

Universidad de La Habana
Facultad de Matemática y Computación



Predicción de la movilidad humana utilizando redes neuronales

Autor:

Brian Ameht Inclán Quesada

Tutores:

Alejandro Lage

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

10/10/2024

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

Agradecimientos

Opinión del tutor

Opiniones de los tutores

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. Introducción al Aprendizaje Supervisado y No Supervisado	6
1.2.2. Redes Neuronales Recurrentes (RNN), LSTM y GRU	6
1.2.3. Redes Neuronales Convolucionales (CNN)	9
1.2.4. Redes Gráficas (GNN)	9
1.2.5. 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	14
1.3.3. Tendencias Emergentes	14
1.4. Análisis Comparativo	15
1.4.1. Comparación de Metodologías	15
1.4.2. Identificación de Oportunidades	16
1.5. Discusión	17
2. Propuesta	19
2.1. Visión general	19
2.2. Descripción y estructura del conjunto de datos	20
2.2.1. Estructura del conjunto de datos	20
2.2.2. Procedencia y recopilación de los datos	21
2.2.3. Organización y características del conjunto de datos	21
2.3. Diseño y Construcción de los Modelos	22
2.3.1. Implementación del baseline	22
2.3.2. Redes neuronales simples	23

2.3.3.	Desarrollo de variantes GRU	26
2.3.4.	Ajuste y personalización de BERT	29
2.4.	Métricas de evaluación	29
2.4.1.	Error Medio Absoluto (MAE)	29
2.4.2.	Precisión de Predicción de Ubicación (LPP)	30
2.4.3.	GeoBLEU	30
2.4.4.	Dynamic Time Warping (DTW)	31
2.4.5.	Análisis cualitativo de las métricas	31
3.	Detalles de Implementación y Experimentos	32
3.1.	Marco experimental	32
3.1.1.	Corpus de datos	32
3.1.2.	Hiperparámetros	34
3.1.3.	Recursos computacionales y hardware utilizado	34
3.2.	Resultados	35
3.3.	Discusión	35
	Conclusiones	36
	Bibliografía	37

Índice de figuras

1.1.	Esquema de una celda LSTM (Long Short-Term Memory) [6].	7
1.2.	Arquitectura interna de una celda GRU (Gated Recurrent Units) [2] .	8
1.3.	Esquema de un modelo híbrido CNN-LSTM para predicción espacio-temporal [25].	11
1.4.	La arquitectura general del modelo <i>Transformer</i> [21].	13
2.1.	Arquitectura del modelo <i>vanilla</i> para predecir un día siguiente a partir de la semana previa.	24
2.2.	Arquitectura del modelo <i>vanilla</i> para predecir la semana siguiente a partir de la semana previa.	26
2.3.	Arquitectura del modelo para predecir la semana siguiente a partir de la semana previa basado en GRU.	27

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 [18, 11]. 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 [9, 23]. 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 [19]. 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 [16]. 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 [19, 18]. 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 [16].

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 [16, 23].

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 [11, 9]. 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 [11].

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. Introducción al Aprendizaje Supervisado y No Supervisado

En el **aprendizaje supervisado**, los modelos son entrenados con datos etiquetados, donde cada entrada tiene una salida conocida. Este enfoque se aplica comúnmente en tareas como la predicción de la próxima ubicación de un individuo con base en su historial de movimientos. Algoritmos como árboles de decisión, SVM y redes neuronales profundas son ampliamente utilizados en este contexto [19, 16].

El **aprendizaje no supervisado**, por otro lado, busca descubrir patrones ocultos sin la necesidad de etiquetas. Técnicas como k-means y PCA han sido empleadas para agrupar trayectorias humanas similares o identificar áreas de alta densidad de movilidad urbana [11, 9].

1.2.2. 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 [12].

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 [4].

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.

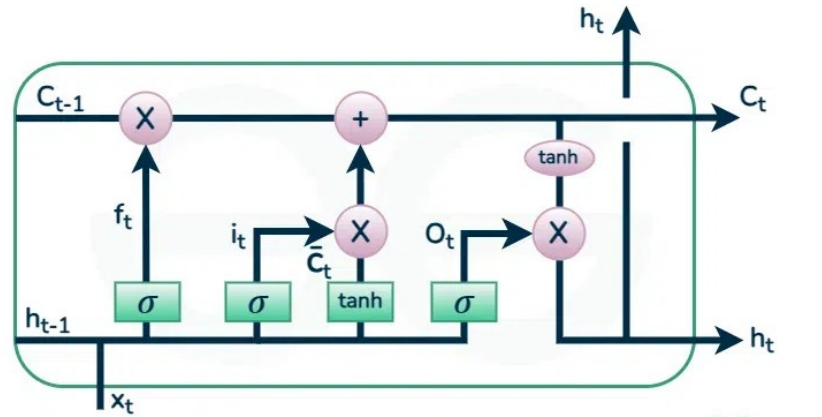


Figura 1.1: Esquema de una celda LSTM (Long Short-Term Memory) [6].

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 [18].

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 [3].

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.

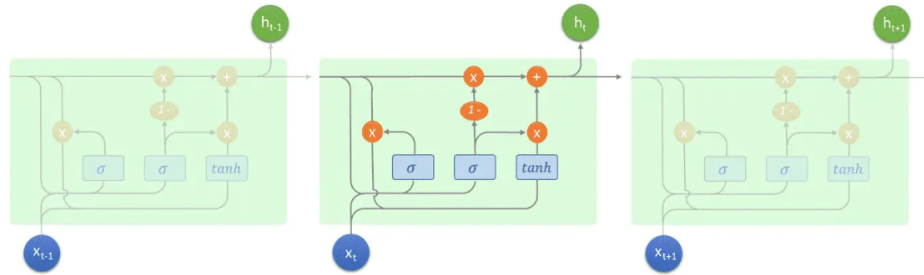


Figura 1.2: Arquitectura interna de una celda GRU (Gated Recurrent Units) [2]

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 [16].

1.2.3. 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 [15].

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 [16, 11]. 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 [24].

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.

1.2.4. 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 [7]. Además, las GNN han mejorado la precisión en la predicción de flujos de tráfico al capturar dependencias espaciales y temporales en sistemas de transporte [13]. También se han empleado en la predicción de movimientos humanos en espacios compartidos, facilitando la navegación de robots en entornos donde interactúan con personas [10].

1.2.5. 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 [19, 18].

Por ejemplo, el modelo híbrido propuesto en [25] utiliza una arquitectura CNN-LSTM para predecir flujos de tráfico en entornos urbanos. 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.

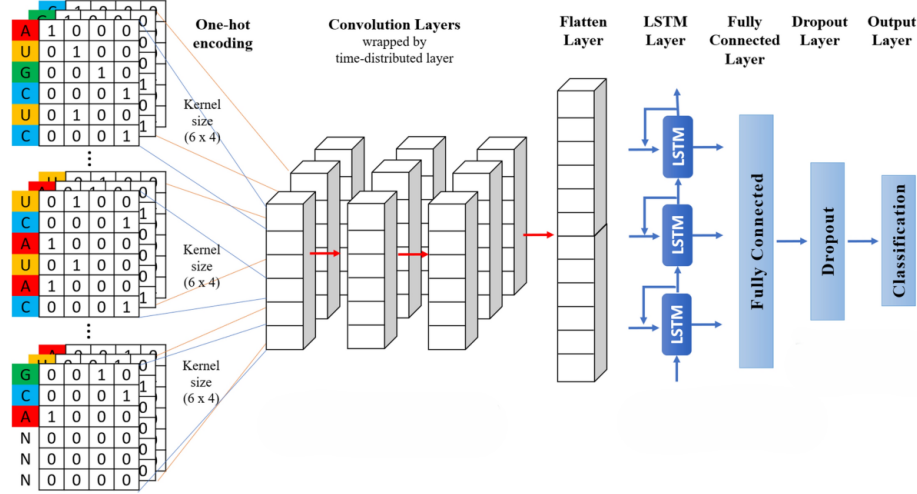


Figura 1.3: Esquema de un modelo híbrido CNN-LSTM para predicción espacio-temporal [25].

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 [11, 9]. 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 [11].

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 [23].

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 [21, 11].

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 [11]. 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.

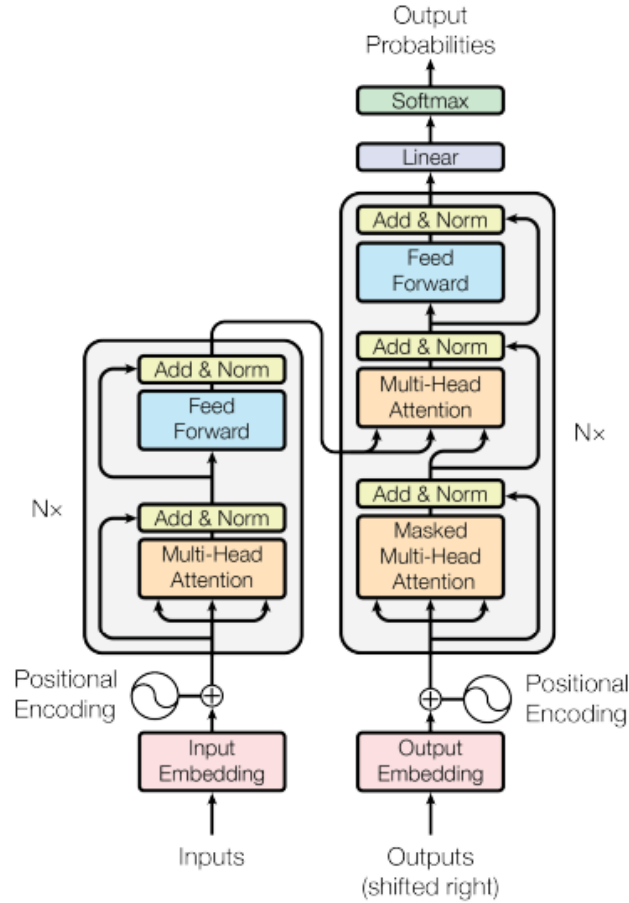


Figura 1.4: La arquitectura general del modelo *Transformer* [21].

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 [9, 23]. 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 [21]. 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 [20].

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 [14].

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 [8].

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 [8].

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 [11]. 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 [8]. 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:

- **Modelos Clásicos (ARIMA, Regresión, etc.):**
 - **Fortalezas:** Simplicidad, bajo costo computacional y eficacia en escenarios con datos lineales y estables.
 - **Debilidades:** Limitaciones para manejar relaciones no lineales y datos heterogéneos [1].
- **Redes Neuronales Recurrentes (RNN, LSTM, GRU):**
 - **Fortalezas:** Capacidades robustas para capturar dependencias temporales y patrones recurrentes.
 - **Debilidades:** Problemas con secuencias largas debido al desvanecimiento del gradiente y altas demandas computacionales [20].

- **Transformers:**

- **Fortalezas:** Procesamiento paralelo, atención global y capacidad para integrar datos heterogéneos.
- **Debilidades:** Requieren grandes volúmenes de datos y recursos computacionales significativos [21].

- **Modelos Híbridos (CNN-LSTM, GNN-Transformer):**

- **Fortalezas:** Combinan las ventajas de arquitecturas especializadas, como la captura de relaciones espaciales y temporales complejas.
- **Debilidades:** Complejidad arquitectónica y mayor costo de implementación [11].

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) [9].
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 [11].
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 [21].

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 [8].
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 [23].

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. La Sección 2.2 describe la estructura del conjunto de datos usado para el problema planteado. En la Sección 2.3 se detallan las tareas específicas realizadas durante el desarrollo, como la construcción y validación de múltiples modelos predictivos. Finalmente en la Sección 2.4 se exponen las métricas de evaluación

2.1. Visión general

En esta propuesta se desarrolló un *pipeline* 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 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 del *pipeline*, integrando las lecciones aprendidas en los pasos previos y ofreciendo un enfoque generalizable para resolver el problema planteado.

2.2. Descripción y estructura 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. Este conjunto de datos es de acceso abierto y contiene información detallada sobre trayectorias humanas en cuatro áreas metropolitanas, recopiladas durante un período de 75 días, con entre 10,000 y 100,000 individuos por ciudad.

2.2.1. 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. Cada usuario tiene un ID numérico que permite rastrear sus movimientos a lo largo del tiempo.
- **d:** Día de la observación, numerado secuencialmente desde el inicio del periodo de estudio.

- **t:** *Timeslot* o intervalo de tiempo dentro del día. Cada día está dividido en 48 *timeslots* de 30 minutos cada uno, cubriendo las 24 horas del día.
- **x:** Coordenada X que representa la ubicación espacial del usuario en un sistema de referencia definido.
- **y:** Coordenada Y que complementa a x , formando un par (x, y) que indica la posición del usuario en el espacio.

2.2.2. Procedencia y recopilación de los datos

Los datos fueron proporcionados por LY Corporation (anteriormente Yahoo Japan) y son de naturaleza sintética, generados específicamente para reflejar patrones de movilidad humana en entornos urbanos. La generación del *dataset* consideró factores como:

- Distribuciones espaciales y temporales de actividades humanas.
- Patrones de movilidad típicos en áreas urbanas.
- Características demográficas de las regiones simuladas.

2.2.3. Organización y características del conjunto de datos

Cada registro en el conjunto de datos 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, desde movimientos diarios hasta tendencias que abarcan semanas o meses.

Tamaño del conjunto de datos: El conjunto incluye millones de registros, lo que garantiza una representación robusta de las dinámicas urbanas. Además, los datos están completamente limpios y bien estructurados. 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. Esta organización facilita la implementación de métodos computacionales para el análisis de trayectorias.

Salto en los datos: Es importante destacar que, debido a la naturaleza del *dataset*, existen saltos entre los *pings* de los usuarios, ya que no siempre se generan observaciones en cada *timeslot*. Esto introduce valores faltantes que reflejan la intermitencia en la movilidad humana y requieren un manejo adecuado durante el preprocesamiento y el modelado.

Usos en esta tesis: En esta tesis se utilizará una de las ciudades, denominada Ciudad A, para realizar todas las tareas de predicción. Este enfoque permitirá un análisis profundo y detallado de la predicción de movilidad humana en un entorno específico. Además, se plantea como trabajo futuro la posibilidad de aplicar *transfer learning* para extrapolar los conocimientos obtenidos en la Ciudad A a las otras ciudades del conjunto de datos.

2.3. 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.

2.3.1. Implementación del baseline

El modelo *baseline* se implementó como una referencia inicial en el *pipeline*, 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.

2.3.2. 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.

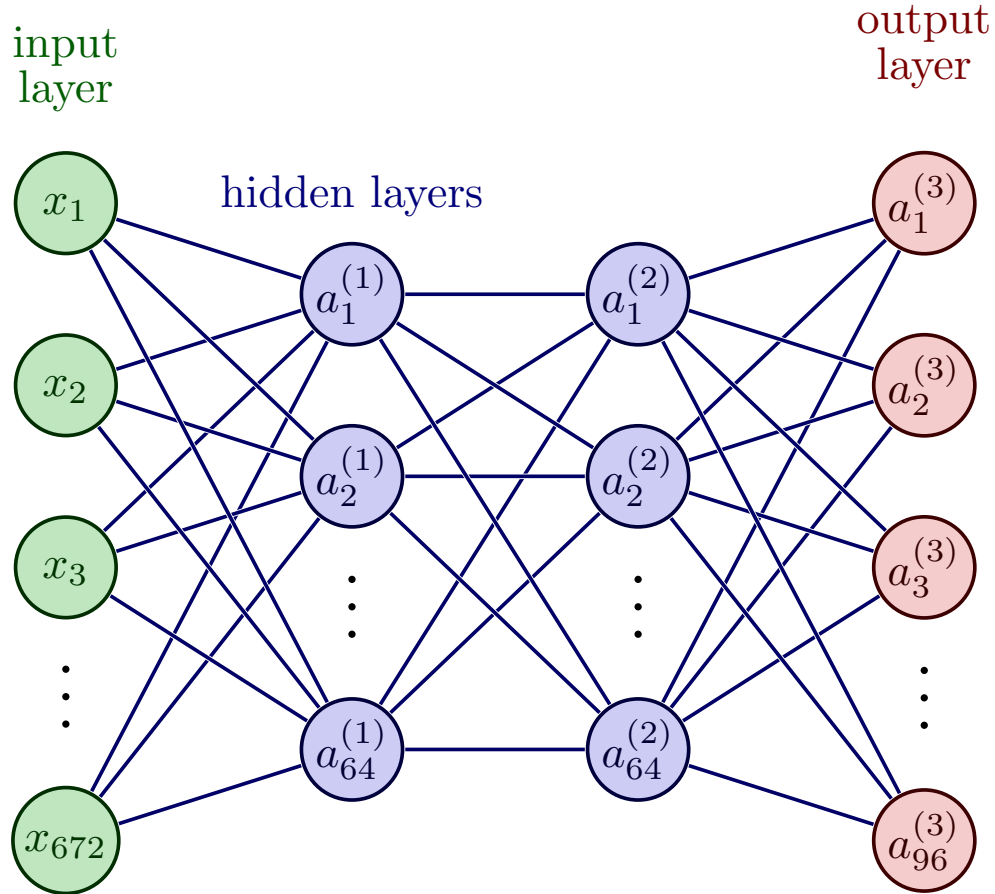


Figura 2.1: Arquitectura del modelo *vanilla* para predecir un día siguiente a partir de la semana previa.

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.

Este modelo aprovecha las secuencias largas de datos históricos para capturar relaciones temporales y patrones complejos, mejorando la capacidad de predicción en comparación con los enfoques más simples. La combinación de máscaras y regularización garantiza que el modelo sea robusto frente a datos incompletos y reduzca el sobreajuste.

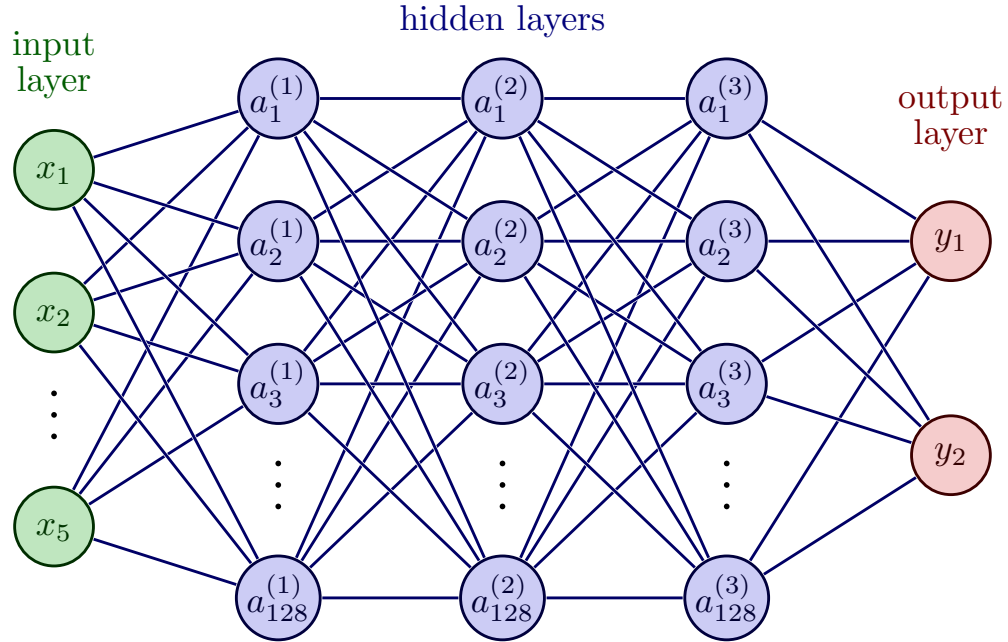


Figura 2.2: Arquitectura del modelo *vanilla* para predecir la semana siguiente a partir de la semana previa.

2.3.3. 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.

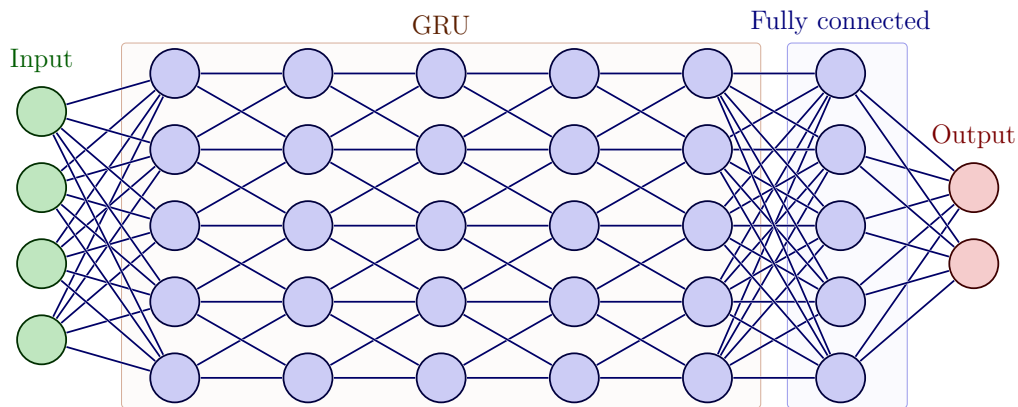


Figura 2.3: Arquitectura del modelo para predecir la semana siguiente a partir de la semana previa basado en GRU.

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 permite al modelo capturar patrones recurrentes específicos de los días de la semana, como diferencias entre días laborables y fines de semana, que suelen influir significativamente en la movilidad humana. Esto hace que esta variante sea más adecuada para conjuntos de datos donde los patrones semanales tienen un impacto importante en las predicciones.

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 $delta_t$, con una dimensión de entrada total de 3.
- **Tamaño del estado oculto:** 128 neuronas en cada capa GRU, ajustado para manejar estas características.

- **Justificación:** *delta_t* proporciona información adicional sobre las interrupciones en la movilidad, lo que permite al modelo interpretar mejor los patrones temporales irregulares. Esto es particularmente útil en datos donde los pings no ocurren en intervalos regulares, capturando mejor las dinámicas temporales complejas.

Cuarta variante: GRU basada en características *x*, *y*, *delta_t*, y *day_of_week*

La cuarta variante combina las características *delta_t* 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*, *delta_t*, 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:** Al incluir *delta_t*, el modelo puede capturar interrupciones temporales en los datos, mientras que *day_of_week* permite capturar patrones recurrentes específicos de días de la semana. Esta combinación hace que el modelo sea más robusto frente a variaciones tanto temporales como semanales en los datos de movilidad.

2.3.4. Ajuste y personalización de BERT

En proceso...

2.4. 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.

2.4.1. 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 [22], según la fórmula:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i^{pred} - y_i^{real}|,$$

donde y_i^{pred} y y_i^{real} 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.

2.4.2. 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:

$$LPP = \frac{\text{Número de predicciones correctas}}{\text{Número total de predicciones válidas}} \times 100.$$

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

2.4.3. 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. GeoBLEU considera subtrayectorias locales, lo que permite una evaluación detallada de segmentos específicos dentro de una trayectoria completa.

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. Esta métrica es particularmente útil para capturar la precisión en segmentos específicos de la trayectoria, proporcionando una evaluación detallada de la calidad de las predicciones.

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 [5].

2.4.4. 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 17.

2.4.5. 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.

Capítulo 3

Detalles de Implementación y Experimentos

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 el corpus de datos, el ajuste de los parámetros y la infraestructura computacional empleada en los experimentos.

3.1.1. Corpus de datos

El conjunto de datos utilizado en esta tesis es extenso y detallado, proporcionando información sobre la movilidad humana en una escala masiva. En total, el *dataset* cuenta con **111,535,175 registros**, organizados en una estructura tabular con cinco columnas principales como se detalló en la Sección 2.2.1.

El *dataset* ocupa aproximadamente **4.2 GB de memoria** y contiene datos bien estructurados y preprocesados. En términos estadísticos, se presentan las siguientes métricas sobre la distribución de los datos:

Tabla 3.1: Estadísticas descriptivas del conjunto de datos.

Métrica	uid	d	t	x	y
Media	49,174.61	37.40	25.96	122.90	85.65
Desviación estándar	28,811.83	21.68	11.17	42.10	42.82
Mínimo	0	0	0	1	1
Percentil 25	23,932	19	18	95	58
Mediana (50 %)	49,258	38	26	127	84
Percentil 75	73,929	57	35	155	111
Máximo	99,999	74	47	200	200

Reducción del tamaño del conjunto de datos

Por cuestiones de *hardware*, no se utilizó la totalidad del conjunto de datos, ya que para todos los modelos era necesario crear máscaras o rellenar con *padding* los conjuntos de datos, lo cual incrementaba significativamente la cantidad de memoria utilizada. Para garantizar la viabilidad del entrenamiento y evaluación de los modelos sin comprometer la representatividad del *dataset*, se decidió utilizar únicamente **3,000 usuarios**, lo que resultó en un total de **3,642,173 registros** antes de aplicar técnicas de preprocesamiento.

Además, dado que cada usuario tiene datos distribuidos en 75 días con 48 *timeslots* por día, al incorporar *padding*, el conjunto de datos extendido resultó en:

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

registros, asegurando que cada usuario tuviera secuencias de datos completas para el modelado.

Preprocesamiento del conjunto de datos

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

- **Generación de secuencias de ventanas deslizantes:** Se crearon secuencias temporales utilizando una estrategia de *sliding window*, lo que permitió capturar la evolución de la movilidad a lo largo del tiempo y mejorar la capacidad de generalización de los modelos.
- **Normalización de los datos:** Todas las características fueron normalizadas utilizando un *scaler* ajustado a los datos de entrenamiento, asegurando que los valores estuvieran dentro de un rango entre 0 y 1 para facilitar el aprendizaje del modelo.

■ **División en conjuntos de entrenamiento, validación y prueba:**

- Los primeros **45 días** se utilizaron como conjunto de **entrenamiento** (*train*).
- Los siguientes **15 días** se reservaron para **validación** (*validation*).
- Los últimos **15 días** se emplearon como conjunto de **prueba** (*test*) para evaluar la capacidad de generalización del modelo.

Esta organización del conjunto de datos permitió asegurar que los modelos fueran entrenados en datos suficientemente amplios y variados, validando su rendimiento en un conjunto separado y evaluando su capacidad de generalización con datos completamente no vistos.

3.1.2. Hiperparámetros

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

Debido a la ausencia de aceleración por GPU, el entrenamiento de estos modelos tomó un tiempo considerablemente mayor en comparación con lo que pudiera haber sido si se empleara una Unidad de procesamiento gráfico.

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 **4 horas**

3.2. Resultados

3.3. Discusión

Conclusiones

Conclusiones

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. 15).
- [2] 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).
- [3] 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. 8).
- [4] 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).
- [5] 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. 31).
- [6] GeeksforGeeks. *Deep Learning / Introduction to Long Short-Term Memory*. Último acceso: 11 de enero de 2025. 2023. URL: <https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/> (vid. pág. 7).
- [7] 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. pág. 10).
- [8] 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. 14, 15, 17).

- [9] 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, 6, 11, 13, 16).
- [10] 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).
- [11] 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, 6, 9, 11, 12, 15, 16).
- [12] 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).
- [13] 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).
- [14] 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. 14).
- [15] 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).
- [16] 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, 6, 8, 9).
- [17] 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. pag. 31).
- [18] 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).

- [19] 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, 6, 10).
- [20] 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. 14, 15).
- [21] 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. 12-14, 16).
- [22] Wikipedia. *Error absoluto medio*. Último acceso: 26 de enero de 2025. 2024. URL: https://es.wikipedia.org/wiki/Error_absoluto_medio (vid. pág. 30).
- [23] 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, 13, 17).
- [24] 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).
- [25] 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. págs. 10, 11).