y 64 el tamaño del lote.

	Arq. 1	Arq. 2	Arq. 3	Arq. 4
Tipo de capas	LSTM x 2 - FC x 1	No aplica	Conv1D x 1 - Flatten x 1 - FC x 1	Dense x 3
Neuronas por capa	64-32-1	No aplica	32 - No aplica - 1	64 - 32 - 1
Técnica de re- gularización	Dropout	Dropout	Dropout y Batch Nor- malization	Dropout
Número de bloques resi- duales	No aplica	1	No aplica	No aplica
Lista de dila- taciones	No aplica	1-2-4	No aplica	No aplica
Tamaño del kernel	No aplica	2	2	No aplica
Número de filtros	No aplica	6	32	No aplica

FC = Fully Connected

Tabla 1. Descripción de las arquitecturas propuestas

5 Evaluación Experimental

En esta Sección, primero se presenta el *dataset* utilizado para evaluar las diferentes arquitecturas propuestas para la tarea de FSBP. Luego, se describe el pre-procesamiento realizado y las características extraídas del *dataset*. A continuación, se exponen los usuarios del *dataset* seleccionados como casos de estudio. Finalmente, se muestran y

se discuten los resultados obtenidos. Para la implementación de las arquitecturas se utilizó la librería *Keras*³ de *Python 3*. Los *scripts* implementados para el procesamiento del dataset y el entrenamiento de los modelos se dejan disponibles públicamente⁴.

5.1 Descripción del Dataset

Para validar la eficiencia de las arquitecturas propuestas fue utilizado el *dataset StudentLife*⁵ [33]. Este *dataset* contiene datos de los sensores de los *smartphone* de 48 estudiantes a lo largo de 10 semanas en la primavera de 2013.

El dataset se recogió de 30 estudiantes de pregrado y 18 graduados. De todo el grupo de estudiantes, 38 eran varones y 10 mujeres. Del grupo de estudiantes de pregrado, dos eran de primer año, 14 de segundo año, 6 de tercer año y ocho de cuarto año. También hubo 13 estudiantes de maestría de primer año y 1 de segundo año, y 3 estudiantes de doctorado. Los participantes fueron racialmente diversos, con 23 caucásicos, 23 asiáticos y 2 afroamericanos. Los datos disponibles incluyen: datos de actividad física, datos de conversación, datos de ubicación, entre otros.

Diferentes tipos de datos fueron registrados con diferentes frecuencias, dependiendo del tipo de sensor y la carga de trabajo global del *smartphone*. Por ejemplo, la actividad física se registró cada 2–3 segundos en 1 de cada 4 minutos, el uso de aplicaciones se registró cada 20 minutos, y los registros de ubicación cada 10 minutos.

5.2 Pre-procesamiento y selección de características

Cada registro del *dataset StudentLife* posee un *timestamp* asociado. Se decidió discretizar las series de tiempo en intervalos de una hora ya que esta es la granularidad utilizada en la mayor parte del trabajo relacionado para la predicción de sedentarismo. Por lo tanto, todas las características que se generaron corresponden a una combinación de usuario/hora en particular y son algún cómputo que resume algún aspecto de los datos de algún usuario/hora particular. Por ejemplo, un intervalo específico puede corresponder al usuario 10 y la hora 2013-04-24 19: 00–20: 00. Los intervalos para los cuales no se tenía información sobre la actividad física del usuario fueron eliminados. Además, se eliminó la información de un usuario en particular que presentaba inconsistencias en sus registros de actividad física. En total, se obtuvieron 60.819 intervalos para los 48 usuarios.

Los registros de actividad de cada usuario están clasificados en el dataset como estacionario, caminando o corriendo. A partir de estos registros se calculó el valor del gasto energético para cada intervalo. Tal como fue explicado anteriormente, el gasto energético es comúnmente medido de términos de *Metabolic Equivalent of Tasks* (MET). A cada tipo de actividad se le asignó un valor estático de MET de acuerdo al

³ <https://keras.io/>

⁴ <https://github.com/sailaperc/sedentarism-prediction>

⁵ <http://studentlife.cs.dartmouth.edu/>

Compendio de Actividades Físicas⁶. Finalmente, se calculó el promedio de MET para cada intervalo.

A continuación, definimos las 20 características generadas a partir del *dataset StudentLife*. Cabe destacar que algunas de estas características fueron seleccionadas en base a investigaciones previas [11, 34], mientras que otras fueron definidas específicamente en este trabajo.

Características obtenidas a partir del GPS:

- locationVariance: utilizado para medir la variabilidad en la ubicación GPS de un participante en cada intervalo, computado como se describe en [35].

Características basadas en el tiempo:

- hourSine: transformación de seno de la hora;
- hourCosine: transformación de coseno de la hora;
- dayOfWeek: día de la semana;
- pastMinutes: el número de minutos transcurridos desde el comienzo del día;
- remainingMinutes: el número de minutos que quedan para terminar el día;

Características de actividad física:

- Stationary level: el porcentaje de instancias de actividad física clasificadas como 'estacionario/a' en cada intervalo;
- Walking level: el porcentaje de instancias de actividad física clasificadas como 'caminando' en cada intervalo;
- Running level: el porcentaje de instancias de actividad física clasificadas como de 'corriendo' en cada intervalo;
- activityMajor: el tipo de actividad física con más instancias en cada intervalo;

Características de audio

- SilenceLevel: el porcentaje de instancias de audio clasificadas como 'silencio' en cada intervalo;
- voiceLevel: el porcentaje de instancias de audio clasificadas como 'voz' en cada intervalo;
- noiseLevel: el porcentaje de instancias de audio clasificadas como 'ruido' en cada intervalo;
- numberOfConversations: el número de conversaciones que ese estudiante tuvo en cada intervalo;

Otras características:

- isCharging: si el *smartphone* se estaba cargando;
- isLocked: si el *smartphone* estaba bloqueado;
- isInDark: si el *smartphone* estaba en la oscuridad;

⁶ <https://sites.google.com/site/compendiumofphysicalactivities/>

- `hasCalendarEvent`: si tenía un evento programado en el calendario;
- `wifiChanges`: el número de cambio de conexiones wifi en cada intervalo;
- `sLevel`: el valor de MET para la hora actual.

Finalmente, dado que las características `dayOfWeek` y `activityMajor` son categóricas, se realizó *dummy-encoding* sobre ellas. Con esta transformación, se obtuvieron 7 *dummy features* representando las categorías para `dayOfWeek` y 3 para `activityMajor`. Como resultado de esta última transformación, se finalizó el proceso con 28 características que serán luego usadas por los modelos predictivos para predecir el comportamiento sedentario de los usuarios.

El siguiente paso del pre-procesamiento consistió en generar los *time-lags*. La cantidad de *time-lags* depende de cada arquitectura propuesta, siendo en algunos casos cero. Para ello, a una instancia del dataset que corresponde a un tiempo t se le agregan las características de tiempos anteriores, dependiendo de la cantidad de *time-lags* que utilice la arquitectura. Este proceso aumenta la cantidad de características tantas veces como *time-lags* utilice la arquitectura, aumentando así el tamaño del dataset.

El último paso del pre-procesamiento consistió en normalizar los datos. Para este paso se utilizaron normalizadores provistos por la librería *Scikit-learn*⁷. Para evitar la fuga de datos (*data leakage*), al momento de preparar los datos de entrenamiento de cada modelo se realizó primero una normalización de los datos de entrenamiento y, con la distribución aprendida por el normalizador, se normalizaron los datos de testeo. El *data leakage* debe ser tenido en cuenta para evaluar la capacidad de generalización de los predictores.

5.3 Selección de Casos de Estudio

Si bien el *dataset StudentLife* cuenta con gran cantidad de usuarios, para los objetivos de nuestro estudio se analizó el comportamiento de tres de ellos. La razón de esto, es que cada usuario posee su propia rutina y características personales, que no tienen porque parecerse a la de los demás usuarios. Por otro lado, algunos usuarios poseen menos datos que otros.

Los usuarios fueron seleccionados de forma tal de representar patrones de gasto energético diferentes. Los usuarios seleccionados fueron: usuario 50 (bajo MET), usuario 31 (MET medio), y usuario 4 (MET alto). Los usuarios seleccionados poseen, además, un diferente grado rutinario en su actividad diaria. Lo anterior puede corroborarse con el cálculo de la correlación entre los gastos energéticos de las mismas horas de los diferentes días de la semana y su desviación estándar. En Fig. 1 y Fig. 2 se muestran el MET promedio y la desviación estándar, respectivamente, a lo largo de los días de la semana y las horas del día de los 3 usuarios seleccionados.

⁷ <https://scikit-learn.org/>

5.4 Resultados

La Tabla 2 muestra los valores del Error Absoluto Medio (MAE) conseguidos para cada arquitectura propuesta. En todos los casos, se utilizaron $\frac{2}{3}$ de los datos (aproximadamente las primeras 6 semanas) para el entrenamiento de los modelos y $\frac{1}{3}$ (aproximadamente las últimas dos semanas) para testeo.

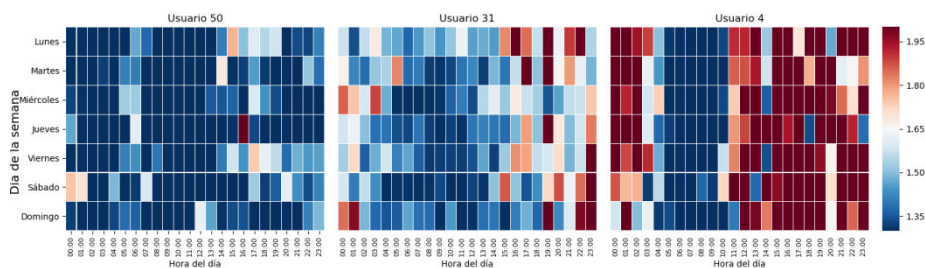


Fig. 1. MET promedio de los tres casos de estudio a lo largo de los días de la semana

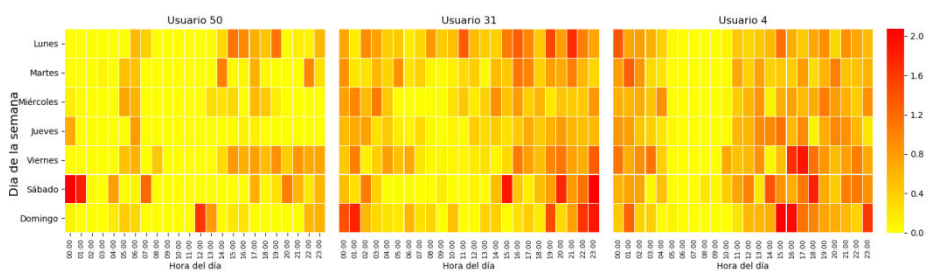


Fig. 2. Desviación estándar del MET de los tres casos de estudio a lo largo de los días de la semana

MAE Test	Lags	Arq. 1	Arq. 2	Arq. 3	Arq. 4
Usuario 50	2	0.102	0.097	0.086	0.109
	4	0.094	0.097	0.113	0.108
	8	0.140	0.090	0.123	0.145
Usuario 31	2	0.306	0.325	0.282	0.285
	4	0.313	0.343	0.296	0.318

	8	0.359	0.331	0.334	0.339
Usuario 4	2	0.528	0.537	0.479	0.529
	4	0.608	0.511	0.496	0.550
	8	0.682	0.530	0.546	0.573

Table 2. Valores de MAE conseguidos para cada arquitectura propuesta

5.5 Discusión

El usuario que presentó mayor desafío para los predictores resultó ser el número 50, ya que fue el que contaba con más casos de comportamiento sedentario. Por lo tanto, resultó difícil que las arquitecturas propuestas aprendieran, a partir de las características del pasado, cuándo iba a tener un comportamiento no sedentario en el futuro. El modelo correspondiente a la arquitectura 3, con 2 lags, fue el que mejor desempeño presentó, con un MAE de 0.086. La Fig. 3 muestra las predicciones de prueba para esta arquitectura.

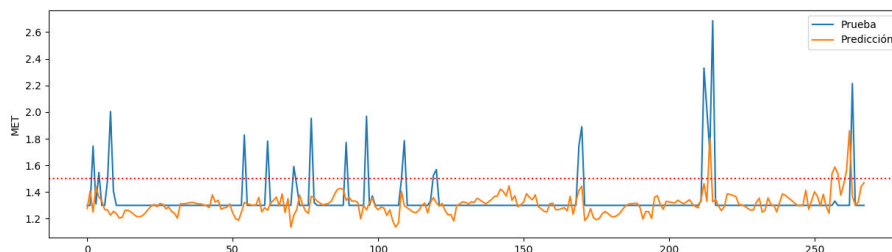


Fig. 3. Predicciones con datos de prueba del usuario 50 con la arquitectura 3 y 2 lags

En el caso del usuario 31, se puede observar un comportamiento más rutinario. Teniendo en cuenta que en el pre-procesamiento se le dio especial atención a las características de tiempo, es entendible que los regresores se desempeñen mejor en estos casos. La arquitectura 3, con 2 lags, fue la que mejor performance tuvo en este caso, con un MAE de prueba de 0.282. La Fig. 4 muestra las predicciones de prueba para esta arquitectura.

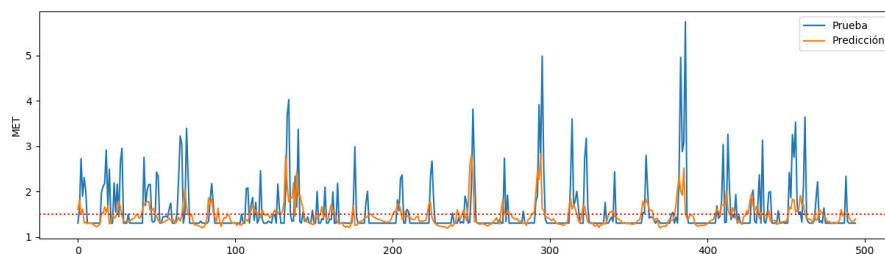


Fig. 4. Predicciones con datos de prueba del usuario 31 con la arquitectura 3

Por último, los modelos sobre el usuario 4 fueron los que mejor desempeño tuvieron, a pesar de tener un mayor MAE que los modelos sobre los otros usuarios. Esto se debe a que este usuario posee muchos más casos de comportamiento no sedentario, por lo que para que el regresor se desempeñe bien debe aprender los patrones de comportamiento del usuario a partir de las características provistas. En el caso del usuario 50, el predictor podría solo aprender a predecir un valor cercano a 1.3 METs para obtener un bajo MAE. La arquitectura que mejor se desempeñó en este caso fue, nuevamente, la número 3, con 2 lags, con un MAE de 0.479. La Fig. 5 muestra las predicciones de entrenamiento y testeo para esta arquitectura.

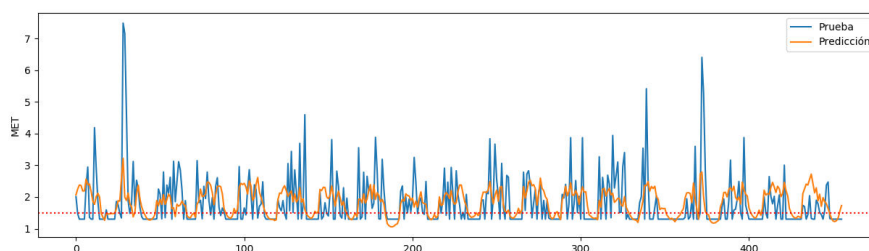


Fig. 5. Predicciones con datos de prueba del usuario 31 con la arquitectura 4

El hecho de que, para todos los modelos, las CNN tuvieron una mejor performance que las RNN, confirma nuestra hipótesis de que las CNN son más adecuadas para problemas de predicción que deban aprender de una secuencia de datos.

6 Conclusiones

En este trabajo nos propusimos comparar diferentes tipos de redes neuronales para evaluar su performance en el problema de la predicción del comportamiento sedentario futuro. Este trabajo es el primero en evaluar el rendimiento de redes neuronales que intenten sacar provecho del carácter secuencial del problema tratado. Se presentaron usuarios con diferentes niveles de gasto energético, lo que hace muy difícil diseñar una arquitectura que tenga una buena performance para todos ellos.

A pesar de haber utilizado un alto grado de regularización, las arquitecturas más complejas son propensas a producir overfitting debido a la poca cantidad de casos de entrenamiento de cada usuario. Este problema podría llegar a ser resuelto aumentando la granularidad en el pre-procesamiento tomando, por ejemplo, intervalos de tiempo de 30 minutos o 15 minutos, resultando así en más casos de entrenamiento.

Otra de las decisiones a enfrentar fue la cantidad de time-lags a tomar. Es lógico pensar que cuanta más información del pasado se le dé a la arquitectura, más información tendrá para hacer mejores predicciones. Esto no es cierto por dos razones. La primera de ellas es que el MET que tendrá un usuario en el tiempo $t+1$ puede no estar relacionado con las características disponibles en el tiempo $t-n$. Al tener pocos casos de entrenamiento, esto puede generar ruido en el modelo. La segunda razón es que, como en muchos casos no se poseen datos sobre todas las horas que duró el experimento, cuantos más time-lags se tengan, mayor es la probabilidad de que un caso de entrenamiento sea descartado. Por ejemplo, si la cantidad de time-lags es m y se tienen los datos del usuario x desde el tiempo t al $t - (m + 1)$, pero no del tiempo $t - m$, el caso de entrenamiento será descartado.

Un punto importante a destacar es que, como los participantes del experimento mantuvieron, por lo general, un comportamiento sedentario, los regresores que intentaban predecir siempre valores cercanos a 1.3 MET fueron descartados, ya que conseguían un MAE bajo pero no resultaban útiles como predictores.

7 Referencias

1. Sedentary Behaviour Research Network. Letter to the editor: standardized use of the terms “sedentary” and “sedentary behaviours.” *Appl Physiol Nutr Metab.* 2012;37: 540–542.
2. Wilmot EG, Edwardson CL, Achana FA, Davies MJ, Gorely T, Gray LJ, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. *Diabetologia.* 2012;55: 2895–2905.
3. Carter S, Hartman Y, Holder S, Thijssen DH, Hopkins ND. Sedentary Behavior and Cardiovascular Disease Risk [Internet]. *Exercise and Sport Sciences Reviews.* 2017. pp. 80–86. doi:10.1249/jes.000000000000106
4. Koster A, Caserotti P, Patel KV, Matthews CE, Berrigan D, Van Domelen DR, et al. Association of sedentary time with mortality independent of moderate to vigorous physical activity. *PLoS One.* 2012;7: e37696.
5. Paing AC, McMillan KA, Kirk AF, Collier A, Hewitt A, Chastin SFM. The associations of sedentary time and breaks in sedentary time with 24-hour glycaemic control in type 2 diabetes [Internet]. *Preventive Medicine Reports.* 2018. pp. 94–100. doi:10.1016/j.pmedr.2018.09.002
6. Felez-Nobrega M, Hillman CH, Dowd KP, Cirera E, Puig-Ribera A. ActivPALTM determined sedentary behaviour, physical activity and academic achievement in college students. *J Sports Sci.* 2018;36: 2311–2316.
7. Magnon V, Vallet GT, Auxiette C. Sedentary Behavior at Work and Cognitive Functioning: A Systematic Review. *Front Public Health.* 2018;6: 239.
8. Falck RS, Davis JC, Liu-Ambrose T. What is the association between sedentary behaviour and cognitive function? A systematic review. *Br J Sports Med.* 2017;51: 800–811.

9. Benatti FB, Ried-Larsen M. The Effects of Breaking up Prolonged Sitting Time: A Review of Experimental Studies. *Med Sci Sports Exerc.* 2015;47: 2053–2061.
10. Atkin AJ, Gorely T, Clemes SA, Yates T, Edwardson C, Brage S, et al. Methods of Measurement in epidemiology: sedentary Behaviour. *Int J Epidemiol.* 2012;41: 1460–1471.
11. He Q, Agu EO. Smartphone usage contexts and sensible patterns as predictors of future sedentary behaviors [Internet]. 2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT). 2016. doi:10.1109/hic.2016.7797695
12. Grundgeiger T, Pichen J, Häfner J, Wallmann-Sperlich B, Löffler D, Huber S. Combating Sedentary Behavior [Internet]. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17. 2017. doi:10.1145/3027063.3053094
13. Fahim M, Baker T, Khattak AM, Alfandi O. Alert me: Enhancing active lifestyle via observing sedentary behavior using mobile sensing systems [Internet]. 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom). 2017. doi:10.1109/healthcom.2017.8210838
14. Ajzen I. The theory of planned behavior [Internet]. *Organizational Behavior and Human Decision Processes.* 1991. pp. 179–211. doi:10.1016/0749-5978(91)90020-t
15. Lecun Y, Bengio Y. Convolutional networks for images, speech, and time series[M]. *The handbook of brain theory and neural networks.* MIT Press, 1998.
16. Palaz, D., Collobert, R., and Magimai-Dos, M. (2013) Estimating Phoneme Class Conditional Probabilities from Raw Speech Signal using Convolutional Neural Networks. INTERSPEECH.
17. Abdel-Hamid, O., Deng, L., Yu, D., and Jiang, H. (2013) Deep Segmental Neural Networks for Speech Recognition. INTERSPEECH.
18. Lee, H., Largman, Y., Pham, P., and Y. Ng, A. (2009) Unsupervised Feature Learning for Audio Classification using Convolutional Deep Belief Networks. NIPS.
19. van den Oord, Aaron, Dieleman, Sander, Zen, Heiga, Simonyan, Karen, Vinyals, Oriol, Graves, Alex, Kalchbrenner, Nal, Senior, Andrew W., and Kavukcuoglu, Koray. WaveNet: A generative model for raw audio. arXiv:1609.03499, 2016
20. Dauphin, Yann N., Fan, Angela, Auli, Michael, and Grangier, David. Language modeling with gated convolutional networks. In ICML, 2017.
21. Kalchbrenner, Nal, Espeholt, Lasse, Simonyan, Karen, van den Oord, Aaron, Graves, Alex, and Kavukcuoglu, Koray. Neural machine translation in linear time. arXiv:1610.10099, 2016A
22. Lea, Colin, Flynn, Michael D., Vidal, Rene, Reiter, Austin, and Hager, Gregory D. Temporal convolutional networks for action segmentation and detection. In CVPR, 2017.
23. Elman, Jeffrey L. Finding structure in time. *Cognitive Science*, 14 (2), 1990.
24. Hermans, Michiel and Schrauwen, Benjamin. Training and analysing deep recurrent neural networks. In NIPS, 2013
25. Bahdanau, Dzmitry, Cho, Kyunghyun, and Bengio, Yoshua. Neural machine translation by jointly learning to align and translate. In ICLR, 2015
26. Pascanu, Razvan, Mikolov, Tomas, and Bengio, Yoshua. On the difficulty of training recurrent neural networks. In ICML, 2013.
27. Hochreiter, Sepp and Schmidhuber, Jurgen. Long short-term memory. *Neural Computation*, 9(8), 1997
28. Cho, Kyunghyun, Van Merriënboer, Bart, Bahdanau, Dzmitry, and Bengio, Yoshua. On the properties of neural machine translation: Encoder-decoder approaches. arXiv:1409.1259, 2014

29. He Q, Agu EO. A frequency domain algorithm to identify recurrent sedentary behaviors from activity time-series data [Internet]. 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). 2016. doi:10.1109/bhi.2016.7455831
30. He Q, Agu EO. Towards sedentary lifestyle prevention: An autoregressive model for predicting sedentary behaviors [Internet]. 2016 10th International Symposium on Medical Information and Communication Technology (ISMICT). 2016. doi:10.1109/ismict.2016.7498879
31. He Q, Agu EO. A Rhythm Analysis-Based Model to Predict Sedentary Behaviors. 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). 2017. doi:10.1109/chase.2017.122
32. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network [Internet]. 2017 International Conference on Engineering and Technology (ICET). 2017. doi:10.1109/icengtechnol.2017.8308186
33. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct. 2014. doi:10.1145/2632048.2632054
34. Cook DJ, Krishnan NC. Activity Learning: Discovering, Recognizing, and Predicting Human Behavior from Sensor Data. John Wiley & Sons; 2015.
35. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *J Med Internet Res*. 2015;17: e175.
36. Bai, Shaojie, Kolter, J. Zico, and Koltun, Vladlen. 2018. "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling". arXiv:1803.01271v2 [cs.LG] 19 Apr 2018