

Descripción de la Arquitectura *Trafficformer*

Componentes principales

1. Traffic Input

- Histórico de velocidades:

$$S_t^I \in \mathbb{R}^{N \times I},$$

donde N es el número de pasos de tiempo y I el número de nodos.

-
- para cada uno de los I nodos/sensores ponemos sus últimas N mediciones de velocidad (u otra variable). Piensa en un cubo de datos tiempo \times nodo.
-
- Máscara espacial basada en distancia y velocidad libre:

$$M^P \in \mathbb{R}^{I \times I}.$$

Matriz binaria o continua que codifica qué pares de nodos pueden influirse. Se calcula con la distancia física y la velocidad de flujo libre (si dos nodos están muy lejos o no comparten vía, la máscara es 0).

2. Temporal Feature Extractor

- Cada secuencia temporal (una columna de S_t^I) pasa por un mismo *Multi-Layer Perceptron* (MLP).
- Salida (embedding temporal por nodo):

$$S_t^{C1} \in \mathbb{R}^{I \times d_t}.$$

3. Traffic Node Feature Interaction

- *Multi-head Self-Attention* con la máscara M^P para limitar la atención a nodos vialmente conectados.
- Estructura estándar ADD NORM \rightarrow Feed-Forward \rightarrow ADD NORM.
- Salida (embedding espaciotemporal):

$$Z_t \in \mathbb{R}^{I \times d_s}.$$

4. Speed Forecasting

- Un segundo MLP por nodo transforma Z_t en las predicciones:

$$\hat{V}_{t+1:t+H} \in \mathbb{R}^{I \times H},$$

donde H es el horizonte de predicción (número de pasos futuros).

Flujo de datos resumido

$$S_t^I \xrightarrow[\text{MLP}]{\text{Temporal}} S_t^{C1} \xrightarrow[M^P]{\text{Self-Attention}} Z_t \xrightarrow[\text{MLP}]{\text{Regresión}} \hat{V}_{t+1:t+H}$$

Tabla de símbolos

Símbolo	Descripción
N	Longitud de la ventana histórica (nº de pasos de tiempo)
I	Número total de nodos/sensores en la red
S_t^I	Matriz de características históricas de tamaño $N \times I$
M^P	Máscara espacial ($I \times I$) que pondera la atención entre nodos
S_t^{C1}	Embedding temporal por nodo ($I \times d_t$)
Z_t	Embedding espaciotemporal final ($I \times d_s$)
$\hat{V}_{t+1:t+H}$	Velocidades predichas para los próximos H pasos

Puntos clave

- El modelo *desacopla* el aprendizaje temporal (MLP verde) del aprendizaje espacial (auto-atención azul).
- La máscara M^P introduce conocimiento a priori de la topología vial, lo que acelera la convergencia y mejora la interpretabilidad.
- Los MLP son ligeros y comparten pesos entre nodos, reduciendo la complejidad frente a arquitecturas basadas en LSTM o CNN.