WinStat: Familia de Codificaciones Posicionales Entrenables para Transformers en el Forecasting de Series Temporales

Cristhian Moya-Mota[1 [0009−0001−4839−469X]], Ignacio Aguilera-Martos[1,2 [0000−0001−9620−5235]], Diego García-Gil[1,3 [0000−0002−1927−8673]], and Julián Luengo[1,2 [0000−0003−3952−3629]]

[1] Instituto Andaluz Interuniversitario en Ciencia de Datos e Inteligencia Computacional (DaSCI), Universidad de Granada, 18012 Granada, España
cris190402@correo.ugr.es, {nacheteam, djgarcia, julianlm}@ugr.es
[2] Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, 18012 Granada, España
[3] Departamento de Lenguajes y Sistemas Informáticos, Universidad de Granada, 18012 Granada, España

Resumen Este resumen está basado en el artículo [4]. Presenta la propuesta *WinStat*, una familia de codificaciones posicionales que utiliza estadísticas de ventanas locales para el *forecasting* de series temporales mediante Transformers. Las codificaciones clásicas carecen de adaptabilidad o imponen costes prohibitivos. WinStat resuelve esto incorporando semántica temporal local explícita en cada representación. Sus variantes híbridas (WinStatFlex, WinStatTPE) aprenden mezclas ponderadas de estadísticas locales junto a señales absolutas y semánticas. La evaluación en cuatro *benchmarks* demuestra que capturar dinámicas de rezago y momentos estadísticos reduce el error cuadrático medio y absoluto medio entre un 10 % y un 30 % frente al estado del arte, validando su robustez mediante pruebas de *shuffling*.

Keywords: Series temporales · Forecasting · Positional Encoding · embedding · contexto local

1. Introducción y Limitaciones del Estado del Arte

El *forecasting* de series temporales (TSF) es vital en sectores como la energía [5] y el clima. Aunque los Transformers [7] mitigaron los problemas de memoria a largo plazo de las redes recurrentes, introdujeron un cuello de botella: la inyección de información posicional [9]. Dado que la auto-atención es invariante a las permutaciones, enfoques previos han empleado codificaciones absolutas (APE) que carecen de adaptabilidad [7], entrenables (LPE) que sobreajustan [1], o relativas que exigen memoria cuadrática. Modelos híbridos como TUPE [3] o RoPE [6] aportan expresividad, pero no asimilan bien las dinámicas locales del TSF multivariante, lo que explica por qué modelos lineales simples a veces superan a Transformers complejos [8].

2. La Familia WinStat: Posición Semántica Local

Para enriquecer la semántica local, *WinStat* inyecta estadísticos extraídos de la propia serie temporal original X. Bajo la premisa de estacionariedad local, y partiendo los datos originales, la variante base extrae ventanas de manera deslizante W_t, con longitud fija w, y centrada con respecto a la posición t. De ella, extrae un vector de momentos $s_t = [\mu_t, \sigma_t, m_t^{\text{mín}}, m_t^{\text{máx}}]$. Para cada (x_t), genera un *embedding* enriquecido: $\tilde{x}_t = [x_t \,\|\, s_t]$, obteniendo así la serie enriquecida $\tilde{X}$. Finalmente, para capturar la velocidad de cambio, la variante *WinStatLag* añade características adicionales mediante diferencias de rezago explícitas $\delta_t^{(\ell_j)} = |x_t - x_{t-\ell_j}|$, donde el parámetro ℓ_j representa la distancia temporal al instante actual t, consolidando toda la información en la serie temporal definitiva X^{final}.

WinStatFlex generaliza esto ponderando múltiples codificaciones simultáneas (X^{PE}, X^{LPE}, X^{tAPE}) mediante una distribución aprendida $\tilde{w} = \text{Softmax}(w)$:

$$X^{final} = \tilde{w}_0 \cdot \tilde{X} + \tilde{w}_1 \cdot X^{PE} + \tilde{w}_2 \cdot X^{LPE} + \tilde{w}_3 \cdot X^{tAPE} \tag{1}$$

WinStatTPE reemplaza X^{tAPE} por un núcleo de similitud gaussiana [2]. Teóricamente, este enriquecimiento altera la geometría de la atención: al depender de distribuciones estadísticas, la formulación introduce un `margen semántico determinista` que separa *queries* y *keys* de posiciones con valores crudos engañosamente idénticos pero regímenes distintos.

3. Resultados y Discusión Analítica

Para aislar empíricamente el efecto de la codificación posicional, se emplea un *backbone* tipo Informer [9] con auto-atención completa (vanilla) $\mathcal{O}(L^2)$, garantizando que no se distorsione la localidad espacial de las estadísticas de ventana.

3.1. Configuración Experimental y Protocolo de Evaluación

El protocolo de validación consiste en una partición tipo *last block* (70/30) aplicada sobre cuatro entornos heterogéneos: HPC, ETT, NYC y TINA. Asimismo, se utiliza la heurística $w \in [\lfloor L/4 \rfloor, \lfloor L/3 \rfloor]$ para la estabilización de la ventana estadística. Bajo estas condiciones, las configuraciones WinStat se evalúan frente a modelos de referencia como el Informer original, APE sinusoidal pura, la omisión de PE (No PE), TUPE [3] y RoPE [6].

3.2. Superioridad Predictiva (MSE/MAE)

En la comparativa global, `WinStatFlex` y `WinStatTPE` emergen como las arquitecturas dominantes. En escenarios como HPC y ETTh1, ambas variantes reducen notablemente el error cuadrático medio (MSE), frente a Informer base, logrando reducciones medias ente 10 % – 30 %, y alcanzando hasta un 90 % de mejora en contextos de alta precisión. Para ETTh2, el *ranking* evidencia las

carencias de otros enfoques del estado del arte: WinStatFlex lidera, mientras que otros como ROPE quedan relegados a la última posición. El comportamiento de la familia WinStat se adapta a la naturaleza subyacente del problema:

- En los registros de movilidad urbana (NYC), `WinStat base` logra el mejor rendimiento, subrayando que este escenario es particularmente sensible a la inyección pura de información estadística local.
- En el entorno industrial de TINA, `WinStatLag` despunta sobre el resto, evidenciando también el potencial bajo alta dimensionalidad.

Para formalizar estadísticamente estos hallazgos, se aplica un Test Bayesiano de Signos y Rangos con un intervalo de equivalencia práctica (ROPE threshold) de 0,05. Este análisis certifica que WinStatFlex marca diferencias significativas a los esquemas *No PE*, *APE* (sinusoidal) y al *Informer* original, desplazando toda la masa de probabilidad fuera de los límites de equivalencia.

3.3. Análisis de Pesos de Ponderación

El mecanismo de atención *Softmax* integrado en WinStatFlex y WinStatTPE permite inspeccionar qué componentes dominan la representación. El análisis arroja un patrón concluyente: el componente estadístico (*Stats*) adquiere consistentemente el mayor peso relativo. En entornos altamente ruidosos y masivos como TINA, el tensor estadístico abarca el `38,77 %` de la importancia en WinStatFlex. En HPC, capturó un `32,09 %`. Este hallazgo ratifica que la distribución estadística del contexto local ofrece descriptores predictivos muy superiores al mero índice temporal geométrico.

3.4. Test de Coste Computacional

Integrar cálculo de ventanas conlleva costes. El perfilado de GPU en las series ETTh demuestra que WinStatFlex y WinStatTPE incrementan el tiempo por época un $\sim 33\,\%$ respecto a Informer. Sin embargo, el consumo de memoria VRAM se mantiene constante, ya que el cálculo dinámico en ventana no exige retención a largo plazo de datos auxiliares.

4. Validación Semántica (Shuffling)

La asimilación del orden temporal se verifica perturbando aleatoriamente las entradas del decodificador [8]. Mientras que un modelo con fuerte dependencia cronológica debería colapsar, los modelos de referencia (Informer base, TUPE y ROPE) apenas sufren fluctuaciones o muestran mejoras anómalas, evidenciando la debilidad de su información secuencial.

Por el contrario, la familia WinStat sufre penalizaciones drásticas. En el dataset NYC, la degradación del MSE para WinStatFlex frente al desorden de la secuencia es de una magnitud de 10^2. Esta vulnerabilidad frente al shuffling demuestra empíricamente que WinStatFlex y WinStatTPE han acoplado profundamente sus predicciones a la estructura ordinal y a la coherencia de la semántica temporal inyectada.

5. Conclusiones

A diferencia de los Transformers tradicionales, cuyos esquemas posicionales ignoran la realidad del *dataset*, la familia *WinStat* corrige este déficit integrando matemáticamente el contexto estadístico local (s_t) y la inercia dinámica (d_t) en el PE. Mediante ponderaciones normalizadas y entrenables, WinStatFlex adapta su foco, produciendo caídas del MSE de entre un 10 % y un 30 % sobre líneas base robustas, consolidando un margen semántico que clarifica el espacio atencional.

Agradecimientos Esta investigación es parte del Proyecto TSI-100927-2023-1, financiado por Plan de Recuperación, Transformación y Resiliencia Europeo, de la Unión Europea-Next Generation EU a través del Ministerio para la Transformación Digital y de la Función Pública y está parcialmente financiada por el Ministerio de Ciencia y Tecnología de España en el marco del proyecto PID2023-150070NB-I00, financiado por MICIU/AEI/10.13039/501100011033 y por FEDER, UE.

Conflicto de Intereses Los autores declaran no tener conflictos de intereses.

Referencias

1. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT. pp. 4171–4186 (2019)
2. Irani, H., Metsis, V.: Positional encoding in transformer-based time series models: A survey. arXiv preprint arXiv:2502.12370 (2025)
3. Ke, G., He, D., Liu, T.Y.: Rethinking positional encoding in language pre-training. In: International Conference on Learning Representations (ICLR) (2021)
4. Moya-Mota, C., Aguilera-Martos, I., García-Gil, D., Luengo, J.: Winstat: A family of trainable positional encodings for transformers in time series forecasting. Machine Learning and Knowledge Extraction **8**(1) (2026)
5. Palma, G., Chengalipunath, E.S.J., Rizzo, A.: Time series forecasting for energy management: Neural circuit policies (ncps) vs. long short-term memory (lstm) networks. Electronics **13**(18), 3641 (2024)
6. Su, J., Lu, Y., Pan, S., Wen, B., Liu, Y.: Roformer: Enhanced transformer with rotary position embedding. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. pp. 683–692. Association for Computational Linguistics (2021)
7. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems. vol. 30 (2017)
8. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? In: AAAI Conference on Artificial Intelligence. vol. 37, pp. 11121–11128. AAAI (2023)
9. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. In: AAAI Conference on Artificial Intelligence. vol. 35, pp. 11106–11115 (2021)