Secuencias temporales

Este es uno de los pasos clave para convertir los datos fotométricos en un formato adecuado para que las redes neuronales recurrentes (como las LSTM) puedan aprender y clasificar las curvas de luz de los diferentes tipos de objetos.

¿Qué es una secuencia temporal?

Una secuencia temporal es una serie de datos organizados en el tiempo. En este caso, existen datos de observaciones de flujo (FLUXCAL) a lo largo del tiempo (MJD) para cada objeto astronómico. Estas observaciones están correlacionadas, ya que el flujo de un objeto varía de forma periódica o con algún patrón a lo largo del tiempo.

Una secuencia temporal puede representarse como una ventana deslizante de un número fijo de observaciones. Por ejemplo, si se decide trabajar con secuencias de 10 observaciones, significa que cada entrada para la red neuronal será un conjunto de 10 puntos consecutivos en el tiempo.

¿Por qué necesitamos secuencias temporales para clasificar las curvas de luz?

Las redes neuronales recurrentes (RNN), y en particular las LSTM (Long Short-Term Memory), están diseñadas para trabajar con secuencias temporales. A diferencia de las redes neuronales tradicionales (perceptrones multicapa o MLP), las RNN pueden mantener información sobre el pasado para mejorar las predicciones futuras. Este comportamiento es ideal para series temporales, donde las observaciones previas afectan las observaciones futuras.

En el caso de las curvas de luz, cada punto en la curva (flujo) está influenciado por los puntos anteriores. Para que la red neuronal pueda aprender los patrones de variación en el tiempo (como los ciclos periódicos de las estrellas Cepheid y RRL), necesita ver un grupo de observaciones secuenciales.

¿Qué hace este paso?

Este paso toma las observaciones de flujo (FLUXCAL) y tiempo (MJD) y las agrupa en secuencias de longitud fija. Por ejemplo, si decides que una secuencia tendrá 10 observaciones, entonces se genera una secuencia con las primeras 10 observaciones, luego otra secuencia con las siguientes 10, y así sucesivamente. Cada secuencia tiene una etiqueta que corresponde al tipo de objeto (0 para Cepheid y 1 para RRL), que será utilizada como la salida para el entrenamiento.

Cómo funcionan las secuencias en el contexto del pipeline

1. Entrada: Un conjunto de observaciones de MJD y FLUXCAL para un objeto astronómico.
2. Salida: Secuencias de longitud fija (por ejemplo, 10 puntos), donde cada secuencia es un conjunto de observaciones consecutivas en el tiempo, junto con su etiqueta de clase.

Cada secuencia temporal que se crea será alimentada a la red neuronal como una entrada. La red procesará la secuencia completa (por ejemplo, una ventana de 10 puntos consecutivos) para tomar una decisión sobre la clase de esa secuencia (si es Cepheid o RRL).

¿Cómo se crean las secuencias temporales?

Vamos a desglosar cómo se crean estas secuencias:

* Datos de entrada: Tienes un conjunto de observaciones que incluyen las columnas MJD, FLUXCAL y label (la etiqueta de clase).
* Secuencia: Una secuencia se construye a partir de un conjunto de observaciones consecutivas en el tiempo.
* Desplazamiento: Las secuencias se desplazan una observación a la vez para generar nuevas secuencias. Esto significa que no estás dividiendo las secuencias de forma disjunta, sino que se están creando ventanas deslizantes.
* Etiqueta de la secuencia: La etiqueta de cada secuencia es la clase del último punto en la secuencia. Por ejemplo, si estás creando una secuencia de longitud 10, la etiqueta de la secuencia será la etiqueta del décimo punto.

¿Cómo las secuencias ayudan a la clasificación?

Cuando se crean las secuencias, la red neuronal podrá ver no solo una observación individual, sino una serie de observaciones que representan el comportamiento del flujo a lo largo del tiempo. Esto es crucial porque:

1. Patrones de variación: Las estrellas Cepheid y RRL tienen diferentes patrones de variación en su luminosidad. Las secuencias temporales permiten que el modelo vea la tendencia a través de múltiples observaciones y no solo un valor puntual.
   * Por ejemplo, las Cepheid suelen tener curvas de luz más regulares y suaves.
   * Las RRL tienen curvas de luz más cortas y abruptas.
2. Dependencia temporal: Los puntos de flujo sucesivos no son independientes entre sí; están correlacionados a través del tiempo. Las redes neuronales recurrentes (RNN o LSTM) pueden aprender dependencias entre los puntos en el tiempo, lo cual es fundamental para clasificar correctamente los diferentes tipos de estrellas basándose en sus curvas de luz.
3. Clasificación basada en secuencias completas: La red neuronal no toma decisiones basándose en observaciones individuales, sino en secuencias de puntos. Esto permite capturar ciclos completos o patrones repetitivos que son característicos de cada tipo de objeto astronómico.

Consideraciones adicionales sobre la longitud de la secuencia

* Longitudes más cortas: Si la secuencia es demasiado corta (por ejemplo, solo 5 puntos), la red neuronal puede no ver suficiente información para detectar patrones significativos en la curva de luz.
* Longitudes más largas: Si la secuencia es demasiado larga (por ejemplo, 100 puntos), el modelo puede tener dificultades para procesar la información o podría necesitar más recursos computacionales para entrenarse.