

Estudio sobre la Efectividad del Positional Encoding en Transformers para Series Temporales y Diseño de Mecanismos Adaptados

Cristhian Moya Mota

Tutor: Julián Luengo Martín

Departamento de Ciencias de la Computación e Inteligencia Artificial

Cotutor: Diego Jesús García Gil

Departamento de Lenguajes y Sistemas Informáticos

8 de septiembre de 2025

Contenido

- 1 Introducción
- 2 Estado del arte
- 3 Propuestas de PE
- 4 Evaluación de PE mediante conjuntos de datos
- 5 Conclusiones
 - Conclusiones
 - Trabajos futuros

Contenido

1 Introducción

2 Estado del arte

3 Propuestas de PE

4 Evaluación de PE mediante conjuntos de datos

5 Conclusiones

- Conclusiones
- Trabajos futuros

Introducción

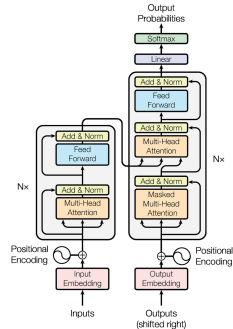
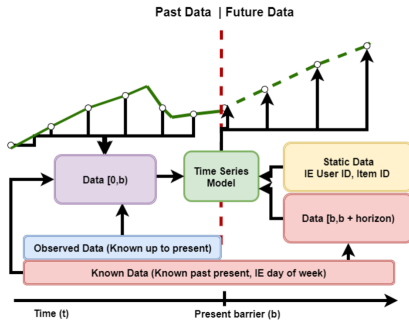


Figura: Forecasting en Series Temporales¹. Transformers²

¹<https://developer.nvidia.com/blog/time-series-forecasting-with-the-nvidia-time-series-prediction-platform-and-triton-inference-server/>

²<https://doi.org/10.48550/arXiv.1706.03762>

Introducción. Justificación

Aspectos que justifican la realización del proyecto:

- Falta de captura de la estructura. Ausencia de información semántica.
- Dificultad para adaptarse a diferentes escalas temporales y falta de semántica.
- Complejidad computacional y falta de interpretabilidad en la metodología.

Introducción. Objetivos

Este proyecto persigue:

- Comprender y exhibir las carencias de los métodos actuales.
- Proponer nuevos encodings para series temporales empleando Transformers.
- Evaluar la efectividad de las nuevas propuestas en diferentes ámbitos y conjuntos de datos.
- Identificar y establecer la base para nuevos métodos de codificación.

Contenido

- 1 Introducción
- 2 Estado del arte
- 3 Propuestas de PE
- 4 Evaluación de PE mediante conjuntos de datos
- 5 Conclusiones
 - Conclusiones
 - Trabajos futuros

Estado del arte. Arquitecturas y PE

- Métodos estadísticos: STL, ARIMA, Prophet.
- Métodos basados en Transformer:
 - Informer (2021)
 - Autoformer (2021)
 - FEDformer (2022)

Problemas:

- 1 Escasa adición de información local.
- 2 Mecanismos de atención poco cercanos a la semántica del dato.

¿Solución? → Crear una nueva familia de codificaciones posicionales, capaz de captar información local y global.

Estado del arte. Caso particular de Informer

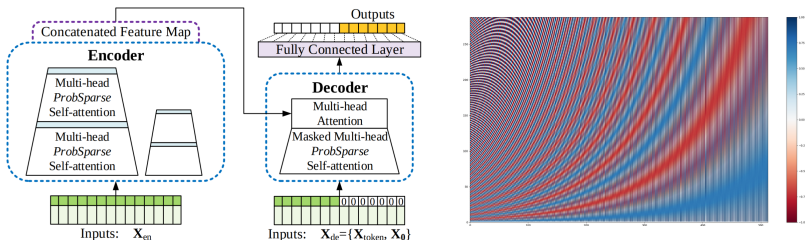


Figura: Informer: arquitectura y encoding empleado³

³<https://doi.org/10.48550/arXiv.2012.07436>

Contenido

1 Introducción

2 Estado del arte

3 Propuestas de PE

4 Evaluación de PE mediante conjuntos de datos

5 Conclusiones

- Conclusiones
- Trabajos futuros

Encodings propuestos: Aspectos clave

Para mejorar la calidad del encoding, hay dos alternativas:

- Modificar el mecanismo de atención → pérdida de información local y complejidad añadida
- Modificar únicamente el PE → permite aprovechar arquitecturas existentes y trabajar directamente sobre el dato

Resultado: familia de modelos WinStat:

- 1 Sólo modificar la codificación posicional
- 2 *Concatenar* información en lugar únicamente sumarla
- 3 *Combinar PE existentes de manera ponderada* (normalizada con Softmax)

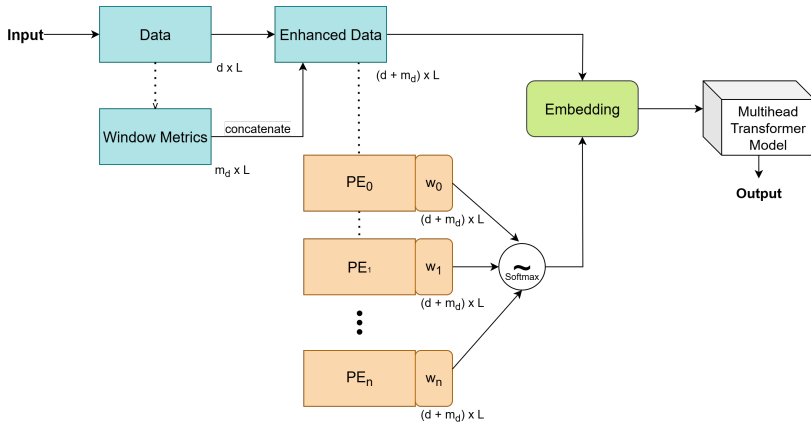


Figura: Arquitectura y proceso de adición del encoding

Encodings propuestos: **WinStat** (Variante base)

Ventana local de tamaño W fijo, en cada posición t :

- *Media*: $\mu_t = \frac{1}{|\mathcal{W}_t|} \sum_{x \in \mathcal{W}_t} x$
- *Desviación estándar*: $\sigma_t = \sqrt{\frac{1}{|\mathcal{W}_t|} \sum_{x \in \mathcal{W}_t} (x - \mu_t)^2}$
- *Mínimo*: $m_t^{\min} = \min_{x \in \mathcal{W}_t} x$
- *Máximo*: $m_t^{\max} = \max_{x \in \mathcal{W}_t} x$

$$s_t = [\mu_t, \sigma_t, m_t^{\min}, m_t^{\max}] \in \mathbb{R}^4$$

El embedding enriquecido de la posición t es:

$$\tilde{x}_t = [x_t \parallel s_t] \in \mathbb{R}^{d+4}$$

Encodings propuestos: WinStatLag

Parte de WinStat, añadiendo $|\mathcal{L}|$ retardos especificados:

$$s_t = [\mu_t, \sigma_t, m_t^{\text{mín}}, m_t^{\text{máx}}] \in \mathbb{R}^4$$

Definimos para cada $\ell_j \in \mathcal{L}$:

$$\delta_t^{(\ell_j)} = |x_t - x_{t-\ell_j}|, \quad \text{si } t - \ell_j \geq 1$$

El *embedding* enriquecido es la concatenación de s_t y $\delta_t^{(\ell_j)}$

$$\tilde{x}_t = [x_t \parallel s_t \parallel \delta_t^{(\ell_1)} \parallel \dots \parallel \delta_t^{(\ell_p)}] \in \mathbb{R}^{d+4+p}$$

Encodings propuestos: WinStatFlex

Basado en WinStatLags, añadiendo la información ponderada de:

- 1 Encoding sinusoidal original
- 2 LPE:

$$LPE_{(pos)} = W_{pos}, \quad W_{pos} \in \mathbb{R}^d$$

$$X_{LPE} = \tilde{x}_t + LPE[pos]$$

- 3 TAPE:

$$\omega_k^{new} = k \cdot \frac{d_{model}}{L} \quad (1)$$

$$\begin{aligned} \text{TAPE}_{(pos,2i)} &= \sin(pos \cdot \omega_i^{new}) \\ \text{TAPE}_{(pos,2i+1)} &= \cos(pos \cdot \omega_i^{new}) \end{aligned} \quad (2)$$



Encodings propuestos: WinStatTPE

- Emplea WinStatFlex como base
- Sustituye TAPE por TPE:

$$S(i, j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)$$

Añadiéndose al encoding sinusoidal original la nueva componente:

$$T-PE(i) = PE(i) + S(i, j)$$

$$X_{TPE} = \tilde{x}_t + T - PE[pos]$$

Contenido

- 1 Introducción
- 2 Estado del arte
- 3 Propuestas de PE
- 4 Evaluación de PE mediante conjuntos de datos
- 5 Conclusiones
 - Conclusiones
 - Trabajos futuros

Evaluación. Conjuntos de datos

Conjuntos de datos para la evaluación:

- Household Power Consumption (HPC)
- ETTh1
- ETTh2
- Yellow Trip Data
- TINA

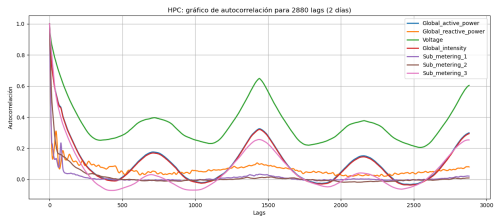


Figura: Autocorrelación de Household Power Consumption

Evaluación. Condiciones de entrenamiento

- 1 Resultado a partir de ejecuciones múltiples promedio (semilla *random*)
- 2 Ejecución de encoding barajado para comprobar alcance de su aportación en cada modelo.
- 3 Métricas: MSE y MAE
- 4 Heurística para el tamaño de ventana: entre $1/3$ y $1/4$ de longitud de secuencia.

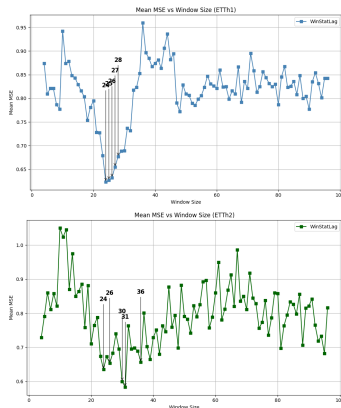


Figura: Estudio tamaño de ventana

Evaluación. Household Power Consumption (I)

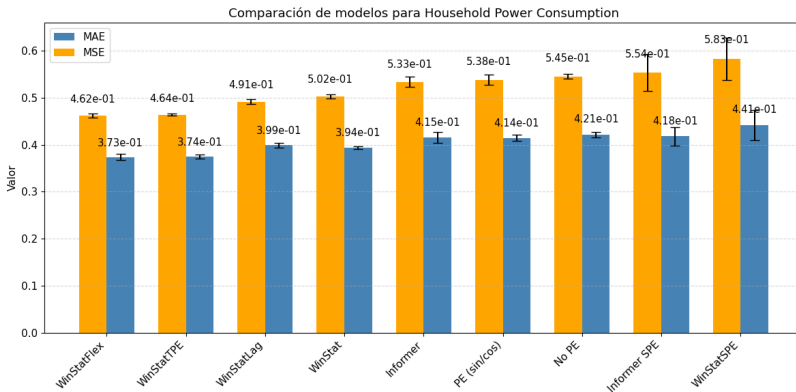


Figura: Resultados de Household Power Consumption

Evaluación. Household Power Consumption (II)

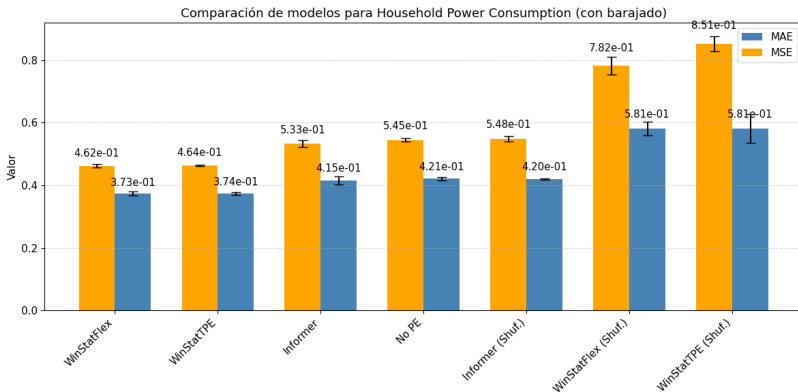


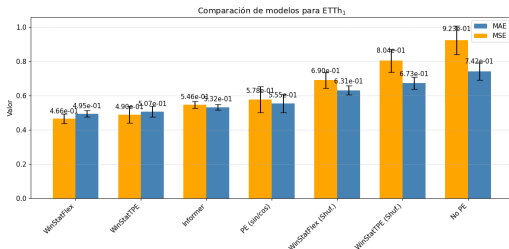
Figura: Resultados de HPC, con barajado en el PE

Evaluación. Household Power Consumption (III)

Modelo	MAE (Media)	MAE (STD)	MSE (Media)	MSE (STD)
WinStatFlex	0,373488	0,006858	0,461892	0,004841
WinStatTPE	0,374392	0,004132	0,463651	0,001917
Informer	0,415272	0,012216	0,532965	0,010976
No PE	0,421343	0,005856	0,544844	0,005447
Informer (Shuffled)	0,419960	0,002769	0,548443	0,009148
WinStatFlex (Shuffled)	0,581124	0,021991	0,782092	0,027368
WinStatTPE (Shuffled)	0,581001	0,046717	0,851115	0,024083

Cuadro: Resultados de las métricas de HPC

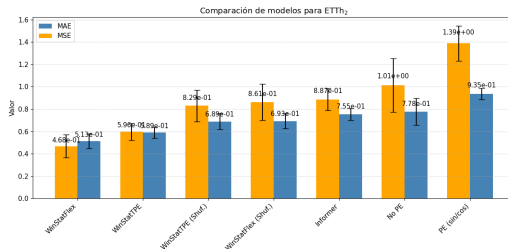
Evaluación. ETTh1 y ETTh2 (I)



Modelo	MAE (Media)	MAE (STD)	MSE (Media)	MSE (STD)
WinStatFlex	0,4947	0,0195	0,4660	0,0277
WinStatTPE	0,5072	0,0312	0,4896	0,0482
Informer	0,5324	0,0171	0,5458	0,0204
PE (sin/cos)	0,5552	0,0531	0,5776	0,0755
WinStatFlex (Shuf.)	0,6313	0,0276	0,6897	0,0461
WinStatTPE (Shuf.)	0,6728	0,0358	0,8036	0,0667
No PE	0,7419	0,0509	0,9235	0,0841

Cuadro: Resultados obtenidos en ETTh1

Evaluación. ETTh1 y ETTh2 (II)



Modelo	MAE (Media)	MAE (STD)	MSE (Media)	MSE (STD)
WinStatFlex	0,5128	0,0651	0,4677	0,1018
WinStatTPE	0,5889	0,0496	0,5964	0,0760
WinStatTPE (Shuf.)	0,6892	0,0714	0,8294	0,1418
WinStatFlex (Shuf.)	0,6929	0,0671	0,8608	0,1625
Informer	0,7549	0,0540	0,8866	0,1001
No PE	0,7777	0,1199	1,0132	0,2417
PE (sin/cos)	0,9355	0,0514	1,3881	0,1582

Cuadro: Resultados obtenidos en ETTh2

Evaluación. Yellow Trip Data (I)

Modelo	MSE (Media)	MSE (STD)	MAE (Media)	MAE (STD)
WinStatTPE	$1,0 \times 10^{-5}$	$5,0 \times 10^{-6}$	$2,547 \times 10^{-3}$	$9,37 \times 10^{-4}$
WinStatFlex	$1,7 \times 10^{-5}$	$8,0 \times 10^{-6}$	$3,529 \times 10^{-3}$	$1,027 \times 10^{-3}$
Informer	$1,9 \times 10^{-5}$	$7,0 \times 10^{-6}$	$3,987 \times 10^{-3}$	$7,95 \times 10^{-4}$
PE (sin/cos)	$2,4 \times 10^{-5}$	$1,0 \times 10^{-5}$	$4,307 \times 10^{-3}$	$1,044 \times 10^{-3}$
No PE	$4,6 \times 10^{-5}$	$6,0 \times 10^{-5}$	$4,626 \times 10^{-3}$	$3,954 \times 10^{-3}$
WinStatTPE (Shuf.)	$5,6 \times 10^{-5}$	$6,3 \times 10^{-5}$	$5,688 \times 10^{-3}$	$3,983 \times 10^{-3}$
WinStatFlex (Shuf.)	$1,302 \times 10^{-1}$	$1,838 \times 10^{-1}$	$7,203 \times 10^{-2}$	$9,583 \times 10^{-2}$

Cuadro: Métricas obtenidas para Yellow Trip

Evaluación. Yellow Trip Data (II)

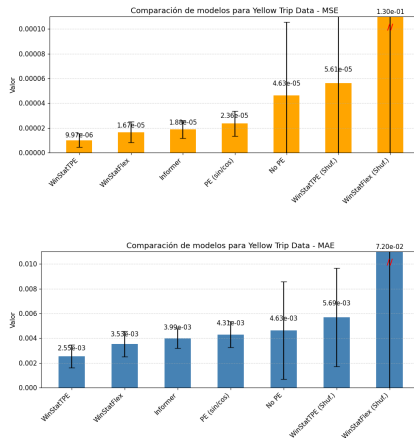
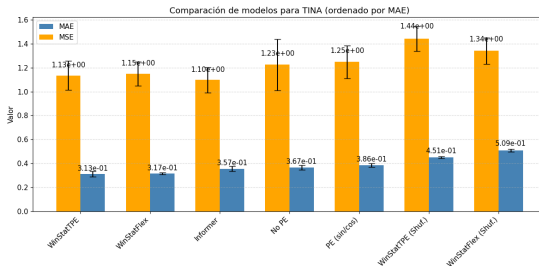


Figura: MSE (arriba) y MAE (abajo) obtenido para cada modelo

Evaluación. TINA



Modelo	MSE (Media)	MSE (STD)	MAE (Media)	MAE (STD)
Informer	1,0965	0,1064	0,3566	0,0200
WinStatTPE	1,1334	0,1216	0,3126	0,0215
WinStatFlex	1,1481	0,1012	0,3170	0,0070
No PE	1,2251	0,2137	0,3665	0,0179
PE (sin/cos)	1,2470	0,1375	0,3861	0,0164
WinStatTPE (Shuf.)	1,3399	0,1100	0,5094	0,1025
WinStatFlex (Shuf.)	1,4412	0,1025	0,4511	0,0061

Cuadro: Resultados para dataset TINA

Evaluación. Resultados finales

Modelo Métrica	Informer		PE (sin/cos)		WinStatFlex		WinStatTPE		No PE	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
HPC	0,5329	0,4152	0,5377	0,4140	0,4618	0,3734	0,4636	0,3743	0,5448	0,4213
ETTh1	0,5458	0,5324	0,5776	0,5552	0,4660	0,4947	0,4896	0,5072	0,9235	0,7419
ETTh2	0,8866	0,7549	1,3881	0,9355	0,4677	0,5128	0,5964	0,5889	1,0132	0,7777
Yellow Trip	$1,9 \times 10^{-5}$	$3,987 \times 10^{-3}$	$2,4 \times 10^{-5}$	$4,307 \times 10^{-3}$	$1,7 \times 10^{-5}$	$3,529 \times 10^{-3}$	$1,0 \times 10^{-5}$	$2,547 \times 10^{-3}$	$4,6 \times 10^{-5}$	$4,626 \times 10^{-3}$
TINA	1,0965	0,3566	1,2470	0,3861	1,1481	0,3170	1,1334	0,3126	1,2251	0,3665
Count	1		0		6		3		0	

Cuadro: Resumen final de resultados

Contenido

- 1 Introducción
- 2 Estado del arte
- 3 Propuestas de PE
- 4 Evaluación de PE mediante conjuntos de datos
- 5 Conclusiones**
 - Conclusiones
 - Trabajos futuros

El proyecto ha conseguido:

- Detectar el principal problema de la falta de localidad, asociada al encoding escaso proporcionado por el enfoque tradicional.
- Demostrar empíricamente los resultados en varios conjuntos de datos sobre los que han sido evaluados.
- Proponer un enfoque sencillo de comprender, mediante estadísticos básicos y ponderaciones normalizadas.
- Obtener, del conjunto de propuestas, un conjunto de PE alternativos que mejoren los resultados, como WinStatFlex, gracias a la creación de un entorno de pruebas estable.

- Aplicación de las técnicas expuestas en la clasificación de series temporales.
- Adaptación de los mecanismos de localidad para su uso en la detección de anomalías⁴.
- Mejorar el aprovechamiento de recursos: aprendizaje federado para la optimización de recursos y privacidad del dato.

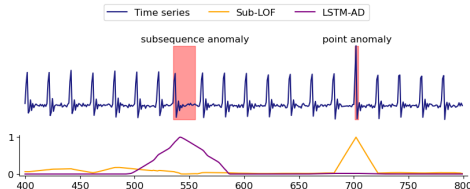


Figura: Detección de anomalías en series temporales

⁴<https://timeeval.github.io/evaluation-paper/>

Gracias



¿Preguntas?