

Universidad de La Habana
Facultad de Matemática y Computación



Predicción de la movilidad humana utilizando redes neuronales

Autor:

Brian Ameht Inclán Quesada

Tutores:

Dr. Milton García Borroto

Dr. Alejandro Lage Castellanos

Trabajo de Diploma
presentado en opción al título de
Licenciado en Ciencia de la Computación

10 de febrero del 2025

Human Mobility Prediction

A mi abuelo Roberto, cuya alegría y amor han sido una guía en mi vida. Aunque ya no estás físicamente, tu legado vive en cada paso que doy. Gracias por inspirarme siempre.

Agradecimientos

A lo largo de este camino de descubrimiento y aprendizaje, cada logro alcanzado ha sido el resultado de un esfuerzo colectivo. Esta tesis es un reflejo no solo de mi trabajo, sino también del apoyo, la guía y el amor de muchas personas a quienes quiero expresar mi más sincero agradecimiento.

En primer lugar, quiero agradecer a mis tutores Milton y Alejandro por guiarme en todo momento, sin ellos esto no pudo ser posible, a todos los profesores que han contribuido a mi formación desde el inicio de mi carrera. Su dedicación y compromiso han sido fundamentales en mi crecimiento académico y personal, permitiéndome superar desafíos y desarrollarme como profesional.

A mi familia, mi más profundo agradecimiento, pues sin su apoyo incondicional, nada de esto habría sido posible. A mi madre, por ser un pilar inquebrantable en mi vida, por su constante aliento y por enseñarme que con esfuerzo y perseverancia se pueden alcanzar los sueños. A mi esposa, Amalia, mi sostén, compañera y amiga, quien ha estado siempre a mi lado, brindándome su amor y apoyo incondicional, siendo clave en mi desempeño y crecimiento personal en esta etapa.

A mi abuela Reina, siempre pendiente de mí, siempre con cariño y sabios consejos a lo largo del camino. A mi abuelo Roberto, cuya alegría y amor han sido una luz en mi vida, gracias por inspirarme siempre. A mi padre, por sus palabras de aliento en los momentos difíciles, por enseñarme a mirar hacia adelante con determinación y nunca rendirme ante los obstáculos. Gracias a mi tía Yaisel, por preocuparse, por querer hacerme mejor persona, a mi prima Jeniffer por la alegría, por la energía, por el cariño. Gracias a Iram por sus consejos, por los debates, por influir en mi vida profesional.

A mis amigos, que son más que amigos, mis hermanos: Mauro, Adrián, Maikol, Brianci, Ariel y Orestes. Con ustedes he compartido mucho más que momentos, hemos construido una hermandad que trasciende el tiempo y la distancia. Gracias por su cariño incondicional, por apoyarme sin dudar, por recordarme siempre quién soy y quién quiero ser. Su confianza en mí ha sido un motor en los momentos de incertidumbre, y su presencia ha sido mi refugio en las buenas y en las malas. Por las interminables conversaciones, las risas, los consejos, las aventuras y hasta los silencios compartidos, les estaré eternamente agradecido. No hay palabras suficientes para

describir lo importante que son en mi vida.

También quiero agradecer a Dariel, Eric, Nahomi y Leo, con quienes compartí no solo largas horas de estudio, sino también risas y momentos inolvidables. Gracias por la compañía y la hermandad que hemos construido en este viaje. A mis amigos del fútbol, por cada aprendizaje, cada victoria y cada derrota, y por ayudarme a fortalecer valores que llevaré conmigo para siempre.

Un agradecimiento especial a todas aquellas personas cercanas a mi familia, aquellos que, sin ser parte del núcleo familiar, han estado pendientes de mí con preocupación genuina y un cariño que se siente como hogar. Su apoyo, sus palabras de aliento y su calidez han sido un bálsamo en este camino, demostrando que la familia no siempre se define por la sangre, sino por el amor y la conexión sincera.

Finalmente, a todos mis compañeros de carrera, quienes han sido parte de esta etapa, compartiendo desafíos y logros.

A todos ustedes, gracias por estar, por creer en mí y por ser parte de este logro.

Opinión del tutor

Como tutores de la tesis de grado de Brian Ameht Inclán Quesada, queremos resaltar su destacada labor en el desarrollo de su investigación sobre la predicción de la movilidad humana mediante redes neuronales. Desde el inicio de su trabajo, Brian ha demostrado una gran iniciativa personal e independencia, abordando un tema complejo y moderno con un enfoque riguroso y creativo. Su proyecto incluye la implementación de diversas arquitecturas neuronales, desde redes simples hasta modelos avanzados como GRU y Transformer-BERT, lo que evidencia su capacidad para manejar técnicas de vanguardia en el campo del aprendizaje profundo. Además, su metodología basada en la segmentación temporal y la exploración de enfoques híbridos demuestra un sólido criterio investigativo y un compromiso con la innovación. En estos meses de trabajo, Brian ha mostrado una actitud proactiva, resolviendo desafíos con autonomía y buscando siempre la mejora de sus modelos. Su capacidad analítica y su disposición para la experimentación han sido claves en el desarrollo de esta investigación, sentando una base sólida para futuras contribuciones en este campo. Consideramos que su desempeño ha sido excelente y que su trabajo representa una contribución valiosa al área de modelización de patrones espacio-temporales con redes neuronales.

Dr. Milton García Borroto Dr. Alejandro Lage Castellanos
Universidad de La Habana

Resumen

En esta tesis se aborda el desafío de la predicción de la movilidad humana mediante el uso de redes neuronales. Se implementaron diferentes arquitecturas, incluyendo una red simple para la predicción de días futuros, una red para estimar patrones semanales y una red GRU para capturar las relaciones temporales complejas en los datos. Además, se exploró el ajuste de modelos preentrenados basados en *Transformer-BERT* para evaluar su aplicabilidad en la modelización de patrones espacio-temporales.

El enfoque seguido se basa en la segmentación temporal de los datos en ciclos semanales y horarios, con predicciones realizadas a nivel individual. La investigación busca explorar diferentes vías para el modelado de estos problemas, sentando las bases para futuras exploraciones con arquitecturas híbridas que combinen redes recurrentes con modelos basados en atención.

Palabras clave: *Predicción de movilidad humana, redes neuronales, GRU, Transformer, BERT, patrones temporales, aprendizaje espacio-temporal.*

Abstract

This thesis addresses the challenge of human mobility prediction using neural networks. Different architectures were implemented, including a simple network for predicting future days, a network to estimate weekly patterns, and a GRU network to capture complex temporal relationships in the data. Additionally, the fine-tuning of pretrained models based on *Transformer-BERT* was explored to assess their applicability in modeling spatiotemporal patterns.

The approach is based on the temporal segmentation of data into weekly and hourly cycles, with predictions made at the individual level. The research seeks to explore different approaches for modeling these problems, laying the foundation for future explorations with hybrid architectures that combine recurrent networks with attention-based models.

Keywords: *Human mobility prediction, neural networks, GRU, Transformer, BERT, temporal patterns, spatiotemporal learning.*

Índice general

Introducción	1
1. Revisión de las Metodologías en Predicción de Movilidad Humana	4
1.1. Métodos Tradicionales de Predicción	5
1.1.1. Modelos Estadísticos	5
1.1.2. Modelos Basados en Grafos	5
1.2. Aprendizaje Automático para la Predicción de Movilidad	6
1.2.1. Redes Neuronales Recurrentes (RNN), LSTM y GRU	6
1.2.2. Redes Neuronales Convolucionales (CNN)	8
1.2.3. Redes Gráficas (GNN)	9
1.2.4. Modelos Híbridos	10
1.3. Avances Recientes y <i>Transformers</i>	11
1.3.1. <i>Transformers</i>	11
1.3.2. Comparación con Modelos Clásicos	13
1.3.3. Tendencias Emergentes	13
1.4. Análisis Comparativo	14
1.4.1. Comparación de Metodologías	14
1.4.2. Identificación de Oportunidades	15
1.5. Discusión	16
2. Propuesta	17
2.1. Visión general	17
2.2. Diseño y Construcción de los Modelos	18
2.2.1. Implementación del baseline	18
2.2.2. Redes neuronales simples	19
2.2.3. Desarrollo de variantes GRU	22
2.2.4. Ajuste y personalización de BERT	26
3. Experimentación y resultados	29
3.1. Marco experimental	29
3.1.1. Métricas de Evaluación	29

3.1.2. Estructura y procesamiento del conjunto de datos	31
3.1.3. Recursos computacionales y hardware utilizado	33
3.1.4. Hiperparámetros	34
3.2. Resultados	35
3.3. Discusión	38
Conclusiones	40
Recomendaciones	42
Bibliografía	44

Índice de figuras

1.1.	Esquema de una celda LSTM (Long Short-Term Memory) [21].	7
1.2.	Arquitectura interna de una celda GRU (Gated Recurrent Units) [4] .	8
1.3.	Esquema de un modelo híbrido CNN-LSTM para predicción espacio-temporal [30].	10
1.4.	La arquitectura general del modelo <i>Transformer</i> [27].	12
2.1.	Arquitectura del modelo <i>vanilla</i> para predecir un día siguiente a partir de la semana previa.	20
2.2.	Arquitectura del modelo <i>vanilla</i> para predecir la semana siguiente a partir de la semana previa.	22
2.3.	Arquitectura del modelo para predecir la semana siguiente a partir de la semana previa basado en GRU.	23
2.4.	Arquitectura del modelo LP-BERT, variante basada en BERT para la predicción de coordenadas futuras a partir de secuencias temporales de movilidad.	28

Introducción

En las últimas décadas, los avances en inteligencia artificial (IA) y aprendizaje automático han transformado profundamente el análisis de grandes volúmenes de datos, proporcionando herramientas innovadoras para resolver problemas complejos. Entre los campos más beneficiados se encuentra la predicción de movilidad humana, una disciplina fundamental para la planificación urbana, la gestión de emergencias y el diseño de ciudades inteligentes. La presente investigación pretende contribuir a este ámbito mediante el desarrollo de modelos avanzados basados en redes neuronales.

El contexto histórico de la movilidad humana muestra una evolución marcada por el aumento en la disponibilidad de datos geoespaciales gracias a dispositivos móviles y sistemas GPS. Sin embargo, esta abundancia de datos plantea desafíos significativos. Entre ellos, la naturaleza espacio-temporal de los datos de movilidad, la irregularidad en los registros y la necesidad de capturar patrones no lineales complejos son obstáculos recurrentes. En particular, el análisis de trayectorias humanas requiere enfoques capaces de manejar interrupciones en los datos y modelar comportamientos tanto regulares como anómalos. Estos desafíos no solo han motivado la investigación científica en el área, sino que también han llevado al desarrollo de metodologías más robustas y versátiles.

La literatura existente demuestra avances notables en el campo, destacando la efectividad de arquitecturas de aprendizaje profundo como las redes neuronales recurrentes (RNN), especialmente las variantes GRU y LSTM, que abordan las dependencias temporales a corto-medio plazo. Estas arquitecturas, más concretamente las LSTM se han combinado con redes convolucionales (CNN) para capturar relaciones espaciales y temporales de manera simultánea [24, 14]. Más recientemente, los modelos basados en *transformers* han surgido como una solución prometedora para manejar datos de movilidad de largo alcance, gracias a su capacidad para procesar relaciones complejas a gran escala [12, 28]. Además, los *transformers* han mostrado ser eficaces en la generalización a diferentes contextos geográficos, lo que los posiciona como una herramienta clave en la investigación futura de movilidad humana.

El problema científico abordado en este trabajo se centra en mejorar la precisión y generalización de los modelos predictivos para movilidad humana en diversos contextos urbanos y rurales. Esto implica desarrollar un enfoque que combine técni-

cas avanzadas de aprendizaje automático con una segmentación temporal detallada, organizando los datos en ciclos semanales y horarios [25]. Tal enfoque permitirá modelar patrones específicos de individuos y abordar las variaciones significativas en las trayectorias humanas.

La hipótesis principal de este estudio es que las arquitecturas RNN como las GRU y una forma correcta en sus configuraciones, y la incorporación de *transformers* mejora la capacidad de predicción de patrones espacio-temporales en datos de movilidad humana.

Objetivo General

Proponer, desarrollar y evaluar una metodología integral para la predicción de patrones de movilidad humana, utilizando modelos basados en redes neuronales recurrentes y *transformers*. Esta metodología busca capturar patrones espacio-temporales complejos, mejorar la precisión y la capacidad de generalización en diversos contextos y facilitar la aplicación de los modelos a distintas áreas urbanas.

La propuesta incluye un componente principal:

- **Desarrollo y evaluación de modelos predictivos avanzados:** Se explorarán enfoques como redes GRU y *transformers*, optimizados para manejar datos secuenciales y abordar desafíos como datos faltantes y patrones temporales irregulares.

Objetivos específicos

- Estudiar la literatura sobre las técnicas empleadas la predicción de patrones espacio-temporales en datos de movilidad humana.
- Desarrollar modelos complementarios para la predicción diaria y semanal.
- Implementar arquitecturas GRU para capturar patrones temporales complejos.
- Evaluar distintas configuraciones de características en la arquitectura GRU para mejorar las predicciones
- Explorar el uso de *transformers* para manejar patrones de largo alcance en datos de movilidad.
- Validar los modelos desarrollados en diferentes escenarios utilizando métricas específicas.

El impacto de este trabajo es significativo, ya que sus contribuciones abarcan aplicaciones prácticas en planificación urbana, gestión de emergencias y personalización de servicios comerciales. Además, aborda cuestiones éticas al garantizar la privacidad y el anonimato de los datos utilizados [19]. También se considera la aplicabilidad de estos modelos en contextos multiculturales, donde las dinámicas de movilidad varían drásticamente según las costumbres locales y las infraestructuras disponibles.

Este documento está estructurado en cuatro capítulos principales. El primer capítulo introduce los conceptos fundamentales y la revisión del estado del arte en predicción de movilidad humana. El segundo capítulo describe la metodología empleada, incluyendo la estructura de los datos y los modelos desarrollados. El tercer capítulo presenta los resultados experimentales y su análisis, y el cuarto capítulo ofrece las conclusiones y propuestas para trabajos futuros.

Capítulo 1

Revisión de las Metodologías en Predicción de Movilidad Humana

La predicción de la movilidad humana se ha convertido en un área de investigación crucial debido a su impacto en múltiples dominios, como la planificación urbana, la gestión del tráfico y la respuesta a desastres. Este campo se beneficia de los avances en tecnologías de recopilación de datos, como los dispositivos GPS y los teléfonos inteligentes, que han facilitado el registro masivo de trayectorias humanas. A pesar de estos avances, la predicción precisa y confiable de movimientos humanos enfrenta desafíos significativos, entre ellos:

- **Variabilidad espaciotemporal:** Los patrones de movilidad pueden variar ampliamente entre individuos y regiones, haciendo difícil generalizar los modelos.
- **Datos incompletos:** Frecuentemente, los datos recolectados contienen lagunas temporales o espaciales que complican los análisis.
- **Generalización en múltiples ciudades:** La capacidad de un modelo para adaptarse a contextos urbanos diversos sigue siendo un reto pendiente.

Este documento explora las principales técnicas y metodologías empleadas en la predicción de movilidad humana, destacando los avances recientes en modelos basados en redes neuronales, *transformers* espaciotemporales y métricas específicas diseñadas para evaluar el rendimiento de estos modelos en escenarios reales.

1.1. Métodos Tradicionales de Predicción

1.1.1. Modelos Estadísticos

Los modelos estadísticos son un enfoque tradicional ampliamente utilizado para la predicción de series temporales. Métodos como ARIMA (*Auto-Regressive Integrated Moving Average*) han demostrado ser efectivos en escenarios con datos estacionales y lineales, mientras que las regresiones lineales y logísticas han sido aplicadas para modelar relaciones directas en conjuntos de datos más simples [25, 24]. Sin embargo, estos modelos presentan limitaciones significativas al enfrentarse a datos espaciotemporales complejos, especialmente debido a su incapacidad para manejar patrones no lineales y relaciones espaciales [19].

1.1.2. Modelos Basados en Grafos

La modelación basada en grafos ha ganado atención en el análisis de movilidad humana debido a su capacidad para representar relaciones espaciales y temporales de manera estructurada. En este enfoque, los nodos representan ubicaciones geográficas, mientras que las aristas reflejan movimientos entre estas ubicaciones, permitiendo un análisis detallado de trayectorias y patrones de movimiento. Algoritmos clásicos como K-Vecinos más Cercanos (KNN) y análisis de centralidad han sido fundamentales para identificar patrones recurrentes y áreas críticas en el transporte urbano [19, 28].

Los principales beneficios de los modelos basados en grafos incluyen su flexibilidad para adaptarse a diferentes contextos y su capacidad para representar conexiones complejas entre ubicaciones. Sin embargo, presentan limitaciones como su alta dependencia en la calidad de los datos y el costo computacional asociado al análisis de grandes redes.

En los últimos años, las Redes Neuronales Gráficas (GNNs) han revolucionado este campo al integrar aprendizaje automático con estructuras de grafos. Estas redes permiten la captura de patrones dinámicos, como fluctuaciones temporales en las trayectorias humanas, ofreciendo un rendimiento notable en tareas como la predicción de tráfico urbano y la planificación de rutas óptimas [14, 12]. Por ejemplo, técnicas modernas como ST-MoE-BERT han demostrado ser especialmente útiles en contextos multiculturales, donde los patrones de movilidad pueden variar drásticamente entre ciudades [14].

A pesar de estos avances, aún existen desafíos pendientes, como la integración de grafos dinámicos que reflejen cambios en tiempo real y la reducción de la complejidad computacional, lo que limita su escalabilidad en aplicaciones masivas.

1.2. Aprendizaje Automático para la Predicción de Movilidad

El aprendizaje automático ha revolucionado la predicción de la movilidad humana al permitir modelar patrones complejos en datos espaciotemporales. A continuación, se detallan las técnicas principales utilizadas en este ámbito, integrando los avances más relevantes de la literatura y los proyectos actuales.

1.2.1. Redes Neuronales Recurrentes (RNN), LSTM y GRU

Las **Redes Neuronales Recurrentes** (RNN) son una clase de redes neuronales diseñadas para procesar datos secuenciales, donde la salida de una neurona en un instante de tiempo se retroalimenta como entrada en el siguiente. Esta característica permite a las RNN capturar dependencias temporales y contextuales en secuencias de datos, siendo especialmente útiles en tareas como el modelado del lenguaje, reconocimiento de voz y predicción de series temporales [15].

Sin embargo, las RNN tradicionales enfrentan desafíos significativos al intentar aprender dependencias a largo plazo debido al problema del desvanecimiento del gradiente. Este fenómeno dificulta la capacidad de la red para ajustar adecuadamente los pesos durante el entrenamiento, limitando su eficacia en secuencias largas.

Para abordar esta limitación, se introdujeron las **Long Short-Term Memory** (LSTM), una variante de las RNN que incorpora una estructura de memoria diseñada para retener información relevante durante períodos prolongados. Las LSTM utilizan una serie de "puertas" (de entrada, olvido y salida) que regulan el flujo de información, permitiendo a la red aprender y recordar dependencias a largo plazo de manera más efectiva [7].

La formulación básica de una LSTM se puede expresar mediante las siguientes ecuaciones:

$$\begin{aligned}
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
 \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\
 C_t &= f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \\
 o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
 h_t &= o_t \cdot \tanh(C_t)
 \end{aligned}$$

Donde: - f_t es la puerta de olvido. - i_t es la puerta de entrada. - \tilde{C}_t es la nueva información candidata para la memoria. - C_t es el estado de la memoria en el tiempo t . - o_t es la puerta de salida. - h_t es el estado oculto en el tiempo t . - x_t es la entrada

en el tiempo t . - W y b son los pesos y sesgos aprendidos durante el entrenamiento. - σ representa la función sigmoide y \tanh la tangente hiperbólica.

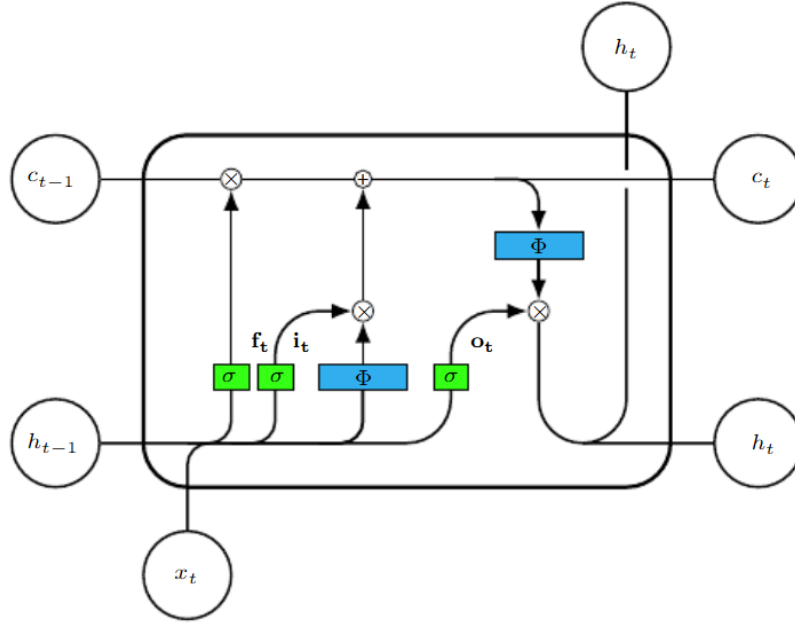


Figura 1.1: Esquema de una celda LSTM (Long Short-Term Memory) [21].

Las LSTM han demostrado ser altamente efectivas en la predicción de trayectorias humanas, permitiendo capturar patrones temporales complejos y mejorar la precisión en escenarios dinámicos. Por ejemplo, se han empleado utilizando características periódicas y de decaimiento temporal, logrando mejoras significativas en la precisión de las predicciones [24].

Otra variante destacada de las RNN son las **Gated Recurrent Units** (GRU), diseñadas como una simplificación de las LSTM para manejar eficientemente dependencias temporales en secuencias de datos. Las GRU combinan las funcionalidades de las puertas de entrada y olvido en una sola, denominada puerta de actualización, y prescinden de un estado de memoria independiente. Esto resulta en una arquitectura más ligera, que reduce la complejidad computacional manteniendo un rendimiento competitivo en muchas tareas [6].

El funcionamiento de una GRU se describe mediante las siguientes ecuaciones:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h)$$

Donde:

- z_t es la puerta de actualización, que controla cuánto de la información pasada se transfiere al estado actual.
- r_t es la puerta de reinicio, que regula cuánto de la información previa debe ignorarse.
- h_t representa el estado oculto en el tiempo t .
- W, U, b son los parámetros aprendidos de la red.
- σ y \tanh son funciones de activación sigmoide y tangente hiperbólica, respectivamente.

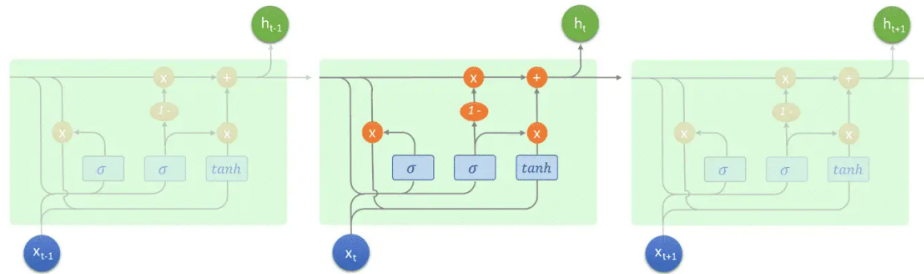


Figura 1.2: Arquitectura interna de una celda GRU (Gated Recurrent Units) [4]

Las GRU han demostrado ser especialmente útiles en aplicaciones donde la capacidad de modelar dependencias a largo plazo es crítica, pero los recursos computacionales son limitados. En particular, han sido empleadas con éxito en tareas de predicción de movilidad humana, donde es necesario capturar patrones temporales en datos espaciotemporales. Su simplicidad y eficiencia las hacen adecuadas para escenarios con datos ruidosos o incompletos, como trayectorias con puntos faltantes o datos irregulares [19].

1.2.2. Redes Neuronales Convolucionales (CNN)

Las **Redes Neuronales Convolucionales** (CNN) son una arquitectura de redes neuronales profundas diseñada para procesar datos con una estructura de cuadrícula,

como imágenes. Su capacidad para capturar características espaciales las hace especialmente útiles en tareas de visión por computadora, como el reconocimiento de objetos y la clasificación de imágenes [18].

En el contexto de la predicción de movilidad, las CNN han sido adaptadas para analizar patrones espaciales en datos representados como mapas de calor o matrices de adyacencia. Al aplicar filtros convolucionales, las CNN pueden identificar características locales y relaciones espaciales que son cruciales para comprender la dinámica de movimiento en un entorno determinado.

Una operación típica de convolución en una CNN se define como:

$$y_{ij} = f \left(\sum_{m,n} x_{i+m,j+n} \cdot w_{mn} + b \right)$$

Donde: - y_{ij} es el valor de la salida en la posición (i, j) . - x es la entrada (por ejemplo, una imagen o un mapa de calor). - w es el filtro (o kernel) aplicado. - b es el sesgo. - f es la función de activación, como la ReLU (Unidad Lineal Rectificada).

Estas arquitecturas son particularmente útiles en escenarios donde la dinámica espacial juega un rol crucial, como la identificación de rutas de alta actividad [19, 14]. Además, en la predicción de flujos de tráfico, las CNN pueden aprender patrones de congestión en diferentes áreas de una ciudad, lo que permite anticipar y gestionar mejor el tráfico vehicular [29].

También pueden combinarse con otras arquitecturas, como LSTM, para modelar datos espaciotemporales. En estas combinaciones, las CNN se utilizan para extraer características espaciales, mientras que las LSTM se encargan de capturar dependencias temporales.

1.2.3. Redes Gráficas (GNN)

Las **Redes Neuronales Gráficas** (GNN) han emergido como herramientas poderosas para modelar relaciones espaciales y temporales en datos de movilidad. Estas redes representan trayectorias como grafos, donde los nodos son ubicaciones y las aristas representan movimientos. Una operación clave es la agregación de características de los vecinos de un nodo:

$$h_v^{(k+1)} = \sigma \left(W \cdot \text{AGG} \left(\{h_u^{(k)} : u \in N(v)\} \right) \right)$$

donde: - $h_v^{(k+1)}$ es la representación del nodo v en la iteración $k+1$. - $N(v)$ representa el conjunto de nodos vecinos de v . - AGG es una función de agregación, como suma o promedio.

Las GNN se han aplicado con éxito en la predicción de movilidad humana y tráfico, permitiendo modelar relaciones espaciales y temporales complejas. Por ejemplo, se

han utilizado GNN para predecir flujos de movilidad a nivel nacional en España, integrando datos de viajes y condiciones meteorológicas [10]. Además, las GNN han mejorado la precisión en la predicción de flujos de tráfico al capturar dependencias espaciales y temporales en sistemas de transporte [16]. También se han empleado en la predicción de movimientos humanos en espacios compartidos, facilitando la navegación de robots en entornos donde interactúan con personas [13].

1.2.4. Modelos Híbridos

Los **modelos híbridos** han demostrado ser particularmente efectivos al combinar diversas arquitecturas para capturar patrones complejos en datos espaciotemporales. Una integración común es la de redes convolucionales (CNN) con redes LSTM, donde las CNN se encargan de extraer características espaciales mientras que las LSTM modelan las dependencias temporales. Este enfoque permite analizar simultáneamente las relaciones entre ubicaciones geográficas y la evolución de patrones en el tiempo [25, 24].

Por ejemplo, un modelo híbrido basado en una arquitectura CNN-LSTM ha sido propuesto para predecir flujos de tráfico en entornos urbanos [30]. En este enfoque, las CNN extraen mapas de características a partir de datos históricos de tráfico, mientras que las LSTM procesan estas características para capturar dinámicas temporales. Este modelo logró mejoras significativas en precisión al incorporar información contextual, como patrones de tráfico horarios y estacionales. La Figura 1.3 ilustra este esquema híbrido.

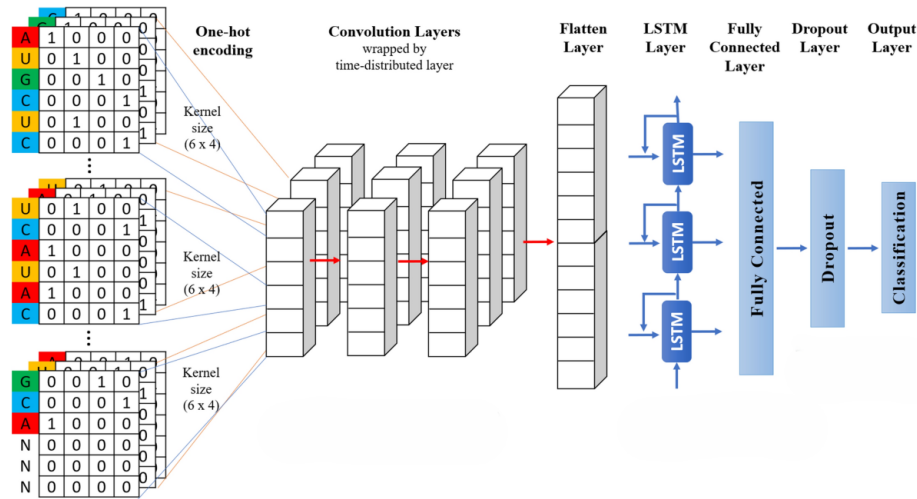


Figura 1.3: Esquema de un modelo híbrido CNN-LSTM para predicción espacio-temporal [30].

Además, se han explorado arquitecturas más avanzadas que incorporan redes neuronales gráficas (GNN) y *transformers*, extendiendo las capacidades de los modelos híbridos. Estas arquitecturas aprovechan la representación estructurada de las GNN para modelar relaciones espaciales complejas entre ubicaciones y las capacidades de atención de los *transformers* para identificar patrones de largo alcance [14, 12]. Por ejemplo, el modelo ST-MoE-BERT utiliza una combinación de expertos espaciales y temporales para mejorar la generalización en diversos escenarios culturales y geográficos [14].

La integración de arquitecturas híbridas no solo mejora la precisión de las predicciones, sino que también permite capturar características únicas de los datos de movilidad, como anomalías y patrones recurrentes. Estos avances hacen que los modelos híbridos sean esenciales para aplicaciones como la planificación de tráfico, la gestión de emergencias y el diseño de sistemas de transporte eficientes [28].

1.3. Avances Recientes y *Transformers*

1.3.1. *Transformers*

Los *transformers* han revolucionado el campo del aprendizaje profundo debido a su capacidad para modelar dependencias a largo plazo en datos secuenciales y estructurados. Introducidos originalmente para tareas de procesamiento del lenguaje natural (NLP), su arquitectura basada en mecanismos de atención ha demostrado ser altamente eficaz en una variedad de dominios, incluyendo la predicción de movilidad humana [27, 14].

A diferencia de las RNN y LSTM, que procesan datos de manera secuencial, los *transformers* procesan las secuencias de datos en paralelo, lo que reduce significativamente los tiempos de entrenamiento. El núcleo de los *transformers* es el mecanismo de atención, que permite al modelo asignar diferentes pesos a cada elemento de la secuencia, enfocándose en las partes más relevantes para la tarea en cuestión. Esto se define mediante las siguientes fórmulas:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Donde: - Q (query), K (key) y V (value) son matrices derivadas de la entrada.
- d_k es la dimensionalidad de las claves. - La función softmax asegura que los pesos asignados sumen 1, lo cual se puede interpretar como probabilidades.

En el contexto de la predicción de movilidad, los *transformers* han sido aplicados para modelar relaciones complejas entre ubicaciones geográficas y temporalidades, como lo demuestra el modelo ST-MoE-BERT [14]. Este modelo combina expertos espaciales y temporales, utilizando el mecanismo de atención para identificar patrones

recurrentes en datos multicapa de trayectorias humanas. La Figura 1.4 ilustra el esquema general de un *transformer* el cual se puede adaptar para tareas de movilidad humana.

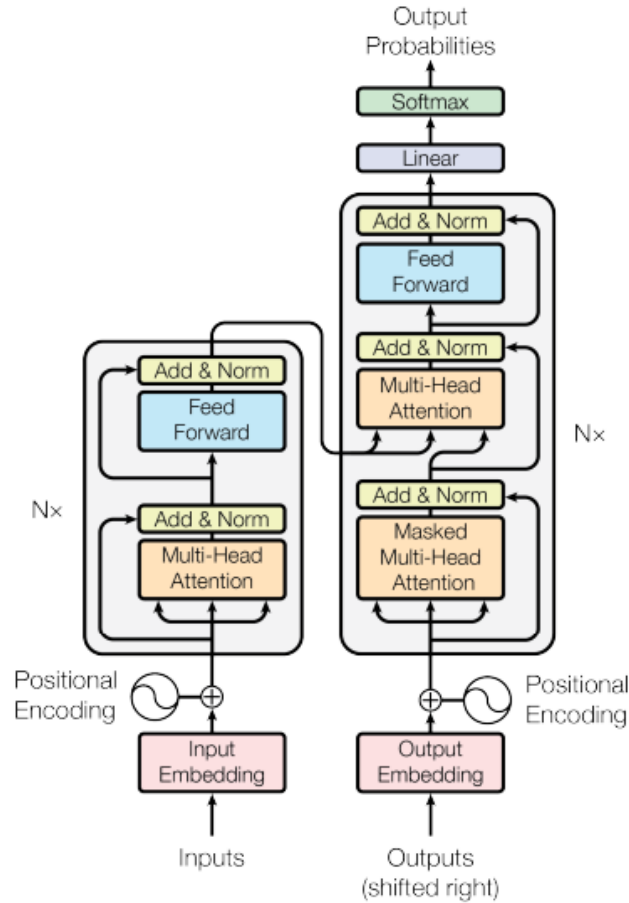


Figura 1.4: La arquitectura general del modelo *Transformer* [27].

Los *transformers* también permiten la integración de datos heterogéneos, como datos meteorológicos, eventos locales y patrones históricos, lo que los hace ideales para abordar los desafíos de predicción en contextos de alta variabilidad [12, 28]. Su capacidad para manejar secuencias de gran longitud y múltiples fuentes de información ha demostrado ser un punto clave para superar las limitaciones de los modelos clásicos basados en RNN y CNN.

Este enfoque representa una evolución significativa en el modelado de patrones espacio-temporales, ampliando las capacidades predictivas y permitiendo un análisis más robusto y generalizado.

1.3.2. Comparación con Modelos Clásicos

Los *transformers* han demostrado ser superiores a los modelos clásicos en diversos aspectos clave, especialmente en tareas que involucran datos espacio-temporales complejos. A diferencia de las redes neuronales recurrentes (RNN) y sus variantes, como LSTM y GRU, que procesan secuencias de datos de manera secuencial, como dijimos anteriormente, los *transformers* operan de forma paralela, lo que reduce significativamente los tiempos de entrenamiento y aumenta la capacidad de capturar relaciones a largo plazo [27]. Este enfoque no solo optimiza el rendimiento computacional, sino que también mejora la capacidad del modelo para manejar secuencias largas, un desafío persistente en las RNN debido al problema del desvanecimiento del gradiente [26].

En comparación con las redes neuronales convolucionales (CNN), ampliamente utilizadas para capturar patrones espaciales locales, los *transformers* han demostrado una capacidad superior para modelar relaciones globales en los datos. Las CNN, debido a su estructura basada en filtros convolucionales con campos receptivos limitados, son altamente efectivas para identificar características locales. Sin embargo, esta limitación en la ventana receptiva puede dificultar la detección de interacciones a gran escala, especialmente en tareas que requieren una comprensión contextual más amplia [17].

Aunque los modelos clásicos siguen siendo relevantes en escenarios específicos, los *transformers* representan una evolución significativa al abordar limitaciones clave de las RNN y CNN, posicionándose como la arquitectura preferida para tareas que demandan precisión y escalabilidad en datos heterogéneos y espacio-temporales.

1.3.3. Tendencias Emergentes

Este campo continúa evolucionando rápidamente, impulsado por los avances en arquitecturas de aprendizaje profundo como los *transformers* y su aplicación a datos espacio-temporales. Una de las tendencias más notables es la integración de *transformers* con arquitecturas híbridas, como redes neuronales gráficas (GNN), para capturar tanto relaciones espaciales como temporales en los datos. Este enfoque permite modelar interacciones complejas y dinámicas, especialmente en contextos donde las relaciones espaciales varían con el tiempo [11].

Modelos avanzados, como el *Spacetimeformer*, representan una evolución significativa al introducir un mecanismo de atención espacio-temporal capaz de aprender relaciones dinámicas entre variables y momentos temporales. A diferencia de las arquitecturas tradicionales que alternan entre capas espaciales y temporales, el *Spacetimeformer* crea un gráfico dinámico donde cada variable y cada instante temporal son tratados como nodos independientes, permitiendo conexiones directas entre ellos. Este enfoque mejora la precisión en tareas como la predicción de tráfico, la demanda eléctrica y el flujo de pasajeros [11].

Además, los *transformers* han demostrado ser herramientas efectivas para manejar datos heterogéneos. Modelos como el *ST-MoE-BERT* integran múltiples fuentes de datos, incluidas condiciones meteorológicas, eventos locales y patrones históricos de movilidad, logrando una representación más rica y contextualizada de los datos [14]. Estos avances destacan por su capacidad para superar las limitaciones de los modelos clásicos en escenarios con alta variabilidad.

Otra tendencia emergente es el desarrollo de *transformers* optimizados para secuencias largas. Técnicas como la atención eficiente y los enfoques híbridos, que combinan *transformers* con capas convolucionales, han permitido ampliar el alcance de estos modelos a problemas de alta dimensionalidad y larga duración, como la predicción espacio-temporal a nivel de ciudades completas [11]. Estas técnicas no solo mejoran el rendimiento computacional, sino que también reducen los requisitos de memoria, haciéndolos más accesibles para aplicaciones en tiempo real.

1.4. Análisis Comparativo

El análisis comparativo en el campo de la predicción de movilidad humana permite identificar fortalezas y debilidades en las metodologías existentes, así como proponer enfoques que aborden las brechas actuales.

1.4.1. Comparación de Metodologías

Las metodologías para la predicción de movilidad humana han evolucionado significativamente, incorporando desde modelos clásicos hasta arquitecturas avanzadas basadas en aprendizaje profundo. Las fortalezas y debilidades de estas metodologías pueden resumirse de la siguiente manera:

Tabla 1.1: Comparación de metodologías para la predicción de movilidad humana

Metodología	Fortalezas	Debilidades
Modelos Clásicos (ARIMA, Regresión)	Simplicidad, bajo costo computacional, eficacia en datos lineales y estables.	Limitaciones para manejar relaciones no lineales y datos heterogéneos [1].
Redes Neuronales Recurrentes (LSTM, GRU)	Capturan dependencias temporales y patrones recurrentes.	Problemas con secuencias largas (desvanecimiento del gradiente) y altas demandas computacionales [26].
<i>Transformers</i>	Procesamiento paralelo, atención global, integración de datos heterogéneos.	Requieren grandes volúmenes de datos y alto costo computacional [27].
Modelos Híbridos (CNN-LSTM)	Combinan ventajas de diferentes arquitecturas, capturan relaciones espaciales y temporales complejas.	Complejidad arquitectónica y mayor costo de implementación [14].

La elección de la metodología adecuada depende del caso de uso, las características de los datos y los recursos disponibles. Sin embargo, los modelos híbridos y basados en *transformers* han demostrado ser los más prometedores en escenarios de alta complejidad.

1.4.2. Identificación de Oportunidades

A pesar de los avances logrados, existen brechas importantes que deben abordarse para mejorar la precisión y aplicabilidad de los modelos actuales:

1. **Adaptación a Escenarios con Datos Limitados:** Muchas metodologías, especialmente los *transformers*, dependen de grandes volúmenes de datos para alcanzar un rendimiento óptimo. Una oportunidad clave es el desarrollo de modelos preentrenados específicos para movilidad humana, similares a los que se usan en el procesamiento del lenguaje natural (NLP) [12].
2. **Integración de Datos Contextuales:** La incorporación de datos adicionales, como condiciones meteorológicas, eventos locales y flujos de tráfico en tiempo real, podría mejorar significativamente la precisión de las predicciones. Modelos como ST-MoE-BERT ya han demostrado los beneficios de integrar múltiples fuentes de datos [14].
3. **Modelos más Livianos y Escalables:** Existe una necesidad de desarrollar modelos más eficientes en términos de memoria y cómputo para aplicaciones en tiempo real y dispositivos con recursos limitados. Las variantes ligeras de *transformers*, como los *transformers* eficientes, representan una solución prometedora [27].
4. **Predicción Multiciudad y Transferencia de Conocimiento:** Los enfoques actuales a menudo se limitan a contextos específicos. La implementación de técnicas de aprendizaje por transferencia para generalizar entre ciudades con infraestructuras de datos variadas es una dirección emergente [11].
5. **Aspectos Éticos y de Privacidad:** Garantizar el anonimato y la protección de datos en modelos predictivos es fundamental para aumentar la aceptación social de estas tecnologías. Las técnicas de privacidad diferencial y los sistemas de anonimización deben integrarse como parte estándar de los *pipelines* de aprendizaje automático [28].

Estas oportunidades identificadas apuntan a la necesidad de desarrollar modelos más adaptables, eficientes y éticos, mientras se integran múltiples fuentes de datos y se mejoran las capacidades de generalización a nuevos contextos. Estos avances

no solo aumentarían la precisión predictiva, sino que también ampliarían el impacto práctico de las tecnologías de predicción de movilidad humana.

1.5. Discusión

Todo el análisis realizado refleja una transición significativa desde los modelos clásicos hacia métodos basados en aprendizaje profundo, impulsados por la creciente disponibilidad de datos espacio-temporales y avances tecnológicos. Los modelos clásicos, como ARIMA y regresiones lineales, aunque eficaces en contextos con datos estables y lineales, presentan limitaciones evidentes al enfrentarse a relaciones no lineales y datos heterogéneos. Este vacío ha sido abordado por arquitecturas más avanzadas, como las redes neuronales recurrentes (RNN) y sus variantes LSTM y GRU, que permiten modelar patrones temporales más complejos. No obstante, estas arquitecturas enfrentan desafíos inherentes, como el manejo de secuencias largas y altos costos computacionales.

En este contexto, los *transformers* han emergido como una solución dominante debido a su capacidad para procesar secuencias largas de manera paralela y capturar relaciones globales mediante mecanismos de atención. Su capacidad para integrar múltiples fuentes de datos y generalizar a diversos contextos multiculturales ha sido demostrada en modelos como ST-MoE-BERT y *Spacetimeformer*, que han redefinido el alcance de las aplicaciones predictivas en movilidad humana. Sin embargo, estas arquitecturas presentan una alta dependencia de grandes volúmenes de datos y recursos computacionales, lo que limita su aplicabilidad en escenarios con recursos limitados.

Además, los modelos híbridos, que combinan arquitecturas como CNN, LSTM, GNN y *transformers*, también han demostrado ser eficaces para capturar relaciones espaciales y temporales complejas de manera simultánea. Aunque prometedores, su complejidad arquitectónica y sus altos costos de implementación plantean desafíos significativos para su adopción generalizada. Una dirección emergente en la investigación es el desarrollo de variantes más ligeras y escalables que mantengan la precisión predictiva mientras optimizan el uso de recursos.

En conclusión, el panorama actual muestra avances significativos en precisión, adaptabilidad y manejo de datos dinámicos. Sin embargo, persisten retos importantes, como la dependencia de datos extensos, la necesidad de optimizar arquitecturas complejas y la integración de datos contextuales adicionales. Estos desafíos abren nuevas oportunidades para la investigación y el desarrollo, con el objetivo de maximizar el impacto práctico de estas tecnologías en la planificación urbana, la gestión de tráfico y otras aplicaciones críticas en la sociedad moderna.

Capítulo 2

Propuesta

En este capítulo se presenta la metodología para la predicción de la movilidad humana, con el objetivo de abordar las complejidades y patrones inherentes en los datos espaciales y temporales. La Sección 2.1 proporciona una visión general de la propuesta, explicando las bases teóricas y conceptuales que guían el enfoque. En la Sección 2.3 se detallan las tareas específicas realizadas durante el desarrollo, como la construcción y validación de múltiples modelos predictivos.

2.1. Visión general

En esta propuesta se desarrolló un esquema para analizar cómo los resultados de predicción de movilidad humana pueden mejorar o variar considerablemente con diferentes modelos y configuraciones. Inicialmente, se diseñó un modelo *baseline*, el cual consiste en utilizar los datos de una semana previa y proyectarlos directamente en la semana siguiente, partiendo del principio de que los patrones semanales de movilidad suelen ser recurrentes. Este modelo *baseline* es una buena aproximación inicial, ya que proporciona un estándar de desempeño que los demás modelos deben intentar superar. Sin embargo, no siempre es trivial mejorar el rendimiento del *baseline*, pues en ciertos casos los patrones de movilidad son tan predecibles que resulta difícil obtener mejores resultados.

Posteriormente, se desarrollaron modelos basados en redes neuronales *vanilla*. El primer modelo predictivo utiliza los datos de una semana previa para estimar la movilidad de un día siguiente, mientras que el segundo modelo *vanilla* predice toda la semana siguiente a partir de los datos de la semana anterior. Estos enfoques iniciales establecen una base para explorar la capacidad de las redes neuronales simples al abordar el problema de predicción.

A continuación, se implementaron modelos basados en arquitecturas de series temporales, específicamente utilizando redes GRU (*Gated Recurrent Unit*). Estas ar-

arquitecturas son especialmente adecuadas para manejar secuencias temporales debido a su capacidad para capturar dependencias a largo plazo de manera más eficiente que las redes *vanilla*. Se diseñaron cuatro variantes del modelo GRU, todas enfocadas en predecir una semana completa a partir de la anterior, pero con pequeñas modificaciones en las características de entrada utilizadas en cada caso. Esto permitió evaluar cómo diferentes configuraciones de entrada pueden impactar en el desempeño del modelo.

Finalmente, se realizó un ajuste (*fine-tuning*) de un modelo *transformer* preentrenado, conocido como BERT, adaptando sus entradas y salidas al problema de predicción de movilidad humana. La elección de utilizar un *transformer* después de las GRU se debe a que los *transformer*, gracias a su capacidad de atención, son capaces de capturar relaciones complejas en los datos y manejar de manera efectiva dependencias a largo plazo. Este modelo representa la etapa más avanzada, integrando las lecciones aprendidas en los pasos previos y ofreciendo un enfoque generalizable para resolver el problema planteado.

2.2. Diseño y Construcción de los Modelos

Esta sección detalla las actividades clave del desarrollo de las propuestas realizadas y las arquitecturas que conforman los modelos de predicción.

2.2.1. Implementación del baseline

El modelo *baseline* se implementó como una referencia inicial, con el objetivo de establecer un estándar mínimo de desempeño para los modelos más complejos. Este modelo se basó en la hipótesis de que los patrones de movilidad humana tienden a repetirse semanalmente, utilizando los datos de una semana anterior para predecir el comportamiento en la semana siguiente.

El procedimiento seguido incluyó los siguientes pasos:

- Limitar los datos a un conjunto específico de días, definiendo un periodo de entrenamiento y validación.
- Generar predicciones replicando los datos correspondientes al mismo día de la semana anterior.
- Completar valores faltantes en las predicciones utilizando interpolación y extrapolación para garantizar trayectorias consistentes. Aunque este enfoque puede introducir errores debido a la interpolación, en este caso no es relevante debido a la naturaleza propia del “modelo”, basado en patrones semanales repetitivos.

- Validar las predicciones asegurando la coherencia y calidad de los resultados obtenidos.

La simplicidad de este modelo *baseline* proporciona una solución inicial que permite establecer un punto de comparación para modelos más avanzados. Aunque no captura patrones complejos, es particularmente efectivo en contextos donde los patrones semanales son consistentes y predecibles.

2.2.2. Redes neuronales simples

Predicción del día siguiente a partir de la semana previa

El primer modelo *vanilla* desarrollado tiene como objetivo predecir las coordenadas de movilidad humana para un día siguiente, utilizando como entrada las secuencias temporales de la semana previa. Este modelo está diseñado para capturar patrones semanales de movilidad, aprovechando la recurrencia y regularidad de las actividades de las personas.

El modelo sigue los siguientes pasos principales:

- **Preparación de los datos:** Se generaron secuencias de 7 días consecutivos por usuario a partir de los datos disponibles, garantizando que cada día tuviera 48 *timeslots*. Los días con datos faltantes fueron completados utilizando ceros para mantener la consistencia en las dimensiones de entrada.
- **Normalización:** Los datos fueron normalizados entre 0 y 1 para asegurar una mejor convergencia durante el entrenamiento del modelo.
- **Máscaras para valores faltantes:** Durante el entrenamiento, se utilizó una máscara para ignorar los valores faltantes en las secuencias objetivo. Esto permitió que el modelo se enfocara únicamente en los datos válidos y evitara que los valores incompletos afectaran negativamente la actualización de los parámetros.
- **Arquitectura del modelo:** Se utilizó una red neuronal *fully connected* con capas densas, diseñada para procesar secuencias de entrada y mapearlas a las coordenadas objetivo del día siguiente. La arquitectura incluyó mecanismos de regularización, como *dropout*, para evitar el sobreajuste.
- **Entrenamiento:** El modelo fue entrenado utilizando un criterio de pérdida diseñado para ignorar las coordenadas faltantes en las secuencias objetivo. Se evaluó el desempeño en un conjunto de validación separado, midiendo la capacidad del modelo para generalizar a datos no vistos.

Este modelo representa un enfoque simple pero efectivo para aprovechar la estructura temporal de los datos de movilidad humana. Si bien no considera dependencias más complejas, establece una base inicial sólida para comprender cómo los patrones semanales influyen en las predicciones a corto plazo.

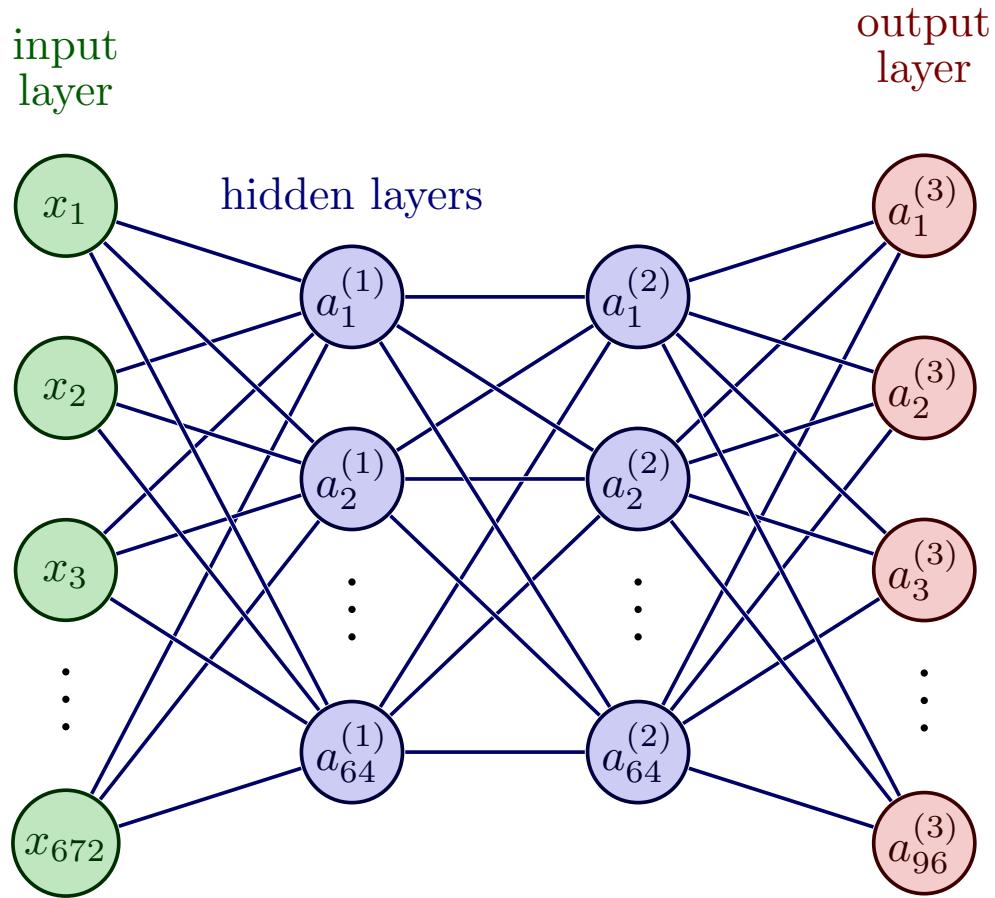


Figura 2.1: Arquitectura del modelo *vanilla* para predecir un día siguiente a partir de la semana previa.

Predicción de 7 días a partir de los 7 días previos

Este modelo está diseñado para predecir la movilidad humana durante 7 días consecutivos, utilizando como entrada los 7 días previos de datos históricos. Se emplearon todas las características disponibles en los datos (uid , d , t , x , y) para capturar patrones temporales más complejos.

El modelo sigue los siguientes pasos principales:

- **Preparación de los datos:** Los datos se dividieron en conjuntos de entrenamiento, validación y prueba, asegurando que los días de entrada y objetivo fueran contiguos y no se solaparan entre los conjuntos. Cada usuario tiene secuencias completas generadas mediante la imputación de valores faltantes con un valor de *padding* definido.
- **Normalización:** Las características de entrada (d , t , x , y) fueron normalizadas utilizando un *scaler* global, ajustado con todos los datos disponibles, para mejorar la convergencia durante el entrenamiento del modelo.
- **Máscaras para valores faltantes:** Se generaron máscaras para las secuencias de entrada y objetivo, identificando los valores válidos que no contenían *padding*. Estas máscaras fueron utilizadas tanto para la normalización como para la pérdida durante el entrenamiento.
- **Arquitectura del modelo:** Se implementó una red neuronal densa con tres capas ocultas, cada una con 128 neuronas y activaciones ReLU. Además, se aplicó *dropout* en cada capa para reducir el riesgo de sobreajuste. La última capa está diseñada para mapear las secuencias de entrada en coordenadas objetivo (x , y) de 7 días consecutivos.
- **Entrenamiento:** El modelo fue entrenado utilizando un criterio de pérdida que ignora los valores con *padding*, permitiendo al modelo enfocarse únicamente en los datos válidos. Se utilizó un conjunto de validación para monitorear el desempeño y seleccionar el mejor modelo basado en la pérdida de validación.

La combinación de máscaras y regularización garantiza que el modelo sea robusto frente a datos incompletos y reduzca el sobreajuste.

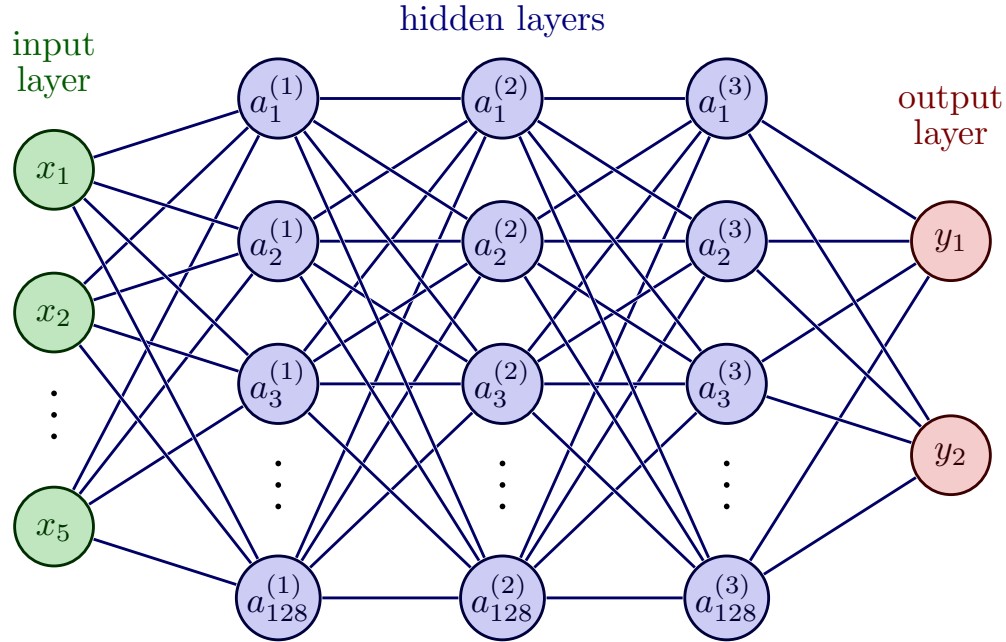


Figura 2.2: Arquitectura del modelo *vanilla* para predecir la semana siguiente a partir de la semana previa.

2.2.3. Desarrollo de variantes GRU

En esta sección se describe el modelo base GRU utilizado para todas las variantes desarrolladas, así como las diferencias específicas de cada variante en las características de entrada y su justificación.

Modelo base GRU

- **Capa de entrada:** La dimensión de entrada varía según las características utilizadas en cada variante, pero siempre corresponde al número de características por *timeslot* (e.g., x , y , t , etc.).
- **Capas GRU:** Cinco capas recurrentes, cada una con un tamaño de estado oculto especificado (e.g., 128 o 256). Estas capas están diseñadas para capturar patrones temporales complejos y mantener dependencias a largo plazo entre los días y los *timeslots*.
- **Capa completamente conectada (*fully connected*):** Esta capa transforma la salida del estado oculto de las GRU (e.g., dimensión 128 o 256) en predicciones de coordenadas x y y para cada *timeslot*. La salida final tiene siempre

una dimensión de $7 \times 48 \times 2$, correspondiente a las predicciones para 7 días consecutivos, 48 timeslots por día y 2 coordenadas por *timeslot*.

- **Regularización:** *Dropout* con una probabilidad de 0.2 aplicado después de cada capa GRU para reducir el riesgo de sobreajuste y garantizar una mejor generalización.

El modelo fue entrenado utilizando el Error Cuadrático Medio (MSE), modificado para ignorar los valores con *padding*. Este criterio de pérdida permite que el modelo se enfoque únicamente en los valores válidos de las secuencias de entrada y objetivo. Además, se utilizó el optimizador Adam con una tasa de aprendizaje inicial de 0.001, y el entrenamiento se realizó durante múltiples épocas, con evaluación en un conjunto de validación para seleccionar la mejor versión del modelo.

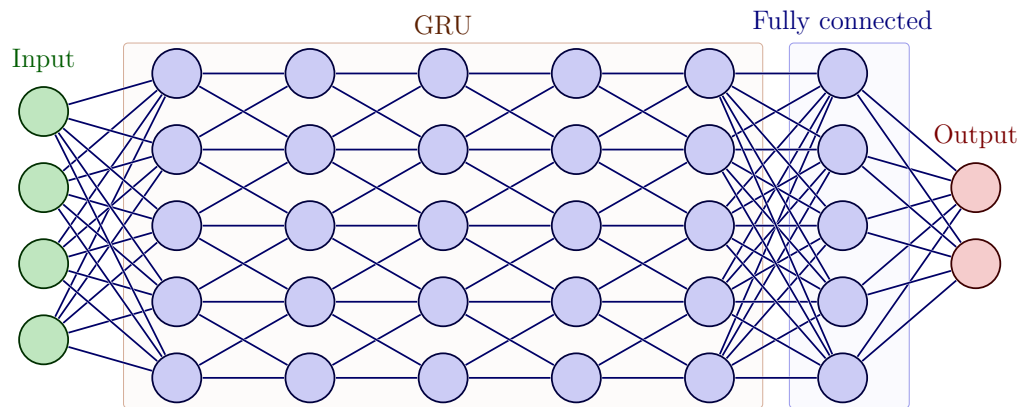


Figura 2.3: Arquitectura del modelo para predecir la semana siguiente a partir de la semana previa basado en GRU.

Primera variante: GRU basada en características x , y , t

La primera variante utiliza las características x , y , y t como entrada, lo que reduce la dimensionalidad inicial del modelo y se centra en capturar las relaciones espacio-temporales más directas en los datos.

Diferencias específicas de esta variante

- **Características de entrada:** x , y , y t , con una dimensión de entrada total de 3.
- **Tamaño del estado oculto:** 128 neuronas en cada capa GRU, optimizado para manejar esta combinación de características.

- **Justificación:** Estas características proporcionan una representación mínima pero suficiente de la posición y el contexto temporal, lo que permite al modelo enfocarse en patrones espaciales y temporales esenciales. Sin embargo, esta variante no toma en cuenta el contexto semanal de los datos, lo que podría limitar su capacidad para capturar patrones recurrentes asociados a días específicos, como diferencias entre días laborables y fines de semana.

Segunda variante: GRU basada en características x , y , t , day_of_week

La segunda variante amplía las características de entrada para incluir el día de la semana, lo que permite al modelo capturar patrones semanales recurrentes además de las relaciones espacio-temporales básicas.

Diferencias específicas de esta variante

- **Características de entrada:** x , y , t , y day_of_week , con una dimensión de entrada total de 4.
- **Tamaño del estado oculto:** 256 neuronas en cada capa GRU, ajustado para manejar el aumento en la dimensionalidad de entrada.
- **Justificación:** La incorporación de day_of_week en el modelo tiene como objetivo capturar patrones temporales recurrentes en la movilidad humana, particularmente aquellos que dependen de la estructura semanal. En muchos entornos urbanos, los movimientos de las personas están fuertemente influenciados por la diferencia entre días laborables y fines de semana, reflejando rutinas como desplazamientos al trabajo o la escuela de lunes a viernes, y patrones más irregulares de ocio y actividades personales los fines de semana. Al incluir esta variable, el modelo puede aprender estas diferencias y ajustarse mejor a las dinámicas semanales, mejorando la precisión en la predicción de trayectorias futuras. Además, permite que el modelo discrimine entre días con comportamientos similares, evitando que realice predicciones erróneas que no consideren la estructura cíclica de la movilidad.

Tercera variante: GRU basada en características x , y , y $delta_t$

La tercera variante incluye la característica $delta_t$, que representa el tiempo transcurrido en términos de *timeslots* entre observaciones consecutivas. Esto permite al modelo capturar la irregularidad en los datos y comprender mejor las dinámicas de movilidad en casos en los que no hay observaciones constantes.

Diferencias específicas de esta variante

- **Características de entrada:** x , y , y Δt , con una dimensión de entrada total de 3.
- **Tamaño del estado oculto:** 128 neuronas en cada capa GRU, ajustado para manejar estas características.
- **Justificación:** La inclusión de Δt en el modelo permite capturar la irregularidad en los *pings* de los usuarios, proporcionando información adicional sobre los intervalos de tiempo entre observaciones sucesivas. En muchos escenarios reales, la movilidad humana no ocurre en intervalos de tiempo fijos, sino que presenta variaciones debido a patrones individuales de comportamiento, disponibilidad de red o eventos externos. Incorporar Δt ayuda al modelo a diferenciar entre movimientos frecuentes y pausas prolongadas, permitiéndole ajustar mejor la predicción en función del tiempo transcurrido entre observaciones. Esto es particularmente relevante para modelos basados en redes neuronales recurrentes, ya que pueden aprender patrones temporales más flexibles y adaptarse mejor a secuencias de datos con espaciamientos variables.

Cuarta variante: GRU basada en características x , y , Δt , y day_of_week

La cuarta variante combina las características Δt y day_of_week para capturar tanto las irregularidades temporales como los patrones semanales recurrentes. Esta variante busca aprovechar las fortalezas de ambas características adicionales para modelar escenarios más complejos.

Diferencias específicas de esta variante

- **Características de entrada:** x , y , Δt , y day_of_week , con una dimensión de entrada total de 4.
- **Tamaño del estado oculto:** 128 neuronas en cada capa GRU, ajustado para manejar la combinación de estas características.
- **Justificación:** La combinación de Δt y day_of_week en esta variante permite capturar tanto la variabilidad temporal irregular de los desplazamientos como la estructura recurrente semanal de la movilidad humana. En conjunto, estas dos características permiten que la arquitectura GRU modele mejor las relaciones temporales complejas, combinando información sobre interrupciones en los registros con la estructura cíclica de la movilidad humana, lo que contribuye a una mayor precisión en la predicción de trayectorias.

2.2.4. Ajuste y personalización de BERT

En esta sección se propone el modelo **LP-BERT** (**Location Prediction BERT**) para la predicción de trayectorias de usuarios a partir de secuencias de datos temporales y espaciales. Este modelo adapta la arquitectura de **BERT** para manejar secuencias de datos de movilidad y generar predicciones de coordenadas futuras en base a datos históricos.

La arquitectura del modelo **LP-BERT** está diseñada para capturar patrones de movilidad en función de información temporal y espacial. Se compone de cuatro módulos principales: (1) representación de entradas mediante *embeddings* y proyecciones, (2) normalización y transformación de *tokens*, (3) modelado con **BERT**, y (4) regresión para la predicción de coordenadas.

El modelo recibe como entrada una secuencia de eventos de movilidad de un usuario, donde cada evento está definido por un conjunto de atributos temporales y espaciales. Estos atributos incluyen:

- **Fecha** (*date* o *d*): Representa el día.
- **Timeslot** (*timeslot*): Indica la franja horaria dentro del día.
- **Día de la semana** (*day_of_week*): Representa la periodicidad semanal del evento.
- **Ubicación** (*x*, *y*): Coordenadas espaciales del usuario en ese momento.
- **Intervalo temporal** (Δt): Diferencia de tiempo con respecto al *ping* anterior.

Dado que los modelos basados en BERT requieren representaciones numéricas con dimensiones compatibles, cada atributo se transforma mediante *embeddings* en el caso de variables categóricas (*date*, *timeslot*, *day_of_week*) y mediante proyecciones lineales en el caso de variables continuas (Δt , (*x*, *y*)).

Para garantizar una representación coherente de los datos antes de ser procesados por el modelo, los *embeddings* generados se normalizan y combinan en una única representación de *token*:

$$e_{token} = e_{date} + e_{timeslot} + e_{dayofweek} + e_{\Delta t} + e_{location} \quad (2.1)$$

Donde cada e_{var} representa el *embeddings* correspondiente a cada variable de entrada. Posteriormente, estos vectores son normalizados mediante *Layer Normalization* para estabilizar la distribución de los valores.

El modelo BERT es la pieza central de la arquitectura y se encarga de capturar relaciones temporales y espaciales a lo largo de la secuencia de eventos. A diferencia del BERT tradicional para procesamiento de lenguaje natural, aquí se emplea como

un codificador de secuencias de movilidad, donde cada *token* representa un evento de movilidad y la máscara de atención se utiliza para identificar valores reales y de relleno (*padding*).

Dado un conjunto de *tokens* de entrada $T = \{t_1, t_2, \dots, t_n\}$, la salida del modelo BERT es un conjunto de representaciones de alta dimensión, de las cuales se extrae el vector correspondiente al *token* $[CLS]$:

$$h_{cls} = \text{BERT}(T, \text{mask}) \quad (2.2)$$

Este vector representa la información agregada de toda la secuencia y es utilizado para la predicción final.

La capa de salida es una regresión lineal que mapea la representación obtenida de BERT a coordenadas espaciales futuras:

$$\hat{Y} = Wh_{cls} + b \quad (2.3)$$

Donde:

- \hat{Y} es la predicción de coordenadas futuras (x, y) en los próximos 7 días con resolución de 48 *timeslots* por día.
- W y b son los parámetros de la capa lineal de regresión.

Finalmente, la salida se reorganiza en una matriz $(7, 48, 2)$, que representa la predicción de la trayectoria futura en función del tiempo.

La Figura 2.4 presenta una representación esquemática de la arquitectura del modelo LP-BERT, destacando cada una de sus etapas clave. Se observa cómo las diferentes características de entrada (*day_of_week*, *timeslot*, *delta t*, coordenadas espaciales) son transformadas en representaciones numéricas mediante *embeddings* y proyecciones lineales. Estas representaciones son luego procesadas por el modelo BERT, que aprende patrones de movilidad al capturar dependencias temporales y espaciales en la secuencia de eventos. Finalmente, una capa de regresión lineal convierte la salida del modelo en predicciones de coordenadas futuras. Esta arquitectura permite modelar la movilidad humana con una mayor capacidad de abstracción y generalización en comparación con enfoques tradicionales.

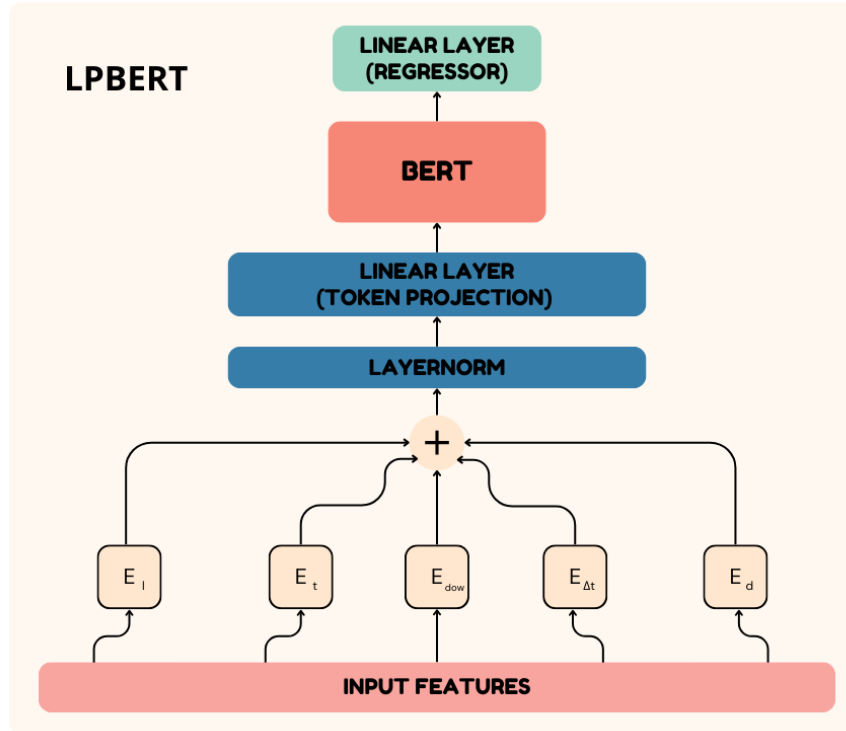


Figura 2.4: Arquitectura del modelo LP-BERT, variante basada en BERT para la predicción de coordenadas futuras a partir de secuencias temporales de movilidad.

Capítulo 3

Experimentación y resultados

En este capítulo, se presentan las técnicas y escenarios de evaluación que componen el marco experimental. Para validar la metodología propuesta, se implementa un prototipo computacional y se evalúa su rendimiento en la predicción de la movilidad humana. En las siguientes secciones, se describen las configuraciones y parámetros utilizados en la validación y evaluación de la metodología. Además, en el capítulo se muestran los resultados y se concluye con una discusión sobre los elementos expuestos.

3.1. Marco experimental

Con el propósito de validar la metodología propuesta, se realizan una serie de experimentos. En esta sección, se describe el marco experimental del análisis efectuado. Para la tarea, se detallan las métricas de evaluación, el corpus de datos, el ajuste de los parámetros y la infraestructura computacional empleada en los experimentos.

3.1.1. Métricas de Evaluación

En esta sección se presentan las métricas utilizadas para evaluar el desempeño de los modelos predictivos implementados en esta tesis. Estas métricas permiten cuantificar la precisión, la calidad y la capacidad de generalización de los modelos.

Error Medio Absoluto (MAE)

El Error Medio Absoluto (MAE) es una métrica estándar para evaluar la precisión de predicciones numéricas. Se calcula como el promedio de las distancias absolutas entre las predicciones y los valores reales [5], según la fórmula:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_{pred}^i - y_{real}^i| \quad (3.1)$$

donde y_{pred}^i y y_{real}^i son las coordenadas predichas y reales, respectivamente, y N es el número total de predicciones válidas. En el contexto de esta tesis, el MAE proporciona una medida directa de la precisión espacial de las predicciones realizadas por los modelos.

Precisión de Predicción de Ubicación (LPP)

La Precisión de Predicción de Ubicación (LPP) evalúa qué porcentaje de las predicciones se encuentra dentro de un margen de tolerancia predefinido alrededor de las posiciones reales. Esta métrica considera:

- Un umbral de tolerancia (e.g., 4 unidades de distancia).
- Las distancias euclidianas entre las coordenadas predichas y las reales.

Se define como:

$$LPP = \frac{\text{Número de predicciones correctas}}{\text{Número total de predicciones válidas}} \times 100 \quad (3.2)$$

Esta métrica es particularmente útil para evaluar la precisión espacial en aplicaciones donde pequeños errores son aceptables.

GeoBLEU

GeoBLEU es una métrica inspirada en el puntaje BLEU utilizado en el procesamiento de lenguaje natural, adaptada para evaluar la similitud entre trayectorias espaciales. Esta métrica se centra en comparar secuencias de ubicaciones predichas y reales, evaluando la coherencia y precisión de las trayectorias generadas por los modelos predictivos.

Cálculo de GeoBLEU: Analiza las subtrayectorias de las rutas predichas y de referencia, comparando n-gramas espaciales para evaluar la similitud local. Un puntaje de GeoBLEU cercano a 1 indica una alta similitud entre la trayectoria predicha y la real, mientras que un puntaje cercano a 0 sugiere una baja similitud.

Implementación: Se utiliza una implementación de GeoBLEU disponible en GitHub, desarrollada por Yahoo Japan Corporation, que facilita su aplicación en proyectos de predicción de movilidad humana [8].

Dynamic Time Warping (DTW)

El Dynamic Time Warping (DTW) es una técnica que mide la similitud entre dos secuencias temporales, permitiendo alineaciones no lineales para capturar variaciones en el tiempo. En el contexto de la predicción de movilidad humana, DTW se utiliza para comparar trayectorias predichas y reales, evaluando la correspondencia entre las secuencias de ubicaciones a lo largo del tiempo.

Cálculo del DTW: DTW encuentra la alineación óptima entre dos secuencias al estirar o comprimir segmentos de tiempo según sea necesario, minimizando la distancia acumulada entre puntos correspondientes. Esto permite comparar trayectorias que pueden variar en velocidad o duración, proporcionando una medida robusta de similitud incluso cuando las trayectorias no están perfectamente sincronizadas.

Aplicaciones: DTW ha sido ampliamente utilizado en el reconocimiento de actividades humanas y en el análisis de series temporales multivariadas, demostrando su eficacia en la comparación de secuencias complejas [22].

Análisis cualitativo de las métricas

Cada una de las métricas descritas aborda diferentes aspectos del problema de predicción de movilidad humana:

- **Error Medio Absoluto (MAE):** Proporciona una evaluación directa de la precisión espacial de las predicciones, midiendo el error promedio entre las ubicaciones predichas y las reales.
- **Precisión de Predicción de Ubicación (LPP):** Evalúa la proporción de predicciones que caen dentro de un margen de tolerancia predefinido, ofreciendo una perspectiva sobre la exactitud de las predicciones en un contexto práctico.
- **GeoBLEU:** Mide la coherencia local de las trayectorias predichas, comparando subtrayectorias y capturando la similitud en patrones de movimiento específicos.
- **Dynamic Time Warping (DTW):** Analiza la similitud global de las trayectorias, permitiendo alineaciones flexibles en el tiempo y evaluando la correspondencia general entre las secuencias de movimiento.

3.1.2. Estructura y procesamiento del conjunto de datos

El conjunto de datos utilizado en esta tesis proviene del Human Mobility Prediction Challenge (HuMob Challenge) 2024, una competencia diseñada para evaluar modelos predictivos de patrones de movilidad humana. Es un conjunto de datos de acceso abierto, proporcionado por LY Corporation, que contiene información sobre

trayectorias humanas en cuatro áreas metropolitanas, recopiladas durante un período de 75 días [2].

Estructura del conjunto de datos

El conjunto de datos está organizado en registros individuales, cada uno representando una observación puntual de un usuario en un lugar y momento específicos. Los registros incluyen las siguientes características:

- **uid:** Identificador único del usuario.
- **d:** Día de la observación, numerado secuencialmente.
- **t:** *Timeslot* dentro del día (48 intervalos de 30 minutos cada uno).
- **x:** Coordenada X de ubicación espacial.
- **y:** Coordenada Y complementaria.

Características del conjunto de datos

Cada registro representa un evento de movilidad, detallando la presencia de un usuario en una ubicación específica durante un *timeslot* particular en un día determinado. La estructura temporal y espacial de los datos permite analizar patrones de movilidad a diferentes escalas.

Tamaño del conjunto de datos: El conjunto incluye **111,535,175 registros**, organizados en una estructura tabular que ocupa aproximadamente **4.2 GB de memoria**. Las coordenadas x e y están acotadas en un rango de 0 a 200, lo que permite interpretar la movilidad como una matriz espacial definida.

Saltos en los datos: Existen valores faltantes debido a la intermitencia en la captura de datos, lo que requiere estrategias específicas de preprocesamiento.

Reducción y preprocesamiento del conjunto de datos

Por restricciones de *hardware*, se utilizó una muestra de **3,000 usuarios**, resultando en **3,642,173 registros** antes del preprocesamiento. Al aplicar *padding*, el total de registros aumentó a:

$$3,000 \times 75 \times 48 = 10,800,000$$

Esto garantizó secuencias completas para el modelado sin comprometer la representatividad de los datos.

Para mejorar la calidad y compatibilidad con los modelos de predicción, se aplicaron los siguientes procedimientos:

- **Generación de secuencias de ventanas deslizantes:** Creación de secuencias temporales con *sliding window*.
- **Normalización:** Uso de un *scaler* ajustado a los datos de entrenamiento para escalar los valores entre 0 y 1.
- **División en conjuntos de entrenamiento, validación y prueba:**
 - **Entrenamiento:** Primeros 45 días.
 - **Validación:** Sigüientes 15 días.
 - **Prueba:** Últimos 15 días.

Esta organización permitió entrenar modelos en un conjunto suficientemente amplio y evaluar su capacidad de generalización en datos no vistos.

3.1.3. Recursos computacionales y hardware utilizado

El entrenamiento y evaluación de los modelos en esta tesis requirieron recursos computacionales significativos debido al tamaño del conjunto de datos y la complejidad de los modelos implementados. A continuación, se detallan las especificaciones del *hardware* utilizado para cada tipo de modelo.

Ejecución de los primeros modelos

Los primeros modelos desarrollados basados en redes neuronales simples fueron ejecutados en una máquina local con las siguientes especificaciones:

- **Sistema operativo:** Windows 11
- **Procesador:** Intel Core i7 de 13^a generación
- **Memoria RAM:** 32 GB
- **Unidad de procesamiento gráfico (GPU):** No disponible

La falta de aceleración por GPU prolongó significativamente el tiempo de entrenamiento en comparación con su uso.

Ejecución de los modelos basados en GRU

Los modelos basados en **GRU**, incluyendo las cuatro variantes desarrolladas, fueron ejecutados en servidores en línea a través de Google Colab, utilizando la GPU **A100** de NVIDIA con las siguientes especificaciones:

- **GPU:** NVIDIA A100 con 40 GB de memoria VRAM
- **Memoria RAM:** 83.5 GB
- **Tiempo total de ejecución:** Aproximadamente **2 horas** para las cuatro variantes de GRU

Ejecución del modelo de *fine-tuning* de BERT

Para el modelo basado en *transformers* (*fine-tuning* de BERT adaptado para la predicción de movilidad humana), se utilizó el mismo entorno de Google Colab con la GPU **A100** y 83.5 GB de RAM. Sin embargo, debido a la complejidad del modelo y la necesidad de ajustar parámetros finos en la red preentrenada, el tiempo de entrenamiento fue significativamente mayor, con un tiempo total de ejecución de aproximadamente **6 horas**

3.1.4. Hiperparámetros

En esta sección, se presentan los valores utilizados para cada modelo desarrollado, abarcando desde redes neuronales simples hasta modelos basados en GRU y transformers.

Los hiperparámetros definidos incluyen el número de capas, la cantidad de neuronas o celdas ocultas, la tasa de aprendizaje utilizada en el optimizador, el tamaño del *batch* para la actualización de los pesos, la tasa de *dropout* aplicada para regularización, el número total de épocas de entrenamiento y el optimizador usado junto a la función de pérdida en cada caso.

Tabla 3.1: Hiperparámetros utilizados en cada modelo

Modelo	Capas	Tasa de Aprendizaje	Batch Size	Dropout	Épocas	Optimizador / Criterio de Pérdida
<i>Baseline</i>	-	-	-	-	-	-
NN _{1day}	3	0.001	64	0.2	100	Adam / <i>Masked</i> MSE
NN _{1week}	3	0.001	64	0.2	100	Adam / <i>Masked</i> MSE
GRU _{x,y,t}	3	0.001	32	0.2	30	Adam / <i>Masked</i> MSE
GRU _{x,y,t,dw}	5	0.001	32	0.2	30	Adam / <i>Masked</i> MSE
GRU _{x,y,dt}	3	0.001	32	0.2	30	Adam / <i>Masked</i> MSE
GRU _{x,y,dt,dw}	3	0.001	32	0.2	30	Adam / <i>Masked</i> MSE
LPBert (<i>Transformer</i>)	12	1e-4	32	-	10	AdamW / <i>Masked</i> MSE

Justificación Los valores seleccionados para los hiperparámetros han sido definidos con base en prácticas recomendadas en la literatura y pruebas experimentales previas realizadas durante el desarrollo del modelo.

El valor de *dropout* se estableció en **0.2** en las redes neuronales y modelos GRU para reducir el sobreajuste sin afectar significativamente la capacidad del modelo de aprender patrones complejos. Estudios previos han demostrado que aplicar valores de *dropout* en el rango de 0.1 a 0.3 es efectivo en modelos de aprendizaje profundo recurrente [23].

En cuanto al número de **épocas**, se utilizó **30** para los modelos GRU, ya que se observó que el entrenamiento convergía en este número de iteraciones sin evidencia de sobreajuste. Para las redes neuronales simples, se emplearon **100 épocas**, dado que estos modelos requieren más iteraciones para alcanzar un rendimiento óptimo. En el caso del modelo basado en *transformers*, se utilizó un ajuste fino con 10 épocas. Si bien estudios previos recomiendan entre 2 y 4 épocas para evitar la sobreadaptación [9], en este trabajo se exploró un mayor número de iteraciones para analizar su efecto en la convergencia del modelo y su capacidad de generalización.

3.2. Resultados

En esta sección se presentan los resultados obtenidos a partir de la evaluación de los modelos predictivos desarrollados en esta tesis. Para medir el desempeño de cada modelo, se han utilizado varias métricas expuestas en la sección 3.1.1 para evaluar diferentes aspectos de la calidad de las predicciones.

A continuación, se presentan diversas tablas que muestran los diferentes modelos y sus variantes. Finalmente, se incluye una tabla comparativa general en la que se sintetizan los resultados obtenidos para todos los modelos evaluados.

Tabla 3.2: Resultados del modelo Baseline

Modelo	MAE ↓	LPP ↑	GeoBLEU ↑	DTW ↓
Baseline	25.3098	41.1507	0.1665	137.6086

La Tabla 3.2 muestra los resultados del modelo *Baseline*, el cual establece una referencia mínima de rendimiento. Los resultados obtenidos indican que este modelo presenta un MAE que sugiere una precisión razonable en la predicción de la movilidad humana. Sin embargo, destaca especialmente en las métricas GeoBLEU y DTW, alcanzando valores competitivos en comparación con modelos avanzados utilizados en el *HuMob Challenge* [3]. Esto confirma que, a pesar de su simplicidad, el enfoque basado en la repetición de patrones semanales es un punto de partida robusto.

Tabla 3.3: Resultados de las Redes Neuronales Simples

Modelo	MAE ↓	LPP ↑	GeoBLEU ↑	DTW ↓
NN _{1day}	23.1377	3.6097	0.0053	138.1401
NN _{1week}	44.3860	0.4534	0.0004	320.9399

La Tabla 3.3 muestra los resultados de las Redes Neuronales Simples en sus dos variantes. El modelo diseñado para predecir un día (NN (1 día)) obtuvo el mejor desempeño, en cuanto a todas las métricas, lo que indica su efectividad a corto plazo. En contraste, el modelo para una semana (NN (1 semana)) presentó un MAE mucho mayor y menor precisión, evidenciando dificultades para capturar patrones semanales de movilidad.

Este comportamiento era esperable, ya que la movilidad humana sigue patrones con variaciones diarias, influenciadas por factores externos como horarios laborales, días festivos y actividades específicas. Intentar predecir una semana completa con una red neuronal simple resulta ineficiente, ya que el modelo no logra retener información suficiente para anticipar los cambios dinámicos en los patrones de movimiento. Por lo tanto, para este tipo de modelo, la predicción a corto plazo es la estrategia más efectiva.

El uso de redes GRU (*Gated Recurrent Unit*) se planteó como una mejora con respecto a las Redes Neuronales Simples, dado que estas últimas mostraron limitaciones al capturar dependencias temporales de medio plazo. Aunque las redes neuronales simples lograron buenos resultados en predicciones a corto plazo, su desempeño se degradó notablemente al intentar predecir secuencias más largas, como una semana completa. Las redes GRU, al ser una variante de las redes recurrentes, fueron incorporadas en la metodología con el propósito de mejorar la capacidad del modelo para manejar información temporal, permitiendo que el estado oculto retenga información clave sobre la movilidad del usuario en el tiempo.

Tabla 3.4: Resultados de las variantes de GRU

Modelo	MAE ↓	LPP ↑	GeoBLEU ↑	DTW ↓
GRU _{x,y,t}	33.4553	1.2345	0.0013	315.7651
GRU _{x,y,t,dw}	35.4411	1.2345	0.0006	332.1241
GRU _{x,y,dt}	32.0062	0.6172	0.0002	302.6173
GRU _{x,y,dt,dw}	32.9777	0.6172	0.0014	314.2238

En la Tabla 3.4 se muestran los resultados obtenidos por las variantes de GRU con

diferentes configuraciones de características de entrada. Se observa que la mejor variante en términos de MAE y DTW fue la configuración $\text{GRU}_{x,y,dt}$, lo que indica que la inclusión del *delta t* como característica ayudó a capturar mejor la dinámica temporal de los datos. Esto sugiere que conocer el tiempo transcurrido entre observaciones permite al modelo ajustar mejor sus predicciones, reduciendo el error acumulado en la trayectoria.

Por otro lado, la variante $\text{GRU}_{x,y,dt,dw}$ logró el mejor desempeño en la métrica GeoBLEU, lo que indica que esta configuración permitió generar trayectorias más similares en términos de secuencias espaciales a las trayectorias reales. Sin embargo, en términos de MAE y DTW, su desempeño fue ligeramente inferior a $\text{GRU}_{x,y,dt}$, lo que sugiere que, aunque el *day of the week* aporta información sobre patrones recurrentes semanales, su impacto puede no ser tan significativo como el *delta t* en términos de reducción de error absoluto.

En cuanto a la métrica LPP, ambas variantes que incluyen *day of the week* alcanzaron el mismo valor máximo (**1.2345**), indicando que el conocimiento del día de la semana podría ayudar a reducir la desviación de las predicciones en ubicaciones clave, aunque su impacto general en otras métricas fue menor.

En general, se puede concluir que el uso de GRU permitió capturar dependencias temporales más complejas en comparación con las redes neuronales simples, logrando un desempeño significativamente mejor que la variante de una semana, lo que indica que la capacidad de memoria y procesamiento secuencial de GRU ayuda a mitigar las limitaciones observadas en las redes densas. Sin embargo, los modelos GRU no lograron superar el rendimiento obtenido por la red neuronal simple en la predicción de un solo día, lo que sugiere que, aunque las GRU mejoran la captura de patrones temporales, su capacidad para predecir trayectorias a corto plazo aún presenta limitaciones frente a modelos más simples y eficientes en esta tarea específica.

No obstante, el rendimiento global aún se encuentra limitado en comparación con el modelo *Transformer*, como se verá en la siguiente tabla.

Tabla 3.5: Resultados del modelo Transformer

Modelo	MAE ↓	LPP ↑	GeoBLEU ↑	DTW ↓
LPBert	31.7024	32.3529	0.0501	72.9249

El modelo tuneado basado en el *Transformer* BERT para la predicción de movilidad, mostró mejoras notables en varias métricas clave en comparación con los modelos anteriores. Como se observa en la Tabla 3.5, LPBert logró reducir significativamente la métrica DTW, lo que indica que las trayectorias predichas por este modelo tienen

una estructura más similar a las trayectorias reales en términos de distancia acumulada. Este resultado sugiere que la capacidad de los *Transformers* para capturar relaciones complejas en las secuencias temporales permite una mayor coherencia en la predicción de trayectorias a largo plazo.

Además, en la métrica GeoBLEU, *LPBert* alcanzó a superar ampliamente a las redes neuronales simples y los modelos GRU. Esto demuestra que la capacidad de atención del *Transformer* permite capturar mejor la estructura global de las trayectorias, generando predicciones más cercanas a los patrones de movilidad humana reales.

En cuanto a MAE y LPP, aunque *LPBert* no logra superar al modelo *Baseline* en estos aspectos, sí muestra mejoras respecto a los modelos GRU y a las redes neuronales simples, lo que indica que los *Transformers* pueden aprender patrones de movilidad más complejos y generar predicciones más estables.

En general, estos resultados destacan la efectividad del enfoque basado en *Transformers* para la predicción de movilidad humana, validando su uso como una estrategia robusta para modelar secuencias espacio-temporales con una mayor capacidad de generalización.

Tabla 3.6: Comparación de todas las arquitecturas

Modelo	MAE ↓	LPP ↑	GeoBLEU ↑	DTW ↓
NN _{1week}	44.3860	0.4534	0.0004	320.9399
GRU _{x,y,t}	33.4553	1.2345	0.0013	315.7651
GRU _{x,y,t,dw}	35.4411	1.2345	0.0006	332.1241
GRU _{x,y,dt}	32.0062	0.6172	0.0002	302.6173
GRU _{x,y,dt,dw}	32.9777	0.6172	0.0014	314.2238
Baseline	25.3098	41.1507	0.1665	137.6086
LPBert	31.7024	32.3529	0.0501	72.9249

La Tabla 3.6 presenta los resultados comparativos de todos los modelos evaluados en esta tesis, incluyendo el *Baseline*, las Redes Neuronales Simples, las variantes de GRU y el modelo *Transformer*.

3.3. Discusión

Los resultados muestran que el *Baseline* estableció un punto de referencia sólido, con métricas competitivas en comparación con arquitecturas más complejas. Esto sugiere que los patrones de movilidad humana presentan una recurrencia semanal

significativa, lo que permite que un enfoque simple basado en la repetición de la última semana alcance un rendimiento notable en ciertas métricas, como **LPP** y **GeoBLEU**.

El uso de redes GRU permitió abordar este problema al incorporar memoria en la modelación de secuencias temporales. Dentro de las variantes evaluadas, la inclusión de la característica *delta t* demostró ser clave para mejorar la precisión del modelo, reflejando la importancia de capturar las irregularidades en la movilidad de los usuarios. No obstante, aunque las GRU lograron mejorar la capacidad del modelo para aprender dependencias temporales, su desempeño en comparación con el *Baseline* no fue uniformemente superior, destacando la dificultad de predecir trayectorias de movilidad en un entorno realista.

Finalmente, el modelo basado en *Transformers* logró los mejores resultados generales, superando significativamente a las demás arquitecturas en **DTW**, lo que sugiere que este enfoque es más adecuado para capturar patrones de movilidad complejos y generalizar a distintos escenarios. La capacidad de atención de **LPBert** permitió una mejor representación de la información secuencial, ofreciendo predicciones más precisas y consistentes a lo largo del tiempo.

Sin embargo, a pesar de su superioridad en ciertas métricas, el modelo basado en *Transformers* no logró superar al *Baseline* en **MAE**, **LPP** y **GeoBLEU**. Esto podría deberse a varios factores, entre ellos la posible necesidad de un mayor volumen de datos de entrenamiento para aprovechar completamente su capacidad de generalización. Los modelos basados en *Transformers* suelen requerir grandes cantidades de datos para su ajuste óptimo, y el tamaño del conjunto utilizado podría haber limitado su rendimiento. Además, la optimización de hiperparámetros podría desempeñar un papel crucial en su desempeño; técnicas como *Random Search* o *Bayesian Optimization* podrían haber permitido encontrar configuraciones más eficientes que mejoraran sus predicciones.

Por último, una exploración más profunda en arquitecturas híbridas que combinen la capacidad de atención de los *Transformers* con la memoria a largo plazo de las GRU podría representar un camino prometedor para mejorar aún más la capacidad de predicción de movilidad humana en futuros trabajos.

Conclusiones

El presente trabajo tuvo como objetivo desarrollar y evaluar diferentes arquitecturas de modelos de aprendizaje profundo para la predicción de movilidad humana. Para ello, se implementó un esquema que permitió comparar múltiples enfoques, desde modelos simples hasta arquitecturas basadas en *Transformers*, aplicados sobre un conjunto de datos representativo de trayectorias humanas. A continuación, se detallan las conclusiones alcanzadas en relación con cada uno de los objetivos específicos de la investigación.

En primer lugar, se llevó a cabo una revisión exhaustiva de la literatura sobre predicción de movilidad humana, explorando diversas técnicas utilizadas para modelar patrones espacio-temporales. Se analizaron enfoques tradicionales basados en reglas, así como métodos de aprendizaje profundo, incluyendo redes recurrentes y *Transformers*. Este estudio permitió definir una metodología clara para el desarrollo de los modelos y comprender los desafíos inherentes a la predicción de trayectorias humanas.

A partir de esta revisión, se desarrollaron modelos complementarios para la predicción de movilidad en diferentes horizontes temporales. Se implementaron redes neuronales simples para la predicción diaria y semanal, lo que permitió establecer una primera aproximación al problema. Los resultados demostraron que la predicción diaria es más efectiva que la semanal, ya que las redes neuronales simples mostraron un desempeño significativamente inferior cuando se intentó predecir una semana completa, lo que indica que este tipo de modelo tiene dificultades para capturar patrones a largo plazo.

Para abordar estas limitaciones, se implementaron arquitecturas basadas en **GRU** con el objetivo de capturar dependencias temporales más complejas. Estas redes demostraron ser más efectivas que las redes neuronales simples para la predicción de movilidad, logrando mejorar la captura de patrones espacio-temporales en los datos. Asimismo, se exploraron distintas configuraciones de características dentro de la arquitectura GRU, incluyendo la incorporación de *day_of_week* y *delta_t*. Se encontró que la variable *delta_t*, que representa el intervalo de tiempo entre registros, fue clave para mejorar la precisión en la predicción, mientras que la inclusión del día de la semana ayudó a capturar patrones recurrentes específicos.

Posteriormente, se exploró el uso de *Transformers* para mejorar la captura de

patrones de largo alcance en los datos de movilidad. La implementación de **LPBert** permitió evaluar la efectividad de los mecanismos de atención en la predicción de trayectorias. Los resultados indicaron que este modelo superó a todas las demás arquitecturas en términos de **DTW**, lo que sugiere una mejor capacidad para modelar trayectorias en escenarios complejos. Sin embargo, en métricas como **MAE** y **LPP**, el modelo *Baseline* obtuvo mejores resultados, lo que podría deberse a factores como la cantidad de datos utilizados y la configuración de hiperparámetros del *Transformer*. Es posible que un ajuste más exhaustivo de estos hiperparámetros o el uso de un volumen de datos mayor permitiera mejorar aún más su desempeño.

Finalmente, todos los modelos fueron validados utilizando un conjunto de métricas específicas, incluyendo **MAE**, **LPP**, **GeoBLEU** y **DTW**. Esta evaluación permitió comparar de manera objetiva el desempeño de cada arquitectura y destacar las fortalezas y debilidades de cada enfoque. La comparación final evidenció que, si bien el modelo **LPBert** mostró un rendimiento superior en términos de **DTW**, el *Baseline* aún se mantiene como una alternativa sólida debido a su buen desempeño en otras métricas y su menor complejidad computacional.

A partir de los hallazgos obtenidos, se identifican varias oportunidades para continuar con esta línea de investigación. Se recomienda implementar estrategias de *transfer learning* para evaluar la capacidad de generalización del modelo *Transformer* en diferentes ciudades o contextos geográficos. También sería pertinente explorar arquitecturas híbridas que combinen redes GRU con mecanismos de atención, con el objetivo de obtener modelos que aprovechen lo mejor de ambos enfoques y logren un mejor balance entre precisión y eficiencia computacional. Asimismo, se propone mejorar los métodos de preprocesamiento y manejo de datos faltantes, incorporando técnicas de imputación avanzadas que permitan reducir la pérdida de información en los registros de movilidad. Optimizar la implementación del modelo *Transformer* para reducir sus requerimientos computacionales y permitir su aplicación en tiempo real o en dispositivos con menor capacidad de procesamiento es otra línea de investigación futura.

En conclusión, esta investigación ha demostrado que el uso de modelos de aprendizaje profundo es una herramienta efectiva para abordar el problema de la predicción de movilidad humana. Si bien se han identificado desafíos en términos de generalización y eficiencia computacional, los resultados obtenidos con *Transformers* marcan un punto de partida prometedor para el desarrollo de modelos cada vez más precisos y aplicables en escenarios reales. El conocimiento adquirido en esta tesis puede servir como base para futuras investigaciones en la intersección entre movilidad humana, aprendizaje automático y optimización de sistemas urbanos, contribuyendo así al desarrollo de soluciones más eficientes para la planificación y gestión de la movilidad en entornos dinámicos.

Recomendaciones

A partir de los hallazgos obtenidos en esta investigación, se proponen las siguientes recomendaciones para futuros trabajos en el ámbito de la predicción de la movilidad humana:

- **Exploración de arquitecturas híbridas:** Integrar modelos GRU con arquitecturas avanzadas de *transformers* puede mejorar la precisión en la predicción de series temporales complejas. Estudios recientes han demostrado que la combinación de GRU con *transformers* es efectiva en la predicción de datos secuenciales, como los provenientes de estaciones meteorológicas [20].
- **Optimización de hiperparámetros:** Implementar técnicas de optimización automática de hiperparámetros, como la búsqueda en cuadrícula, búsqueda aleatoria u optimización bayesiana, puede mejorar significativamente el rendimiento de los modelos de aprendizaje automático. Estas metodologías permiten explorar de manera sistemática el espacio de hiperparámetros para identificar las configuraciones óptimas que maximizan la precisión predictiva. Para modelos más complejos, como los basados en *transformers*, se recomienda el uso de herramientas avanzadas como *Optuna* o *Ray Tune* para realizar una búsqueda eficiente y escalable de hiperparámetros.
- **Incorporación de atención temporal:** Implementar mecanismos de atención que permitan al modelo enfocarse en momentos específicos de la secuencia puede mejorar la captura de patrones relevantes en los datos de movilidad.
- **Uso de datos contextuales:** Integrar información adicional, como datos geospaciales, condiciones climáticas o eventos sociales, podría enriquecer el modelo y proporcionar predicciones más precisas.
- **Evaluación en diferentes contextos:** Probar los modelos en diversos entornos urbanos y rurales permitirá evaluar su generalización y adaptabilidad a distintos patrones de movilidad.

- **Actualización continua del modelo:** Implementar técnicas de aprendizaje en línea o incremental para que el modelo se adapte a cambios en los patrones de movilidad a lo largo del tiempo.
- **Consideración de la eficiencia computacional:** Al desarrollar modelos más complejos, es importante balancear la precisión con la eficiencia computacional, asegurando que las soluciones sean viables para aplicaciones en tiempo real.

Al seguir estas recomendaciones, se espera avanzar en el desarrollo de modelos más robustos y precisos para la predicción de la movilidad humana, aprovechando las ventajas de las arquitecturas híbridas y los *transformers*.

Bibliografía

- [1] ¿Qué son los procesos de media móvil integrada autorregresiva (Arima)? Consultado el 12 de enero de 2025. URL: <https://es.statisticseasily.com/glosario/%C2%BFQu%C3%A9-son-los-procesos-ARIMA-de-media-m%C3%B3vil-integrada-autorregresiva%3F/> (vid. pág. 14).
- [2] HuMob Challenge 2024. *Human Mobility Prediction Challenge 2024*. 2024. URL: <https://wp.nyu.edu/humobchallenge2024/> (vid. pág. 32).
- [3] HuMob Challenge 2024. *Resultados Finales del Human Mobility Prediction Challenge 2024*. 2024. URL: <https://wp.nyu.edu/humobchallenge2024/final-results/> (vid. pág. 35).
- [4] Diego Calvo. *Red neuronal recurrente: Qué son y cómo funcionan*. Accedido: 12 de enero de 2025. s.f. URL: <https://www.diegocalvo.es/red-neuronal-recurrente/> (vid. pág. 8).
- [5] J. A. Camarillo-Peñaranda y A. J. Saavedra-Montes. «Recomendaciones para Seleccionar Índices para la Validación de Modelos». En: *TecnoLógicas* 16.30 (2013), págs. 109-123. URL: <https://www.redalyc.org/pdf/3442/344234341009.pdf> (vid. pág. 29).
- [6] Kyunghyun Cho et al. «Learning phrase representations using RNN encoder-decoder for statistical machine translation». En: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014, págs. 1724-1734 (vid. pág. 7).
- [7] J. Chung et al. «Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling». En: *arXiv preprint arXiv:1412.3555* (2014). URL: <https://arxiv.org/pdf/1412.3555> (vid. pág. 6).
- [8] Yahoo Japan Corporation. *geobleu: Python implementation of GEO-BLEU, a similarity evaluation method for trajectories*. Último acceso: 26 de enero de 2025. 2024. URL: <https://github.com/yahoojapan/geobleu> (vid. pág. 30).
- [9] Jacob Devlin et al. «BERT: Pre-training of deep bidirectional transformers for language understanding». En: *arXiv preprint arXiv:1810.04805* (2019). URL: <https://arxiv.org/abs/1810.04805> (vid. pág. 35).

- [10] J. González et al. «Nation-wide human mobility prediction based on graph neural networks». En: *Applied Intelligence* 51 (2021), págs. 8891-8904. URL: <https://link.springer.com/article/10.1007/s10489-021-02645-3> (vid. pag. 10).
- [11] Jake Grigsby et al. «Long-Range Transformers for Dynamic Spatiotemporal Forecasting». En: *arXiv preprint arXiv:2109.12218* (2023). URL: <https://arxiv.org/abs/2109.12218> (vid. págs. 13-15).
- [12] J. Gunkel, A. Tundis y M. Mühlhäuser. «The Story of Mobility: Combining State Space Models and Transformers for Multi-Step Trajectory Prediction». En: *HuMob'24, October 29-November 1, 2024, Atlanta, GA, USA*. 2024 (vid. págs. 1, 5, 11, 12, 15).
- [13] M. Gunkel, A. Tundis y M. Mühlhäuser. «Graph network-based human movement prediction for socially-aware robot navigation in shared workspaces». En: *Neural Computing and Applications* (2024). URL: <https://link.springer.com/article/10.1007/s00521-024-10369-x> (vid. pag. 10).
- [14] H. He, H. Luo y Q. R. Wang. «ST-MoE-BERT: A Spatial-Temporal Mixture-of-Experts Framework for Long-Term Cross-City Mobility Prediction». En: *HuMob'24, October 29-November 1, 2024, Atlanta, GA, USA*. 2024 (vid. págs. 1, 5, 9, 11, 14, 15).
- [15] S. Hochreiter y J. Schmidhuber. «Long Short-Term Memory». En: *Neural Computation*. Vol. 9. 8. MIT Press, 1997, págs. 1735-1780. URL: <https://www.cs.cmu.edu/~bhiksha/courses/deeplearning/Fall.2018/papers/LSTM.pdf> (vid. pag. 6).
- [16] H. Jiang et al. «Graph Neural Network for Traffic Forecasting: The Research Progress». En: *ISPRS International Journal of Geo-Information* 12.3 (2023), pag. 100. URL: <https://www.mdpi.com/2220-9964/12/3/100> (vid. pag. 10).
- [17] Asifullah Khan et al. «A survey of the Vision Transformers and their CNN-Transformer based Variants». En: *Artificial Intelligence Review* (2024). URL: <https://arxiv.org/abs/2305.09880> (vid. pag. 13).
- [18] Yann LeCun et al. «Gradient-Based Learning Applied to Document Recognition». En: *Proceedings of the IEEE* 86.11 (1998), págs. 2278-2324. URL: <http://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf> (vid. pag. 9).
- [19] A.-S. Li et al. «Using the Temporal-Trajectory-based K Nearest Neighbor Algorithm to Predict Human Mobility Patterns». En: *HuMob'24, October 29-November 1, 2024, Atlanta, GA, USA*. 2024 (vid. págs. 3, 5, 8, 9).

- [20] Juan Pérez. «Aplicación de Transformers en la Predicción de Series Temporales». En: *Ingenius* 29.1 (2023), págs. 15-25. DOI: 10.17163/ings.n29.2023.02. URL: <https://revistas.ups.edu.ec/index.php/ingenius/article/view/9557/8811> (vid. pág. 42).
- [21] J. A. Rodríguez, L. G. Castellanos y M. T. Ramírez. «CNN-LSTM con mecanismo de atención suave para el reconocimiento de acciones humanas en videos». En: *ResearchGate* (2021). URL: https://www.researchgate.net/publication/357233318_CNN-LSTM_con_mecanismo_de_atencion_suave_para_el_reconocimiento_de_acciones_humanas_en_videos (vid. pág. 7).
- [22] Skyler Seto, Wenyu Zhang y Yichen Zhou. «Multivariate Time Series Classification Using Dynamic Time Warping Template Selection for Human Activity Recognition». En: *arXiv preprint arXiv:1512.06747* (2015). URL: <https://arxiv.org/abs/1512.06747> (vid. pág. 31).
- [23] Nitish Srivastava et al. «Dropout: A simple way to prevent neural networks from overfitting». En: *The Journal of Machine Learning Research* 15.1 (2014), págs. 1929-1958 (vid. pág. 35).
- [24] Y. Sun. «Trajectory Prediction Using Random Forests with Time Decay and Periodic Features». En: *HuMob'24, October 29-November 1, 2024, Atlanta, GA, USA*. 2024 (vid. págs. 1, 5, 7, 10).
- [25] M. Suzuki. «Human Mobility Prediction using Personalized Spatiotemporal Models». En: *HuMob'24, October 29-November 1, 2024, Atlanta, GA, USA*. 2024 (vid. págs. 2, 5, 10).
- [26] Yuan Tang et al. «Resurrecting Recurrent Neural Networks for Long Sequences». En: *arXiv preprint arXiv:2303.06349* (2023). URL: <https://arxiv.org/abs/2303.06349> (vid. págs. 13, 14).
- [27] A. Vaswani et al. «Attention Is All You Need». En: *Advances in Neural Information Processing Systems* 30 (2017). URL: <https://arxiv.org/abs/1706.03762> (vid. págs. 11-15).
- [28] K. Yasuda et al. «Multiple Systems Combination to Improve Human Mobility Prediction». En: *HuMob'24, October 29-November 1, 2024, Atlanta, GA, USA*. 2024 (vid. págs. 1, 5, 11, 12, 15).
- [29] J. Zhang, Y. Zheng y D. Qi. «Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction». En: *Proceedings of the AAAI Conference on Artificial Intelligence* 31 (2017). URL: <https://ojs.aaai.org/index.php/AAAI/article/view/10724> (vid. pág. 9).

- [30] Y. Zhang, Z. Zheng y H. Peng. «Hybrid deep learning model based on CNN and LSTM for urban traffic flow prediction». En: *Scientific Reports* 11 (2021), pág. 93656. URL: <https://www.nature.com/articles/s41598-021-93656-0> (vid. pág. 10).