利用道路进出时间戳学习交通流动力学与轨迹的物理约束神经常微分方程*

Dennis Bogatov Wilkman[1],Karl Henrik Johansson[1],和 Matthieu Barreau[1]

arXiv:2611.30577v1 [cs.LG] 23 Nov 2026

摘要—**我们研究如何利用摄像机记录的路段进出时间戳（EET），以及领跑车队的探测车辆的位置，学习微观交通动力学并重建车辆轨迹。由于每辆车的位置仅在两个点上测得，其间的轨迹未知，因此动力学难以恢复。通过将神经常微分方程与局部密度近似相结合，我们学习一种跟驰模型，其中速度–密度关系由神经网络表示。对于如此稀疏的数据，无约束网络可能学到不符合物理规律的关系，例如负速度，或道路完全拥堵时仍存在非零速度。因此，我们引入一种归一化方法，将其应用于带有学习得到的状态变化的单调神经网络，从而强制满足端点条件，保证速度–密度关系可容许，同时还能在理论上证明，该方法保留了逼近任意可容许关系的能力。我们在一阶和二阶仿真上验证了该方法：成功恢复一阶模型的速度–密度关系，并重建两种模型的轨迹；在长 125 m 的路段上，两者的平均最大位置误差分别为 2.50 m 和 2.51 m。在真实环形道路数据上，所提出的模型仍能得到可容许关系，而两种基线方法均违反约束；然而，数据的高度离散也凸显了一阶模型在驾驶员行为均一假设下的局限性。**

I. 引言

交通拥堵会增加排放和能源消耗，并浪费行程时间。这些损失大多发生在拥堵状态下，此时交通速度下降，扰动会被放大 [1]。诸如使用自动驾驶车辆等控制策略可以抑制这些扰动 [2], [3], ，但这需要准确估计交通状态及其动力学。探测车辆和新型基础设施传感器等近期技术进展，可用于实现这一目标。

从方法论角度来看，交通流动力学和轨迹可通过基本图（FD）重构 [4], 、离散时间或连续时间循环神经网络 [5], [6],，以及结合基于模型和基于数据的表述的物理信息机器学习 [7], [8]来估计。物理信息方法尤其灵活，能够融合来自多种不同来源的测量值 [9]。

这些算法依赖传感器数据。几十年来，人们一直利用环形线圈检测器的流入和流出测量值来重建高速公路交通密度 [1]。近年来，研究人员开始探索利用探测车辆、网联车辆或自动驾驶车辆，从交通流内部获取车辆级测量，以实现更准确的重建 [3], [10]。

研究界迅速采用探测车辆，部分原因是与标准传感器相比，其安装成本较低。然而，要实时收集和处理探测车辆的数据，需要复杂的通信架构 [11]。本文提出使用一辆位置随时间已知的探测车辆，并结合已安装用于测速执法的传感器，重建道路路段上的交通状况和动力学。这些路段由自动车牌识别（ANPR）摄像机划定，摄像机记录每辆车辆的进出时间戳（EET） [12], [13]。这些时间戳包含行程时间信息，我们利用这些信息推断交通流动力学。

本文使用神经常微分方程（神经常微分方程） [14]学习跟车模型的速度–密度关系，并根据EET数据和一辆探测车辆的轨迹重建每辆车辆的轨迹。我们的贡献有三点：(i) 将神经常微分方程训练调整为适用于EET测量，在这种情况下，每辆车辆仅在进入和离开道路路段时被观测到；(ii) 提出一种速度-密度网络，通过构造满足物理约束，将归一化单调网络与学习得到的状态变化相结合，同时保留近似任意可容许关系的能力；(iii) 在一阶和二阶交通仿真以及真实环形道路数据 [2]上验证该方法。之后

我们在第二节和第三节介绍预备知识与问题，在第四节提出（i）和（ii），在第五节提出（iii），并在第六节总结。

II. 预备知识

在陈述问题之前，我们将介绍微观交通流建模、神经常微分方程和受约束神经网络的基础知识，以清晰地阐述我们的贡献。

A. 微观交通流建模

对交通流建模最精细的方法是微观建模：每辆车都有自己的轨迹，该轨迹是常微分方程的解 [1, 第 5 节。2]在一阶假设下，车辆 $i \in \{1, \dots, N\}$ 的速度直接取决于其周围的局部密度$\rho_i(t)$，由此得出跟驰模型

*本研究由 Knut 和 Alice Wallenberg 基金会资助的 Wallenberg人工智能、自主系统与软件计划（WASP）支持。[1]作者来自瑞典皇家理工学院电气工程与计算机科学学院决策与控制系统系、Digital Futures，地址：瑞典 100-44 {dwilkman, kalle, barreau}@kth.se

$\dot{x}_i(t) = v(\rho_i(t))$，$x_i(0) = x_i^0$，其中 x_i 是车辆 i 的位置，x_i^0 是其初始位置。全局密度是道路上的车辆数除以道路长度，与占有率成正比；局部密度则是路段长度趋近于 0 时的极限。如果 $x_0^0 > x_1^0 > \cdots > x_N^0$，车辆 i 周围的局部密度近似为：

$$\rho_i(t) \simeq \frac{L_{car}}{x_{i-1}(t) - x_i(t)}, \tag{1}$$

其中 L_{car} 表示车辆的长度加上车辆之间的最小间距 [15]。因此，密度被归一化为：$\rho_i \to 0$ 为 $x_{i-1} - x_i \to \infty$，而 $\rho_i \to 1$ 为 $x_{i-1} - x_i \to L_{car}$，即车辆首尾相接时。此时，跟驰模型变为：

$$\begin{cases} \dot{x}_i(t) = v\left(\frac{L_{car}}{x_{i-1}(t) - x_i(t)}\right), & x_i(0) = x_i^0, \\ x_0 \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}_+), \end{cases} \tag{2}$$

其中 x_0 表示领航车辆的已知轨迹。

速度–密度关系 v 源自基本图 [1, 第 2.1 节3]，通常部分未知且具有随机性 [16]。不过，可以推断出一些基本性质。首先，当道路几乎为空时，即$\rho \approx 0$，车辆以自由流速度 v_{free}行驶，不受周围车辆影响，如自由流状态 [17]所示。其次，当车辆首尾相接时，即 $\rho \approx 1$，它们无法移动，否则就会发生碰撞。当密度超过临界密度时，车辆会受到前方车辆的影响而减速；这种拥堵状态的建模要复杂得多。此外，车辆的速度平均而言会随局部密度 [7]增加而降低：道路越不拥堵，速度越高。我们将 v 的这些性质总结如下：

$$v(0) = v_{free}, \tag{3a}$$
$$v(1) = 0, \tag{3b}$$
$$v(\rho_1) \geq v(\rho_2), \quad 0 \leq \rho_1 \leq \rho_2 \leq 1. \tag{3c}$$

这些观察结果促成了对稳态均匀行为的各种模型。Greenshields [18]提出了线性模型 $v(\rho) = v_{free}(1 - \rho)$，该模型具有便于分析的性质，但无法捕捉状态转变。Newell 和 Daganzo [19], [20] 随后提出了一种通过三角形流量–密度关系 $q(\rho) = \rho v(\rho)$来描述状态转变的模型，该关系对应的速度–密度关系为

$$v_{ND}(\rho) = v_{free} \min\left(1,\, w\left(\frac{1}{\rho} - 1\right)\right), \tag{4}$$

其中 $w > 0$，该关系在临界密度 $\rho_c = w/(1 + w)$处存在一个折点，将自由流状态与拥堵状态分隔开来。无论采用何种模型，拥堵状态都难以捕捉，因为反应时间等因素可能导致同一密度对应不同的速度，这表明一阶模型——即速度瞬时适应密度的模型——对于现实交通而言是过于粗略的近似 [21]。

二阶模型通过对加速度而非速度进行建模，避免速度发生瞬时变化，其依据是车辆自身的速度、间距 $s_i(t) = x_{i-1}(t) - x_i(t)$，以及与前车的相对速度 $\Delta v_i(t) = v_{i-1}(t) - v_i(t)$：

$$\begin{cases} \dot{x}_i(t) = v_i(t), & x_i(0) = x_i^0, \\ \dot{v}_i(t) = a\left(v_i(t), s_i(t), \Delta v_i(t)\right), & v_i(0) = v_i^0, \\ x_0, v_0 \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}_+), \end{cases} \tag{5}$$

其中 x_0, v_0 是已知的前车轨迹和速度，而加速度函数 a 是连续的。尽管二阶模型无法捕捉现实交通固有的随机性，但它们能够再现更真实的行为，因为速度可以偏离基本图，同时在平均意义上保持接近该基本图 [1,第 5.2 节。1]。其缺点是，如果缺乏谨慎的建模，可能会发生碰撞，因为当间距趋近于 0 时，速度仍可能保持非零。在现有的二阶模型中，智能驾驶员模型（IDM） [22] 将使车辆趋向其期望速度的自由道路项，与维持其与前车之间安全间距并预判前车速度变化的交互项相结合，从而避免碰撞。

B. 使用神经常微分方程学习动力学

我们将一个包含 [23] 层、每层有 n_ϕ 个神经元的多层感知机（MLP） [23] with $l_\phi \geq 1$ 定义为

$$\phi(\rho; \theta) := W_{l_\phi+1}\big(H_{l_\phi} \circ \cdots \circ H_1\big)(\rho) + b_{l_\phi+1}, \tag{6}$$

其中，$H_i(z) := \sigma(W_i z + b_i)$ 表示第 i 个隐藏层，该层采用逐元素非线性激活函数 σ $\theta = \{W_i\ b_i\}_{i=1}^{l_\phi+1}$；$W_1 \in \mathbb{R}^{n_\phi \times 1}$ 而 (,) 汇集了权重，$W_{l_\phi+1} \in \mathbb{R}^{1 \times n_\phi}$ $W_i \in \mathbb{R}^{n_\phi \times n_\phi}$ $b_{l_\phi+1} \in$，以及其他项，偏置则为 $\mathbb{R}$ 和 $b_i \in \mathbb{R}^{n_\phi}$，以及其他项。神经元数量足够多时，多层感知机可以在紧集上以任意精度逼近任意连续函数 [24]。

考虑一个状态为 $x(t) \in \mathbb{R}^D$ 的动力系统，

$$\dot{x}(t) = f\big(t, x(t)\big), \qquad x(0) = x^0, \tag{7}$$

其动力学 f arepartially unknown，例如式 (2) 中的速度–密度关系或式 (5) 中的加速度。神经ˆ常微分方程 [14] replaces f，采用可训练动力学 $f(\cdot, \cdot; \theta)$，通常由参数为 θ 的多层感知机构成，并通过从相同初始条件对其积分来预测状态：

$$\hat{x}(t; \theta) = x^0 + \int_0^t \hat{f}\big(\tau, \hat{x}(\tau; \theta); \theta\big)\, d\tau, \tag{8}$$

其中，积分使用数值求解器计算，例如四阶龙格–库塔法（RK4）。通过最小化比较ˆx 与 x测量值的损失，可以求得参数 θ；这些测量值可能仅在少数几个时刻可用，例如终端代价 $\Phi(\hat{x}(T; \theta)) = \|\hat{x}(T; \theta) - x(T)\|^2$。学习动力学而非轨迹，能够推广到采样点之外，因为模型捕捉了生成数据的机制。它ˆ还允许通过在已知物理模型 [25] 中加入学习得到的修正项，或像本文这样约束其结构，将已知物理规律融入 f。

通过对求解器进行反向传播来计算损失的梯度，需要存储每个中间状态，因此内存占用会随求解器步数增加，从而限制问题的时间窗口。伴随灵敏度方法 [26], [14]通过反向求解第二个常微分方程来避免这一问题，并且适用于还包含运行代价 ℓ [27]的损失:

$$\mathcal{L}(\theta) = \Phi\big(\hat{x}(T;\theta)\big) + \int_0^T \ell\big(t, \hat{x}(t;\theta)\big)\, dt. \tag{9}$$

梯度为

$$\nabla_\theta \mathcal{L}(\theta) = \int_0^T \left(\frac{\partial \hat{f}}{\partial \theta}\right)^\top \lambda(t)\, dt, \tag{10}$$

其中，伴随变量 λ 通过向后求解以下方程得到:

$$\begin{aligned} \dot{\hat{x}}(t) &= \hat{f}\big(t, \hat{x}(t);\theta\big), \qquad \hat{x}(T) \text{ given by (8)}, \\ \dot{\lambda}(t) &= -\left(\frac{\partial \hat{f}}{\partial \hat{x}}\right)^\top \lambda(t) - \left(\frac{\partial \ell}{\partial \hat{x}}\right)^\top, \\ \lambda(T) &= \left(\frac{\partial \Phi}{\partial \hat{x}}\big(\hat{x}(T)\big)\right)^\top. \end{aligned} \tag{11}$$

所有偏导数均在 $(t,\ \hat{}x(t;\ \theta);\theta)$ 处计算。为简洁起见，省略了 $\hat{}x$ 对 θ 的依赖。随后将梯度 (10) 输入基于梯度的优化器（如 Adam [28], ），以更新 θ。

C. 受约束神经网络

将先验知识融入多层感知机的一种方法是通过定制架构施加约束，确保这些约束对所有参数值都成立。边界条件和端点条件通常以这种方式强制实现 [29], [30]。对于 (3) 中的端点条件， [8] 将多层感知机 (6) 乘以一个在 $\rho = 0$ 和 $\rho = 1$ 处为零的因子，并加上一个在 $\rho = 0$ 处等于 v_{free}、在 $\rho = 1$ 处等于 0 的项:

$$\tilde{\phi}(\rho;\theta) := \rho(1-\rho)\,\phi(\rho;\theta) + (1-\rho)\, v_{free}, \tag{12}$$

我们将其称为固定端点模型。于是，对于任意 θ, ~ ~$\phi(0;\ \theta) = v_{free}$，且 $\phi(1;\theta) = 0$。这种硬约束的主要优点是可靠：无论数据是否包含端点信息，或训练是否受到噪声干扰，学得的关系都满足物理性质。然而，(12) 并未˜强制保证单调性，因此 ϕ 在 (0，1) 内可能超过 v_{free} 或变为负值。

也可以通过构造来强制满足单调性：如果式 (6) 中所有权重 W_i 均为正，且 σ 严格递增（例如双曲正切），那么 ϕ 就严格递增；将网络的输出取负，即可得到递减网络。这类网络能够以任意精度逼近 $[0,1]$ 上的任意连续非递减函数 [31]。然而， ϕ 不能简单地与单调性构造相结合。首先，单调的 $\tilde{\phi}$并不能使 ϕ 单调：例如，恒定的 $\phi \equiv c > v_{free}$会使 ϕ 在 $\rho = 0$附近增大。其次，单调的ϕ 无法表示所有可容许关系：对于˜而言，若 ϕ等于 Newell–Daganzo 关系，则对应的 ϕ在临界密度以下递增，并在临界密度以上不久便开始递减。第四节提出了一种构造方法，既能强制满足这两项性质，又能保留通用逼近能力。

III. 问题描述

我们考虑一段单车道道路区段 $[x_e, x_f]$，一辆领航探测车辆沿其行驶，其轨迹为 $t \mapsto x_0(t)$，另有 N辆跟驰车辆 x_i, $i \in \{1, \dots, N\}$，其位置依次为 $x_i(t) > x_{i+1}(t)$，对所有 $t \geq 0$ 均如此（车辆不会相互超越）。

我们假设该路段由入口和出口摄像头界定，摄像头会记录每辆车驶入和驶出该路段的时间 [12]。由此得到数据集

$$\mathcal{D} = \{(t_{e,i}, t_{f,i})\}_{i=1}^N \tag{13}$$

其中 $x_i(t_{e,i}) = x_e$，且 $x_i(t_{f,i}) = x_f$。我们还假设已知探测车辆的信息，该车辆会持续向基础设施传送其位置 $x_0(t)$。图 1 展示了这一场景。利用这些数据，我们有两个目标：（i）确定速度–密度关系；（ii）重建车辆轨迹。

为实现这些目标，我们使用跟驰模型 (2) 对车辆进行建模，其中速度–密度关系是一个参数为θ的神经网络 $\rho \mapsto \hat{}v(\rho;\theta)$，重建轨迹 $\hat{}x_i(t;\theta)$ 则通过对 $\hat{}v$ 积分得到。这两个目标归结为寻找能够最小化轨迹误差的参数 θ

$$\mathcal{L}_x(\theta) := \frac{1}{N}\sum_{i=1}^N \int_{t_{e,i}}^{t_{f,i}} |\hat{x}_i(t;\theta) - x_i(t)|^2\, dt, \tag{14}$$

以及真实轨迹上的速度–密度误差,

$$\mathcal{L}_v(\theta) := \frac{1}{N}\sum_{i=1}^N \int_{t_{e,i}}^{t_{f,i}} |\hat{v}(\rho_i(t);\theta) - \dot{x}_i(t)|^2\, dt, \tag{15}$$

其中 $\rho_i(t)$ 是真实轨迹的局部密度 (1)。由于 (14) 是对时间和车辆取平均，我们还考虑最坏情况轨迹误差，即最大偏差

$$\mathcal{L}_{\max}(\theta) := \max_{i\in\{1,\dots,N\}} \max_{t\in[t_{e,i},t_{f,i}]} |\hat{x}_i(t;\theta) - x_i(t)|. \tag{16}$$

然而，这三种误差都取决于路段内的轨迹，而这些轨迹无法观测；因此，我们最小化出口误差，它只需要 $\mathcal{D}$ 和 x_0:

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^N (\hat{x}_i(t_{f,i};\theta) - x_f)^2. \tag{17}$$

该问题是 II-B 节所述的神经常微分方程。由于我们有 $\mathcal{L}_x(\theta) = 0 \Rightarrow \mathcal{L}(\theta) = 0$，但反之则不成立，因此我们希望通过强制满足交通流理论约束 (3)，尽可能多地融入物理知识，从而缩小可能解的范围，使其与数据一致。下一节将提出一种经过调整的训练算法，以有效地最小化 (17)，并在 $\hat{}v$ 中原生地强制满足这些约束。

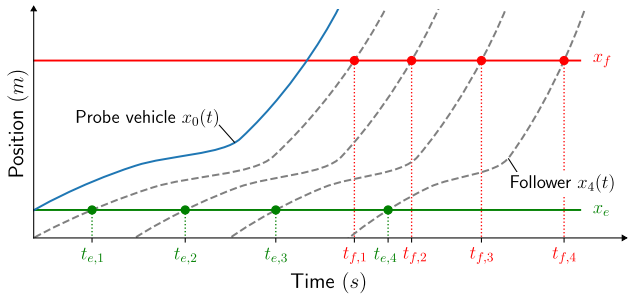

图 1. 包含 $N = 4$ 辆跟车车辆的问题设定。探测车辆轨迹$x_0(t)$（实线）已知，而跟车车辆的轨迹 $x_i(t)$（虚线）未知；仅测量它们在 x_e 处的进入时间$t_{e,\,i}$ 和在 x_f 处的出口时间 $t_{f,\,i}$ （圆点）。

IV. 方法

本节首先针对 EET测量调整神经常微分方程的训练，然后构建一个速度-密度网络，使其在设计上满足约束 (3)，并捕捉临界密度处的状态变化。

A. 用于 *EET*测量的神经常微分方程

与 II-B节中的设定相比，跟车车辆通过其密度相互耦合，而车辆 i仅在其进入和离开时间$t_{e,\,i}$ 和$t_{f,\,i}$被观测到，且在 $t_{e,\,i}$之前没有关于其位置的信息。因此，我们让每辆跟车车辆在进入路段之前保持在入口位置，之后再按照学习得到的动力学行驶。将估计得到的跟车车辆位置记为 $\hat{x}(t;\theta) = (\hat{x}_1(t;\,\theta)\ldots,\hat{x}_N(t;\,\theta))^\top$，并令 $\hat{x}_0(t;\,\theta) := x_0(t)$ 表示已知的探测车辆轨迹，则跟车车辆 $i \in \{1,\ldots,N\}$ 的估计动力学为

$$\frac{d\hat{x}_i(t;\theta)}{dt} = \hat{f}_i(t,\hat{x};\theta) := \begin{cases} \hat{v}\big(\hat{\rho}_i(t);\theta\big), & t \geq t_{e,i}, \\ 0, & \text{otherwise}, \end{cases} \tag{18}$$

其中 $\hat{x}_i(0;\theta) = x_e$，其中 $\hat{\rho}_i(t) := L_{car}/(\hat{x}_{i-1}(t;\,\theta) - \hat{x}_i(t;\theta))$ 是沿估计轨迹评估的局部密度（1）。由于车辆不会超车，车辆 $i-1$ 在车辆 i之前进入，因此只需为 $t \geq t_{e,\,i}$ 定义 $\hat{\rho}_i$。堆叠的右端项 $\hat{f} := (\hat{f}_1,\ldots,\hat{f}_N)^\top$ 是神经常微分方程（8）的可训练动力学。由于已经驶出的车辆仍会影响其后的车辆，所有轨迹都积分至最后一个出口时间 $T := t_{f,\,N}$。

出口误差（17）在每个跟驰车辆自身的出口时间，而不是单一的终止时间，对其进行评估。因此，它具有形式（9），其中

$$\Phi(\hat{x}(T;\theta)) := \frac{1}{N}\big(\hat{x}_N(T;\theta) - x_f\big)^2, \tag{19}$$

$$\ell(t,\hat{x}(t;\theta)) := \frac{1}{N}\sum_{i=1}^{N-1}\big(\hat{x}_i(t;\theta) - x_f\big)^2\,\delta(t - t_{f,i}), \tag{20}$$

其中，δ是狄拉克δ函数，即对于$\tau \in (0,\ T)$以及在 τ处连续的任意 φ，$\int_0^T \varphi(t)\,\delta(t-\tau)\,dt = \varphi(\tau)$。将（19）和（20）代入（11），伴随变量从 $\lambda_N(T) = \frac{2}{N}(\hat{x}_N(T;\,\theta) - x_f)$开始并且 $\lambda_i(T) = 0$对于 $i < N$成立，在离开时间之间遵循齐次伴随动力学，并在每个离开时间发生跳跃：

$$\lambda_i(t_{f,i}^-) = \lambda_i(t_{f,i}^+) + \frac{2}{N}\big(\hat{x}_i(t_{f,i};\theta) - x_f\big) \tag{21}$$

对于 $i \in \{1,\ldots,N-1\}$，而其他分量是连续的。梯度随后由 II-B节得出；由于跟驰车辆 i 的动力学在 $t_{e,\,i}$之前消失，因此它仅在进入之后对梯度产生贡献。由于 $\hat{\rho}_{i+1}$取决于 $\hat{x}_i$，车辆 i 在驶出后仍会影响其跟驰车辆。这就是轨迹要积分至时间T的原因，也是探测轨迹必须保持已知直到最后一个出口的原因。

B. 通过硬约束强制满足交通流动力学

如第 III 节所述，出口误差趋于零并不意味着轨迹误差也趋于零：可能存在多个速度–密度关系，它们都能同样准确地复现进入和离开时间，而仅凭现有数据无法区分这些解。为帮助学习并防止学习得到的模型不切实际，我们将 $\hat{v}$ 限制为物理上可容许的关系，即满足 (3) [7]的关系。连续的 v 满足 (3) 当且仅当 v/v_{free} 属于归一化可行速度–密度关系的集合：

$$\mathcal{A} := \left\{ g \in \mathcal{C}([0,1]) : \begin{array}{l} g \text{ non-increasing}, \\ g(0) = 1,\ g(1) = 0 \end{array} \right\}, \tag{22}$$

其中，$\mathcal{C}([0,1])$表示定义在$[0,1]$, 上的连续实函数，并配备范数 $\|\cdot\|_\infty$。我们通过三个组成部分构造 $\hat{v}$，使得对于每个 θ，$\hat{v}/v_{free} \in \mathcal{A}$：单调网络（第 IV-B.1 节）、固定端点的归一化（第 IV-B.2 节），以及在临界密度处的状态变化（第 IV-B.3 节）；这三者在第 IV-B.4 节中组合起来。

1）强制单调性：如第 II-C 节所述，具有正权重和严格递增激活函数的多层感知机是严格递增的。将权重参数化为其绝对值可以保证权重为正，但在 0 处不可微，这可能会导致训练期间出现问题；因此，我们改用平滑重参数化

$$(\mathcal{P}_\epsilon\phi)(\rho;\theta) := \phi(\rho;\theta_\epsilon), \quad \theta_\epsilon := \left\{\left(\sqrt{W_i^2 + \epsilon},\, b_i\right)\right\}_{i=1}^{l_\phi+1}, \tag{23}$$

其中，平方和平方根均逐元素应用，且 $\epsilon > 0$。$\mathcal{P}_\epsilon\phi$ 的权重下界为 $\sqrt{\epsilon}$，因此使用 $\sigma =$ 双曲正切时，对于每个 θ，$\mathcal{P}_\epsilon\phi$ 都严格递增。反过来，任何权重 W_i 均不小于 $\sqrt{\epsilon}$ 的多层感知机，都等于逐元素~~使用~~ $W_i = \sqrt{\tilde{W}_i^2 - \epsilon}$ 的 $\mathcal{P}_\epsilon\phi$。因此，由于 $\epsilon \to 0$，$\mathcal{P}_\epsilon\phi$ 可以表示任何具有正权重的多层感知机，从而能够以任意精度逼近 $[0,1]$ 上的任意连续非递减函数（第 II-C 节）。然而，$\mathcal{P}_\epsilon\phi$ 是递增而非递减的，且不满足端点条件；接下来引入的归一化将解决这两个问题。

2) 固定端点：由于 (12) 无法与单调网络结合（第 II-C 节），我们转而使用网络自身在端点处的值对网络输出进行归一化。对于满足 $\phi(0) \neq \phi(1)$ 的 $\phi \in \mathcal{C}([0,1])$，我们定义如下算子：

$$(\mathcal{T}\phi)(\rho) := \frac{\phi(\rho) - \phi(1)}{\phi(0) - \phi(1)}, \qquad \rho \in [0,1]. \tag{24}$$

根据构造，$\mathcal{T}\phi$ 是连续的，且 $(\mathcal{T}\phi)(0) = 1$ 以及 $(\mathcal{T}\phi)(1) = 0$；由于对于 $g \in \mathcal{A}$，$g(0) - g(1) = 1$ 且 $g(1) = 0$不会改变可容许关系，即 $\mathcal{T}g = g$。此外，$\mathcal{T}$对其自变量的仿射变换保持不变：对于任意标量 $a \neq 0$和 b，$\mathcal{T}(a\phi + b) = \mathcal{T}\phi$，因为 b 会约去，而 a会从式 (24) 的分子和分母中同时提取出来。

定理 1。设 $\phi \in \mathcal{C}([0,1])$ 满足 $\phi(0) \neq \phi(1)$。则：

(i) 如果 ϕ 是非递减或非递增的，则 $\mathcal{T}\phi \in \mathcal{A}$；(ii) 如果 $\eta = \|\phi - g\|_\infty < \frac{1}{2}$ $g \in \mathcal{A}$ $\|\mathcal{T}\phi -$: $g\|_\infty \leq \frac{2\eta}{1-2\eta} \xrightarrow[\eta\to 0]{} 0$ - 对某个 成立，则 。

证明。(i) 对于 $\rho_1 \leq \rho_2$，$(\mathcal{T}\phi)(\rho_\mathrm{T})$ − $\phi\ \rho_2\ )/(\phi\ 0\ -\phi\ 1\ )$ ϕ $(\mathcal{T}\phi)(\rho_2) = (\phi(\rho_1)$ ( ) ( ) ( )。若 为非递增函数，则分子非负且分母为正；若 ϕ 为非递减函数，则分子非正且分母为负。在这两种情况下，$\mathcal{T}\phi$ 均为非递增函数，并且由于它是连续的，且 $(\mathcal{T}\phi)(0) = 1$ 和 $(\mathcal{T}\phi)(1) = 0$，$\mathcal{T}\phi \in \mathcal{A}$。

(ii) 令 $e := \phi - g$，使得对所有 ρ都有 $|e(\rho)| \leq \eta$。由于$g(0) = 1$ 且 $g(1) = 0$，(24) 的分母为 $d :=\phi(0) - \phi(1) = 1 + e(0) - e(1)$，因此 $|d| \geq 1 - 2\eta > 0$。利用$\phi = g + e$，

$$\big((\mathcal{T}\phi)(\rho) - g(\rho)\big)\, d = e(\rho) - \big(1 - g(\rho)\big)e(1) - g(\rho)\, e(0).$$

由于 $0 \leq g(\rho) \leq 1$，右端项的绝对值有界于 $\eta\big(1+(1-g(\rho))+g(\rho)\big) = 2\eta$，再除以 $|d| \geq 1 - 2\eta$ 即可得到该界。 □

根据定理 1(i)，$\mathcal{T}$将任意严格单调函数映射为可容许关系；根据定理 1(ii)，它能以约两倍的因子保持逼近精度；第 IV-B.4 节在所提出的模型中结合了这两种性质。

3) 状态变化：实测速度–密度数据表明，自由流状态与拥堵状态由一个临界密度分隔，在该密度处，关系的斜率会突然变化，如 Newell–Daganzo 关系 (4) 所示。使用双曲正切等光滑激活函数时，$\mathcal{T}\mathcal{P}_\epsilon\phi$ 需要许多神经元才能逼近这样的折点。为此，我们先对递增网络 $\psi := \mathcal{P}_\epsilon\phi$进行整流，再进行归一化，从而将折点纳入模型。如果 $\psi(0) \leq 0 < \psi(1)$,

$$\big(\mathcal{T}\max(0,\psi)\big)(\rho) = 1 - \frac{\max\big(0,\psi(\rho)\big)}{\psi(1)}, \tag{25}$$

在临界密度 $\hat{}\rho_c$ 之前，其值等于 1；在该点，$\psi(\hat{}\rho_c) = 0$，之后则严格递减，形成一个自由流平台段，随后是拥堵分支。因此，临界密度由网络自身设定，并与网络一同学习，无需额外参数。

然而，采用 max(0, ·) 时，(25) 在 $\hat{}\rho_c$ 以下恰好为 1，因此其关于 θ 的梯度在该处为零，导致自由流状态下的车辆不会参与训练；如果所有估计密度都低于 $\hat{}\rho_c$，训练便会停滞。此外，如果 $\psi(1) \leq 0$，则 $\mathcal{T}\max(0,\psi)$ 未定义，因为此时 $\max(0,\psi(0)) = \max(0,\psi(1)) = 0$。因此，我们用 softplus函数 替换 max(0, ·)

$$\mathcal{S}_\beta(z) := \frac{1}{\beta}\log\left(1 + e^{\beta z}\right), \tag{26}$$

一种光滑、正且严格递增的 $\max(0,z)$ 近似，其尖锐度为 $\beta > 0$，满足对于所有 z，$0 < \mathcal{S}_\beta(z) - \max(0,z) \leq \log(2)/\beta$，因此当 $\beta \to \infty$ 时误差趋于零。逐点应用时，$\mathcal{S}_\beta\psi$ 对每个 θ 都为正且严格递增，因此 $\mathcal{T}\mathcal{S}_\beta\psi$ 始终有良好定义，并且其梯度在 $\hat{}\rho_c$ 以下不再消失。此外，对于 $c > 0$，$\mathcal{S}_\beta(cz) = c\,\mathcal{S}_{c\beta}(z)$，因此根据 $\mathcal{T}$ 的仿射不变性，

$$\mathcal{T}\mathcal{S}_\beta(c\psi) = \mathcal{T}\big(c\,\mathcal{S}_{c\beta}\psi\big) = \mathcal{T}\mathcal{S}_{c\beta}\psi. \tag{27}$$

因此，将网络输出缩放 c 等价于以相同因子增强 softplus函数的尖锐度，所以网络可以自行增强折点的尖锐度；尖锐度实际上是学习得到的，而 β 主要决定其在初始化时的取值。在实践中，$\mathcal{S}_\beta\mathcal{P}_\epsilon\phi$ 就是一个带有 Softplus 输出激活函数的单调网络。最后，整流不会降低表达能力：对于任意 $g \in \mathcal{A}$，递增函数 $1 - g$ 都是非负的，因此经过 max(0, ·) 后保持不变；由于 $\mathcal{T}$ 消除了式 (27) 中的输出缩放，对于任意固定的 β，softplus 误差都可以任意小（命题 1）。

4) 所提出的模型：结合单调网络、状态变化和归一化，所提出的模型为

$$\hat{v}(\rho;\theta) := v_{free}\left(\mathcal{T}\mathcal{S}_\beta\mathcal{P}_\epsilon\phi\right)(\rho;\theta). \tag{28}$$

由于 $\mathcal{S}_\beta\mathcal{P}_\epsilon\phi$ 严格递增，定理 1(i) 得出 $\hat{}v/v_{free} \in \mathcal{A}$，即 $\hat{}v$ 对任意 θ、$\epsilon > 0$和$\beta > 0$均满足 (3)；又由于最后应用了 $\mathcal{T}$，端点条件得以精确满足。此外，命题 1 表明，对于任意固定的 β，所提出的模型都能以任意高的精度逼近任何可容许关系。

命题 1。设 $g \in \mathcal{A}$ 和 $\beta > 0$。对于任意 $\delta > 0$ 及任意充分小的 $\epsilon > 0$，存在一个多层感知机 ϕ，使得 $\|\mathcal{T}\mathcal{S}_\beta\mathcal{P}_\epsilon\phi - g\|_\infty \leq \delta$。

证明。设 $h := 1 - g$，它是连续、非递减且非负的，并令 $\gamma > 0$和 $c > 0$。由于 ch连续且非递减，对于足够宽的 ϕ和充分小的 ϵ，存在 θ，使得 $\psi := \mathcal{P}_\epsilon\phi$满足 $\|\psi - ch\|_\infty \leq c\gamma$（IV-B.1 节）；由 (27) 得，$\mathcal{T}\mathcal{S}_\beta\psi = \mathcal{T}\mathcal{S}_{c\beta}(\psi/c)$。由于 $\mathcal{S}_{c\beta}$ 是 1-Lipschitz 的，且 $h = \max(0, h)$，从而 $0 < \mathcal{S}_{c\beta}(h) - h \leq \log(2)/(c\beta)$，由三角不等式可得

$$\|\mathcal{S}_{c\beta}(\psi/c) - h\|_\infty \leq \gamma + \log(2)/(c\beta) =: \eta.$$

根据 $\mathcal{T}\ \mathcal{T}\mathcal{S}_{c\beta}\ \psi/c\ = \mathcal{T}(1 -$ 的仿射不变性，( ) $\mathcal{S}_{c\beta}\ \psi/c\ )$ $\|1 - \mathcal{S}_{c\beta}\ \psi/c\ - g\|_\infty = \|\mathcal{S}_{c\beta}\ \psi/c\ -$ ( )，其中 ( ) ( )

$h\|_\infty \le \eta$。因此，如果 $\eta < 1/2$，定理 1(ii) 给出

$$\|\mathcal{T}\mathcal{S}_\beta\mathcal{P}_\epsilon\phi - g\|_\infty \le \frac{2\eta}{1-2\eta},$$

其至多为 δ if $\eta \le \delta/(2+2\delta)$，例如，对于 $\gamma \le \delta/(4+4\delta)$和 $c \ge (4+4\delta)\log(2)/(\beta\delta)$。 □

V. 数值结果

我们在模型失配程度逐渐增加的三个场景中评估该方法。首先，使用一阶 Newell–Daganzo (ND) 模型模拟数据。该模型的真实速度–密度关系已知，且所提出的模型可以很好地表示该关系，因此可以将学习到的关系与真实关系直接比较。其次，使用二阶 IDM 模拟数据。在该模型中，车辆会响应相对速度，且无法瞬时改变速度，因此任何一阶模型都无法精确复现其轨迹。第三，使用来自 [2]，的真实环形道路数据，其中驾驶员的行为各不相同，且车辆长度未知。

在所有场景中，$N = 9$ 辆跟车车辆在探测车辆后方 $[x_e, x_f] = [0, 125]$ m处驶过该路段；探测车辆的速度为环形道路数据中测得的领航车辆速度。两种模拟均采用 $v_{free} = 50$ km/h（≈ 13.9 m/s）的自由流速度和 $l = 4.6$ m的车辆长度，其余参数则经过设置，使其行为与环形道路数据相匹配。ND [19] 采用 $s_0 = 0.6$ m的最小间距，IDM [22] 采用 $s_0 = 1.0$ m的最小间距、$T_h = 1.0$s的车头时距、$a_{max} = 2.5$ m/s的最大加速度、$b_{comf} = 2.0$ m/s的舒适减速度以及 4 的加速度指数。在密度 (1) 中，我们设定 $L_{car} = l + s_0$，即 ND 为 5.2 m，IDM 为 5.6 m：由于驾驶员至少保持 s_0的间距，它相当于有效车辆长度，其中 ρ =1 对应完全拥堵。ND 遵循三角形关系 $v(\rho) = v_{free}\min(1, w(1/\rho - 1))$，其中 $w = 0.36$ 对应临界密度 $\rho_c = w/(1+w) \approx 0.26$。

对于环形道路，领航车辆是实验A中 [2]，的自动驾驶车辆，其后是紧跟在它后方的 9 辆车辆；时间范围从实验的 $t = 104$ 秒开始，直到最后一辆跟驰车辆在 $t \approx 140$ 秒离开路段。我们使用数据中提供的平滑位移和速度。我们对时间和空间进行平移，使领航车辆从 $x_0(0) = 0$ 开始。进入和离开时间分别是每辆跟驰车辆穿过 x_e 和 x_f 的时刻。由于车辆长度和间距未知，我们对环形道路采用与 IDM 仿真相同的 $L_{car} = 5.6$ 米。

我们比较三个模型，它们共用相同的多层感知机 (6) 骨干网络 ϕ，包含 $l_\phi = 3$ 个隐藏层，每层有 $n_\phi = 16$ 个神经元，并使用双曲正切作为激活函数。第 IV 节中的所提出的模型 通过 (23) 对权重进行重参数化，并采用 $\epsilon = 10^{-8}$ 来强制满足单调性；此外，还使用尖锐度为 $\beta = 30$ 的 softplus 状态变化 (26)。由于初步实验表明，在 (0,1) 内没有折点的初始化效果较差，因此我们使用 PyTorch 的默认初始化方式初始化权重和偏置，并反复重新抽取，直到 $\psi(0) < 0 < \psi(1)$，以确保初始临界密度 $\hat{\rho}_c$ 位于 (0,1) 内。在五次训练运行中，初始 $\hat{\rho}_c$ 的范围为 0.08 至 0.82。固定端点模型在 (12) 中定义，而无约束模型 则简单地为 $\hat{v} = v_{free}\,\phi$，因此三个模型使用的自由流速度均为 $v_{free} = 13.9$ m/s。

我们使用步长为 $\Delta t = 0.1$ 秒的四阶龙格–库塔法进行积分，并在准确的离开时间处计算每个出口位置 $\hat{x}_i(t_{f,\,i};\theta)$；具体做法是从 $t_{f,\,i}$ 之前的最后一个时间步开始，追加一个最终的部分步长。由于速度–密度关系仅在 $[0,1]$，上有定义，因此在积分过程中，我们将估计密度 $\hat{\rho}_i$ 限制在 $[0,1]$ 范围内，因为训练期间估计间距可能低于 L_{car}。我们使用 Adam [28]，训练 1000 次迭代，学习率为 10^{-3}，一阶矩估计和二阶矩估计的衰减率分别为 0.9 和 0.99。前 900 次迭代中学习率保持恒定，随后在最后 100 次迭代中按余弦调度衰减至零。每个模型均使用不同的随机初始化训练五次。

除作为训练损失的出口误差 (17) 外，我们还报告轨迹误差 (14)、速度误差 (15) 和最大偏差 (16)。由于这些指标需要真实轨迹，而训练过程中并未使用真实轨迹，因此它们作为验证指标。学习得到速度–密度关系后，还可以将其用于根据同一路段上新的进入和离开时间重建轨迹，前提是这些时间覆盖相似的密度范围。

A. 学习得到的速度–密度关系

图 2 展示了学习得到的速度–密度关系。对于 ND，所提出的模型与真实关系高度吻合，包括平台段以及临界密度$\rho_c \approx 0.26$附近的折点，尽管数据仅覆盖$\rho \approx 0.3$以上的密度；因此，平台段和折点是由模型结构而非数据提供的。对于 IDM，速度–密度轨迹形成狭窄的环路，因为车辆无法瞬间改变速度，而所提出的模型穿过这些环路的中心。

固定端点模型对观测到的密度范围拟合得相当不错，尽管它平滑了 (a) 中的折点；而且，由于式 (12) 未强制满足单调性，该模型在所有三种情况下的高密度区间都会出现负值，在环形道路的低密度区间则会超过v_{free}。在高密度下，真实速度接近于零，且可用数据点很少，因此微小偏差就足以导致负速度。无约束模型在所有情况下都收敛到几乎恒定的速度，约为 6.5 m/s，接近该路段的平均速度。这样的关系已经能将出口位置拟合到几米以内（表 I）；而在缺乏先验知识的情况下，无约束模型无法摆脱这一较差的局部最小值。

B. 轨迹重建

图 3 展示了使用所提出模型重建的轨迹。对于 ND 和 IDM，重建轨迹几乎与真实轨迹重合，平均最大偏差分别为 2.50 m 和 2.51 m（表 I）。对于环形道路，前几辆跟车车辆的重建精度较高，但沿车队向后逐渐下降，平均最大偏差为 17.1 m。由于每辆车都跟随其前车的重建轨迹，因此误差

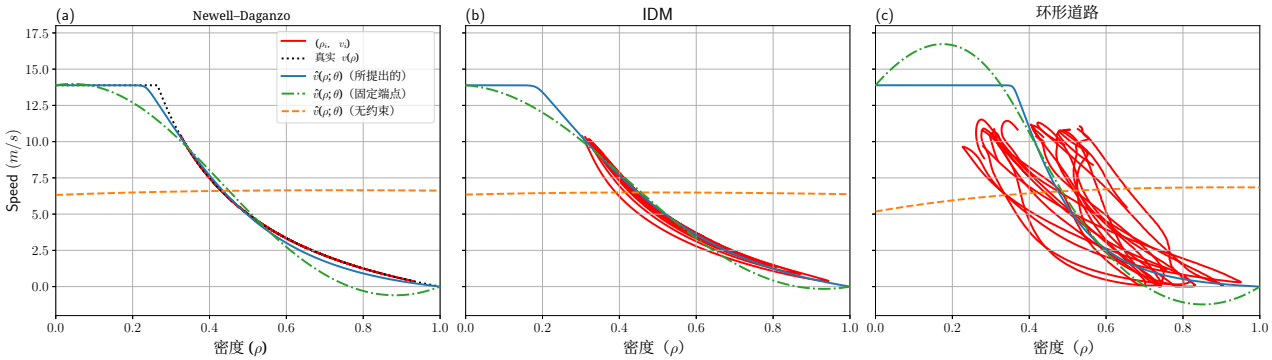

图 2. 表 I 中五次训练运行之一所学习的速度–密度关系 $\hat{v}(\rho;\theta)$，分别对应 (a) Newell–Daganzo 一阶模型、(b) IDM，以及 (c) 环形道路数据。图中比较了三种模型：所提出的模型（蓝色实线）、固定端点模型 (12)（绿色点划线）和无约束模型（橙色虚线）。(a) 中的黑色点线表示真实关系。红色曲线表示路段内真实跟车车辆的速度–密度轨迹（$\rho_i(t)$ $v_i(t)$），仅供参考，未用于训练；它们在 (a) 和 (b) 中覆盖 $\rho \in [0.3, 0.95]$，在 (c) 中覆盖 $\rho \in [0.23, 0.95]$。在此范围内，所提出的模型在 (a) 中紧密贴合真实关系，在 (b) 和 (c) 中紧密贴合数据；但在 (c) 的最低密度处除外，此时其位于 v_{free} 的平台段高于观测速度。无约束模型则收敛到几乎恒定的速度。在此范围之外，学习到的关系仅由模型结构决定：无约束模型违反了全部三项约束；固定端点模型在所有情况下都在 $\rho = 1$ 附近变为负值，并在 (c) 中超过 v_{free} 约 3 m/s；而所提出的模型通过构造满足 (3)。

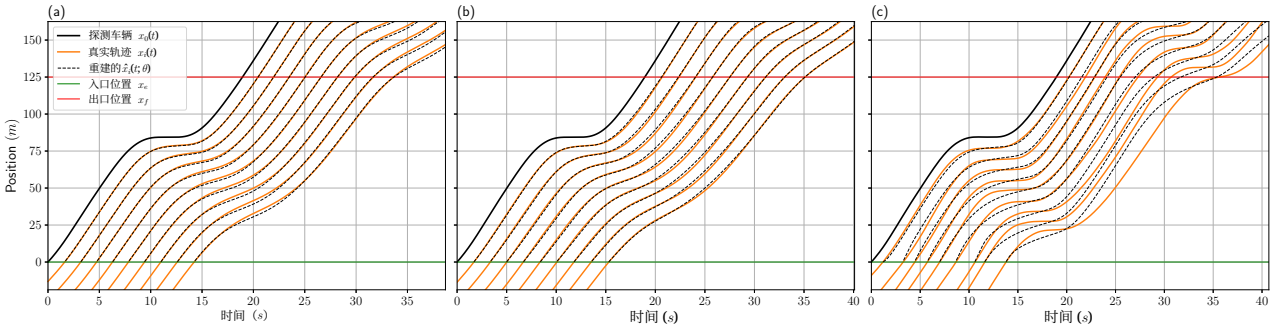

图 3. 使用所提出的模型重建的轨迹（黑色虚线）与真实轨迹（橙色），数据来自与图 2 相同的训练轮次：(a) Newell–Daganzo，(b) IDM，(c) 环形道路数据。每条重建轨迹均从跟驰车辆在进入时间处的 x_e 开始。可用数据仅包括探测车辆的轨迹（黑色实线），以及跟车车辆经过进入位置 x_e（绿色）和出口位置 x_f（红色）的时间；真实轨迹仅供参考。

表 I

学习得到的模型的误差。出口误差 (17) 即训练损失，在离开时间处评估；轨迹误差 (14) 和速度误差 (15) 对 $N = 9$ 辆跟车车辆取平均值，最大偏差 (16) 则取所有这些车辆中的最大值，计算区间均为各跟车车辆的进入和离开时间之间。除出口误差外，其他误差均使用训练期间不可获得的真实轨迹。 $\pm$ 为 5 次训练运行的样本标准差，每次运行均采用不同的随机初始化，并训练 1000 次迭代。每种情景下的最低平均值以粗体标出。

Data	模型	出口误差 $\mathcal{L}$ (m^2)	轨迹误差 $\mathcal{L}_x$ (m^2s)	速度误差 $\mathcal{L}_v$ (m^2/s)	最大偏差 $\mathcal{L}_{\max}$ (m)
Newell–Daganzo	所提出的	**0.0378 ± 0.0056**	**12.6 ± 1.6**	**0.783 ± 0.105**	**2.50 ± 0.15**
	固定端点	0.192 ± 0.055	54.4 ± 12.7	4.46 ± 1.46	4.70 ± 0.44
	无约束	2.12 ± 0.08	1685 ± 5	155 ± 1	21.9 ± 0.2
IDM	所提出的	0.0744 ± 0.0031	**21.9 ± 0.3**	**1.55 ± 0.02**	**2.51 ± 0.00**
	固定端点	**0.0674 ± 0.0013**	45.5 ± 2.8	3.60 ± 0.25	2.99 ± 0.10
	无约束	2.79 ± 0.55	1543 ± 57	128 ± 12	19.9 ± 1.8
环形道路	所提出的	4.54 ± 0.05	**381 ± 1**	**238 ± 2**	17.1 ± 0.1
	固定端点	**4.50 ± 0.54**	385 ± 14	433 ± 130	**16.9 ± 0.3**
	无约束	17.3 ± 2.0	3176 ± 53	307 ± 3	27.9 ± 0.3

沿车队累积。此外，与仿真不同，图3(c)中的真实轨迹在进入该路段之前就已呈现出不均匀的间距，这表明驾驶员在相近速度下保持着不同的间距。这种异质性，与加上未知的车辆长度，这违反了具有共同 L_{car}的单一速度–密度关系这一假设，并导致用于学习动力学的局部密度 (1) 出现误差。这也解释了

图 2(c) 中速度–密度轨迹的巨大分散。

表 I 证实了这些观察结果。对于 ND，所提出的模型捕捉到了真实关系的状态变化，其轨迹误差约为固定端点模型的四分之一，也比无约束模型低两个数量级。对于 IDM，固定端点模型的出口误差最低，但其轨迹误差是所提出模型的两倍，这说明出口误差低并不意味着轨迹准确。对于环形道路，固定端点模型的出口误差和最大偏差略低于所提出的模型，差距不到 2%，轨迹误差也相近，但其速度误差几乎是所提出模型的两倍，甚至高于无约束模型：它虽然复现了离开时间，却采用了错误的速度–密度关系。除此之外，无约束模型在所有指标上的误差都是最大的。因此，这两种受约束模型的局限源于模型失配，而非约束。

总体而言，与无约束模型相比，施加端点条件本身就能显著改善学习效果，因为它排除了平凡的常数解。单调性并不一定能改善模型在观测到的密度范围内的拟合，但它能保证学习到的关系在该范围之外仍然符合要求，而固定端点模型在图 2 的三个子图中都未能做到这一点。此外，单调性通过限制能够复现离开时间的关系，使较低的出口误差更能体现动力学的准确性：对于 IDM 和环形道路，固定端点模型的出口误差低于所提出的模型，但其速度误差约为后者的两倍。另一方面，ND 的精度提升很可能源于状态变化。

C. 讨论

学习动力学需要数据对其进行激励。在我们的场景中，领航车辆速度的变化源自环形道路实验中自然出现的启停波，这使跟车车辆经历了广泛的密度范围，从约 0.23 到 0.95。若没有这样的变化，所有车辆都将以近乎恒定的速度和密度行驶，而进入和离开时间只能提供速度–密度关系上单个点的信息。按照持续激励条件的思路，此时系统便不可辨识。反过来，已访问密度范围之外的关系不会影响轨迹，因此无法从数据中学习。在该范围内，关系完全由模型结构决定，这就是为什么硬约束对于在整个密度范围内获得可容许关系至关重要。

环形道路的结果也凸显了一阶模型在驾驶员同质时的局限性：相同的密度可能对应差异很大的速度，而且重建误差会沿车队逐渐增大。

VI. 结论

我们研究了这样一个问题：根据路段的进入和离开时间，以及引领车队的探测车辆的轨迹，学习微观交通动力学并重建车辆轨迹。

我们针对这些EET测量调整了神经常微分方程训练，并引入了一种归一化方法。该方法与单调网络和学习得到的状态变化相结合，能够保证对于每个参数值，都得到符合可容许关系的速度–密度关系，同时保留逼近任意可容许关系的能力。对于仿真一阶和二阶交通，所提出的模型恢复了一阶模型的速度–密度关系，并成功重建了两者的轨迹；而无约束模型未能重建基本图和轨迹。在真实环形道路数据上，所提出的模型仍能得到符合可容许关系的关系，但异质驾驶行为限制了重建精度。

未来的工作包括：在保持车辆间耦合关系的同时，学习个体驾驶员行为和多种车辆类别（如轿车和卡车）；将该方法扩展到二阶模型和随机模型，以捕捉现实世界数据中观察到的离散程度；并推导领航车辆的激励条件，在这些条件下，可以根据进入和离开时间识别速度–密度关系。

致谢

作者使用 Claude（Anthropic）根据作者的草稿修订了所有章节的文本，以改进语言和表达，并检查和润色了定理 1（第 IV-B.2 节）和命题 1（第 IV-B.4 节）的陈述与证明。所有由 AI 生成的内容均由作者审阅和编辑，作者对本文承担全部责任。

参考文献

[1] A. Ferrara、S. Sacone 和 S. Siri，《高速公路交通建模与控制》。Springer，2018年。[2] R. E. Stern 等，“通过控制自动驾驶车辆消除启停波：现场实验”，《交通研究：*C*辑——新兴技术》，第89卷，第205–221页，2018年。[3] J. W. Lee 等，“通过网联自动驾驶车辆（CAV）进行交通控制：一项包含100辆CAV的开放道路现场实验”，《*IEEE*控制系统杂志》，第45卷，第28–60页，2025年。[4] R. Shi、Z. Mo、K. Huang、X. Di 和 Q. Du，“用于交通状态和基本图估计的物理信息深度学习范式”，《*IEEE*智能交通系统汇刊》，第23卷，第11688–11698页，2021年。[5] X. Ma、Z. Tao、Y. Wang、H. Yu 和 Y. Wang，“利用远程微波传感器数据进行交通速度预测的长短期记忆神经网络”，《交通研究：*C*辑——新兴技术》，第54卷，第187–197页，2015年。[6] Y. Liang、K. Ouyang、H. Yan、Y. Wang、Z. Tong 和 R. Zimmermann，“利用神经常微分方程建模轨迹”，载于*IJCAI*会议论文集，第1498–1504页，2021年。[7] M. Barreau、M. Aguiar、J. Liu 和 K. H. Johansson，“用于交通密度辨识与状态重构的物理信息学习”，载于*IEEE*决策与控制会议（*CDC*）论文集，第2653–2658页，2021年。[8] D. Wilkman、K. Morozovska、K. H. Johansson 和 M. Barreau，“利用物理信息神经网络进行在线交通密度估计”，载于*IEEE*决策与控制会议（*CDC*）论文集，第236–242页，2025年。[9] X. Di、R. Shi、Z. Mo 和 Y. Fu，“用于交通状态估计的物理信息深度学习：综述与展望”，《算法》，第16卷，第305页，2023年。[10] N. Baloul、A. Hayat、T. Liard 和 P. Lissy，“基于有限采集数据的交通流重构”，载于*IEEE*决策与控制会议（*CDC*）论文集，第7981–7986页，2025年。

[11] J. C. Herrera、D. B. Work、R. Herring、X. J. Ban、Q. Jacobson 和 A. M. Bayen，“利用支持 GPS 的移动电话获取的交通数据评估：Mobile Century 实地实验”，《交通研究：C辑——新兴技术》，第18卷，第568–583页，2010年。[12] F. M. Oliveira-Neto、L. D. Han 和 M. K. Jeong，“利用车牌识别设备和新型加权编辑距离进行在线车牌匹配”，《交通研究：C辑——新兴技术》，第21卷，第306–320页，2012年。[13] S´ecurit´e routi`ere，“固定式测速雷达。”https://www.securite-routiere.gouv.fr/radars/differents-types-de-radars/radars-fixes/radars-de-controle-de-la-vitesse-fixes。访问日期：2026年11月。[14] R. T. Chen、Y. Rubanova、J. Bettencourt 和 D. K. Duvenaud，“神经常微分方程”，《神经信息处理系统进展》，第31卷，2018年。[15] M. Di Francesco、S. Fagioli、M. D. Rosini 和 G. Russo，“车辆流与行人流宏观模型的领航车辆近似”，载于《活性粒子》第1卷，第333–378页，Springer，2017年。[16] X. Qu、J. Zhang 和 S. Wang，“高速公路交通随机基本图：模型开发、解析性质、验证与广泛应用”，《交通研究：B辑——方法论》，第104卷，第256–271页，2017年。[17] B. S. Kerner，《现代交通流理论与控制导论》。Springer，2009年。[18] B. D. Greenshields、J. R. Bibbins、W. S. Channing 和 H. H. Miller，“交通容量研究”，载于《公路研究委员会会议论文集》，第14卷，1935年。[19] G. F. Newell，“公路交通运动波的简化理论，第一部分：一般理论”，《交通研究：B辑——方法论》，第27卷，第281–287页，1993年。[20] C. F. Daganzo，“元胞传输模型：与流体动力学理论一致的公路交通动态表征”，《交通研究：B辑——方法论》，第28卷，第269–287页，1994年。[21] M. L. Delle Monache 等，“基于三维交通数据的三相基本图”，《Axioms》，第10卷，第17页，2021年。[22] M. Treiber、A. Hennecke 和 D. Helbing，“实证观测和微观仿真中的拥堵交通状态”，《物理评论E》，第62卷，第1805页，2000年。[23] I. Goodfellow、Y. Bengio 和 A. Courville，《深度学习》。MIT出版社，2016年。[24] K. Hornik，“多层前馈网络的逼近能力”，《神经网络》，第4卷，第251–257页，1991年。[25] C. Rackauckas 等，“面向科学机器学习的通用微分方程”，arXiv:2001.04385，2020年。[26] L. S. Pontryagin，《最优过程的数学理论》。Routledge，2018年。[27] K. Miao、L. Zhao、H. Wang、K. Gatsis 和 A. Papachristodoulou，“Opt-ODENet：利用可微优化层进行神经常微分方程控制器设计，以保障安全性和稳定性”，载于学习动力学与控制会议（L4DC）论文集，第1217–1229页，2025年。[28] D. P. Kingma 和 J. Ba，“Adam：一种随机优化方法”，arXiv:1412.6980，2014年。[29] I. E. Lagaris、A. Likas 和 D. I. Fotiadis，“用于求解常微分方程和偏微分方程的人工神经网络”，《IEEE神经网络汇刊》，第9卷，第987–1000页，1998年。[30] S. Chen、Z. Liu、W. Zhang 和 J. Yang，“用于求解偏微分方程正问题中多个情形的硬约束宽体物理信息神经网络模型”，《应用科学》，第14卷，第189页，2023年。[31] H. Daniels 和 M. Velikova，“单调与部分单调神经网络”，《IEEE神经网络汇刊》，第21卷，第906–917页，2010年。