

Análisis de partículas PM 2.5 mediante redes LSTM

Zúñiga González Daniel Iván

Junio 2020

1 Introducción

La actividad diaria de la Ciudad de México genera una gran cantidad de sustancias que modican la composición natural del aire. La quema de combustibles fósiles para el transporte y la generación de energía, tanto a nivel industrial como doméstico, produce miles de toneladas de contaminantes que diariamente son emitidos a la atmósfera. Una manera de proteger la salud de la población es a través del monitoreo y la difusión continuos del estado de la calidad del aire.

En la Ciudad de México, el Sistema de Monitoreo Atmosférico (SIMAT) es el responsable de la medición permanente de los principales contaminantes del aire. La materia particulada o PM 2.5 (por sus siglas en inglés), son partículas muy pequeñas en el aire que tiene un diámetro de 2.5 micrómetros o menos. La materia particulada, uno de los seis criterios de contaminantes del aire de la U.S. EPA, es una mezcla que puede incluir sustancias químicas orgánicas, polvo, hollín y metales. Estas partículas pueden provenir de los automóviles, camiones, fábricas, quema de madera y otras actividades, y son particularmente dañinas para la salud, ya que se ha demostrado que causa muchos efectos serios, incluyendo enfermedades cardíacas y pulmonares pues pueden desplazarse más profundamente dentro de los pulmones cuando respiramos.

Las LSTM son un tipo especial de redes recurrentes. La característica principal de las redes recurrentes es que la información puede persistir introduciendo bucles en el diagrama de la red, por lo que, básicamente, pueden "recordar" estados previos y utilizar esta información para decidir cuál será el siguiente. Esta característica las hace muy adecuadas para manejar series cronológicas. Mientras las redes recurrentes estándar pueden modelar dependencias a corto plazo, las LSTM pueden aprender dependencias largas, por lo que se podría decir que tienen una "memoria" a más largo plazo. El presente trabajo es importante en el sentido en que muestra la utilidad de las redes LSTM como una alternativa al estudio de series temporales con modelos de la estadística clásica y proporciona una predicción de los niveles de PM 2.5 en el año 2020 para la toma de decisiones más informada en políticas ambientales en la Ciudad de México.

2 Objetivos

El presente documento tiene como objetivos:

- Promover el uso de redes neuronales artificiales como alternativa a los modelos estadísticos clásicos de análisis de series temporales;
- Proporcionar una predicción sobre el comportamiento de la materia particulada 2.5 en el año 2020 para la toma de decisiones en materia ambiental.

3 Materiales y Métodos

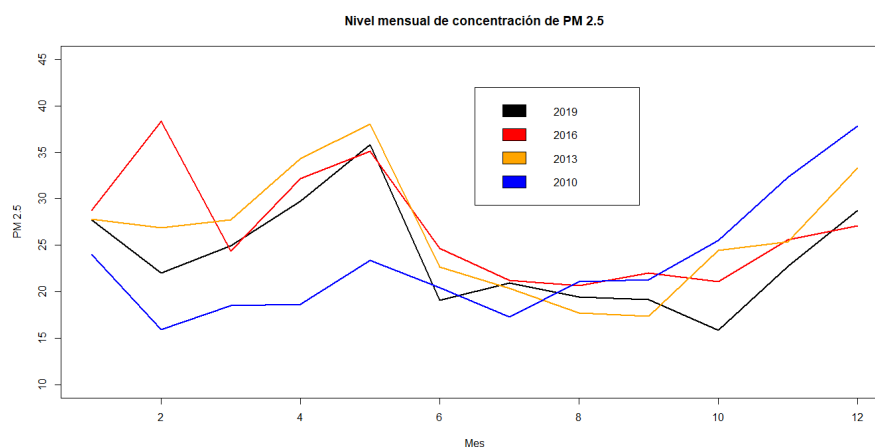
Se descargaron las bases de datos de acceso libre a través de la página web del gobierno de la Ciudad de México (fuente:). Dichas bases se encuentran en formato excel, de manera anualizada. Cada registro corresponde al nivel de partículas PM 2.5 suspendidas en el aire, captadas en cada una de las 24 horas del día, de manera automatizada, por las 34 estaciones de monitoreo de la Ciudad de México (21 estaciones) y del Estado de México (13 estaciones). Se consideraron los datos desde el primer día de enero del año 2010 hasta el último día de diciembre del año 2019 (10 años), de la información recabada por la estación de monitoreo MER, ubicada en el centro histórico de la ciudad. Esta estación permite el uso de datos del período de interés para el presente trabajo, no siendo posible en el caso de otras estaciones, ya que algunas de ellas entraron en función en años posteriores al 2010. Además, los datos provenientes de la misma, se encuentran en el rango de aquellos con menor cantidad de datos faltantes. Se realizó un análisis exploratorio. Los datos faltantes se sustituyeron por la media anual del año en cuestión, resaltando la importancia de poder completar dichos vacíos por la dependencia temporal de la serie, pues estos presentan estacionalidad en períodos de un año. Se promedió el nivel de PM 2.5 de manera mensual, obteniendo 120 registros correspondientes al período de análisis 2010-2019. Mediante el uso del software libre Python, se ajustó un modelo de red neuronal artificial Long-Short Term Memory (LSTM) con distintas estructuras, habiéndose considerado como la mejor de estas las de características siguientes:

- Doce entradas y una salida
- Una capa LSTM con 200 unidades y función de activación relu;
- Optimizador Adam
- Error cuadrático medio como función de pérdida
- Nueve años de entrenamiento (2010-2018) y un año para validación (2019)
- Ciento ochenta épocas

Tras el ajuste, se procedió a la predicción del nivel de PM 2.5 mensual del año 2020.

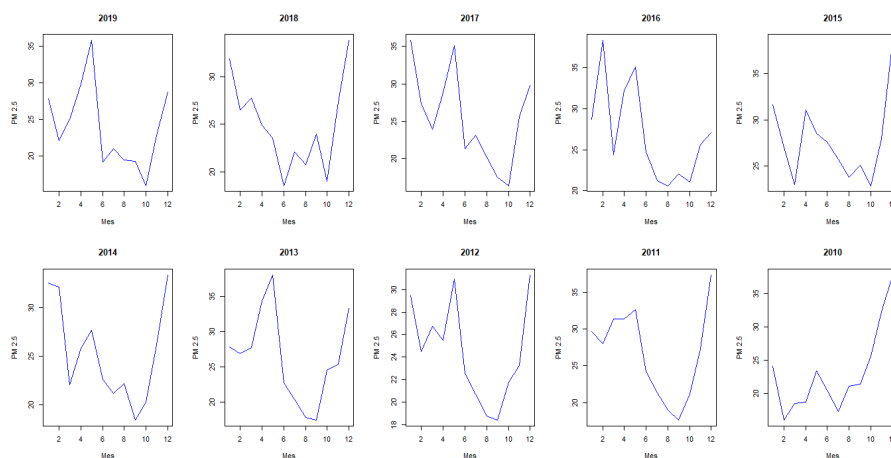
4 Resultados

En la siguiente gráfica se muestra el promedio mensual de PM 2.5 en los años 2010, 2013, 2018 y 2019.

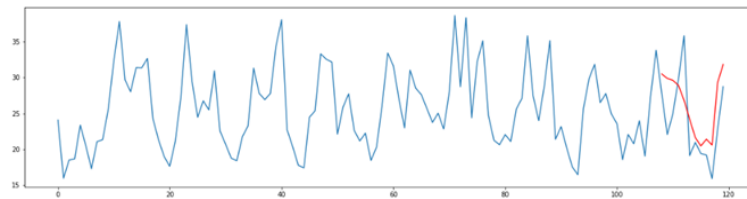
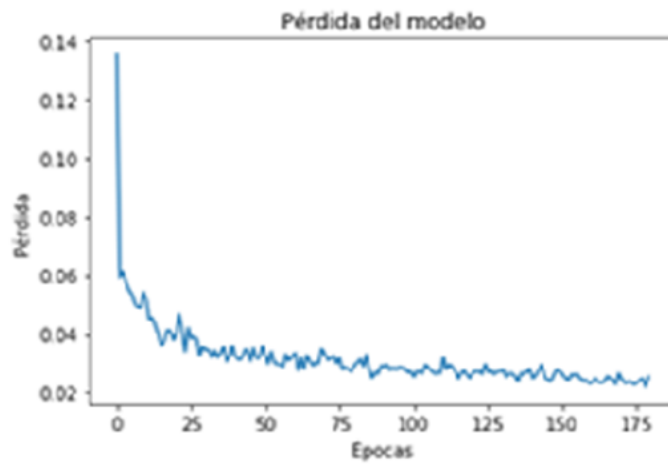


Como se puede observar, existe estacionalidad anual en la serie de datos. Hay una mayor acumulación de partículas en épocas de frío debido a la poca capacidad de dispersión de estas por causa de la temperatura, mientras que en verano, la lluvia distribuye o discipa los contaminantes.

Gráficamente, la serie temporal para cada año se ve de la siguiente manera:



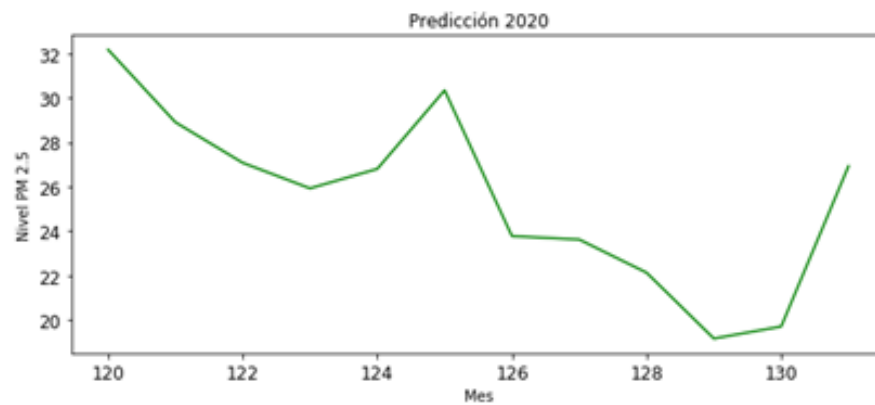
Tras el ajuste y compilación del modelo, se obtuvieron las siguientes gráficas de pérdida y de comparación entre el conjunto de validación original y estimación del modelo entrenado:

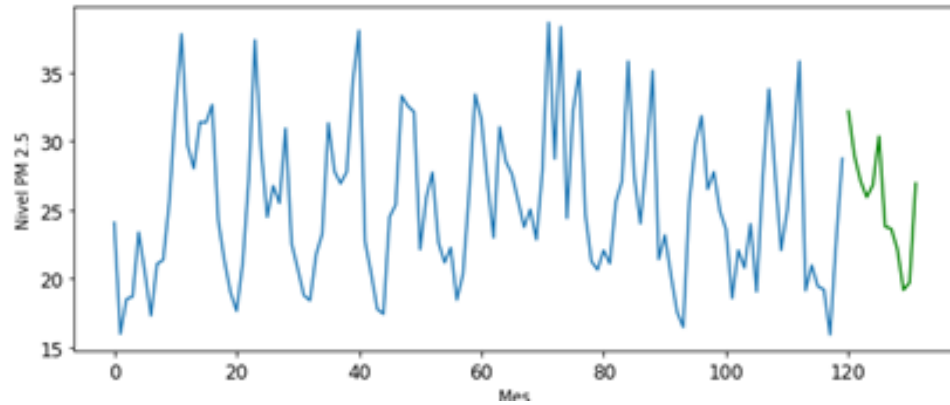


Finalmente, se obtuvieron las siguientes estimaciones mensuales para el año 2020:

Mes	Enero	Febrero	Marzo	Abril	Mayo	Junio
PM 2.5	32.1	28.9	27.0	25.9	26.7	30.3

Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
23.7	23.6	22.1	19.1	19.6	26.9





5 Conclusiones

A diferencia de los problemas más simples de clasificación y regresión, los problemas de series temporales agregan la complejidad del orden o la dependencia temporal entre las observaciones. Tradicionalmente, el pronóstico de series de tiempo ha estado dominado por métodos lineales como ARIMA porque se conocen bien y son efectivos en muchos problemas. Pero estos métodos tradicionales sufren algunas limitaciones.

Las redes neuronales aproximan una función de mapeo desde las variables de entrada a las variables de salida. Esta capacidad general es valiosa para series temporales por varias razones, por ejemplo:

- Las redes neuronales son resistentes al ruido en los datos de entrada y en la función de mapeo e incluso pueden soportar el aprendizaje y la predicción en presencia de valores perdidos;
- No hacen suposiciones fuertes sobre la función de mapeo y aprenden fácilmente relaciones lineales y no lineales;
- Se puede especificar un número arbitrario de características de entrada, lo que proporciona soporte directo para pronósticos multivariados.

Convirtiendo a las redes neuronales artificiales en una buena alternativa para el tratamiento de series temporales.

6 Bibliografía

Red Automática de Monitoreo Atmosférico (RAMA). (2020). Bases de datos. 10/06/2020, de Gobierno de la Ciudad de México Sitio web: <http://www.aire.cdmx.gob.mx/default.php?opc=%27aKBh%27>

González-Avella, J. C.. (2020). ANÁLISIS DE SERIES TEMPORALES USANDO REDES NEURONALES RECURRENTE. 10/06/20, de APSL Sitio

web: <https://www.apsl.net/blog/2017/06/14/analisis-de-series-temporales-usando-redes-neuronales-recurrentes/>

Dirección de Monitoreo Atmosférico. (2020). Estaciones de monitoreo. 10/06/20, de Gobierno de la Ciudad de México Sitio web: <http://www.aire.cdmx.gob.mx/default.php?opc=>

Anónimo. (S/F). Red Automática de Monitoreo Atmosférico (RAMA). 10/06/20, de Anónimo Sitio web: <https://www.eumed.net/tesis-doctorales/2011/smv/11-UnEncrypted.pdf>

Torres, J. (2020). Redes Neuronales Recurrentes. 12/06/20, de Jordi TORRES.AI Sitio web: <https://torres.ai/redes-neuronales-recurrentes/>

Calvo, D. (2018). Red Neuronal Recurrente – RNN. 12/06/20, de DIEGO CALVO Sitio web: <https://torres.ai/redes-neuronales-recurrentes/>

Centro de Monitoreo de la Calidad del Aire del Estado de Querétaro. (2017). Material Particulado. 09/06/20, de CemCAQ Sitio web: <http://www.cemcaq.mx/contaminacion/particulas-pm>

Anónimo. (S/F). Partículas suspendidas. 10/06/20, de Anónimo Sitio web: <http://www2.inecc.gob.mx/publicaciones2/libros/517/cap4.pdf>

RED AUTOMÁTICA DE MONITOREO ATMOSFÉRICO. (S/F). BASE DE DATOS DE LA RED AUTOMÁTICA DE MONITOREO ATMOSFÉRICO. 09/06/20, de Gobierno de la Ciudad de México Sitio web: <http://www.aire.cdmx.gob.mx/descargas/datos/excel/RAMAxls.pdf>

Espinosa Guzmán, A. A., et al. (2017). MODELADO DE PARTÍCULAS PM10 Y PM2.5 MEDIANTE REDES NEURONALES ARTIFICIALES SOBRE CLIMA TROPICAL DE SAN FRANCISCO DE CAMPECHE, MÉXICO. 10/06/20, de CONACYT Sitio web: <https://www.scielo.br/pdf/qn/v40n9/0100-4042-qn-40-09-1025.pdf>