

ANTEPROYECTO DE TRABAJO DE FIN DE MÁSTER



AUTOR: CHRISTIAN MARISCAL CALVO

TUTOR: DAVID FERNÁNDEZ BARRERO

COTUTOR: MARIO COBOS MAESTRE

TITULACIÓN: MÁSTER EN CIENCIA Y TECNOLOGÍA DESDE EL ESPACIO

TÍTULO: INTERPRETACIÓN MEDIANTE REGRESIÓN SIMBÓLICA DE UNA RED NEURONAL PROFUNDA PARA LA PREDICCIÓN DE LA VELOCIDAD DEL VIENTO SOLAR

Índice

1.	INTRODUCCIÓN	3
2.	OBJETIVOS	4
3.	DESCRIPCIÓN DEL TRABAJO	4
4.	METODOLOGÍA.....	5
5.	MEDIOS	6
6.	BIBLIOGRAFÍA.....	6

1. INTRODUCCIÓN

Aunque en las últimas décadas el estudio de la atmósfera solar y sus problemas relacionados, como el calentamiento coronal o la aceleración del viento solar y su modelización ha experimentado grandes avances [1] todavía no se dispone de un conocimiento completo de los procesos físicos implicados o se tiene un censo cuantitativo de qué estructuras corales contribuyen a procesos específicos de viento solar. Los flujos de viento solar rápido son conocidos por estar conectados con las regiones centrales de grandes agujeros corales. Sin embargo, los flujos de viento solar lento parecen proceder de un amplio rango de regiones, como los límites de los agujeros corales o los *streamers*.

Además, el viento solar puede tener efectos dañinos en diversos campos del ámbito terrestre. Por ejemplo, la llegada de radiación electromagnética tras fulguraciones solares o erupciones de filamentos aumentan el nivel de ionización en el lado iluminado de la Tierra y produce un riesgo de pérdidas de transmisiones de radio de alta frecuencia. La llegada de partículas energéticas supone un riesgo para satélites [2], astronautas y la aviación, sobre todo en vuelos a alta altitud y latitud. Por último, la llegada de plasma procedentes de eyecciones de masa corales energiza el conjunto magnetosfera-termosfera-ionosfera-atmósfera neutra energizándola, lo cual pone en riesgo las redes de potencia eléctrica y los sistemas de comunicaciones por radio y GPS entre otros.

En conclusión, tanto la importancia por entender el proceso físico-matemático que afecta al viento solar desde su salida desde el Sol como las afectaciones que puede derivar en Tierra hacen necesaria la investigación del fenómeno [3].

Por otro lado, al auge de la Inteligencia Artificial en los últimos años también ha afectado al sector espacial [4]. En este contexto, el empleo de redes neuronales permite conseguir modelos predictivos fiables. Sin embargo, evaluar el proceso seguido por estas para obtener los resultados sigue siendo un reto. Por fortuna, existe una rama de los Algoritmos Evolucionarios, denominada Programación Genética, que permite obtener modelizaciones de estos procesos y que será la que empleemos en este proyecto. Esta permite la resolución de problemas no lineales basándose en lenguaje simbólico, representándose mediante estructuras de árboles y ha permitido un gran número de avances recientes en el sector espacio, con sistemas de navegación autónomos [5] o dinámicas galácticas [6].

En este sentido, la red neuronal autoregresiva SWAN (*Solar Wind Attention Network*) [7] desarrollada recientemente permite la predicción de la velocidad del viento solar de una manera interpretable. Sin embargo, esta interpretabilidad es limitada y, para mejorarla, el trabajo propuesto pone el enfoque en la Programación Genética. En concreto, la Regresión Simbólica.

2. OBJETIVOS

Objetivo principal:

Desarrollo de modelos simbólicos sobre la predicción de la velocidad del viento solar de SWAN.

Objetivos secundarios/hitos:

1. Aprender programación genética y herramientas asociadas.

Para llevar a cabo el objetivo general se empleará programación genética. Para desarrollarla, será necesario emplear las librerías disponibles en Python para el desarrollo de este tipo de algoritmos evolucionarios, como DEAP. Además, se hará uso de la librería SymPy para la simplificación matemática de las expresiones obtenidas.

2. Modelización de la predicción de viento solar en calma con objetivo simple.

En primera instancia, el objetivo sería predecir el viento solar en calma, empleando una técnica de objetivo simple en la que generemos un modelo con una precisión aceptable.

3. Modelización de la predicción de viento solar en calma con técnicas multiobjetivo.

Consistiría en una ampliación de la técnica de optimización de un objetivo, a multiobjetivo, donde intentaríamos minimizar la profundidad del árbol generado (*code bloat*) y optimizar las métricas del algoritmo.

4. Modelización de la predicción de viento solar con distintos eventos.

Se ampliará la modelización matemática de viento solar lento a otros fenómenos como viento rápido recurrente, viento rápido no recurrente y CMEs (*Coronal Mass Ejections*).

5. Análisis del resultado.

Por último, se realizaría un análisis del resultado, con el objetivo de esclarecer cuáles son los parámetros de mayor relevancia que afecten a las predicciones.

3. DESCRIPCIÓN DEL TRABAJO

El trabajo consiste en relacionar la salida de SWAN con su entrada a través de una expresión analítica, que permita modelizar las predicciones de viento solar obtenidas por la red para los distintos fenómenos nombrados en el apartado anterior.

Para ello se contará con un conjunto de datos conformado por las predicciones de SWAN junto con sus entradas y que será aportado por los propios creadores de la red.

El enfoque seguido será iterativo, añadiendo poco a poco mayor complejidad a la regresión. Es decir, primero se hará la modelización para el caso más sencillo (viento solar lento) con técnicas mono-objetivo. Posteriormente se ampliará a técnicas multiobjetivo (profundidad del árbol generado y precisión obtenida) y por último se ampliarán a otros fenómenos solares más complejos.

4. METODOLOGÍA

Se llevará a cabo la metodología CRISP-DM (*Cross-Industry Standard Process for Data Mining*). CRISP-DM es un método probado para orientar trabajos de minería de datos.

Como metodología, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas.

Como modelo de proceso, CRISP-DM ofrece un resumen del ciclo vital de minería de datos (figura 1).

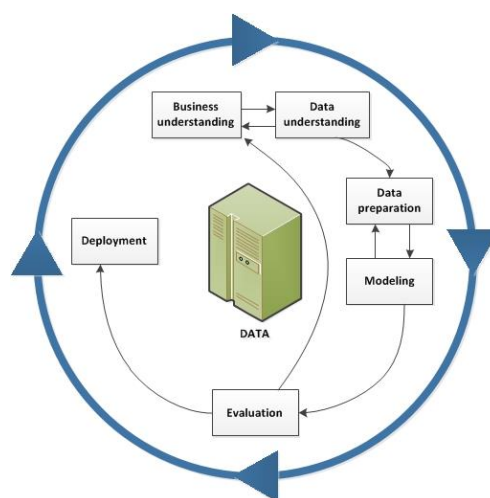


Figura 1: Ciclo de vida de minería de datos según CRISP-DM.

El ciclo vital del modelo contiene seis fases con flechas que indican las dependencias más importantes y frecuentes entre fases. La secuencia de las fases no es estricta. De hecho, la mayoría de los proyectos avanzan y retroceden entre fases si es necesario.

A continuación, se realiza una breve descripción de las fases de la metodología:

1. Entendimiento del problema y los datos.

Consiste en la etapa de estudio de los datos y el problema asociado. Es necesario comprender primero de dónde partimos para obtener un mejor resultado final.

Para esta etapa, se realizará primero un estudio de los datos proporcionados y se estudiará el proceso de llegada del viento solar.

2. Preparación de los datos

Esta etapa consiste en realizar la limpieza de los datos, eliminar duplicados o emplear técnicas de ingeniería de atributos o reducción de dimensionalidad. Se trata de una etapa previa para proporcionar datos de calidad al modelo.

3. Modelización

Aquí se genera la arquitectura del algoritmo. Se definen sus características e hiperparámetros y todo el flujo del algoritmo.

4. Evaluación

Se definen las métricas consideradas de interés que nos proporcionarán una medida de lo preciso que es nuestro algoritmo.

Además, si en esta fase los resultados no son los previstos, se deberá volver a otras anteriores, constituyendo por tanto un proceso abierto y cíclico, basado en el enfoque a resultados.

En el marco de este proyecto, no será necesario realizar la etapa de despliegue.

5. MEDIOS

Se empleará un ordenador que permita el uso de Notebooks para Python, así como el empleo de las principales librerías asociadas al proyecto, como Pandas, Numpy, DEAP o SymPy.

En caso necesario se ampliarán estos recursos con otros más potentes, por ejemplo empleando aceleradores por GPU u otros ordenadores disponibles.

6. BIBLIOGRAFÍA

[1]: *Origins of the Ambient Solar Wind: Implications for Space Weather*, Cranmer et al., 2017.

[2]: *Solar-wind damage to spacecraft thermal control coatings*, Jorgenson, 2012.

[3]: *El soplo del viento solar: Las manchas solares, los ciclos solares y la vida en la Tierra*, Arsentiev et al., 2014.

[4]: *Evolutionary Computation in Astronomy and Astrophysics: A Review*, García Gutiérrez et al., 2011.

[5]: *Incremental evolution of autonomous controllers for unmanned aerial vehicles using multi-objective genetic programming*, *Cibernetics and intelligent System*, Grant, 2004.

[6]: *Genetic Programming and other fitting techniques in Galactic Dynamics*, *Data Analysis Software and Systems (ADASS)*, Teuben, 2004.

[7]: *SWAN: A multihead autoregressive attention model for solar wind speed forecasting*. *Space Weather*, Cobos et al., under review.

[8]: *Genetic Programming. On the programming of computers by means of natural selection*, Koza, 1992.