

DOI: 10.12382/bgxb.2025.0576

基于改进神经常微分方程网络的动态多目标时空预测

王艺萌^{1,2}, 程岳³, 邢薇薇^{3*}, 刘渭滨^{2,4**}, 康晓⁵

(1. 中国地质大学(北京) 数理学院, 北京 100083; 2. 北京交通大学 视觉智能交叉创新教育部国际合作联合实验室, 北京 100044; 3. 北京交通大学 软件学院, 北京 100044; 4. 北京交通大学 计算机科学与技术学院, 北京 100044; 5. 中兵智能创新研究院有限公司, 北京 100072)

摘要: 多目标时空预测是动态场景中目标演化与决策的关键。传统静态方法难以捕捉动态环境下多目标间的动态时空演变过程。而神经常微分方程网络虽适配动态系统, 但在多目标时空关联方面仍存局限。为此提出一种基于改进神经常微分方程网络的动态多目标时空预测算法, 将神经常微分方程与稀疏图卷积网络相结合, 建立动态多目标时空预测架构。首先利用稀疏图卷积网络从历史观测轨迹中提取多目标之间的稀疏时空交互特征, 实现对多目标时空关联间的建模; 然后利用神经常微分方程灵活高效的时序建模能力对目标的高维隐藏状态进行时序建模, 最终得到多目标的时空预测轨迹。实验结果表明, 所建模型在 ETH/UCY 数据集上的平均位移误差和最终位移误差分别为 0.36 和 0.56, 相比于基准模型神经常微分方程分别降低了 3% 和 10%, 预测结果准确度更高。

关键词: 多目标预测; 时空关联建模; 神经常微分方程; 时空演变

中图分类号: TP391 **文献标志码:** A

文章编号: 1000-1093(2026)11-250576-xxx

Dynamic Multi-Target Spatio-Temporal Prediction Based on Improved Neural Ordinary Differential Equations Network

WANG Yimeng^{1,2}, CHENG Yue³, XING Weiwei^{3*}, LIU Weibin^{2,4**}, KANG Xiao⁵

(1. School of Science, China University of Geosciences, Beijing 100083, China; 2. Visual Intelligence +X International Cooperation Joint Laboratory of MOE, Beijing Jiaotong University, Beijing 100044, China; 3. School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China; 4. School of Computer Science and Technology, Beijing Jiaotong University, Beijing 100044, China; 5. China North Artificial Intelligence and Innovation Research Institute, Beijing 100072, China)

Abstract: Spatio-temporal prediction of multi-targets is key to target evolution and decision-making in dynamic scenarios. Traditional static methods struggle to capture the dynamic spatiotemporal evolution of multiple objectives in dynamic environments. While neural networks based on stochastic differential equations are well-suited for dynamic systems, they still have limitations when it comes to modeling spatiotemporal correlations among multiple objectives. Therefore, this paper proposes a dynamic multi-target spatio-temporal prediction algorithm based on the improved neural ordinary differential equation network. It establishes a dynamic multi-target spatio-temporal prediction architecture by combining neural ordinary differential equations and sparse graph convolutional network. Firstly, the sparse graph convolutional network is used to extract the sparse spatio-temporal interaction features between multi-targets from historical observation trajectories to achieve the modelling of multi-target spatio-temporal correlations; then the flexible and efficient temporal modelling capability of the neural ordinary differential equations is used to carry out the temporal modelling of the high-dimensional hidden state of the targets, and finally, the spatio-temporal prediction trajectories of the multiple targets are obtained. Experiments show that the average displacement error (ADE) and final displacement error (FDE) of this model on the ETH/UCY dataset are 0.36 and 0.56, which are decreased by 3% and 10%, respectively, compared with the ordinary differential equations network model, resulting in higher prediction accuracy.

Keywords: multi-target predictions; spatio-temporal correlation modeling; neural ordinary differential

收稿日期: 2025-6-29

基金项目: 北京市自然科学基金-丰台轨道交通前沿研究联合基金重点研究项目(L231005)

*通信作者邮箱: wwxing@bjtu.edu.cn

**通信作者邮箱: wbliu@bjtu.edu.cn

0 引言

动态多目标时空预测是智能系统实现自主感知与决策的基础,广泛应用于自动驾驶、人机交互等^[1-6]领域。在实际应用场景中,目标往往呈现出复杂的运动规律,表现出动态多样且强交互的特性。基于历史观测数据对多目标的未来状态进行合理预测,有助于为系统提供前瞻性决策依据,提高系统整体运行的安全水平。因此,如何在有限的观测数据下实现多目标时空预测成为实现智能系统自主感知与实时决策的关键。

传统的动态多目标时空预测多采用隐马尔可夫模型^[7]、高斯过程回归^[8]等概率统计方法。但这类方法往往依赖预先设定的统计假设,当场景中目标数量增加、交互行为复杂时,其建模能力和泛化性容易受到限制,导致预测性能下降。随着深度学习的快速发展,研究者们多采用循环神经网络(Recurrent Neural Network, RNN)、时间卷积网络(Temporal Convolutional Network, TCN)等深度学习算法^[9-11]来解决此类问题。Dong^[12]提出了一种面向多目标时空预测的混合特征融合方法,将视觉感知信息与地图层面的语义表示进行协同建模,以实现对目标时空特征的统一刻画。然而,该特征融合方式较为固定,对于动态目标间的复杂交互关系建模仍存在一定的局限性。Gupta 等^[13]提出了交互生成对抗网络(Social Generative Adversarial Network, Social-GAN)模型,引入目标间的交互机制,将序列预测与生成对抗网络相结合,提升对目标间的交互行为的建模能力。然而,该方法普遍依赖于静态或隐式的交互建模,难以准确刻画目标间随时间动态变化的空间关系。

为更好地建模多目标间的交互关系,研究者引入图神经网络构建时空图结构。Salzmann 等^[14]提出基于图神经网络的 Trajectron++ 轨迹预测框架,该框架融合了 RNN 和条件变分自编码器进行不确定性建模,同时结合系统动力学约束确保预测轨迹的物理可行性。但该模型结构复杂、训练成本高,难以应对高动态环境中目标间关系的快速变化。Wang 等^[15]提出了动态场景下的多目标轨迹预测模型 Goal-CurveNet。通过异构图注意力机制建模多目标之间的时空交互并利用 B 样条曲线拟合实现连续轨迹生成,模型有效融合了图神经网络与生成建模思想,但在处理非结构化行为时仍存在一定的泛化局限性。Mohamed 等^[16]提出交互时空图卷积神经网络(Social Spatio-Temporal Graph Convolutional Neural Network, Social-STGCNN)模型,将目标之间的交

互关系建模为动态图结构,利用图卷积网络提取空间交互特征,并结合时序卷积网络进行时间维度建模,实现对多目标未来轨迹的准确预测。

尽管上述图结构方法提升了空间交互的建模精度,但其时间维度依然受限于 TCN 或 RNN 等离散建模范式,这类离散时间建模算子在相邻采样时刻之间缺乏显式的数学描述,导致其在时序预测中难以保证轨迹在连续时间尺度上的物理平滑性。为克服传统离散时间序列建模的局限性,Chen 等^[17]提出的神经常微分方程(Neural Ordinary Differential Equation, Neural ODE)通过将神经网络参数化为微分方程的导数函数,在连续时间域内刻画隐藏状态的演化过程,从而实现对系统动态的连续建模。在处理物理约束显著的动力系统以及不规则采样轨迹等方面展现出独特优势。Liang 等^[18]于 2021 年提出了融合时空门控机制的时空常微分方程(Spatio-Temporal Ordinary Differential Equation, ST-ODE)模型,将 Neural ODE 的连续时间建模能力引入随机潜变量空间中,实现对空间轨迹的连续生成。该方法在一定程度上缓解了离散时间建模难以保证轨迹平滑性的问题,但主要针对单目标轨迹建模展开,空间特征处理方式仍依赖向量拼接与线性映射,在涉及多目标协同行为场景中难以充分刻画目标间复杂的交互关系。近年来,相关研究尝试将图结构引入连续时间建模框架以提升多目标交互场景下模型的时空预测能力。2023 年 Xiong 等^[19]提出动态图常微分方程(Dynamic Graph Ordinary Differential Equation, DGODE)模型,通过图结构显式描述目标间的关联关系,构建自适应扩散矩阵挖掘潜在节点依赖,并结合 Neural ODE 对交通网络中的时空动态进行了连续建模。Wen 等^[20]于 2024 年提出了基于智能体图的常微分方程(Agent Graph Ordinary Differential Equation, Agent Graph ODE)模型,利用图神经网络构建动态交互图来表征智能体目标间的空间拓扑结构,并借助 ODE 对多智能体目标交互及动态连续变化进行建模,有效改善了离散模型在模拟智能体间交互行为时物理平滑性不足的问题,实现了对多智能体轨迹的连续预测。

上述工作多依赖于预定义的全局图或密集拓扑结构,未能充分考虑多目标交互中的全连接图结构会引入大量冗余的交互噪声,在后续的 Neural ODE 的数值积分过程中,这些冗余的交互噪声会被导数函数反复迭代并层层累积,不仅显著增加了模型的计算负载,还会使 Neural ODE 捕捉到的导数向量场出现方向性偏差,进而产生显著的轨迹漂移等问题,

导致预测结果的偏差。

鉴于此，本文提出一种基于稀疏图卷积^[21]改进的 Neural ODE 网络模型。针对传统的全连接图结构导致的空间冗余与噪声放大问题，利用稀疏图卷积剔除了多目标密集交互的冗余噪声；针对传统离散时间序列建模离散点处跳跃连接导致的目标物理平滑性不足，利用 Neural ODE 求解器的自适应积分机制，将离散状态映射重构为潜在空间中连续的动力学演化路径，实现了空间拓扑的稀疏性约束与时间维度的连续动力学建模的融合。本文方法同时捕捉多目标间的动态空间交互和连续时序演化特征，提升了模型对动态环境中多目标时空建模能力，提高动态多目标时空预测的精度。

1 Neural ODE 网络

1.1 问题描述与分析

在多目标时空预测任务中，假设存在 n 个目标，其在空间中的位置可表示为二维坐标

$$X_t^n = (x_t^n, y_t^n), t \text{ 为连续时间序列。这里以第 } n$$

个目标为例，希望根据一段 t 时间内的历史轨迹

$$X_{\text{obs}}^n = \{X_{t_0}^n, X_{t_1}^n, \dots, X_{t_n}^n\}$$

$$\text{迹 } X_{\text{pred}}^n = \{X_{t_n+1}^n, X_{t_n+2}^n, \dots, X_{t_n+\tau}^n\}。$$

当前的动态多目标时空预测模型多采用 RNN^[22]、TCN^[23]等方法编码时序信息，主要依赖于隐藏状态在离散时间步上的跳跃式更新。而本文研究的问题聚焦于由离散采样点推演至未来 τ 个时间步内的预测，在离散时间间隔内，如 $t_0 \sim t_1$ 的两个相邻时间采样点区间内，由于 RNN 与 TCN 缺乏连续时间形式的显式数学定义与描述，难以描述区间内状态的演化过程，导致预测结果在本质上表现为由离散点构成的分段式连接，如图 1(a)所示。相比之下，本文采用 Neural ODE 进行连续时间建模，如图 1(b)所示，通过参数化状态导数，将预测任务转化为动力学方程的求导与积分过程，实现状态在时间域上的连续演化，提升模型连续时间建模的能力。

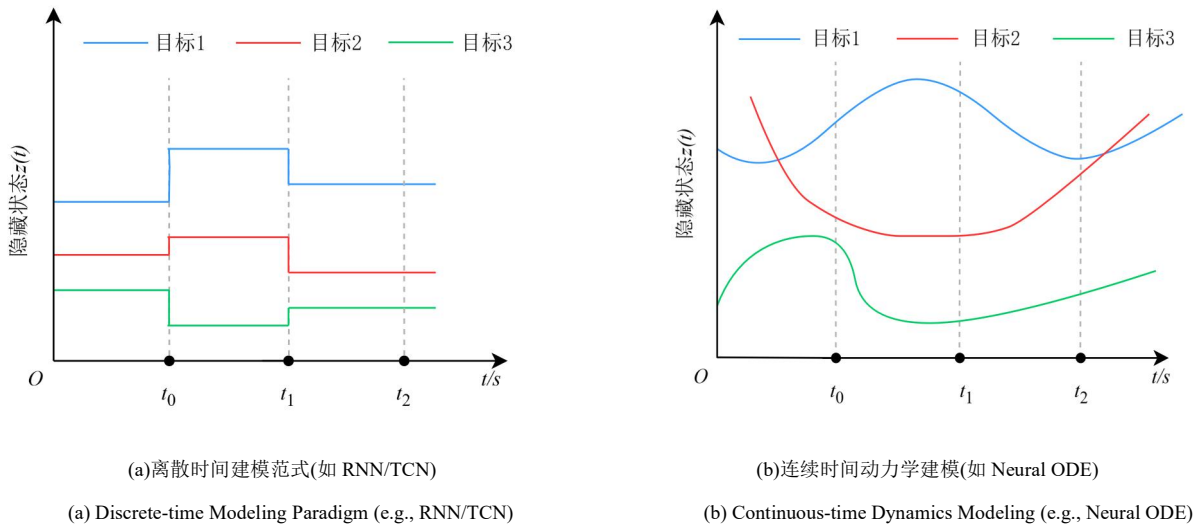


图 1 隐藏状态建模对比图

Fig.1 Comparison diagram of hidden state modeling

因此，本文将多目标时空预测问题建模为一个时空动态系统中的连续状态演化问题，具体而言，将多个目标的状态信息随时间变化的过程统一表示为一段连续状态轨迹，突破传统离散时间建模方法在相邻时间采样点间动态信息缺失与演化不连续的局限。基于这一思想，本文引入 Neural ODE 方法，通过构建一个由神经网络参数化的隐藏状态连续动力学模型，参数化状态导数并在时间域上进行连续积分，从而以连续形式刻画目标在时空中的演变过程，提升模型对动态行为的表达能力。

1.2 模型实现

图 2 展示了 Neural ODE 的基本结构，如图所示， t_0, t_1, \dots, t_n 表示一组连续的时间序列节点； x_0, x_1, \dots, x_n 为对应时间节点上的离散观测点； $z(t)$ 为隐藏状态序列， $z(t_0)$ 为初始隐藏状态。模型在给定一段时间区间内，将历史观测数据编码为 $z(t_0)$ ，然后根据由神经网络定义的向量场函数，通过 ODE 求解器对状态演化过程进行数值积分，得到在任意时间点上的连续隐藏状态表示 $z(t)$ 。最后，模型在指定的离散时间节点上对 $z(t)$ 进行采样，并通过解码器将其映射回观测空间，得到对应的预测结果，并与真实的离散观测数据进行误差计算，从而实现模型参数的全局

优化，以逐步提升模型的预测精度与鲁棒性。

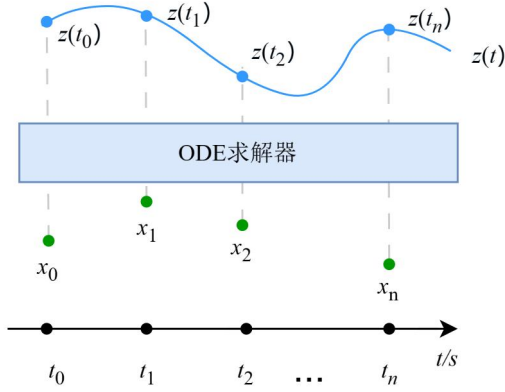


图 2 Neural ODE 网络模型

Fig.2 Network model of Neural ODE

传统的深度神经网络由若干离散的层组成，即 z_{l+1}

式中： $z_l \in \mathbf{R}^d$ 表示神经网络的第 l 层，代表网络前一层层的输出； $f(\cdot, \theta_l)$ 表示非线性变换函数，参数 θ_l 表示第 l 层神经网络中的可学习参数集合。将网络中的各层看作连续时间中的一个隐藏状态，在层数趋近无穷小时，就可以将其隐藏状态视为时间的函数，通过离散观测数据学习得到一个神经网络参数化的向量场函数 f_θ 来描述状态随时间的变化。Neural ODE 对隐藏状态进行连续建模为

$$\frac{dz(t)}{dt} = f_\theta(z(t), t) \quad (2)$$

式中： $z(t)$ 为 t 时刻的隐藏状态； $f_\theta(\cdot)$ 表示由参数 θ 构成的神经网络，拟合隐藏状态演化的连续向量场。

如式(3)所示，给定初始隐藏状态 $z(t_0)$ ，ODE 求解器通过数值积分的方式在时间区间 $[t_0, t_1]$ 上求解出最终状态 $z(t_1)$ 。将向量场函数、初始状态与历史时间序列分别输入求解器中，求解得到高维隐藏状态的张量表示，如式(4)所示。

$$z(t_1) = z(t_0) + \int_{t_0}^{t_1} f_\theta(z(t), t) dt \quad (3)$$

$$\{z(t_1), \dots, z(t_n)\} = \text{ODESolve}(f_\theta, z_0, \{t_1, \dots, t_n\}) \quad (4)$$

式中： $\{z(t_1), \dots, z(t_n)\}$ 表示连续时间上高维隐藏状态。将隐藏状态投影以生成具体的预测值，公式如下：

$$\hat{x}_t = W \cdot z(t) + b \quad (5)$$

式中： \hat{x}_t 为 t 时刻的预测值； W 、 b 分别为映射线性变换时的权重与偏置。

2 稀疏图卷积改进网络

在多目标时空预测中，密集或全连接的图结构难以过滤无关交互，易产生冗余噪声，导致 Neural ODE 在后续的数值积分过程中面临噪声放大与误差累积问题，进而产生显著的轨迹漂移等问题，影响预测精度。因此，本文提出了一种基于稀疏图卷积改进的 Neural ODE 网络模型，模型结构如图 3 所示，模型利用自注意力机制分别得到时空图邻接矩阵，通过交互掩码模块分别构建动态空间与时间稀疏邻接矩阵，并进行双向稀疏图卷积操作对多目标间的交互关系与时间依赖性进行建模，有效过滤冗余交互。随后将所得时空特征输入 Neural ODE 中，利用 Neural ODE 实现连续时间动态演化，从而显著提升了整体时空预测任务的精度与轨迹的物理平滑性。

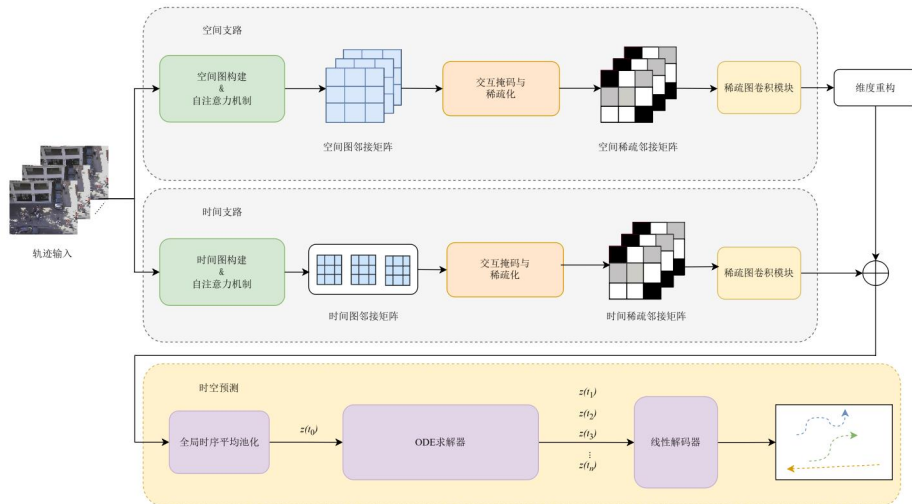


图 3 模型架构图

Fig.3 Model architecture diagram

2.1 稀疏加权邻接矩阵构建

2.1.1 时空信息编码

自注意力能够有效提取多目标间动态交互关系。空间注意力以每一时间步中所有目标为节点，对目标间的空间关联进行建模；时间注意力机制则以同

一目标在不同时间步上的状态为节点，建模目标行为的时间演化过程。通过分别计算空间和时间维度上的多头自注意力权重，得到空间图和时间图邻接矩阵，分别表示所有目标在当前帧的交互强度与连续时间步上的动态相关性，为后续稀疏邻接矩阵的

构建奠定基础。计算公式如下：

$$A_{ij} = \left[\text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) \right]_{ij} \quad (6)$$

式中： A_{ij} 为由注意力机制产生的邻接矩阵代表注意力权重， i, j 为目标； Q, K 分别为通过线性映射生成的查询向量和键向量； d_k 为 Q, K 的向量维度。

2.1.2 邻接矩阵稀疏化

为增强空间注意力图的时序一致性，利用 1×1 卷积和 PReLU 非线性激活函数对不同时间步上的空间注意力图在通道维度上进行融合；而对于时间注意力图则进行特征拼接即可。为了进一步提升对时空注意力图中的有效交互信息的关注度，引入非对称卷积和 Sigmoid 函数分别构建时空图的交互掩码。采用 7 个堆叠的非对称卷积层，分别沿时间维度和空间维度提取局部与全局特征，使用 Sigmoid 函数映射到 $[0,1]$ 区间内，得到交互图的掩码 $M^{(s)}$ 和 $M^{(t)}$ 。设置稀疏化阈值 ε ，将掩码与注意力图相乘，对小于 ε 的结果置零，公式如下：

$$\hat{A}_{ij} = \begin{cases} A_{ij} \cdot M_{ij}, & M_{ij} > \varepsilon \\ 0, & \text{其他} \end{cases} \quad (7)$$

式中： \hat{A}_{ij} 为稀疏邻接矩阵； M_{ij} 为掩码图； ε 为设定的稀疏化阈值。

在稀疏注意力矩阵归一化阶段，采用 ZeroSoftmax 函数^[15]替代传统 Softmax，以适配稀疏交互结构的建模需求。与 Softmax 不同，该函数在对注意力得分进行非线性映射的过程中能够在保持非零权重相对比例关系的同时，将原本为零的权值严格限制为零，保持了稀疏注意力矩阵原有的稀疏性，公式如下：

$$\text{ZeroSoftmax}(\hat{A}_{ij}) = \frac{(\exp(\hat{A}_{ij}) - 1)^2}{\sum_{k=1}^n (\exp(\hat{A}_{ik}) - 1)^2 + \eta} \quad (8)$$

式中： η 为常数，用来防止除 0 错误。

得益于 ZeroSoftmax 函数对零权重连接的抑制作用，模型能够更集中地关注响应值较高的注意力边，在减少交互噪声的同时增强对关键信息的建模能力。

2.2 双向稀疏图卷积

本文采用双向稀疏图卷积网络分别在时间和空间两个维度上提取特征，以全面刻画多目标间的时空交互关系。在空间-时间维度中，基于空间稀疏邻接矩阵对空间图进行图卷积操作。将得到的空间特征在时间维度上进行重排，结合时间邻接矩阵再次进行图卷积操作，使模型能够在时间序列层面捕获目标状态的连续演变规律。在时间-空间维度上与前

者顺序相反，网络更多聚焦于时间演变，在时间维度上应用时间图卷积捕获目标的行为序列演化特征。将提取到的时间特征在空间维度上再次进行图卷积操作，形成多目标时空关联表示。在每个时间步上利用空间稀疏邻接矩阵来聚合具有时间感知的目标特征，实现目标间的交互建模。每个方向上均包含两层图卷积，其形式为

$$H^{(l+1)} = \sigma(\hat{A}_{ij} H^{(l)} W^{(l)}) \quad (9)$$

式中： $H^{(l)}$ 表示 l 层输入特征； σ 为非线性激活函数； $W^{(l)}$ 为可学习权重。

最后在通道维度上使用 1×1 卷积进行整合，实现双路特征融合，得到融合后的时空特征 Z 。双向稀疏图卷积模块通过在两个方向上进行两次图卷积操作，有效捕捉了多目标间的局部交互关系与跨时间的长程依赖，实现目标间的时空关联建模，为后续 Neural ODE 进行连续时间建模提供了重点突出、结构清晰的多目标时空特征。

2.3 Neural ODE 解码

在 Neural ODE 网络的编码部分，采用稀疏图卷积网络实现了动态多目标在历史轨迹中的时空关联建模，而后利用 Neural ODE 网络对持续时序演变的强建模能力，实现稀疏图卷积特征在时序上的建模，以实现动态多目标的时空预测。

本文首先对所得的时空特征张量在时间维度上进行全局平均池化，以获得用于 Neural ODE 建模的二维状态表示。向量场函数采用多层感知机来实现，公式如下：

$$f_{\theta}(t) = \frac{dz(t)}{dt} = W_2 \cdot \tanh(W_1 \cdot t + b_1) + b_2 \quad (10)$$

式中： W_1, W_2 为全连接层的权重参数； b_1, b_2 为偏置项； $\tanh(\cdot)$ 为非线性激活函数。

基于上述向量场定义，将初始状态与预测时间序列输入至 Neural ODE 求解器中，采用 Euler 数值积分方法进行近似求解，得到预测时间序列内的隐藏状态序列。该过程实现了多目标时空轨迹在连续时间上的动态演化，最终通过线性映射将隐藏状态转换为各时间步对应的预测轨迹结果。

3 实验与结果分析

3.1 数据集

本文采用 ETH^[24]/UCY^[25]数据集进行测试评估，涵盖了从稀疏到拥挤的不同空间密度下的多目标轨迹。轨迹数据以连续帧的目标位置序列形式给出，不仅记录了目标在物理空间中的运动过程，更包含了多目标间的交互结构。其中，ETH 包含 ETH 和 HOTEL 2 个子集，UCY 包含 UNIV、ZARA1 和

ZARA2 共 3 个子集。实验中将连续的数据划分为包含 8 帧历史观察序列和 12 帧未来轨迹的预测序列。所有位置坐标均以首帧为参考进行位移表示，并针对不同场景分别进行归一化处理，以提升训练过程的数值稳定性。

3.2 实验设置

本模型在 PyTorch 框架下实现，所有实验验证均在单块 NVIDIA GeForce RTX 3090 GPU 上完成。训练过程中采用单层自注意力机制，嵌入维度设置为 64，注意力头数为 4。非对称卷积网络由 7 层堆叠构成，稀疏阈值设置为 0.5。连续时间建模部分采用由 4 层 MLP 构成的 Neural ODE 模块，数值求解过程选用 Euler 方法。模型训练阶段使用 Adam 优化算法，训练轮次设置为 300，批大小为 128，初始学习率取 0.01。在学习率调度方面采用 MultiStepLR 调度器，并将第 50 和 100 轮设置为关键节点，衰减因子为 0.5。在每一轮训练结束后进行评估，并保存当前最佳模型。

3.3 评价指标

本文采用平均位移误差（Average Displacement Error, ADE）和最终位移误差（Final Displacement Error, FDE）两个指标来对模型的预测结果进行评估。ADE 表示在预测时间范围内，预测轨迹与真实

轨迹在每个时间步之间欧氏距离的平均值。采用 ADE 评价指标以反映模型在全局时间序列上的整体预测精度。FDE 则表示预测的最终位置与真实最终位置之间的欧氏距离，体现的是模型在轨迹终点预测方面的准确性，尤其适用于评估模型在长时间序列推理中对目标最终状态的判断能力。定义如下：

$$ADE = \frac{1}{N \times T_p} \sum_{i=1}^N \sum_{t=1}^{T_p} \|\hat{x}_t^i - x_t^i\|_2 \quad (11)$$

$$FDE = \frac{1}{N} \sum_{i=1}^N \|\hat{x}_{T_p}^i - x_{T_p}^i\|_2 \quad (12)$$

式中： N 为目标个数， T_p 为预测帧数； \hat{x}_t^n 为 t 时刻的预测结果； x_t^n 为 t 时刻的真实结果。

3.4 对比实验

为了验证本文提出模型的有效性，在 ETH/UCY 标准数据集的 5 个子场景 ETH、HOTEL、UNIV、ZARA1 和 ZARA2 上与 Neural ODE、SGCN 等多种预测方法进行了对比实验，结果如表 1 所示。实验结果表明，本模型在所有子集上的平均 ADE 和 FDE 分别为 0.36 m 和 0.56 m，为所有对比方法中的最低值，相较对比模型有了显著提高。

表 1 在 ETH/UCY 数据集上的 ADE 和 FDE 结果对比

Table 1 Comparison of ADE and FDE results on ETH/UCY dataset

m

模型	ADE/FDE					
	ETH	HOTEL	UNIV	ZARA1	ZARA2	平均
Neural ODE ^[17]	0.72/1.13	0.28/0.45	0.44/0.77	0.29/0.52	0.24/0.44	0.39/0.66
Social LSTM ^[26]	1.09/2.35	0.79/1.76	0.67/1.40	0.47/1.00	0.56/1.17	0.72/1.54
Social-STGCNN ^[16]	0.64/1.11	0.49/0.85	0.44/0.79	0.34/0.53	0.30/0.48	0.44/0.75
SGCN ^[21]	0.63 /1.03	0.32/0.55	0.37 /0.70	0.29/0.53	0.25/0.45	0.37/0.65
RSBG w/o context ^[27]	0.80/1.53	0.33/0.64	0.59/1.25	0.40/0.86	0.30/0.65	0.48/0.99
Social-BIGAT ^[28]	0.69/1.29	0.49/1.01	0.55/1.32	0.30/0.62	0.36/0.75	0.48/1.00
本文模型	0.63/0.88	0.27/0.36	0.37/0.68	0.28/0.49	0.23/0.41	0.36/0.56

注：加粗数字表示表中最优结果。

整体来看，在 FDE 指标上，本模型的表现均显著优于 Social-LSTM、Social-STGCNN、SGCN 等基于离散时序建模的模型，充分表明了 Neural ODE 的引入通过对隐藏状态进行连续时间域内的数值积分，有效地抑制了长时预测中的误差累积效应。在保障物理平滑性的前提下，显著增强了模型对复杂时空关联的捕捉能力。

具体而言，模型在各个子集上的性能基本达到最优。其中，在 HOTEL、ZARA1 和 ZARA2 子集上，

两项指标均取得最低数值，表明模型在多样化场景下仍具备良好的稳定性。在目标数量较少且运动清晰的 ETH 子集中，本文模型在 FDE 指标上降幅最大，反映出稀疏图卷积能够精准剔除微小的交互噪声，使 Neural ODE 连续演化过程能够最大限度地逼近物理真实轨迹，实现高精度预测。在目标密集、交互复杂的 UNIV 子集上，稀疏图卷积的引入显著增强了模型的鲁棒性。通过交互掩码的方式剔除空间冗余，有效屏蔽了目标间的交互噪声；利用双向稀疏

图卷积从杂乱的交互信息中精准捕获时空关联特征，为 Neural ODE 提供了更清晰准确的向量场，从而确保 Neural ODE 能够更精准地实现对观测隐藏状态的连续时间建模，有效缓解了数值积分过程中的噪声放大与误差累积效应，使得模型在高密度目标且交互复杂的场景下依然能够维持稳健的预测性能。

表 2 展示了不同交互稀疏化阈值 ϵ 值下模型在 5 个子集场景下的预测结果。其中，阈值为 0 代表多目

标间交互程度最大，即保留全部多目标间的关联；阈值为 1 则代表目标间无交互。由表 2 可以看出，稀疏化阈值选择为 0.5 时模型表现最优，平均 ADE 与 FDE 分别为 0.36 m 和 0.56 m，数值最低，优于全关联 ($\epsilon=0$) 或无交互 ($\epsilon=1$) 的情形。因此，对交互稀疏化程度进行合理约束有助于提升复杂环境下的目标时空预测效果。

表 2 不同稀疏化阈值 ϵ 下 ADE 和 FDE 结果对比

ϵ	ADE/FDE					
	ETH	HOTEL	UNIV	ZARA1	ZARA2	平均
0	0.70/1.30	0.29/0.31	0.38/0.72	0.29/0.51	0.24/0.43	0.38/0.65
0.50	0.63/0.88	0.27/0.36	0.37/0.68	0.28/0.49	0.23/0.41	0.36/0.56
1.00	0.64/1.09	0.32/0.34	0.41/0.67	0.43/0.56	0.28/0.48	0.42/0.63

注：加粗数字表示表中最优结果。

3.5 可视化结果

为了更加直观地展示模型的预测效果，将本文模型与 Neural ODE 模型进行了可视化对比。如图 4

所示，图 4(a)为 Neural ODE 模型可视化结果，图 4 (b) 为本文模型可视化结果。



(a)神经常微分方程模型可视化结果

(a) Visualization results of the Neural ODE model



(b)基于稀疏图卷积改进模型可视化结果

(b) Visualization results of the improved model based on sparse graph convolution

图 4 稀疏 (左) 和密集 (右) 场景下可视化结果图

Fig.4 Visualization results in sparse (left) and dense(right) scenarios

在稀疏场景下，当目标间存在跟随超越行为时，Neural ODE 模型在交互初期产生明显的轨迹偏移，

而本文模型有效修正了因交互噪声导致的预测偏差，使得预测轨迹与真实路径保持较高一致性。在密集

场景下, 多名目标在相近方向上同时运动, Neural ODE 模型出现了轨迹回折及明显偏离的情况, 表明其对无关目标干扰较为敏感。相比之下引入稀疏图卷积后, 本文模型通过动态剔除空间交互冗余, 减弱了冗余噪声在后续数值积分过程中的累积效应。由图 4(b)中可见, 改进后的模型预测路径与真实运动趋势一致性更高, 在维持高精度预测的同时显著增强了模型在复杂动态干扰下的鲁棒性。

为避免仅凭视觉判断带来的主观性, 结合图 4 所示样例的两项误差指标进行了量化分析。如表 3 所示, 在 ZARA2 场景样例中, 本文模型的两项指标分别为 0.22 m/0.41 m, 优于 Neural ODE 模型的 0.31 m/0.58 m。在 UNIV 场景样例中, 本文模型的两项误差指标分别为 0.39 m/0.66 m, 同样优于 Neural ODE 模型, 进一步证明了本文模型在不同目标密度的场景下都保持了良好的预测精度。同时, 相较于 Neural ODE 模型, 本文模型通过引入稀疏图卷积动态剔除空间交互冗余, 减弱了冗余噪声在后续数值积分过程中的累积效应, 显著提升预测精度。

表 3 可视化图片中的 ADE 和 FDE 结果

模型	ADE/FDE	
	ZARA2-222	UNIV-316
Neural ODE ^[17]	0.31/0.58	0.42/0.83
本文模型	0.22/0.41	0.39/0.66

4 结论

本文提出了一种基于改进 Neural ODE 网络的多目标时空预测模型, 以实现动态多目标的时空预测。在模型结构上, 引入的稀疏图卷积网络成功捕捉了多目标间复杂的时空交互关系, 结合 Neural ODE 连续时间建模的强大能力, 显著提升了预测的时序连续性与动态表达能力。通过在 ETH/UCY 标准数据集上的实验验证, 本文模型在 ADE 和 FDE 指标上的平均值分别为 0.36 m 和 0.56 m, 相较于各个对比方法表现最优, 相比于 Neural ODE 模型在 ADE 和 FDE 误差上分别降低了 3%和 10%, 展现出较高的预测准确度。

但模型仍存在一定的局限性, 如未显式引入场景语义信息, 后续将进一步结合高精度地图等以增强模型对复杂环境的理解, 实现准确高效的预测。

参考文献 (References)

[1] CHEN X B, ZHANG H J, ZHAO F, et al. Intention-aware vehicle trajectory prediction based on spatial-temporal dynamic attention network for internet of vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 19471-19483.DOI: 10.1109/TITS.2022.3170551.

[2] MO X Y, HUANG Z Y, XING Y, et al. Multi-agent trajectory prediction with heterogeneous edge-enhanced graph attention network[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(7): 9554-9567. DOI: 10.1109/TITS.2022.3146300.

[3] ZHAI L B, ZHU X M, CHENG C. Energy efficient data evacuation path for multi-UAV system based on multi-objective optimization method[J]. IEEE Transactions on Vehicular Technology, 2025, 74(5): 7364-7377.

[4] FENG Z K, WU D, HUANG M X, et al. Graph-attention-based reinforcement learning for trajectory design and resource assignment in multi-UAV-assisted communication[J]. IEEE Internet of Things Journal, 2024, 11(16): 27421-27434.

[5] WANG Z H, XUE L, LUO X P, et al. Driving state-aware anomaly detection for autonomous vehicles[J]. IEEE Transactions on Information Forensics and Security, 2025, 20:3788-3803.

[6] 崔皓,叶洪涛,罗文广,等. 车辆纵向队列分布式多目标模型预测控制[J]. 控制工程, 2025, 32(12): 2126-2134. Cui H, Ye H T, Luo W G, et al. Distributed multi-objective model predictive control of vehicle longitudinal queues[J]. Control Engineering, 2025,32(12): 2126-2134. (in Chinese)

[7] ZHANG B, LIU X, ZHENG H F, et al. Hidden Markov model based cyberattack prediction in power systems[J]. IEEE Transactions on Smart Grid, 2024, 16(2): 1694-1705.

[8] BECKERS T, HIRCHE S. Prediction with approximated Gaussian process dynamical models[J]. IEEE Transactions on Automatic Control, 2021, 67(12): 6460-6473.

[9] LAI G K, CHANG W C, YANG Y M, et al. Modeling long-and short-term temporal patterns with deep neural networks[C]//Proceedings of the 41st international ACM SIGIR Conference on Research & Development in Information Retrieval. Ann Arbor, MI, US: ACM, 2018: 95-104.

[10] LEA C, FLYNN M D, VIDAL R, et al. Temporal convolutional networks for action segmentation and detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Honolulu, HI, US: IEEE, 2017: 156-165.

- [11] WU H X, XU J H, WANG J M, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. Virtual: Curran Associates, 2021: 22419-22430.
- [12] DONG M. Pedestrian cross forecasting with hybrid feature fusion[C]//Proceedings of the 15th Asian Conference on Machine Learning. Wuhan, China: PMLR, 2024: 327-342.
- [13] GUPTA A, JOHNSON J, LI F F, et al. Social GAN: socially acceptable trajectories with generative adversarial networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, US: IEEE, 2018: 2255-2264.
- [14] SALZMANN T, IVANOVIC B, CHAKRAVARTY P, et al. Trajectron++: dynamically-feasible trajectory forecasting with heterogeneous data[C]//Proceedings of the European Conference on Computer Vision. Glasgow, UK: Springer, 2020: 683-700.
- [15] WANG X C, YANG X, ZHOU D K. Goal-CurveNet: a pedestrian trajectory prediction network using heterogeneous graph attention goal prediction and curve fitting[J]. *Engineering Applications of Artificial Intelligence*, 2024, 133: 108323.
- [16] MOHAMED A, QIAN K, ELHOSEINY M, et al. Social-STGCNN: a social spatio-temporal graph convolutional neural network for human trajectory prediction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, US: IEEE, 2020: 14424-14432.
- [17] CHEN R T Q, RUBANOVA Y, BETTENCOURT J, et al. Neural ordinary differential equations[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates, 2018: 6572- 6583.
- [18] LIANG Y X, OUYANG K, YAN H S, et al. Modeling trajectories with neural ordinary differential equations[C]//Proceedings of the 30th International Joint Conference on Artificial Intelligence. Montréal, Canada: IJCAI, 2021: 1498-1504.
- [19] XIONG N, YANG Y, JIANG Y Q, et al. Diffusion graph neural ordinary differential equation network for traffic prediction[C]//Proceedings of the 2023 International Joint Conference on Neural Networks. Gold Coast, QLD, Australia: IEEE, 2023: 1-8.
- [20] WEN S, WANG H, LIU D, et al. Second-order graph odes for multi-agent trajectory forecasting[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, HI, US: IEEE, 2024: 5101-5110.
- [21] SHI L S, WANG L, LONG C J, et al. SGCN: Sparse graph convolution network for pedestrian trajectory prediction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, US: IEEE, 2021: 8994-9003.
- [22] RANGAPURAM S S, SEEGER M, GASTHAUS J, et al. Deep state space models for time series forecasting[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates, 2018: 7796–7805.
- [23] LUO D H, WANG X. ModernTCN: a modern pure convolution structure for general time series analysis[C]//Proceedings of the Twelfth International Conference on Learning Representations. Vienna, Austria: ICLR, 2024: 1-43.
- [24] PELLEGRINI S, ESS A, SCHINDLER K, et al. You'll never walk alone: modeling social behavior for multi-target tracking[C]//Proceedings of the IEEE 12th International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009: 261-268.
- [25] LERNER A, CHRYSANTHOU Y, LISCHINSKI D. Crowds by example[C]//Computer Graphics Forum. Oxford, UK: Blackwell Publishing Ltd. 2007, 26(3): 655-664.
- [26] ALAHI A, GOEL K, RAMANATHAN V, et al. Social LSTM: human trajectory prediction in crowded spaces[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, US: IEEE, 2016: 961-971.
- [27] SUN J H, JIANG Q H, LU C W. Recursive social behavior graph for trajectory prediction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, US: IEEE, 2020: 660-669.
- [28] KOSARAJU V, SADEGHIAN A, MARTÍN-MARTÍN R, et al. Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, BC, Canada: Curran Associates, 2019, 32: 137-146