





TrAISformer—A Transformer Network with Sparse Augmented Data Representation and Cross Entropy Loss for AIS-based Vessel Trajectory Prediction

一种具有稀疏增强数据表示和交叉熵损失的
变换器网络，用于基于AIS的船舶轨迹预测

2025/7/27

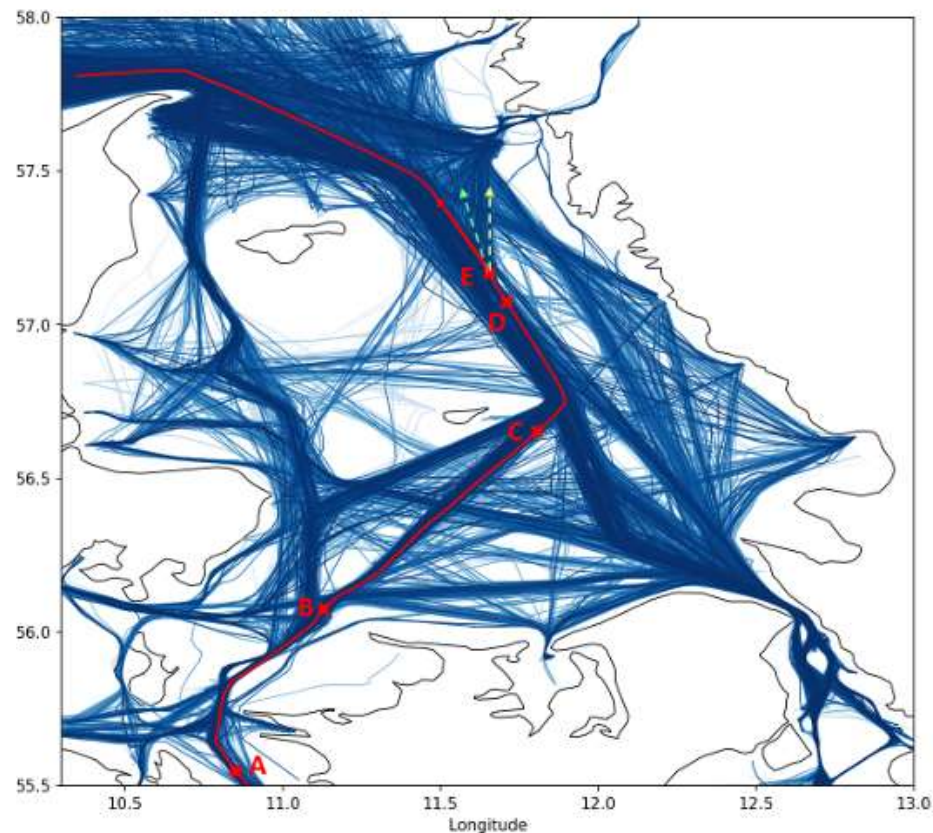


I Introduction

I Introduction

研究背景

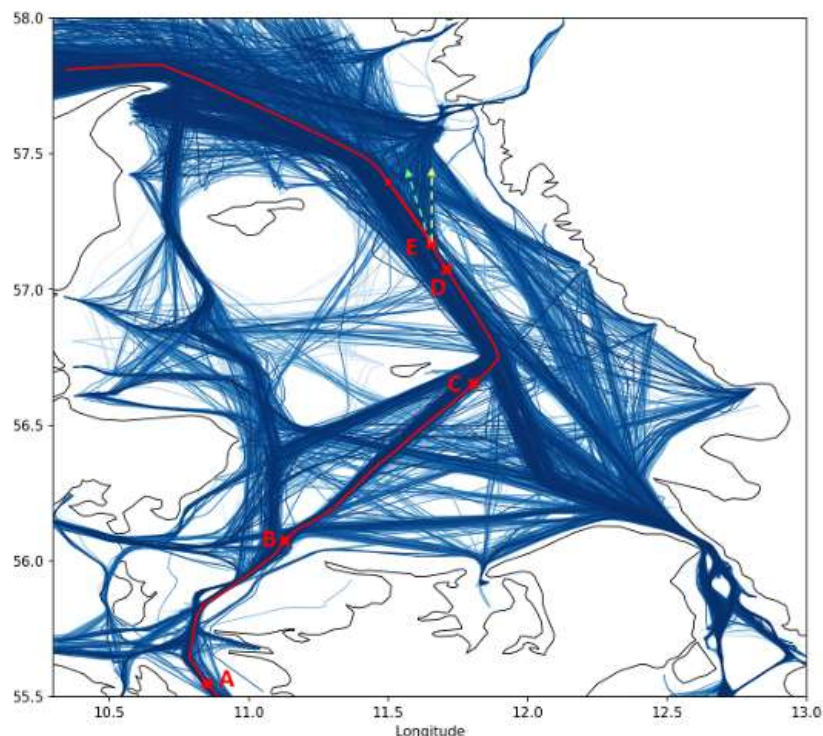
船舶轨迹预测在众多海事应用和服务中发挥着关键作用。但由于船舶运动数据固有的异构性和多模态特性，即便是现代机器学习技术，利用 AIS 数据进行船舶轨迹预测几小时以上的船舶位置仍然极具挑战性。仍面临挑战。



I Introduction

当前挑战

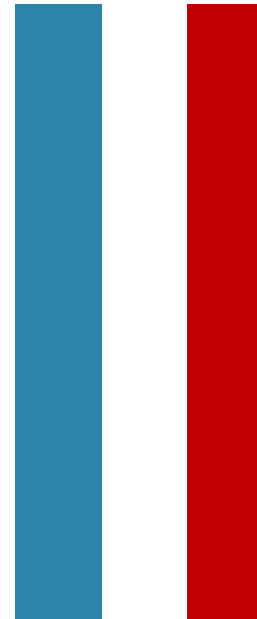
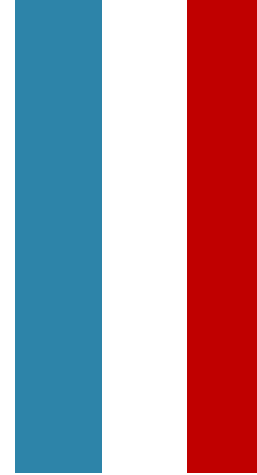
- AIS 数据的异构性
- 船舶轨迹的多模态性
- 长期依赖关系的建模困难



研究贡献

- TrAISformer提出：采用了一种特定的稀疏高维 AIS 数据表示方法，并将预测任务构建为分类问题，以明确建模 AIS 数据的异构性和船舶轨迹的多模态性。
- 设计基于交叉熵的多分辨率分类损失函数摒弃传统 MSE 回归损失（易合并多模态轨迹），将预测任务转化为分类问题
- 长期依赖捕捉：利用概率 Transformer 架构来捕捉船舶 AIS 轨迹中的长期依赖关系。

II PROBLEM STATEMENT AND RELATED WORK



II-A PROBLEM STATEMENT

基于 AIS 的船舶轨迹预测任务被定义为：利用历史 AIS 观测序列，在特定时间范围内预测船舶未来位置。具体表述如下：

AIS 观测值：单个时间步 t 的 AIS 数据为 4 维向量，包含船舶的核心运动信息：[lat, lon, SOG, COG]其中，lat、lon表示位置，SOG、COG表示运动状态。

$$\mathbf{x}_t \triangleq [\textit{lat}, \textit{lon}, \textit{SOG}, \textit{COG}]^T$$

轨迹与预测目标：船舶轨迹由连续时间步的观测值组成。预测目标是基于历史轨迹，推断未来 L 步的轨迹分布

$$p(\mathbf{x}_{T+1:T+L} | \mathbf{x}_{0:T})$$



II- B RELATED WORK

现有船舶轨迹预测方法主要分为两类:

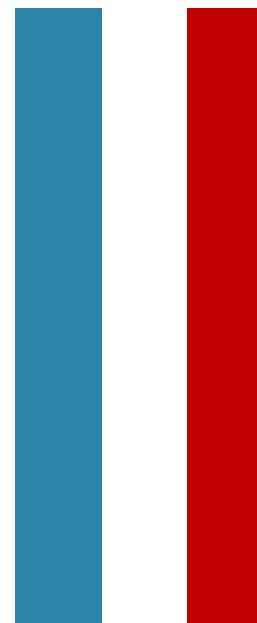
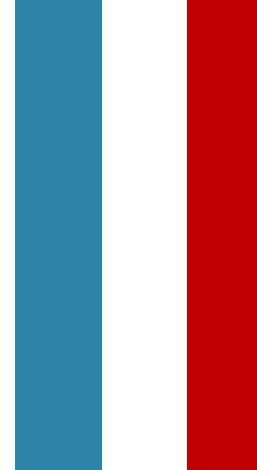
基于状态空间的方法

- 核心思路: 结合船舶运动的动态先验模型 (如曲线运动模型 CMM) 与滤波算法 (如卡尔曼滤波、粒子滤波), 推断后验分布
- 局限: 无法捕捉复杂运动模式
- 滤波方法易受误差累积影响

基于学习的方法:

- 核心思路: 利用深度学习模型 (如 LSTM、GRU) 从历史数据中学习时序模式, 直接预测未来轨迹。
- 局限: 传统 LSTM/GRU 难以处理 AIS 数据的多模态特性 (同一状态对应多条可能轨迹), 易输出 “平均化” 的错误路径, 中长时预测精度有限。

III PROPOSED APPROACH



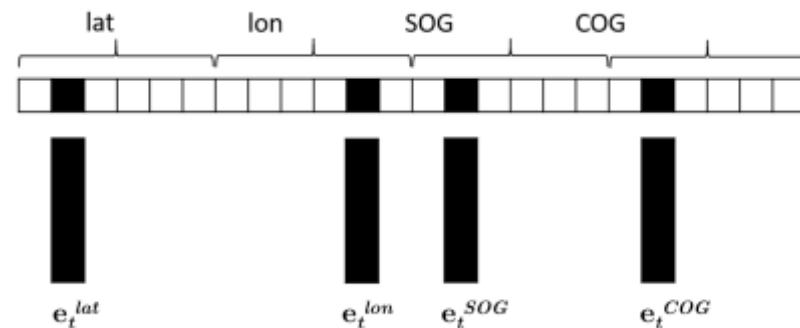
III-A AIS 数据的离散稀疏表示

AIS 消息最常用的表示方式是 4 维实值向量，包含船舶的位置和速度信息。这种特征空间中编码复杂的船舶运动模式存在挑战。一种自然的思路是将特征空间扩展到更高维度。本文将纬度、经度、SOG 和 COG 分别离散化为四个分箱

$$h_t \triangleq [1_t^{lat}, 1_t^{lon}, 1_t^{SOG}, 1_t^{COG}]^T$$

由于分箱向量信息稀疏，再将向量对应一个64维高维嵌入向量，这种方法确保在嵌入空间中，仅使用数量限定个值。船舶的 AIS 轨迹由此表示为：

$$e_{0:T} \triangleq \{e_0, e_1, \dots, e_T\}$$



Algorithm 1: fourhot(x_t, R, SOG_{max}, N).

Description: Create "four-hot" vector.

Input: AIS observation

$$x_t \triangleq [lat, lon, SOG, COG]^T,$$

the limits of the ROI

$$R \triangleq [lat_{min}, lat_{max}, lon_{min}, lon_{max}]^T, \\ SOG_{max},$$

the numbers of bins

$$N \triangleq [N^{lat}, N^{lon}, N^{SOG}, N^{COG}]^T.$$

Output: "Four-hot" vector h_t .

// Create the one-hot vector for each attribute.

$$1_t^{lat} = \text{onehot}(x_t^{lat}, lat_{min}, lat_{max}, N^{lat})$$

$$1_t^{lon} = \text{onehot}(x_t^{lon}, lon_{min}, lon_{max}, N^{lon})$$

$$1_t^{SOG} = \text{onehot}(x_t^{SOG}, 0, SOG_{max}, N^{SOG})$$

$$1_t^{COG} = \text{onehot}(x_t^{COG}, 0, 360, N^{COG})$$

// Concatenate the one-hot vectors

$$h_t = [1_t^{lat}, 1_t^{lon}, 1_t^{SOG}, 1_t^{COG}]^T$$

Return: h_t

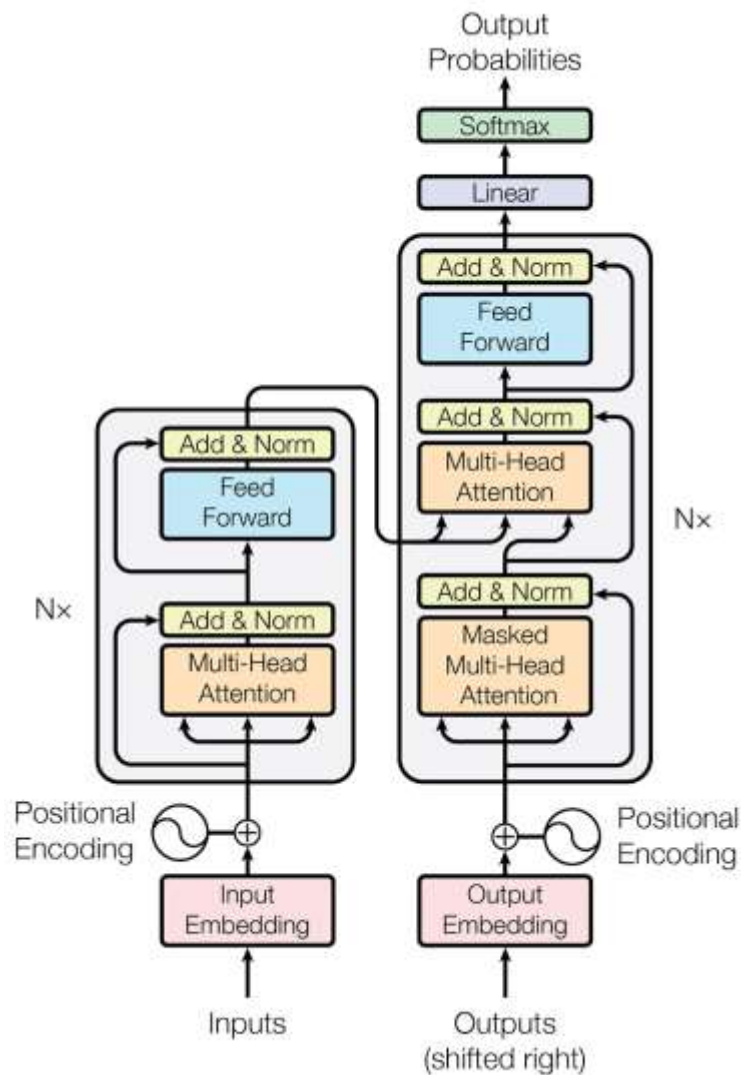
III-B Transformer架构

为了准确预测船舶的轨迹，预测模型需要捕捉历史 AIS 观测数据中可能存在的长期依赖关系，Transformer 神经网络 自然成为非常合适的选择

TrAISformer 中的 Transformer 网络由一系列堆叠的注意力层组成。每一层都作为自回归模型，采用点积多头自注意力机制：

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_e}} \right) V \quad (4)$$

本实验中，Transfoermer输入为高维嵌入向量 e ，经过模型编码解码得到最后一层的输出是一个与 h 维度相同的向量/



III-C学习方法

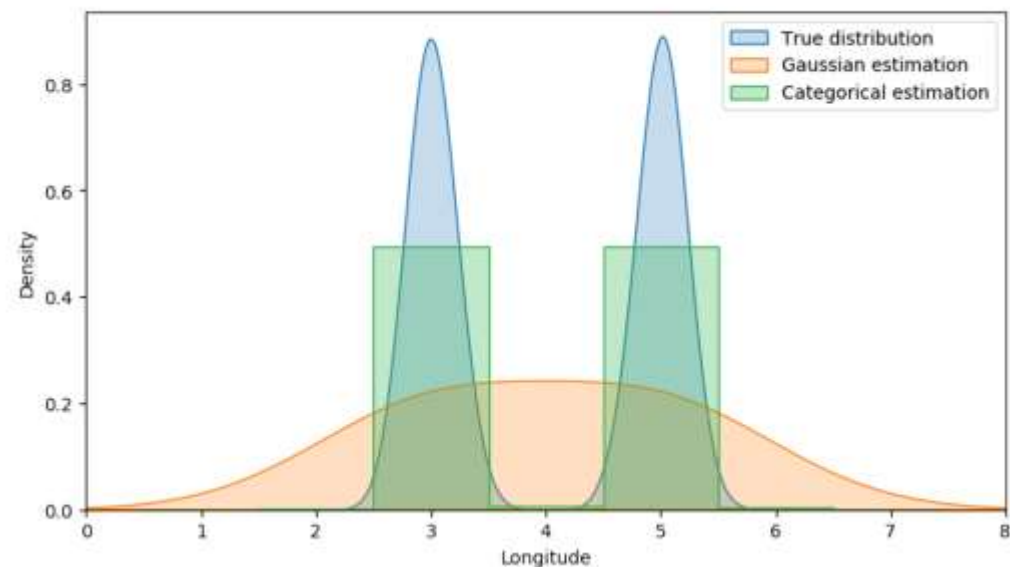
轨迹预测的传统思路是将其视为**回归问题**：模型基于历史观测直接预测未来连续值。损失函数常用**均方误差**，会“平均化”多模态轨迹，输出不符合实际的中间路径

$$\mathcal{L}_{MSE} = \frac{1}{L} \sum_{l=1}^L \|\mathbf{x}_{T+l}^{pred} - \mathbf{x}_{T+l}^{true}\|_2^2,$$

为捕捉多模态特性，文中将预测任务转化为**分类问题**，利用四热向量与嵌入向量的一**对一映射关系**，将未来轨迹的条件概率分布转化为：

$$p(\mathbf{h}_{T+l} | \mathbf{e}_{0:T+l-1}) = p(\mathbf{e}_{T+l} | \mathbf{e}_{0:T+l-1}).$$

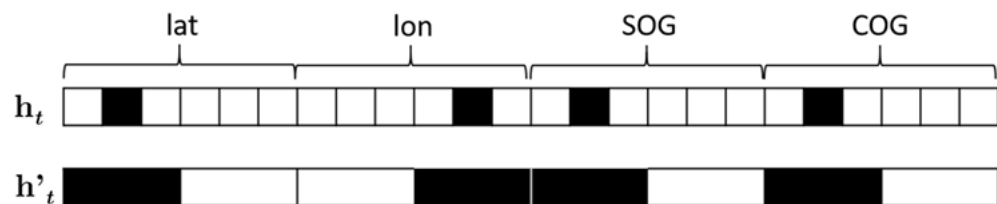
四热向量由 4 个独热向量拼接而成（对应 lat、lon、SOG、COG），因此预测任务被分解为**4 个并行的分类任务**



III-C学习方法

为优化多头部分类任务，设计基于交叉熵的损失函数，直接度量预测类别分布与真实“四热向量”的差异

为进一步提升鲁棒性，引入粗粒度“四热向量”，分箱数量为细粒度的 1/3，分辨率降低。



结合粗细粒度，最终损失函数为：

$$\mathcal{L}_{CE} = \sum_{l=1}^L CE(p_{T+l}, \mathbf{h}_{T+l}) + \beta CE(p'_{T+l}, \mathbf{h}'_{T+l}).$$

Algorithm 2: $ce_loss(\mathbf{h}_t, \mathbf{l}_t, \mathbf{N})$.

Description: Calculate the cross-entropy loss \mathcal{L}_{CE} .

Input: “four-hot” vector \mathbf{h}_t ,
the output of the transformer \mathbf{l}_t ,
the numbers of bins \mathbf{N} .

Output: the cross-entropy $CE(\mathbf{h}_t, \mathbf{l}_t)$.

// Split \mathbf{h}_t back into 4 one-hot
vectors, each corresponding to an
attribute of the AIS observation.

$\mathbf{l}_t^{lat}, \mathbf{l}_t^{lon}, \mathbf{l}_t^{SOG}, \mathbf{l}_t^{COG} = split(\mathbf{h}_t, \mathbf{N})$

// Split \mathbf{l}_t into 4 heads.

$\mathbf{l}_t^{lat}, \mathbf{l}_t^{lon}, \mathbf{l}_t^{SOG}, \mathbf{l}_t^{COG} = split(\mathbf{l}_t, \mathbf{N})$

// Calculate the cross-entropy for
each head.

$p_t^{lat} = CE(\text{Categorical}(\text{logit} = \mathbf{l}_t^{lat}), \mathbf{l}_t^{lat})$

$p_t^{lon} = CE(\text{Categorical}(\text{logit} = \mathbf{l}_t^{lon}), \mathbf{l}_t^{lon})$

$p_t^{SOG} = CE(\text{Categorical}(\text{logit} = \mathbf{l}_t^{SOG}), \mathbf{l}_t^{SOG})$

$p_t^{COG} = CE(\text{Categorical}(\text{logit} = \mathbf{l}_t^{COG}), \mathbf{l}_t^{COG})$

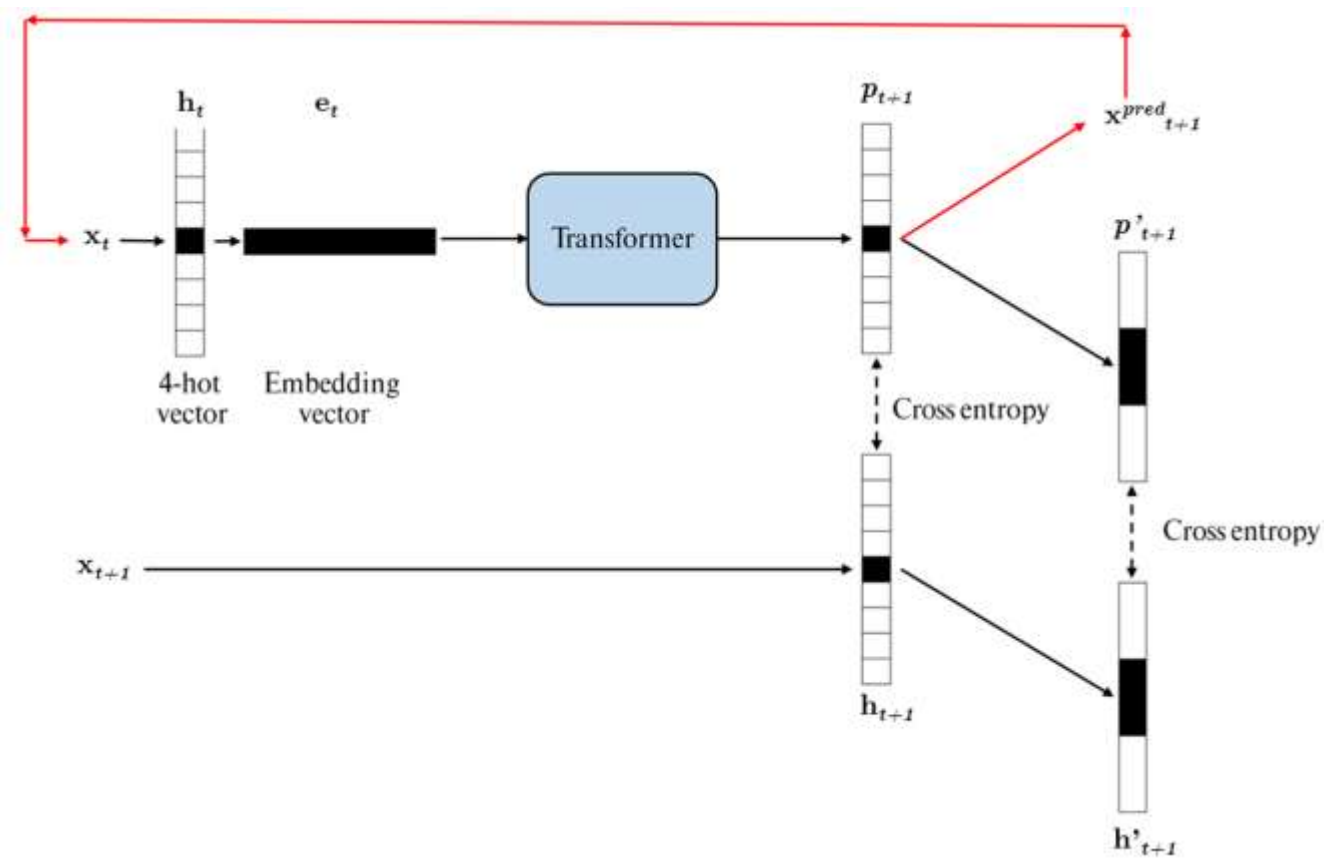
// Calculate “total” cross-entropy.

$ce_with_logit(\mathbf{h}_t, \mathbf{l}_t) = p_t^{lat} * p_t^{lon} * p_t^{SOG} * p_t^{COG}$

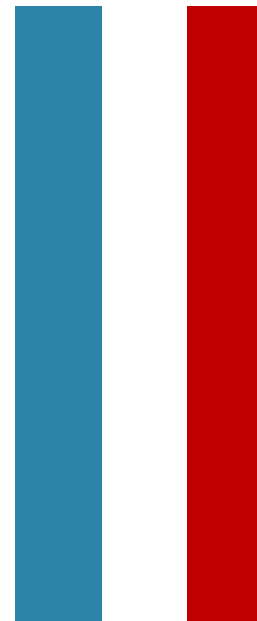
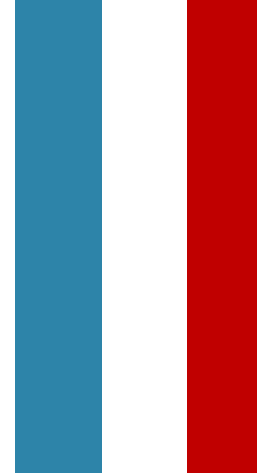
Return: $ce_with_logit(\mathbf{h}_t, \mathbf{l}_t)$

III-C学习方法

核心流程



IV Experiment



IV Experiment

数据集：

- 丹麦海事局 (DMA) 提供的公开 AIS 数据集上测试了 TrAISformer。该数据集包含 2019 年 1 月 1 日至 2019 年 3 月 31 日期间货船和油轮的 AIS 观测数据。感兴趣区域 (ROI) 是一个从 $(55.5^\circ, 10.3^\circ)$ 到 $(58.0^\circ, 13.0^\circ)$ 的矩形区域。

基线模型：

- LSTM seq2seq 序列到序列模型
- Conv_seq2seq 卷积序列到序列模型
- GeoTrackNet 模型

数据预处理：

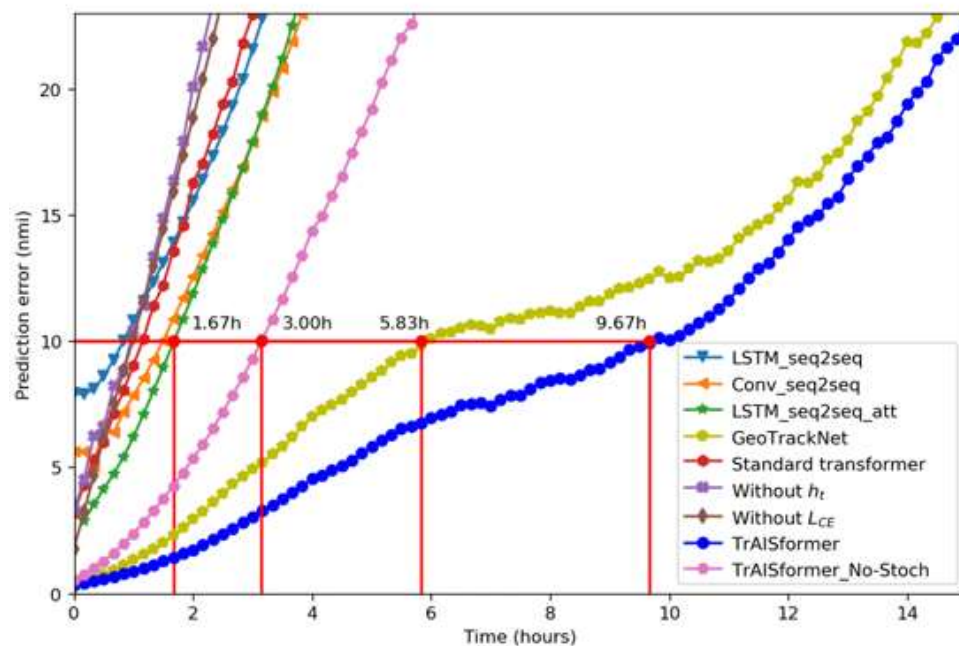
- 移除具有不切实际速度值的 AIS 消息 (对地速度 ≥ 30 节) ;
- 移除系泊或锚泊状态的船舶;
- 移除距离海岸线 1 海里范围内的 AIS 观测值;
- 将非连续航次拆分为连续航次。连续航次指的是两个连续 AIS 消息之间的最大时间间隔小于预设值 (此处为 2 小时) 的航次;
- 移除长度小于 20 (单位未明确) 或持续时间少于 4 小时的 AIS 航次;
- 移除异常消息。如果经验速度 (通过行驶距离除以两个连续消息之间的相应时间间隔计算得出) 不切实际 (此处为超过 40 节), 则该 AIS 消息被视为异常消息;
- 以 10 分钟为采样间隔对 AIS 数据进行下采样

IV Experiment

实验结果:

TABLE I: Mean prediction performance of the benchmarked models (in nautical miles).

Model	1h	2h	3h
LSTM_seq2seq	5.83	8.39	11.64
Conv_seq2seq	4.23	6.77	9.66
LSTM_seq2seq_att	3.35	6.41	9.65
<i>Clustering_LSTM_seq2seq_att</i> ^l [18]	0.78	1.93	3.66
GeoTrackNet [45]	0.72	1.59	2.67
TrAISformer	0.48	0.94	1.64
TrAISformer_No-Stoch	1.28	2.88	5.02



IV Experiment

消融实验:

- 移除了e和h嵌入，以验证高维编码的意义。
- 我们保留e但移除h，以评估稀疏约束的相关性。
- 测试了一个与 TrAISformer 架构相同但使用回归损失作为训练损失的模型，以验证分类损失的关键作用

TABLE II: Mean prediction performance (in nautical miles) of the models in the ablation study.

Model	AIS data representation	Embedding $\mathbf{x}_t \rightarrow \mathbf{e}_k$	Loss function	1h	2h	3h
Without \mathbf{e}_t and \mathbf{h}_t (standard transformer)	$[lat, lon, SOG, COG]^T$	None	\mathcal{L}_{MSE}	4.75	8.36	11.40
Without \mathbf{h}_t	$[lat, lon, SOG, COG]^T \rightarrow \mathbf{e}_t$	MLP	\mathcal{L}_{MSE}	5.02	9.69	15.04
Without the classification loss \mathcal{L}_{CE}	“four-hot” vector $\rightarrow \mathbf{e}_t$	Via \mathbf{h}_t	\mathcal{L}_{MSE}	5.53	10.64	16.06
<i>TrAISformer</i>	“four-hot” vector $\rightarrow \mathbf{e}_t$	Via \mathbf{h}_t	\mathcal{L}_{CE}	0.48	0.94	1.64

结果显示，所有消融模型的性能均显著低于 TrAISformer

V Conclusion



V Conclusion

本文提出了一种基于 Transformer 的船舶轨迹预测模型 TrAISformer，旨在解决 AIS 数据轨迹预测中的多模态性和长期依赖问题。

模型核心设计包括：将船舶的纬度、经度、航速（SOG）、航向（COG）等连续属性通过“四热向量”（four-hot vector）离散化，转化为高维稀疏表示；采用类似 GPT 的 Transformer 架构捕捉轨迹长期依赖；使用多分辨率交叉熵损失函数优化模型，以保留轨迹的多模态特性。

未来工作将探索结合天气、船舶交互等因素，以及模型压缩以提升实用性。

谢谢！

