

Campo de Transición de Markov y Aprendizaje profundo para codificar y Prever datos de Series Temporales Multivariantes de COVID-19

Moroco Ramos, Luis Angel; Velazco Paredes, Yuber¹

Abstract

La modelización de procesos del mundo real, como la calidad del aire, suele ser una tarea difícil debido tanto a su fenómeno caótico y no lineal como a la elevada dimensión del espacio muestral. La información recogida puede ayudar a los organismos gubernamentales, los responsables políticos y los particulares a preparar medidas de precaución para eliminar el impacto de los episodios de contaminación atmosférica. Si bien Deep Learning se desarrolló con éxito en este ámbito, la selección de la arquitectura de red sigue siendo una tarea problemática y que requiere mucho tiempo a la hora de desarrollar un modelo para una situación práctica. Este trabajo plantea el uso de Markov Transition Field (MTF) para codificar los datos de Series Temporales multivariadas, para poder procesarlas y extraer patrones profundos usando Deep Learning. Los primeros resultados, usando técnicas estadísticas como ARIMA, mostraron un bajo rendimiento.

1. Introducción

La previsión de la calidad del aire es uno de los temas de investigación sobre calidad del aire en la actualidad debido a la contaminación atmosférica urbana y, en concreto, a los episodios de contaminación, es decir, a las altas concentraciones de contaminantes que provocan efectos adversos para la salud e incluso muertes prematuras entre grupos sensibles como los asmáticos y las personas mayores (Tiittanen et al., 1999).

Esta previsión es fiable y eficaz para controlar las medidas y puede sugerirse como acción preventiva y evasiva para las normativas que se vayan a aplicar. En las Series Temporales (TS), las observaciones históricas se analizan para desarrollar un modelo que describa la relación entre el tiempo y las variables y que se utilizará para extrapolar las TS en el futuro (Niska et al., 2004).

El problema Multivariate Time Series Forecasting (MTSF) tiene como objetivo pre-dictaminar simultáneamente los valores futuros de N variables dados sus datos de ST pasadas

modelando con precisión las dependencias inter e intra variables (Chauhan et al., 2022). Tiene muchas aplicaciones importantes, como la previsión de la demanda de artículos vendidos por minoristas (Bandara et al., 2019), el flujo de tráfico (Li et al., 2017), la demanda y el suministro de energía (Saxena et al., 2019), o la matriz de covarianza, la volatilidad, la detección de anomalías (Li & Jung, 2023) y las distribuciones de cola larga en finanzas (Yan et al., 2020). Como tal, se trata de un área muy estudiada con su propia comunidad de investigación especializada. El problema fue abordado con modelos estadísticos y algoritmos de Machine Learning (ML) tradicionales; sin embargo, los problemas modernos a menudo presentan grandes paneles de TS relacionadas, todas las cuales deben preverse simultáneamente (Benidis et al., 2021) y deben tener en cuenta aspectos como la tendencia y las variaciones estacionales la correlación entre valores observados cercanos en el tiempo (Lara-Benítez et al., 2021). Esto motivó el desarrollo de modelos especializados que puedan captar los patrones subyacentes de las TS, de modo que puedan extrapolarse al futuro con eficacia. Esto lo hizo susceptible de Deep Learning (DL), aunque no siempre fueron una herramienta estándar para abordar estos problemas. De hecho, su eficacia se ha considerado históricamente desigual (Benidis et al., 2021). Especialmente en datos de TS.

(Wang & Oates, 2015) se inspiraron en los recientes éxitos de las técnicas de aprendizaje supervisado y no supervisado en visión por ordenador, se plantearon el problema de codificar series temporales como imágenes para permitir a las máquinas reconocer, clasificar y aprender "visualmente" estructuras y patrones. La reformulación de las características de las TS como pistas visuales ha suscitado gran interés en la informática y la física. Es por ellos que presentaron tres representaciones para codificar TS como imágenes que denominamos Suma Angular Gramiana/Campo Diferencial (GASF/GADF) y Campo de Transición de Markov (MTF).

1.1. Series Temporales (TS)

Las TS se define como una secuencia de valores ordenados cronológicamente y observados a lo largo del tiempo. Aunque el tiempo es una variable que se mide de forma continua, los valores de una TS se muestrean a intervalos

constantes o irregulares (Torres et al., 2021).

1.2. Replantación Matemática

Las TS pueden ser univariantes (una variable dependiente del tiempo) o multivariantes (más de una variable dependiente del tiempo).

1.2.1. UNIVARIANTE

Se tiene: $y = y(t-L), \dots, y(t-1), y(t), y(t+1), \dots, y(t+h)$ es una TS Univariante con L datos históricos. Donde cada $y(t-i)$, para $i = 0, \dots, L$, representa el valor almacenado de la variable y en el tiempo $t-i$ (Torres et al., 2021). El proceso de previsión consistiría en estimar el valor de $y(t+1)$, denotado como $\hat{y}(t+1)$, buscando minimizar el error, típicamente representada como una función $y(t+1) - \hat{y}(t+1)$. También es posible cuando $h > 1$ y deseamos hallar los h valores siguientes a $y(t)$ que son $y(t+i)$ donde $i = 1, \dots, h$. La función de minimización sería $\sum_{i=1}^h (y(t+i) - \hat{y}(t+i))$ (Torres et al., 2021).

1.2.2. MULTIVARIANTE

Pueden expresarse como una matriz de TS Univariadas indexadas:

$$\begin{pmatrix} y_1(t-L) \dots y_1(t-1), y_1(t), y_1(t+1) \dots y_1(t+h) \\ y_2(t-L) \dots y_2(t-1), y_2(t), y_2(t+1) \dots y_2(t+h) \\ \vdots \\ y_n(t-L) \dots y_n(t-1), y_n(t), y_n(t+1) \dots y_n(t+h) \end{pmatrix} \quad (1)$$

Donde $y_i(t-m)$ identifica a cada conjunto de TS, con $i = \{1, 2, \dots, m\}$, siendo $m = \{0, 1, \dots, m\}$ los datos históricos y muestra actual; y, $m = \{-1, -2, \dots, -h\}$ los valores h futuros, donde $h > 0$ (Torres et al., 2021). En cuanto a la previsión, por lo general, sólo debe predecirse una TS y las demás se denominan TS independientes. Para esto, los métodos clásicos (ARIMA) (Zhang, 2003b), el Vector Auto-regressive Model (VAR) (Fan et al., 2015) y el Gaussian Process (GP) como enfoque bayesiano (puede aplicarse directamente), (Brahim-Belhouari & Bermak, 2004) la aplica para TS no estacionarias; estas, (Wu et al., 2020) asumen la estacionalidad y no se adaptan bien a los datos de series temporales multivariantes.

2. Previsión basada en Aprendizaje Profundo (DL)

Antes del auge de las técnicas de minería de datos, los métodos tradicionales utilizados para Previsión de Series de Tiempo (TSF) se basaban principalmente en modelos

estadísticos, como el suavizado exponencial (ETS) (Hyndman et al., 2008) y los métodos Box-Jenkins como ARIMA (Lara-Benítez et al., 2021). Estos modelos se basan en la construcción de funciones lineales a partir de observaciones pasadas recientes para proporcionar predicciones futuras, y se han utilizado ampliamente para tareas de previsión durante las últimas décadas (Hyndman et al., 2008). ML, por su parte, requiere un esfuerzo considerable por parte del usuario para construir características significativas, lo que suele llevar bastante tiempo y es propenso a errores (van den Hoogen et al., 2020). DL (Bloemheuvel et al., 2022) permite la extracción automática de características con la ayuda de múltiples capas que pueden utilizar procesos no lineales. Cuando se trata de Series Temporales Multivariantes (MTS), se debería poder explotar la estructura de dependencia entre las distintas TS (Benidis et al., 2021).

2.1. Redes Neuronales Convolucionales (CNN)

La CNN es un tipo de red neuronal feedforward capaz de extraer características de los datos con estructuras de convolución (Li et al., 2022). Fueron originalmente previstas para tareas de visión por ordenador (Li et al., 2019). Se consideran la tecnología más avanzada para muchas tareas de clasificación, como el reconocimiento de objetos, (Hochreiter & Schmidhuber, 1997) el reconocimiento del habla (Hinton et al., 2012) y el reconocimiento de patrones. (Acharya et al., 2018) Las CNN pueden extraer automáticamente características de datos brutos de alta dimensión con una topología de cuadrícula, como los píxeles de una imagen, sin necesidad de ingeniería de características (Lara-Benítez et al., 2021).

En cuanto a MTSF, no se han utilizado mucho en la bibliografía, ya que se ha dado mucha más importancia a las RNN. No obstante, varios trabajos han propuesto CNN como extractores de características solos o junto con bloques recurrentes para proporcionar previsiones (Lara-Benítez et al., 2021). Es decir, fueron parte del pipeline de previsión. Sin embargo, Markova (Markova, 2022) estudió la capacidad de las CNN 1D en MTSF. El modelo CNN 1D multivariante de múltiples pasos se elabora y entrena con el tipo de cambio histórico del EUR/USD. Los datos intradía en un formato de marco temporal de 5 minutos se transforman en una estructura tridimensional para preparar los datos para el ajuste de una red neuronal convolucional. Además de Wang (Wan et al., 2022), propone una Multivariate Temporal Convolutional Attention Network (MTCAN) basado en CNN, utilizando la convolución dilatada 1D como unidad básica para construir bloques asimétricos y, a continuación, la extracción de características se realiza mediante el mecanismo de autoatención para obtener finalmente los resultados de predicción.

2.2. Redes Neuronales Recurrentes (RNN)

RNN están diseñadas específicamente para tratar datos secuenciales, como secuencias de palabras en problemas relacionados con la traducción automática, datos de audio en el reconocimiento del habla o ST. Todos estos problemas presentan una característica común, y es que los datos tienen una dependencia temporal entre ellos (Torres et al., 2021). Son más sensible a las TS y tiene memoria en la transmisión de datos, (Liu et al., 2021) la entrada en el tiempo t necesita referirse al resultado en el tiempo $t - 1$, y se obtiene un valor de salida basado en una transformación de función compleja. En el caso de TS, la longitud de la secuencia de datos de entrada suele ser diferente del tamaño de la secuencia de datos de salida, que suele ser el número de muestras que hay que predecir (Torres et al., 2021).

2.2.1. RED DE MEMORIA A LARGO - CORTO PLAZO (LSTM)

LSTM es un tipo especial de RNN, que evita eficazmente el fenómeno de desaparición y explosión del gradiente (Cuanto más larga es la secuencia de entrada, más referencias temporales se necesitan, lo que hace que la red sea más profunda. Cuando la secuencia es larga, es difícil que el gradiente se propague desde la última secuencia a la anterior. Véase (Le & Zuidema, 2016)) de las RNN creando un efecto de retención entre la entrada y la retroalimentación.

En la actualidad, LSTM se utiliza ampliamente en el campo de MTSF (Liu et al., 2021). Shaheen (Alhirmizy & Qader, 2019), utiliza LSTM, motivado por las falencias de los métodos clásicos, para la previsión de España capital Madrid Calidad del aire utilizando un conjunto de datos que informa sobre el tiempo y el nivel de contaminación cada hora durante dos años desde 2015 hasta 2016 los datos incluyen la fecha-hora, la concentración de contaminación de los siguientes: SO₂, NO₂, NO, CO. Xu (Xu & Yoneda, 2021), propone un modelo de aprendizaje multitarea con autoencoder de memoria a corto plazo (LSTM) para predecir las series temporales de PM 2,5 en múltiples ubicaciones de la ciudad; esto, buscaba excavar implícita y automáticamente la relevancia intrínseca entre los contaminantes en diferentes estaciones, además del uso de Autoencoders apilados para codificar el patrón de evolución clave de los sistemas meteorológicos urbanos. Liu (Liu et al., 2019), propone un novedoso modelo llamado Adaptive Noise Reducer-stacked Auto-encoder-validating-AdaBoost-based LSTM (ANR-SAE-VALSTM). Aplicando un novedoso ANR para la eliminación del ruido de las TS, el uso de SAE para extraer características de las series temporales multivariantes eliminadas y el eliminador de características VALSTM para entrenar un predictor conjunto de prevención del ajuste excesivo. Fouladgar (Fouladgar & Främling, 2020), propone Forward and Backward Variable-sensitive LSTM (FBVS-

LSTM) que consta de dos mecanismos de decaimiento y algunos datos informativos. Las entradas del modelo son principalmente el indicador de omisión, los intervalos de tiempo de omisión tanto en la dirección hacia delante como hacia atrás y la tasa de omisión de cada variable. Al aprender por separado las características de cada parámetro, el modelo se adapta para hacer frente a la falta masiva; esto, mejoró la precisión en la Previsión. Said (Said et al., 2021), presenta un enfoque de aprendizaje profundo para predecir el número acumulado de casos de COVID-19 utilizando una Bidirectional LSTM(Bi-LSTM) aplicada a series temporales multivariantes, También en Fake News Detection (Islam et al., 2022). A su vez, Guo (Guo et al., 2021), combina un Multi-scale Residual-CNN (MRC)(Basado en la convolución unidimensional, que detecta de forma adaptativa características de diferentes escalas temporales y fusionarlas. Véase (Tu et al., 2016)) y LSTM (Aprendizaje de dependencias) llamado (MRC-LSTM) para implementar la Previsión del precio de cierre de Bitcoin. Bhanja (Bhanja & Das, 2021), propuso una Stacked LSTM (S-LSTM) para MTSF. Aquí se usaron seis técnicas diferentes de normalización y datos financieros como el mercado de valores. Ouham (Ouham et al., 2021) propone una CNN+LSTM para la previsión de carga de trabajo, uso de CPU, la memoria y el uso de la red. El método de autorregresión vectorial filtra las interdependencias lineales. Luego, se calculan los datos residuales y se introducen en la capa de red neuronal convolucional que extrae características complejas de cada uno de los componente. Finalmente, LSTM modela información temporal de tendencias irregulares en componentes de sST en el modelo híbrido propuesto. Sorkun (Sorkun et al., 2020) propuso una MTSF basado en LSTM. Encontraron el mejor modelo para datos globales de radiación solar basándose en una combinación de diferentes variables meteorológicas (como temperatura, humedad y nubosidad). Los resultados demuestran que el método de MTSF ofrece mejores resultados que el modelo univariante anterior. Y se constata que la temperatura y la nubosidad son los parámetros más eficaces para predecir la radiación solar futura (Liu et al., 2021).

LSTM es popular en MTSF, dado que optimiza el gradiente es adecuado para el análisis de datos de series largas (Liu et al., 2021). Y la investigación muestra que el rendimiento del modelo LSTM integrado suele ser mejor que el de un solo modelo.

3. Imágenes de Series Temporales

En (Wang & Oates, 2015) se presentan dos marcos para codificar TS. EL primero, es un Campo Angular Gramiano (GAF), representa las TS en un sistema de coordenadas polares en lugar de las típicas coordenadas cartesianas. En la matriz gramiana, cada elemento es en realidad el coseno de la suma de ángulos (Campanharo et al., 2011). Por

otro lado, el Campo de Transición de Markov (MTF), que es construir la matriz de Markov de bins cuantílicos tras la discretización y codificar la probabilidad de transición dinámica en una matriz cuasi-gramiana.

3.1. Campo de Transición de Markov (MTF)

(Wang & Oates, 2015) extiende la representación de (Campanharo et al., 2011) representando las ideas de transición de markov secuencialmente para preservar la información en el dominio temporal.

Dada una TS X , identificamos sus Q intervalos cuantílicos y asignamos cada x_i al intervalo q_j correspondiente ($j \in [1, Q]$). Así, construimos una matriz de adyacencia ponderada QQ llamada W contando las transiciones entre los intervalos de cuantiles a la manera de una cadena de Markov de primer orden a lo largo del eje temporal. $w_{i,j}$ viene dada por la frecuencia con la que un punto del cuantil q_j es seguido por un punto del cuantil q_i . Tras la normalización por $\sum_j w_{i,j} = 1$, W es la matriz de transición de Markov. Es insensible a la distribución de X y a la dependencia temporal de los pasos temporales t_i . (Wang & Oates, 2015) propone:

$$M = \begin{bmatrix} w_{ij|x_1 \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_1 \in q_i, x_n \in q_j} \\ w_{ij|x_2 \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_2 \in q_i, x_n \in q_j} \\ \vdots & \ddots & \vdots \\ w_{ij|x_n \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_n \in q_i, x_n \in q_j} \end{bmatrix} \quad (2)$$

Construimos una matriz de transición de Markov QQ llamado W dividiendo los datos (magnitud) en Q intervalos cuantílicos. Los intervalos cuantílicos que contienen los datos en el momento i y j (eje temporal) son q_i y q_j ($q \in [1, Q]$). M_{ij} en el MTF denota la probabilidad de transición de $q_i \rightarrow q_j$. Es decir, extendemos la matriz W que contiene la probabilidad de transición en el eje de magnitud en la matriz MTF considerando las posiciones temporales (Wang & Oates, 2015).

Al asignar la probabilidad del cuantil en el paso temporal i al cuantil en el paso temporal j en cada píxel M_{ij} , la MTF M codifica en realidad las probabilidades de transición de varios intervalos de la serie temporal $M_{i,j||i-j|=k}$ denota la probabilidad de transición entre los puntos con intervalo de tiempo k . Por ejemplo, $M_{ij|j-i=1}$ ilustra el proceso de transición a lo largo del eje temporal con un paso de salto. La diagonal principal M_{ii} , que es un caso especial cuando $k = 0$, captura la probabilidad de cada cuantil hacia sí mismo (la probabilidad de autotransición) en el paso de tiempo i . Para que el tamaño de la imagen sea manejable y el cálculo más eficiente, reducimos el tamaño de la MTF promediando los píxeles de cada parche mm no solapado con el kernel de desenfoque $\{\frac{1}{m^2}\} m.m$. Es decir, se suman

las probabilidades de transición de cada subsecuencia de longitud m .

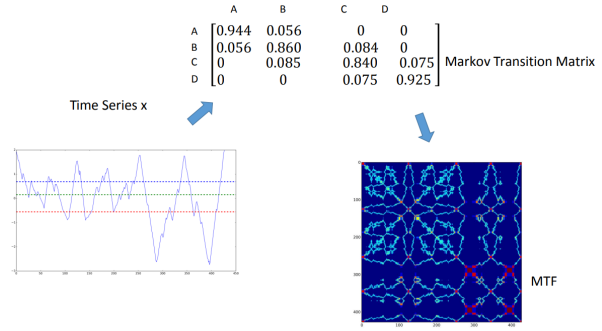


Figure 1. X es una secuencia de TS en el conjunto de datos "ECG". X se divide primero en Q intervalos cuantílicos. A continuación, se calcula la matriz de transición de Markov W y, por último, construimos su MTF con la ecuación (2) (Wang & Oates, 2015).

4. Trabajos Relacionados

(Gamboa, 2017) propone un marco donde las características aprendidas por la CNN se utilizan como entrada a un Modelo de Markov Oculto, consiguiendo una caída en la tasa de error superior al 10%. La aplicación de CNNs en estos trabajos presupone la restricción de que las TS están compuestas por un único canal. Por otro lado (Wang & Oates, 2015) aplica CNN para clasificar TS utilizando representaciones GAF y MTF en 20 conjuntos de datos en diferentes dominios como medicina, entomología, ingeniería, astronomía, procesamiento de señales y otros; luego, compararon la tasa de error de clasificación del enfoque GASF-GADF-MTF donde obtuvieron resultados competitivos para la clasificación en comparación con los mejores métodos recientes. La imputación utilizando GASF logró un rendimiento mejor y más estable que en los datos brutos. En (Gamboa, 2017) también se propone CNN en mosaico para realizar clasificaciones a partir de las imágenes. En los experimentos presentados, ambos métodos (GAF, MTF) se evalúan por separado en 12 conjuntos de datos "difíciles" "en los que la tasa de error de clasificación es superior a 0,1 con el enfoque SAX-BoP de última generación" (Lin et al., 2012), que son 50Words, Adiac, Beef, Coffee, ECG200, Face (all), Lightning-2, Lightning-7, OliveOil, OSU Leaf, Swedish Leaf y Yoga. Ellos usaron los métodos como colores de imágenes, obteniendo resultados competitivos. (Luque Manani & Lopez del Alamo, 2019) propusieron GAFs para codificar TS en imágenes de textura 2D, aprovechando posteriormente la arquitectura CNN-LSTM profunda donde LSTM utiliza una CNN como frontend, como marco novedoso de MTSF. Se obtuvieron resultados prometedores frente a los enfoques más modernos.

5. Metodología

Para probar la efectividad de codificar los datos de TS en un modelo de predicción, debemos conocer:

5.1. Dataset

Air Pollution Forecasting - LSTM Multivariate fue extraído de Kaggle, este dataset comprende más de una variable, lo que contribuirá a la precisión del modelo.

Se trata de un conjunto de datos que informa sobre el tiempo y el nivel de contaminación **cada hora durante cinco años** en la embajada estadounidense de Pekín (China). Comprende:

Table 1. Información de las variables

VARIABLE	NRO ROWS	TYPE
DATE	43800	OBJECT
POLLUTION	43800	FLOAT64
DEW	43800	INT64
TEMP	43800	FLOAT64
PRESS	43800	FLOAT64
WND_DIR	43800	OBJECT
WND_SPD	43800	FLOAT64
SNOW	43800	INT64
RAIN	43800	INT64

Este modelo buscará predecir la variable **pollution**.

5.1.1. EXPLORACIÓN DE DATOS

PLOT de la variable pollution:

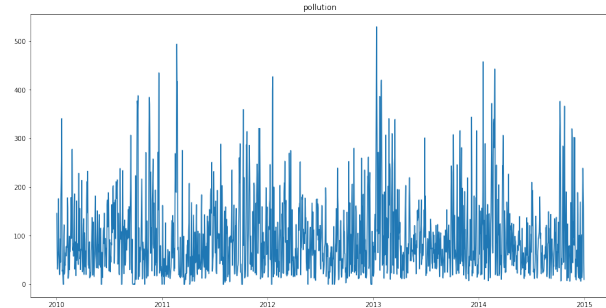


Figure 2. Visualización de datos *POLLUTION* del dataset Air Pollution de Kaggle.

En cuanto a las demás variables, tomando una ventana de datos de 100:

Visualizando los datos completos:

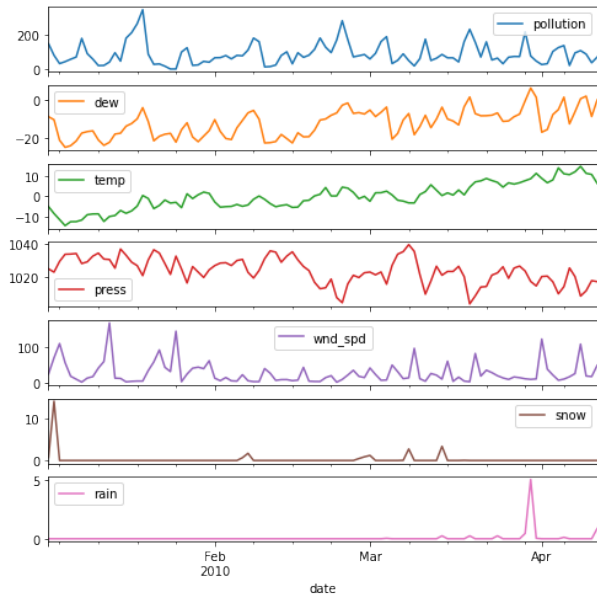


Figure 3. Visualización de todos los datos del dataset Air Pollution de Kaggle. *window* = 100

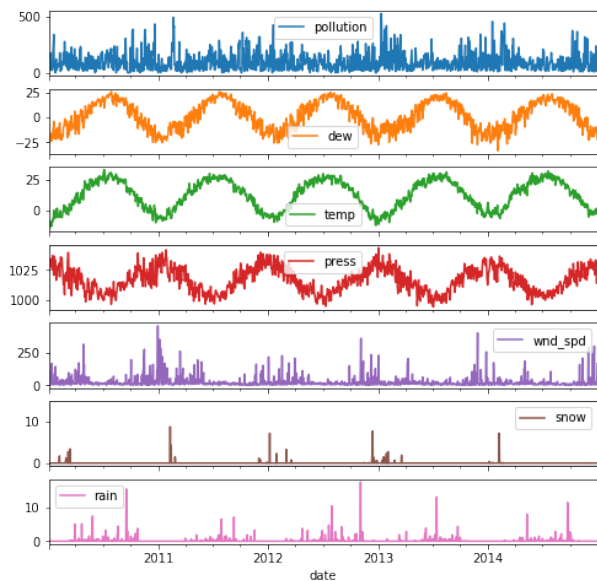


Figure 4. Visualización de datos *POLLUTION* del dataset Air Pollution de Kaggle. Datos completos

Puede verse claramente una estacionalidad en los datos.

5.2. Modelos

5.2.1. ARIMA

ARIMA (Ho & Xie, 1998) puede reflejar los cambios de diferentes patrones de datos y requiere menos parámetros.

Existen muchos métodos para estimar los parámetros del modelo ARIMA (Janacek, 2010), como la estimación por mínimos cuadrados (LSE), la estimación de máxima verosimilitud (MLE), la estimación bayesiana, etc (Liu et al., 2021). Tuvo buenos resultados, como Ho (Ho & Xie, 1998) presenta un ejemplo ilustrativo sobre los fallos de un sistema mecánico. También se realiza una comparación con el modelo Duane tradicional con resultados satisfactorios. Por su parte, (Zhang, 2003a) propone una metodología híbrida que combina los modelos ARIMA y ANN para aprovechar las ventajas de ambos en la modelización lineal y no lineal; la cual, reporta una mejora en conjunto con respecto a los métodos por separado. Volkan (Ediger & Akar, 2007), propone el uso de ARIMA y ARIMA estacional (SARIMA) para estimar la futura demanda de energía primaria de Turquía de 2005 a 2020; esto, concluyó en la recomendación política al gobierno en cuanto a seguridad energética. Valipour (Valipour, 2012), usó ARIMA para predecir las precipitaciones en zonas clave de Irán basándose en 50 años de datos de precipitaciones, y demostró su eficacia anual (Liu et al., 2021).

Este enfoque trabaja muy bien con datos de TS univariantes, es decir, con solo una variable predictoría. En este caso, se usó un paquete de **pmdarima**, que brinda un módulo para hacer MTSE. Los resultados se muestran a continuación.

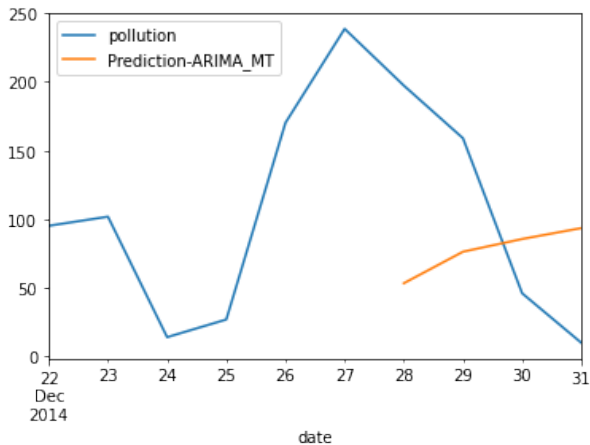


Figure 5. Resultados PMDARIMA en cuanto a MTSE

Como lo dice la literatura, los enfoques tradicionales no poseen la capacidad suficiente para encontrar patrones profundos en los datos multivariados, lo que resulta en un bajo rendimiento.

5.2.2. LSTM

El marco teórico y aplicaciones de LSTM se exponen en el punto 2.2.1. Por ahora, se entrenarán los datos con un modelo únicamente de LSTM.

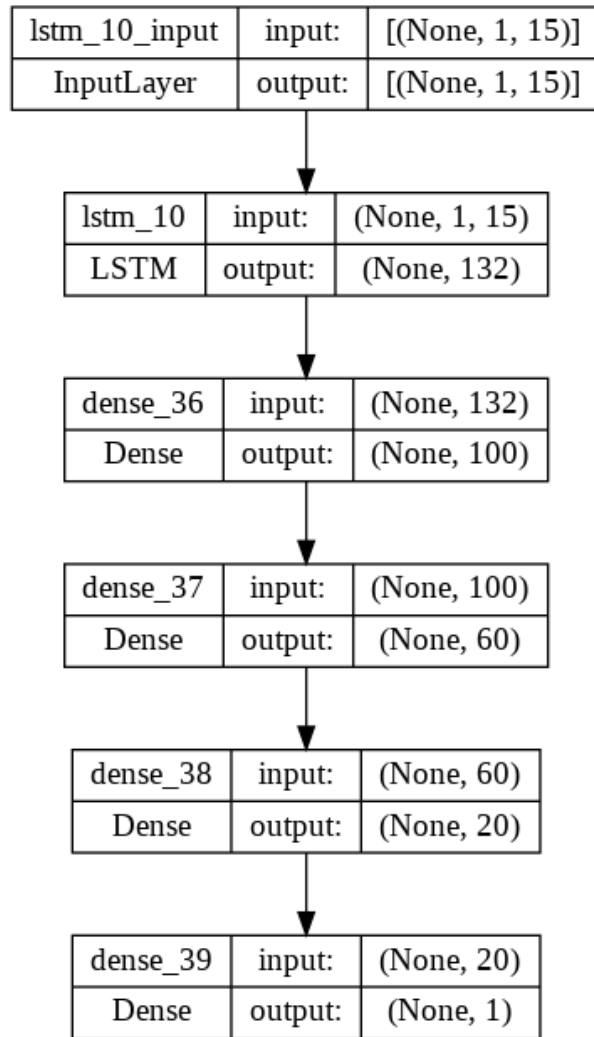


Figure 6. Arquitectura de LSTM

En cuanto al resultado MAE:

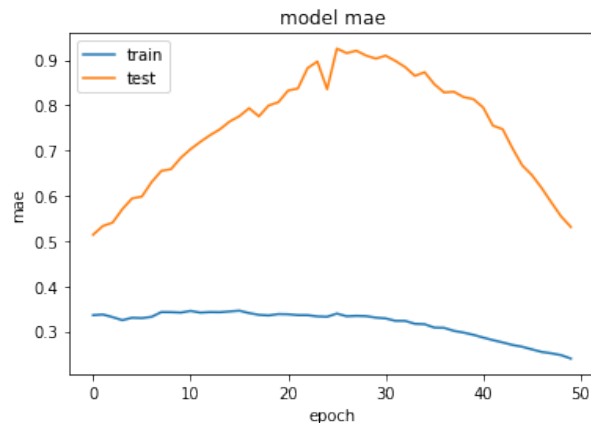


Figure 7. MAE: 0.2752

En cuanto al resultado LOSS:

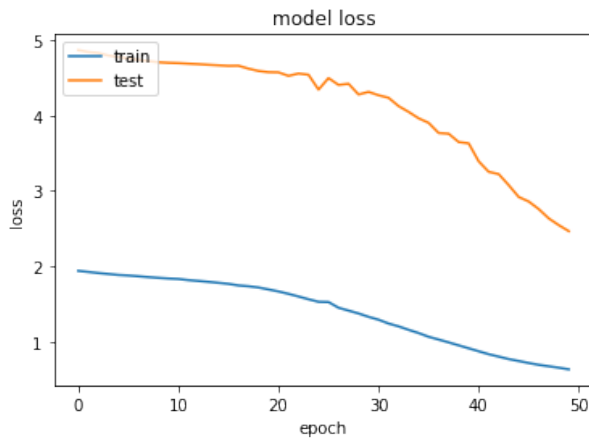


Figure 8. LOSS: 0.4236

6. Propuesta

En base a esto, se propone un marco que tendrá como base los datos de 'Air Pollution Forecasting - LSTM Multivariate' extraído de Kaggle, que será convertido a imágenes por el método de **Campo de Transición de Markov** propuesto por (Wang & Oates, 2015), después se pasará por una CNN 2D para extraer las características profundas y patrones. Finalmente, se usará la función `flatten()` para pasar a datos unidimensionales y ser procesados por **LSTM**.

References

- Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adeli, H., and Puthankattil, S. D. Automated eeg-based screening of depression using deep convolutional neural network. *Computer Methods and Programs in Biomedicine*, 161, 04 2018. doi: 10.1016/j.cmpb.2018.04.012.
- Alhirmizy, S. and Qader, B. Multivariate time series forecasting with lstm for madrid, spain pollution. In *2019 International Conference on Computing and Information Science and Technology and Their Applications (ICCISTA)*, pp. 1–5, 2019. doi: 10.1109/ICCISTA.2019.8830667.
- Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., and Seaman, B. Sales demand forecast in e-commerce using a long short-term memory neural network methodology. pp. 462–474, Berlin, Heidelberg, 2019. Springer-Verlag. ISBN 978-3-030-36717-6. doi: 10.1007/978-3-030-36718-3_39.
- Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, Y. B., Robinson, D., Turkmen, C., Gasthaus, J., Bohlke-Schneider, M., Salinas, D., Stella, L., Aubet, F.-X., Callot,

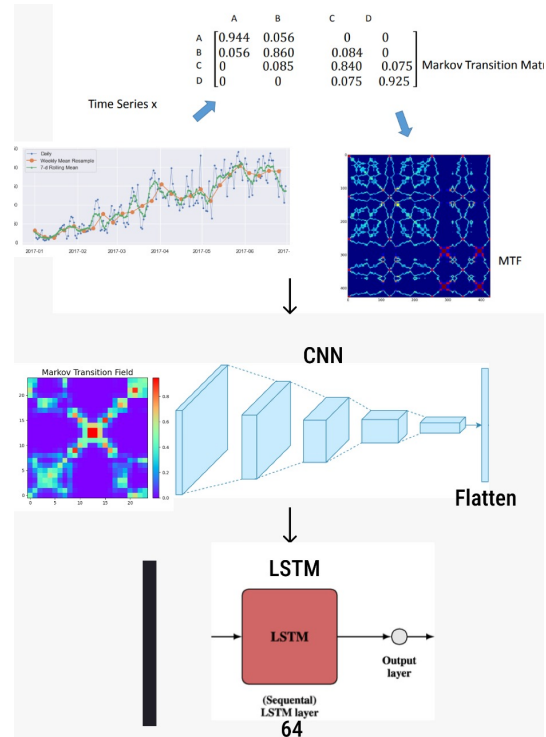


Figure 9. PROPUESTA

L., and Januschowski, T. Deep learning for time series forecasting: Tutorial and literature survey. *ACM Computing Surveys*, 2021.

- Bhanja, S. and Das, A. Deep neural network for multivariate time-series forecasting. In *Proceedings of international conference on frontiers in computing and systems*, pp. 267–277. Springer, 2021.
- Bloemheuvel, S., van den Hoogen, J., Jozinović, D., Michelini, A., and Atzmueller, M. Graph neural networks for multivariate time series regression with application to seismic data. *International Journal of Data Science and Analytics*, aug 2022. doi: 10.1007/s41060-022-00349-6.
- Brahim-Belhouari, S. and Bermak, A. Gaussian process for nonstationary time series prediction. *Computational Statistics & Data Analysis*, 47(4):705–712, 2004.
- Campanharo, A. S. L. O., Sirer, M. I., Malmgren, R. D., Ramos, F. M., and Amaral, L. A. N. Duality between time series and networks. *PLOS ONE*, 6(8):1–13, 08 2011. doi: 10.1371/journal.pone.0023378. URL <https://doi.org/10.1371/journal.pone.0023378>.
- Chauhan, J., Raghuveer, A., Saket, R., Nandy, J., and Ravindran, B. Multi-variate time series forecasting on variable subsets. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, aug 2022. doi: 10.1145/3534678.3539394.

- Ediger, V. Ş. and Akar, S. Arima forecasting of primary energy demand by fuel in turkey. *Energy policy*, 35(3): 1701–1708, 2007.
- Fan, F., Bell, K., Hill, D., and Infield, D. Wind forecasting using kriging and vector auto-regressive models for dynamic line rating studies. In *2015 IEEE Eindhoven PowerTech*, pp. 1–6, 2015. doi: 10.1109/PTC.2015.7232348.
- Fouladgar, N. and Främling, K. A novel lstm for multivariate time series with massive missingness. *Sensors*, 20(10): 2832, 2020.
- Gamboa, J. C. B. Deep learning for time-series analysis, 2017. URL <https://arxiv.org/abs/1701.01887>.
- Guo, Q., Lei, S., Ye, Q., and Fang, Z. Mrc-lstm: A hybrid approach of multi-scale residual cnn and lstm to predict bitcoin price. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, 2021.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., and Kingsbury, B. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012. doi: 10.1109/MSP.2012.2205597.
- Ho, S. L. and Xie, M. The use of arima models for reliability forecasting and analysis. *Computers & industrial engineering*, 35(1-2):213–216, 1998.
- Hochreiter, S. and Schmidhuber, J. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hyndman, R., Koehler, A. B., Ord, J. K., and Snyder, R. D. *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2008.
- Islam, T., Hosen, M. A., Mony, A., Hasan, M. T., Jahan, I., and Kundu, A. A proposed bi-LSTM method to fake news detection. In *2022 International Conference for Advancement in Technology (ICONAT)*. IEEE, jan 2022. doi: 10.1109/iconat53423.2022.9725937.
- Janacek, G. Time series analysis forecasting and control. *Journal of Time Series Analysis*, 31(4): 303–303, July 2010. doi: 10.1111/j.1467-9892.2009. URL <https://ideas.repec.org/a/bla/jtsera/v31y2010i4p303-303.html>.
- Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.
- Lara-Bení tez, P., Carranza-García, M., and Riquelme, J. C. An experimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems*, 31(03):2130001, feb 2021. doi: 10.1142/s0129065721300011.
- Le, P. and Zuidema, W. Quantifying the vanishing gradient and long distance dependency problem in recursive neural networks and recursive lstms, 2016. URL <https://arxiv.org/abs/1603.00423>.
- Li, G. and Jung, J. J. Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. *Information Fusion*, 91:93–102, 2023. ISSN 1566-2535. doi: <https://doi.org/10.1016/j.inffus.2022.10.008>.
- Li, S., Zhao, X., and Zhou, G. Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. *Computer-Aided Civil and Infrastructure Engineering*, 34(7): 616–634, 2019. doi: <https://doi.org/10.1111/mice.12433>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.12433>.
- Li, Y., Yu, R., Shahabi, C., and Liu, Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, 2017. URL <https://arxiv.org/abs/1707.01926>.
- Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):6999–7019, 2022. doi: 10.1109/TNNLS.2021.3084827.
- Lin, J., Khade, R., and Li, Y. Rotation-invariant similarity in time series using bag-of-patterns representation. *J. Intell. Inf. Syst.*, 39(2):287–315, oct 2012. ISSN 0925-9902. doi: 10.1007/s10844-012-0196-5. URL <https://doi.org/10.1007/s10844-012-0196-5>.
- Liu, F., Cai, M., Wang, L., and Lu, Y. An ensemble model based on adaptive noise reducer and overfitting prevention lstm for multivariate time series forecasting. *IEEE Access*, 7:26102–26115, 2019. doi: 10.1109/ACCESS.2019.2900371.
- Liu, Z., Zhu, Z., Gao, J., and Xu, C. Forecast methods for time series data: A survey. *IEEE Access*, 9:91896–91912, 2021. doi: 10.1109/ACCESS.2021.3091162.
- Luque Mamani, E. F. and Lopez del Alamo, C. GAF-CNN-LSTM for Multivariate Time- Series Images Forecasting. *LatinX in AI Research at ICML 2019*, June 2019. URL <https://hal.archives-ouvertes.fr/hal-02266994>. Poster.

- Markova, M. Convolutional neural networks for forex time series forecasting. In *AIP Conference Proceedings*, volume 2459, pp. 030024. AIP Publishing LLC, 2022.
- Niska, H., Hiltunen, T., Karppinen, A., Ruuskanen, J., and Kolehmainen, M. Evolving the neural network model for forecasting air pollution time series. *Engineering Applications of Artificial Intelligence*, 17(2):159–167, 2004. ISSN 0952-1976. doi: <https://doi.org/10.1016/j.engappai.2004.02.002>. URL <https://www.sciencedirect.com/science/article/pii/S0952197604000119>. Intelligent Control and Signal Processing.
- Ouham, S., Hadi, Y., and Ullah, A. An efficient forecasting approach for resource utilization in cloud data center using cnn-lstm model. *Neural Computing and Applications*, 33(16):10043–10055, 2021.
- Said, A. B., Erradi, A., Aly, H. A., and Mohamed, A. Predicting covid-19 cases using bidirectional lstm on multivariate time series. *Environmental Science and Pollution Research*, 28(40):56043–56052, 2021.
- Saxena, H., Aponte, O., and McConky, K. T. A hybrid machine learning model for forecasting a billing period’s peak electric load days. *International Journal of Forecasting*, 35(4):1288–1303, 2019. ISSN 0169-2070. doi: <https://doi.org/10.1016/j.ijforecast.2019.03.025>.
- Sorkun, M., Incel, O., and Paoli, C. Time series forecasting on multivariate solar radiation data using deep learning (lstm). *Turkish Journal of Electrical Engineering and Computer Sciences*, 28:211–223, 01 2020. doi: 10.3906/elk-1907-218.
- Tiittanen, P., Timonen, K. L., Ruuskanen, J., Mirme, A., and Pekkanen, J. Fine particulate air pollution, resuspended road dust and respiratory health among symptomatic children. *Eur Respir J*, 13(2):266–273, February 1999.
- Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., and Troncoso, A. Deep learning for time series forecasting: A survey. *Big Data*, 9(1):3–21, 2021. doi: 10.1089/big.2020.0159. URL <https://doi.org/10.1089/big.2020.0159>. PMID: 33275484.
- Tu, Z., Cao, J., Li, Y., and Li, B. Msr-cnn: applying motion salient region based descriptors for action recognition. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 3524–3529. IEEE, 2016.
- Valipour, M. Critical areas of iran for agriculture water management according to the annual rainfall. *European Journal of Scientific Research*, 84:600–608, 08 2012.
- van den Hoogen, J., Bloemhevel, S., and Atzmueller, M. An improved wide-kernel cnn for classifying multivariate signals in fault diagnosis. In *2020 International Conference on Data Mining Workshops (ICDMW)*, pp. 275–283, 2020. doi: 10.1109/ICDMW51313.2020.00046.
- Wan, R., Tian, C., Zhang, W., Deng, W., and Yang, F. A multivariate temporal convolutional attention network for time-series forecasting. *Electronics*, 11(10):1516, 2022.
- Wang, Z. and Oates, T. Imaging time-series to improve classification and imputation. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.
- Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., and Zhang, C. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’20*, pp. 753–763, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450379984. doi: 10.1145/3394486.3403118.
- Xu, X. and Yoneda, M. Multitask air-quality prediction based on lstm-autoencoder model. *IEEE Transactions on Cybernetics*, 51(5):2577–2586, 2021. doi: 10.1109/TCYB.2019.2945999.
- Yan, X., Zhang, W., Ma, L., Liu, W., and Wu, Q. Parsimonious quantile regression of financial asset tail dynamics via sequential learning, 2020. URL <https://arxiv.org/abs/2010.08263>.
- Zhang, G. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003a. ISSN 0925-2312. doi: [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0).
- Zhang, G. P. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003b.