

Aplicación de la inteligencia artificial al pronóstico de las variables meteorológicas en el aeropuerto de Vigo

Jorge ROBINAT VALENCIANO

Agencia Estatal de Meteorología

jrobinatv@aemet.es

Resumen: En este trabajo se describe cómo la inteligencia artificial (IA) puede mejorar la predicción de las variables meteorológicas en el aeropuerto de Vigo. Se parte de una base de datos histórica con los pronósticos de un modelo meteorológico y una serie de boletines meteorológicos aeronáuticos (METAR) en la que se registran las variables meteorológicas del aeropuerto. Se entrenan algoritmos de IA para pronosticar mejor que el modelo meteorológico las variables meteorológicas. El modelo meteorológico proporcionará las variables de entrada al algoritmo de IA. Al final del artículo se ofrecerá un enlace donde se muestran los pronósticos del algoritmo y los boletines del aeropuerto de Vigo en tiempo real.

Palabras clave: inteligencia artificial, modelo meteorológico, pronóstico aeronáutico.

1. MAPA DE SITUACIÓN Y DATOS DE PARTIDA

Los datos de partida son un modelo meteorológico y los boletines de aeropuerto (METAR). El modelo meteorológico que se quiere mejorar es el WRF con una resolución espacial de 4 km. Un archivo histórico con los pronósticos de este modelo desde el año 2008 se puede encontrar en el siguiente enlace de [Meteogalicia](#) [1]. Se escogen los 10 puntos más cercanos al aeropuerto tal como se detalla en la figura 1. Los puntos se enumeran de forma que el 0 corresponde al punto más cercano y el 9 al más lejano.

Se pueden distinguir dos tipos de variables pronosticadas por el modelo: las variables pronosticadas que se miden directamente en los METAR del aeropuerto como las relacionadas con el viento, la visibilidad horizontal, la cantidad de nubes, la precipitación, la temperatura o la



Figura 1. Puntos más cercanos al aeropuerto de Vigo según el modelo WRF 4 km.

presión; y las variables que no se miden en los boletines como la energía potencial convectiva disponible o la temperatura a 500 milibares de altura.

2. CLASIFICAR LAS VARIABLES METEOROLÓGICAS

Existen dos tipos de problemas en IA: los problemas de regresión y los problemas de clasificación. Este artículo se va a centrar solo en los problemas de clasificación. Para poder clasificar se deben agrupar las observaciones en categorías como en los siguientes ejemplos.

- La variable temperatura se puede agrupar en varias categorías que cubren el conjunto de las temperaturas posibles. Temperaturas: entre 0 y 3 grados, 3 y 5 grados, 6 y 7 grados, etc. Es posible estar interesados solamente en temperaturas que sobrepasen un determinado umbral. Por ejemplo, agrupar los datos en dos categorías: menor o igual de 0 grados y mayor de 0 grados. En algunos casos interesa pronosticar la ocurrencia o no de un determinado fenómeno meteorológico como la niebla o algún tipo de precipitación.
- Los boletines METAR aportan información meteorológica interesante desde el punto de vista aeronáutico como la altura y la cantidad de nubes o la visibilidad horizontal. Se puede clasificar la altura de las nubes más bajas en dos tipos: por encima de 1500 pies o por debajo de 1500 pies. La visibilidad horizontal también se puede dividir en dos apartados: menor o igual de 1000 metros y mayor de 1000 metros. Las condiciones de vuelo del aeropuerto (*Flight Rules*) también pueden ser objeto de su pronóstico por IA. Un artículo al respecto, aplicado al aeropuerto de Vigo, puede verse en el siguiente [enlace](#) [2].
- En los boletines METAR, es posible distinguir dos tipos de variables. Por un lado, las variables que se pronostican directamente por los modelos como las relacionadas con el viento, la temperatura, la cantidad de nubes o la visibilidad; y por otro lado, las variables como el tiempo presente que el modelo meteorológico no pronostica directamente.

En los siguientes apartados se presentan los resultados de dos algoritmos. Un algoritmo que predice una variable no pronosticada directamente por el modelo (la niebla o bruma) y un algoritmo que mejora la predicción del modelo meteorológico (visibilidad). En los dos casos se utilizará el modelo WRF con el análisis de las 00 UTC y un alcance de 24 horas.

3. UN ALGORITMO PARA PRONOSTICAR LA APARICIÓN DE LA NIEBLA O LA BRUMA

El algoritmo utiliza como variables de entrada algunas de las variables del modelo meteorológico. En este caso, el algoritmo usa las variables: visibilidad, intensidad de viento, cobertura nubosa y humedad relativa. Los datos de tiempo presente solo se emplean para entrenar el algoritmo. El algoritmo empleado es el LGB Classifier con 250 estimadores de la librería Sklearn. Se han usado 106 310 datos de los años 2008 a 2020. De estos 106 310 datos, 10 631 (un 10 %) se han usado en las verificaciones; el resto, para entrenar el algoritmo.

Los parámetros que miden la calidad de las predicciones empleados en este artículo serán: la precisión, la exhaustividad, el valor F y el *Heidke skill score* (HSS). La entropía medirá la cantidad de información aportada por los pronósticos.

Para poder medir esos parámetros se necesita construir una matriz de confusión, es decir, una tabla donde las columnas son los pronósticos del modelo y las filas los datos medidos. En la tabla 1 se muestra la matriz de confusión que compara lo pronosticado por un algoritmo sobre la aparición de la niebla o la bruma y el tiempo presente real obtenido de los METAR.

Niebla o bruma	¿Evento pronosticado?	¿Evento pronosticado?	Total
	Sí	No	
¿Evento observado? Sí	578	879	1457
¿Evento observado? No	230	8944	9174
Total	808	9623	10 631

Tabla 1. Matriz de confusión de un algoritmo de IA sobre aparición de niebla o bruma. HSS=0,46.

Una vez establecida la matriz de confusión se pasa a concretar los parámetros de calidad y cantidad de la información suministrada por el modelo meteorológico o el algoritmo de IA.

3.1. Precisión

La precisión indica la relación entre los casos correctamente pronosticados y el total de los pronósticos. En el caso de la tabla 1, el algoritmo ha pronosticado niebla o bruma 808 veces. De esas 808 veces, 578 hubo realmente niebla o bruma y en 230 no se dio el fenómeno. La precisión para el caso de niebla o bruma es 578 dividido por 808, aproximadamente un 72 % como se refleja en la tabla 2.

Niebla o bruma	72 %	40 %	51 %
Sin niebla ni bruma	91 %	97 %	94 %
	Precisión	Exhaustividad	F1

Tabla 2. Informe global del algoritmo de IA que pronostica bruma o niebla. Se miden los parámetros: precisión, exhaustividad y F1.

3.2. Exhaustividad o Recall

Es la relación entre el número de veces que una variable está correctamente pronosticada y el número de veces que es observada. En la matriz de confusión de la tabla 1 la niebla o la bruma se da realmente en 1457 ocasiones. De esas 1457 ocasiones, 578 veces se pronostica de forma correcta y 879 de forma errónea. La exhaustividad será: 578 dividido por 1457. Aproximadamente un 40 % como se ve en la tabla 2.

3.3. El valor F o F1

Es la media armónica entre las dos variables anteriores. Algo menor que la media aritmética. Se muestra en la tabla 2.

3.4. El Heidke skill score (HSS)

Mide la proporción normalizada de pronósticos correctos después de eliminar las previsiones que serían correctas solo por casualidad. Una predicción perfecta implica HSS = 1 y una predicción debida completamente al azar implica HSS = 0. Su valor aparece al pie de las tablas 1, 4 y 5.

3.5. Porcentaje de entropía sobre la máxima entropía posible

Mide la cantidad de información suministrada. No la calidad de la misma. La información puede ser suministrada por el modelo meteorológico, el algoritmo de IA o la propia climatología. Se mide la entropía de la distribución discreta de probabilidad. Por ejemplo en la tabla 3, cuando el algoritmo pronostica niebla o bruma, en el 72 % de los casos se tiene niebla o bruma y en el 28 % no. La entropía para este caso viene dada por la fórmula:

$$\text{Entropía} = -0,72 \times \log_2(0,72) - 0,28 \times \log_2(0,28) = 0,86,$$

donde \log_2 es el logaritmo en base 2.

La entropía máxima se daría para el caso en que el algoritmo acertase o fallase al 50 %. La entropía sería:

$$\text{Entropía} = -0,5 \times \log_2(0,5) - 0,5 \times \log_2(0,5) = 1.$$

El cociente entre ambas es del 86 % según la tabla 3.

Niebla o bruma	72 %	9 %	14 %
Sin niebla ni bruma	28 %	91 %	86 %
% entropía	86 %	43 %	58 %
	Niebla o bruma	Sin niebla ni bruma	Climatología

Tabla 3. Precisiones y entropías para los pronósticos de niebla o bruma. Climatología y su entropía.

4. COMPARANDO ALGORITMOS DE IA CON EL MODELO METEOROLÓGICO: LA VISIBILIDAD

Para poder afirmar que un sistema de IA mejora el pronóstico de un modelo meteorológico se debe emplear el mismo criterio de evaluación para ambos. En este artículo se emplean los criterios de evaluación de la IA y el HSS al modelo meteorológico. El modelo meteorológico no pronostica de forma directa la ocurrencia de la niebla o la bruma pero sí la visibilidad horizontal. Se tendrán tres tipos de datos: la visibilidad horizontal real medida en el aeropuerto, la visibilidad pronosticada por el modelo meteorológico y la visibilidad pronosticada por el algoritmo de inteligencia artificial.

Se dividirá la visibilidad horizontal en dos categorías: visibilidad menor o igual de 1000 metros y visibilidad mayor de 1000 metros. Se escoge el punto más cercano del modelo meteorológico al aeropuerto de Vigo, el p0 de la figura 1. Se han usado datos de los años 2014 a 2020. En total se han comparado 57 481 datos del modelo meteorológico con los datos de tiempo observado. La matriz de confusión con los resultados del modelo meteorológico puede verse en la tabla 4.

Visibilidad menor o igual de 1000 metros	¿Evento pronosticado? Sí	¿Evento pronosticado? No	Total
¿Evento observado? Sí	947	4585	5532
¿Evento observado? No	1564	50 385	51 949
Total	2511	54 970	57 481

Tabla 4. Matriz de confusión del modelo meteorológico. HSS=0,19.

En el caso del algoritmo de IA, de los 57 481 del modelo meteorológico, se han usado el 90 % para entrenar el algoritmo de IA y el 10 % (5749 datos) en la verificación. El algoritmo de IA empleado es el LGB Classifier con 250 estimadores de la librería Sklearn. Los resultados de la verificación pueden verse en la tabla 5.

Visibilidad menor o igual de 1000 metros	¿Evento pronosticado? Sí	¿Evento pronosticado? No	Total
¿Evento observado? Sí	152	396	548
¿Evento observado? No	63	5138	5201
Total	215	5534	5749

Tabla 5. Matriz de confusión del modelo de IA. HSS=0,36.

5. CONCLUSIONES

Parece que es posible mejorar la calidad de los pronósticos de aeropuerto mediante el empleo de algoritmos de IA que tomen como datos de entrada las salidas de los modelos meteorológicos. Para el caso concreto del aeropuerto de Vigo se pueden consultar estos pronósticos para varias variables y alcances en el siguiente [enlace](#) [3]. Todo el proyecto informático del aeropuerto de Vigo se encuentra en el [enlace](#) [4]. En el futuro se procederá a mejorar la calidad de los algoritmos y extender las técnicas de IA a otros aeropuertos y entornos de predicción.

AGRADECIMIENTOS

Agradezco a todas las personas e instituciones que han puesto a mi disposición de forma libre y gratuita los datos meteorológicos necesarios para la realización de este trabajo. En concreto, a AEMET por los boletines METAR, la Universidad Estatal de Iowa por su base de datos de boletines METAR y a Meteogalicia por su base de datos del modelo WRF. También estoy muy agradecido a todos los desarrolladores de librerías de software libre. Mi trabajo nunca hubiera sido posible sin librerías como: Sklearn o Pycaret. No puedo olvidar, en este apartado, a mis compañeros de la Delegación Territorial de Galicia (AEMET) Carlos Moure, José María Pascual y David Urtaza por sus aportaciones a este artículo.

REFERENCIAS

1. https://www.meteogalicia.gal/web/modelos/threddsIndex.action?request_locale=gl
2. <https://medium.com/analytics-vidhya/forecasting-flight-rules-with-ai-at-vigo-airport-levx-ce3e4d12ab84>
3. https://share.streamlit.io/granantuin/levx_st/main/operational_LEVX_st.py
4. https://github.com/granantuin/LEVX_st