arXiv:2303.12740v1 [cs.LG] 21 Mar 2023

反演基本图并预测边界条件：机器学习如何改进交通流宏观模型

Maya Briani[a]、Emiliano Cristiani[a]和 Elia Onofri[a, b]

a意大利国家研究委员会计算应用研究所，意大利罗马b罗马第三大学数学与物理系，意大利罗马

2023年3月23日

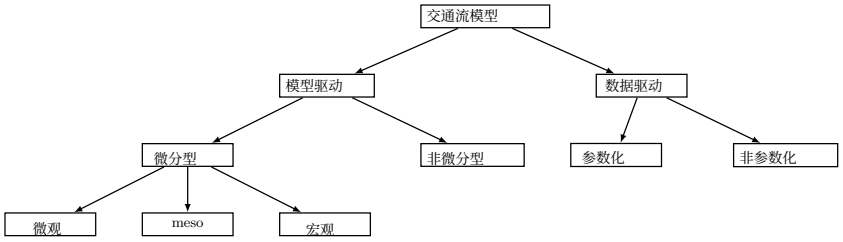
摘要

本文旨在开发新方法，将机器学习技术与宏观微分模型相结合，用于车辆交通的估算和预测。众所周知，数据驱动方法和模型驱动方法各有优缺点，有时还能互为补充。本文采用一个数据集，其中包含高速公路上行驶车辆的流量和速度数据，这些数据由固定传感器采集，并按车道和车辆类别分类。我们利用一种基于长短期记忆网络递归神经网络的机器学习模型，外推两项重要信息：1）传感器下方是否正在出现拥堵；2）近期（30分钟内）将通过传感器下方的车辆总数。随后利用这些信息提高基于 LWR 的一阶多类别模型的准确率，该模型描述传感器之间交通流的动力学。第一项信息用于反演（凹函数）基本图，从而根据通量数据恢复车辆密度，并将该密度数据直接输入模型。这有助于更准确地近似传感器之间的动力学，尤其是在道路未受监测路段发生事故时。第二项信息则作为交通模型所依据方程的边界条件，以便更准确地重建道路上未来任意时刻的车辆总数。本文将讨论一些以真实情景为依据的示例。真实数据由意大利高速公路公司 Autovie Venete S.p.A. 提供。

关键词。 交通流；车辆；基本图；LWR模型；机器学习；长短期记忆网络。
MSC-2020年。 76A30；68T07。

1 引言

交通状态估计（TSE）和交通预测有着悠久而扎实的传统，可追溯至20世纪50年代。下图概述了车辆交通流建模研究方向的一个大致划分：



模型驱动方法和数据驱动方法各有优缺点，相关内容在例如 [43, 52] 中已有详尽阐述：模型驱动方法能够将人类对系统的认知融入模拟器，前提是这些知识至少能够合理地转化为方程。不同主体之间的差异也可以（以随机方式）纳入考量，包括驾驶员的心理因素。尤其是微分宏观模型，能够展现基于偏微分方程（PDE）的各种方法的优势，例如为计算道路网络上交通系统的*Wardrop*均衡 提供合适的工具 [50, 8, 13]。另一方面，这种方法往往会过度简化交通流物理规律，因为模型始终无法捕捉汽车和驾驶员的所有特征。大多数模型难以处理交通传感器采集的噪声和波动数据，众多参数的校准也颇具挑战。此外，用于方程离散化的数值格式（Godunov、Lax-Friedrichs 等）是求解方程所必需的，但会引入额外的、往往不可忽略的近似误差。最后，模型还需要一些真实情景中无法获得的额外输入，例如未来任意时刻的边界条件。

相比之下，数据驱动方法更适合处理交通流中的非线性特征，因此可能比模型驱动方法更加准确，但它们不了解交通流的物理规律，可能导致不可行的估计结果。这些方法通常也缺乏可解释性和鲁棒性。更重要的是，这些模型的泛化能力通常较弱，并且高度依赖训练数据样本。如果训练数据的质量较差（存在缺失、过高估计或过低估计的数据），其预测准确度就会大幅下降。近年来，人们提出了许多机器学习（ML）方法：与偏微分方程（PDE）的复杂系统相比，这些方法通常采用相对简单的结构，因此计算开销较小，更适合实时应用。然而，由于依赖大量历史数据，训练的计算成本可能非常高，也容易导致数据过拟合。

近期研究越来越关注混合方法，它们试图

兼取两种方法之长；然而，这类研究在文献中较为少见，而且每篇论文仅采用两种方法中的个别方面，因此得出的结果差异很大。本文试图推动这一研究领域的发展，提出一种计算方法，利用机器学习技术从数据中外推微分宏观交通流模型所需的信息。

相关文献

基本图首先，我们必须提及基本图。它是所有模型驱动方法的基本组成部分之一，尤其是在宏观尺度上。它定义了流量与车辆密度之间的关系 [26, 19], ，见图1。显然，

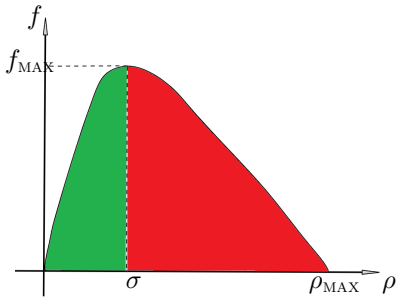


图1：基本图 $f = f(\rho)$。绿色部分对应自由流状态，红色部分对应拥堵状态。密度 σ对应最大流量 f_{MAX} （道路通行能力）。道路空无一车时（$\rho = 0$），以及道路完全拥堵、车辆停滞时（$\rho = \rho_{\mathrm{MAX}}$），流量均为零。

车辆流量在道路空无一车（密度为零）或道路完全拥堵（密度最大，车辆首尾相接且停止）时均为零。对于中间密度水平，真实数据呈现出更复杂的动力学，尤其是在道路通行能力达到最大值附近。事实上，驾驶员对相同交通状况的反应各不相同，此外，加速和减速也远非瞬时发生。因此，交通流会出现一些不稳定性 [6, 35, 46]。在一阶交通模型中，基本图可以用单个函数来定义；而二阶交通模型则允许基本图具有多值性，也就是说，单一密度值可以对应多个流量值，这与现实情况完全一致。近期，文献 [41]提出了一种有趣的基于机器学习的方法，用于估计基本图。

模型驱动的交通流数学模型最早出现于1955—1956年，当时发表的开创性论文 [29, 37], 提出了著名的 LWR模型。它是一种一阶（基于速度）模型，由一个双曲型偏微分方程构成，其中

观测量是车辆密度 ρ。该模型源于车辆动力学与流体动力学之间合理的类比。沿着同一思路，该模型随后扩展为二阶模型，以纳入 [36, 51] 和 [1, 57], 中的惯性（加速度）效应，分别由此诞生了 PW模型和 ARZ模型。与此同时，工程领域的文献中提出了 CTM模型 [14]和 METANET模型 [32] 。这两种模型分别等价于 LWR模型和 PW模型的离散化版本 [49]。关于数学模型的文献浩如烟海；有关微分型模型及其校准，我们请读者参阅综述 [49, 45, 22, 3, 39] ；有关非微分型模型（元胞自动机），请参阅 [30] 。

我们也致力于将数学模型推广到道路网络。由于必须加入交叉口条件，以确保所得偏微分方程组的解具有唯一性，因此这并非易事。关于基本概念，请参阅 [5, 4], 以及书籍 [20] 。

我们感兴趣的另一种推广是多类别模型。在这种情况下，同一道路上有多个车辆类别（例如汽车和卡车）共同通行。每个类别都有特定的动力学，各类别之间以复杂的方式相互作用。关于目前最常用的多类别模型概述，请参阅 [17, 44, 38], 以及近期出版的书籍 [19, 26] 。

数据驱动 交通数据可以通过多种方法和技术采集。通常可以获得欧拉数据，即由沿道路布设的固定传感器（用于统计驶过的车辆）提供的数据；也可以获得拉格朗日数据，即由配备了 GPS 系统等设备的探测车辆提供的数据。关于交通数据的概述，请参阅[9, 10, 53] 。因此，典型的分析对象是流量 f 和速度 v，而不是密度ρ 这一更抽象的概念；密度是模型驱动方法的特征。

早期的数据驱动研究主要依赖基于历史数据的统计方法。近年来，机器学习领域的大量研究为数据驱动方法注入了新的活力，众多研究采用从最简单到更为新颖的各种技术来探讨交通状态估计问题。

本文主要关注用于交通理解、估计和预测的人工神经网络（ANN）方法，参见 [43, 16, 34,54, 58, 60]。人工神经网络主要分为两类：前馈模型，例如著名的单层或多层感知机（S/MLP）；以及反馈模型，例如递归神经网络（RNN）。关于这一主题，读者可参阅近期的综述 [27, 48, 18, 31, 42] ；更具体地说，我们重点关注长短期记忆递归神经网络（LSTM-RNN，简称 LSTM）在交通数据预测中的应用。LSTM 是一种强大的工具，已证明能够捕捉长期时间特征依赖关系，并减少梯度爆炸或梯度消失。在近期文献中，LSTM 经常被用于（既可以单独使用，也可以作为更复杂结构的构建模块）分析和预测 f 与 v：就速度而言，例如， [21]采用融合深度学习方法，预测两分钟后的车道级交通速度； [24] 则在车速与汽车

类型之间的相关性基础上，构建了一个用于预测五分钟后高速公路车速的预测模型（LSTM ＋ 4层 MLP）； [52] 则将 LSTM 与经过小波分析辅助的细致数据预处理相结合，以进行 15 分钟后的车速预测。就流量而言，例如， [59] 将时空相关性整合到二维长短期记忆网络中，以预测 15 分钟后的交通流； [47] 将天气数据与注意力模型相结合，以进行短期交通流量预测； [18] 则介绍了一种新颖的方法，即采用基于长短期记忆网络的注意力模型，根据 120 分钟的数据（每 10 分钟聚合一次）预测即将到来的流量。

混合方法 为克服基于模型和基于数据的方法各自的不足，同时发挥它们的潜力，近期研究引入了耦合方法，使物理学与数据协同发挥作用。文献中实现耦合的方式各不相同，因为尚未形成统一的思路：物理模型可以（i）融入人工神经网络的训练过程，从而得到所谓的物理信息神经网络（PINNs）；（ii）与人工神经网络并行使用，例如 [41], 中同时进行交通状态估计、模型参数辨识和基本图估计；或（iii）在人工神经网络之后 使用，例如本研究以及 [23], 中利用数据为宏观交通流模型提供交叉口处一致的边界条件。

近期大多数研究属于物理信息神经网络（PINNs）这一类，通常通过构建定制代价函数将物理学融入模型，尤其是试图发挥深度学习模型（PIDL）的强大能力；例如， [40]的作者利用环形线圈检测器和探测车辆的观测数据，重点研究高速公路交通状态估计，并构建了一个耦合模型，其中包括物理非知情神经网络（PUNN）和一个采用基于物理偏差度量的定制损失函数的物理信息神经网络（PINN）。同样，论文 [25], 构建了一个基于CTM和LWR、采用不同基本图（Greenshields模型、Daganzo模型和反λ型）的定制损失函数，以解决数据稀疏性和传感器噪声问题。另一个基于多种物理因素构建定制代价函数的例子见 [2], ：作者利用探测车辆数据，在城市环境中进行交通状态估计，并构建了一个六项损失函数，既用于重建道路状态，也用于平滑探测车辆轨迹。值得一提的还有论文 [33],，作者提出了以不同常见基于物理的模型为基础构建的PIDL跟驰模型架构，用于预测每辆车的速度变化。最后，基于物理正则化高斯过程（PRGP）的方法也颇具意义，例如 [56] 提出的方法（随后在 [55]中扩展）：该方法构建了随机物理正则化高斯过程，并使用贝叶斯推断算法估计该物理正则化高斯过程本身的均值和核函数。

论文贡献

本文研究的是来自一系列沿高速公路布设的固定传感器的交通数据。传感器之间的距离约为 1 至 20 千米

这些传感器能够统计从其下方经过的车辆数量、估算车辆速度，并按车辆长度对车辆进行分类（例如将车辆分为轻型车辆和重型车辆）。传感器数据每分钟汇总一次，发送至中央处理单元，并存储在数据库中。速度数据对应于在 1 分钟时间间隔内观测到的各类别车辆速度的平均值。

主要目标是借助一阶多类别 LWR 类宏观微分模型，估计整条道路在当前和未来时刻的交通状况，并以流量 f、密度ρ和速度 v 等宏观量表示。该模型描述轻型与重型交通流的联合动力学，已在 [6] 中介绍。具体而言，我们考虑两个步骤：

临近预报。这是对当前时刻 t_0 （现在）道路上每个位置的交通估计。为此，我们将道路划分为若干连续路段，每个路段的起点和终点都设有一个传感器。然后，我们运行模型，将初始时刻设为 $t = t_0 - \Delta t_{\mathrm{PAST}}$，最终时刻设为 $t = t_0$，其中 Δt_{PAST} 是一个参数。模型在每个路段上独立运行，并给出该路段的密度变化。在 $t_0 - \Delta t_{\mathrm{PAST}}$ 时刻，我们假设道路为空，随后道路开始因传感器数据而逐渐填满，这些数据充当流入和流出的边界条件。如果 Δt_{PAST} 足够大，道路就会完全填满，并可沿每个路段计算出可靠的密度估计。

问题在于如何利用传感器数据来施加边界条件：数学模型通常需要密度数据 作为狄利克雷边界条件，但在我们的案例中，传感器无法提供这些信息。另一种做法是将通量数据 直接输入用于离散化建模方程的数值格式中。遗憾的是，这种方法并非总是可行，因为无法保证传感器数据与数值模型的解相容；此外，模型的解也并非总是“正确”的，尤其是在传感器之间发生拥堵事件时，详情见第6节。因此，我们探索了第三种方法：训练一个基于长短期记忆网络的人工神经网络，以实时检测传感器处的拥堵形成。这本身就是一个有趣而复杂的问题，它还带来一项副产品：一种能够无歧义地反演凹形基本图的工具，可以区分自由流状态和拥堵状态，见图1。随后，我们使用该工具将通量数据转换为密度值，并将其输入模型，从而解决上述不相容问题，同时为模型增加新的物理信息，在大多数情况下使解更加准确。

预测。 这是任意未来时间 $t_0 < t < t_0 + \Delta t_{\mathrm{FUT}}$ 的交通估计，其中 Δt_{FUT} 是模拟的持续时间（在我们的案例中为 30 分钟）。在这种情况下，传感器数据尚不可用，因此我们不考虑传感器，而是将整条高速公路视为一个连续的长路段。我们采用前面使用的相同数学模型，并将临近预报的交通估计作为密度的初始条件。

由于模型需要任意时间 $t_0 < t < t_0 + \Delta t_{\mathrm{FUT}}$ 的边界条件，也就是说，它需要估算将有多少车辆

在时间 $t_0+\Delta t_{\mathrm{FUT}}$ 之前驶入和驶离道路，因此，如何施加这些边界条件就成了一个问题，而传感器显然无法在时间 t 提供这些信息。为此，我们建立了另一个基于 LSTM 的人工神经网络，用于预测时间区间 $[t_0, t_0+\Delta t_{\mathrm{FUT}}]$ 内的传感器数据。更确切地说，人工神经网络的输出不会是逐分钟的通量数据，因为这些数据波动太大，无法保证预测可靠；相反，我们选择预测时间区间 $[t_0, t_0+\Delta t_{\mathrm{FUT}}]$ 内的车辆总数：这是一项更简单但仍有用的信息，因为它可以被视为恒定边界条件，在大多数情况下，能准确反映截至$t=t_0+\Delta t_{\mathrm{FUT}}$整条道路上的总量。

论文结构

本文其余部分安排如下。第2节介绍我们所考虑的交通流数据，并概述 Autovie Venete S.p.A. 提供的基准数据集。

第3节概述将用于增强数据集的人工神经网络结构。

第4节讨论用于提供拥堵事件实时检测信息的网络的训练与验证。本节还讨论拥堵事件的短期（几分钟）预测。

第5节讨论用于提供传感器处预期交通流量中期（30分钟）预测信息的网络的训练与验证。

第6节探讨通过使用前文所述的人工神经网络获得的增强数据集与数学模型之间的联系，涵盖临近预报和预测两种情况。

最后，第 7 节对本文作总结，并给出一些结语。

2 讨论数据

在宏观层面上，交通流由三个状态变量来描述：流量 f，即单位时间内通过某一点的车辆数；密度 ρ，即单位长度内的车辆数；以及 速度 v。在数学领域，这些量是定义在连续时空域上的函数，并满足以下关系

$$f(x,t)=\rho(x,t)\ v(x,t) \tag{1}$$

对于道路上的任意点 x 和时间 t均成立。然而在现实中，这些信息只能以离散样本的形式测量，因此无法在同一时刻精确计算。众所周知，通过对式 (1) 求逆来计算ρ 往往会导致较差的结果，尤其是在 ρ较大时，参见[23]。

我们的基准数据集由意大利高速公路公司Autovie Venete S.p.A.提供，数据采集时间为2020年9月至2021年3月



图2：意大利A4的里雅斯特—威尼斯高速公路及其通往/来自乌迪内、波代诺内和戈里齐亚的支线，由 Autovie Venete S.p.A. 管理。

2022年。该数据集包含来自意大利东北部三条高速公路沿线固定传感器的交通数据，分别是A23、A28和A4（部分路段），见图2。除A4高速公路的部分路段设有三条车道外，所有高速公路每个方向通常均有两条车道。车辆的限速通常根据其长度和重量而有所不同，此外，重型车辆不得使用最左侧车道。

数据由45个传感器组采集，平均每天总计提供301,200条记录。每组由沿道路的位置 x 和行驶方向确定，并由每条车道*1*个传感器 组成（因此每组有2个或3个传感器）。每个传感器都会对经过其前方的车辆进行计数，记录其速度，并根据德国TLS$5{+}1$ 类别标准 [7]对车辆进行分类。计数误差据称为 $\pm 3\%$，速度误差为 $\pm$最大值$\{3$ 公里/小时，$3\%\}$。下文中，我们将类别1和2合并为轻型车辆（摩托车、汽车、厢式货车和汽车挂车），将类别3、4和5合并为重型车辆（卡车、卡车挂车、牵引车和公共汽车）。在第4节和第5节中，我们将分别按类别和传感器组进一步汇总数据。

空间粒度的变化很大，因为传感器之间的距离从 1 到 20 千米不等。相比之下，时间粒度更加规律，因为每个传感器每分钟都会传输一次数据，以汇总测量数据的形式呈现：这意味着数据库会存储该时间间隔内通过的车辆总数，以及各类别的平均速度。这些测量数据会保留 2 小时，用于实时分析，之后便会转移到单独的数据库中。因此，历史数据无法实时获取。

有几点需要说明：

- 如上所述，仅通过反演关系式（1），无法根据流量和速度数据还原较大的密度值。这是测量数据聚合与时间粒度共同作用的结果（低流量需要较长的时间间隔才能被检测到，因为车辆行驶速度较慢）。

- 尽管通量数据在日尺度上呈现出规律性，但其每分钟的波动都很大，见图3。

- 最重要的是，我们的数据无法区分空道路和完全拥堵的道路。在这两种情况下，测量流量都为 0，速度则未定义。

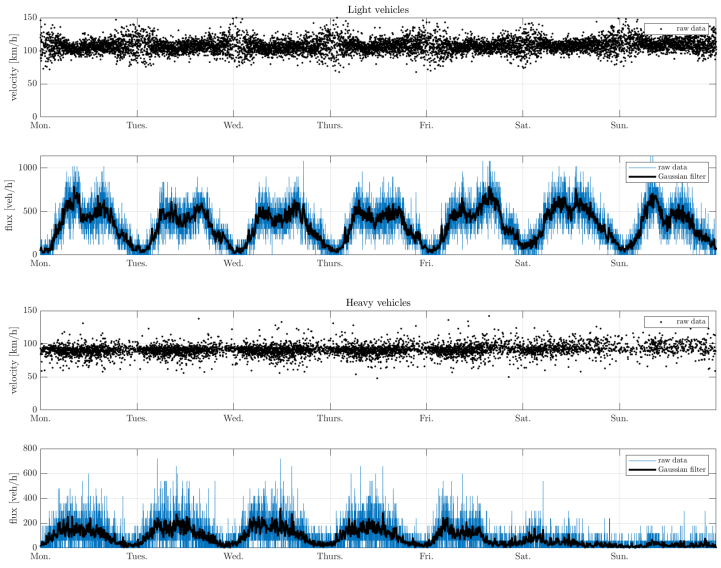


图3：周一至周日整周内轻型车辆（上图）和重型车辆（下图）的速度和流量。我们观察到，原始通量数据每分钟都波动很大，但应用高斯滤波器（黑线）后，可以看出一种每天重复出现的模式。夜间，所有车辆的通量数据都会下降，而轻型车辆的速度数据则变得更加分散。正如预期，周末重型车辆的流量相当低。

这些说明对于理解为何难以实时检测拥堵事件的形成至关重要；而这对于准确估计和预测交通流也必不可少，即使是在距离最先观测到拥堵的传感器很远的地方也是如此。为更好地理解这一点，我们在图4中展示了四起具有不同发展特点的拥堵事件。相比之下，图5展示了两种非常相似的交通状况，它们都以流量下降为特征，却以截然不同的方式演变：一种演变为自由流状态，另一种则演变为拥堵状态。这清楚地表明，尽管通过观察事后数据（例如一整天的数据），相对容易检测出拥堵事件（并将其与空道路状况区分开来），但要在拥堵形成当下做到这一点却非常困难。

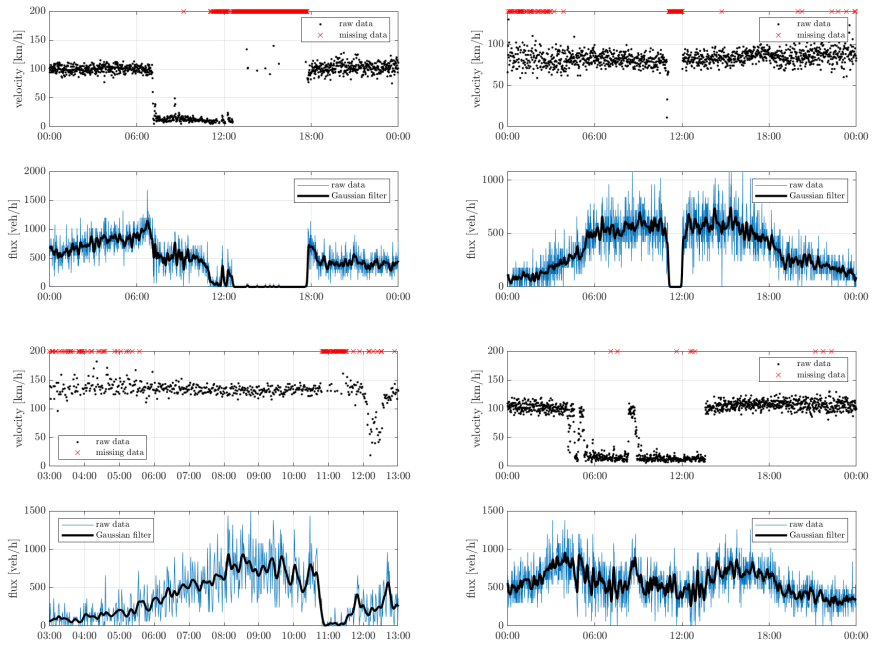


图4：四种具有不同特征的拥堵事件：（左上）我们观察到速度和流量迅速下降，随后流量消失，而速度变得无法确定（仍有部分快速车辆通过的情况除外）；（右上）流量和速度骤降；（左下）流量先下降，随后速度下降；（右下）速度下降，而流量仅部分降低。

需要注意的是，在已部署的45组传感器中，有17组还配备了先进技术[1]，这些技术结合了多普勒雷达、超声波发射器和被动红外辐射探测器，还能够识别停驶车辆和严重拥堵状况，并将其作为布尔标志记录在相应的分钟数据中。下文将这些传感器称为 $3T$ 传感器，并将其作为监督学习的主要数据来源之一。

[1]由 Asim Technologies Ltd 提供，系列 TT295。

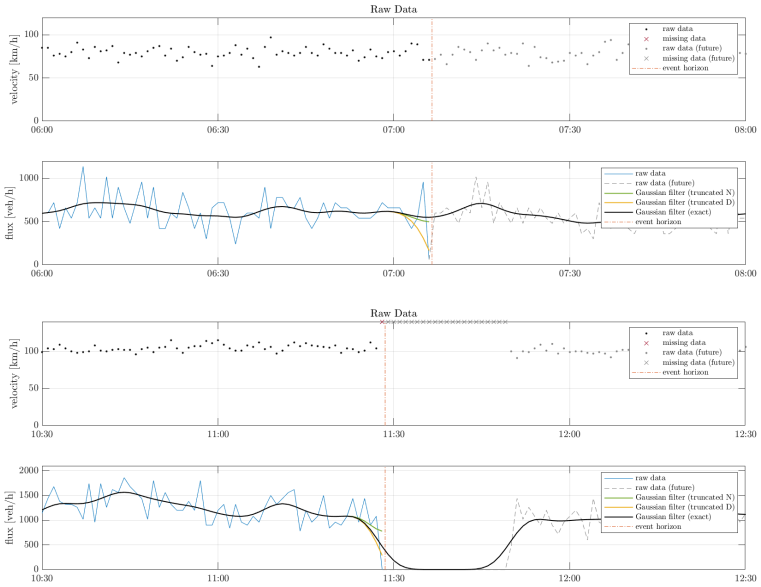


图5：（上）正常交通状况，其特点是流量通常会出现较大波动。7:05，流量突然下降，随后交通流恢复正常；（下）11:29，出现了非常相似的情况，但这次逐渐演变成排队。除了图3中已展示的高斯滤波器外，这里还展示了另外两个未使用未来数据（事件视界之后的数据）得到的高斯滤波器。截断是通过假设狄利克雷型边界条件或诺伊曼型边界条件得到的。我们看到，在事件视界处，无论是高斯滤波器还是原始数据，都不足以区分这两种情景。

3 面向数据集的监督式机器学习方法

在本节中，我们建立了一种独特的基于机器学习的方法，用于解决第 1 节中提出的两个问题，即实时检测拥堵事件，以及预测传感器处预期的平均通量。

正如我们已经提到的，在监督学习技术中，循环神经网络在时间序列分析方面表现得非常有效。事实上，它们是一类专用于分析时间序列$\mathbf{X} = (\mathbf{x}_t)_{t=1,2,\ldots}$的特殊人工神经网络，其特征数量固定为 N_{IN}（即 $|\mathbf{x}_t| = N_{\text{IN}}$）；同一个计算单元会针对序列中的每个步骤迭代应用一次，并将当前步骤的特征与前一步骤的输出一同作为输入。在我们的案例中，可从数据集中提取的特征是按车辆类别和/或车道组织的流量 f和速度 v 。

尤其是，LSTM-RNN 如今已成为许多时间序列相关问题的首选工具，能够捕捉较长时间范围内的时间特征或依赖关系。LSTM 的核心思想是将每个步骤的输出$\mathbf{h}_t$（在 LSTM 术语中称为隐藏状态）与生成该输出的已处理数据分开，从而保留 LSTM 自身的一种“内部记忆”$\mathbf{c}_t$（或细胞状态）。$\mathbf{h}_t$ 和$\mathbf{c}_t$ 的维度相同N_{HID}（即 $|\mathbf{h}_t| = |\mathbf{c}_t| = N_{\text{HID}}$），这也代表了 LSTM 能够捕捉的记忆“长度”（时间窗口）。具体而言，LSTM 的输出始终仅依赖于所分析序列中最后的 N_{HID} 个步骤：如果在少于 N_{HID} 个步骤后就输出结果，答案的准确度会较低。与 N_{IN} 不同，该参数必须通过适当的方法进行调优（超参数调优）。随后，可根据新的输入$\mathbf{x}_t$（可能已归一化）并结合先前的输出$\mathbf{h}_{t-1}$来更新细胞状态。

实际上，长短期记忆网络更新基于四个不同的组件$\mathbf{g}_t^{\text{F}}$、$\mathbf{g}_t^{\text{I}}$、$\mathbf{g}_t^{\text{C}}$和$\mathbf{g}_t^{\text{O}}$（或门，$|\mathbf{g}^{\cdot}| = N_{\text{HID}}$），这些组件分别对$\mathbf{h}_{t-1}$ 和$\mathbf{x}_t$ 的独立加权组合进行计算，如下所示：

$$
\begin{aligned}
\mathbf{g}_t^{\text{F}} &= \text{sigmoid}\Big(\mathbf{b}^{\text{F}} + \mathbf{W}^{\text{F}} \times \mathbf{x}_t + \mathbf{R}^{\text{F}} \times \mathbf{h}_{t-1}\Big),\\
\mathbf{g}_t^{\text{I}} &= \text{sigmoid}\Big(\mathbf{b}^{\text{I}} + \mathbf{W}^{\text{I}} \times \mathbf{x}_t + \mathbf{R}^{\text{I}} \times \mathbf{h}_{t-1}\Big),\\
\mathbf{g}_t^{\text{C}} &= \tanh\Big(\mathbf{b}^{\text{C}} + \mathbf{W}^{\text{C}} \times \mathbf{x}_t + \mathbf{R}^{\text{C}} \times \mathbf{h}_{t-1}\Big),\\
\mathbf{g}_t^{\text{O}} &= \text{sigmoid}\Big(\mathbf{b}^{\text{O}} + \mathbf{W}^{\text{O}} \times \mathbf{x}_t + \mathbf{R}^{\text{O}} \times \mathbf{h}_{t-1}\Big),
\end{aligned}
\tag{2}
$$

其中，$\times$ 表示标准的矩阵向量乘积，$\mathbf{W}^{\cdot}$、$\mathbf{R}^{\cdot}$和$\mathbf{b}^{\cdot}$ 分别称为输入权重、循环权重和偏置。这些权重通常在初始阶段随机赋值，并通过反向传播误差，作为训练阶段的目标进行调整。

每个门都有不同的作用：前三个门控制细胞状态$\mathbf{c}$ 的更新，第四个门则决定下一个隐藏状态$\mathbf{h}$（调节

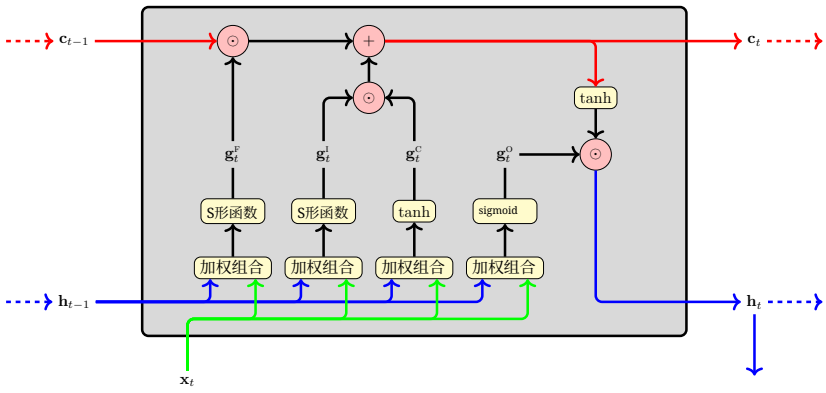


图6：长短期记忆网络计算单元在时间 t处理输入的示意结构。W.Comb. 表示带偏置的加权组合，写作 $\mathbf{b}_{\cdot} + \mathbf{W}_{\cdot} \times \mathbf{x}_t + \mathbf{R}_{\cdot} \times \mathbf{h}_{t-1}$。所有分量都是长度为 N_{HID} 的向量，但 $\mathbf{x}$ 的长度为 N_{IN}。我们回顾一下，$\mathbf{h}_t$ 也是网络的输出。

细胞状态的贡献）。更正式地说：

$\mathbf{g}^{\mathrm{F}}$ 或遗忘门：通过在 (0，1) 范围内进行变换来减弱$\mathbf{c}_{t-1}$。$\mathbf{g}^{\mathrm{I}}$ 或输入门：决定候选项对$\mathbf{c}_t$的影响程度，其变换范围为 (0，1)。$\mathbf{g}^{\mathrm{C}}$ 或候选门：表示用于调节 c 的细胞输入激活，其取值范围为 $(-1, 1)$。$\mathbf{g}^{\mathrm{O}}$ 或输出门：决定$\mathbf{c}_t$ 如何构成输出，变换范围为 (0，1)。

因此（图示见图6），更新规则如下：

$$\mathbf{c}_t = \mathbf{g}_t^{\mathrm{F}} \odot \mathbf{c}_{t-1} + \mathbf{g}_t^{\mathrm{I}} \odot \mathbf{g}_t^{\mathrm{C}}, \qquad \mathbf{h}_t = \mathbf{g}_t^{\mathrm{O}} \odot \tanh(\mathbf{c}_t), \tag{3}$$

其中，$\odot$ 表示 Hadamard（逐元素）乘积。

特别需要注意的是，长短期记忆网络的输出 $\mathbf{h}$ 是一个 N_{HID} 向量，因此需要对其进行处理，以便生成预测或分类。我们选择的工具是标准单层感知机前馈网络，它将 N_{HID} 特征压缩为

- 预测任务：一个长度为 N_{PRED} 的输出 $\mathbf{o}$，其中每个元素代表一项独立的预测。- 分类任务：一个长度为 N_{CLASS} 的特征向量 $\mathbf{z}$，其中每个元素代表问题中的一个类别，然后将其输入 softmax层进行转换

为概率向量 $\hat{\mathbf{o}}$ （表示属于特定类别的概率）。回顾一下，softmax 函数按分量定义如下：

$$\mathrm{softmax}(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=0}^{N_{\mathrm{CLASS}}-1} e^{z_j}}, \qquad i = 0, \ldots, N_{\mathrm{CLASS}} - 1. \tag{4}$$

最终的输出可以是概率向量 $\hat{\mathbf{o}}$，也可以是概率最高的类别的索引 o，即 $o = \mathrm{argmax}(\hat{\mathbf{o}})$。请注意，$\hat{\mathbf{o}}$ 可作为预测的置信度 指标：$\max(\hat{\mathbf{o}})$ 越接近 1，人工神经网络对预测就越有把握。

我们回顾一下，单层感知机是一个常规的带偏置加权组合，即长度为 N_{PRED} 或 N_{CLASS} 的输出 $\mathbf{z}$ 可表示为 $\mathbf{z}_t = \mathbf{b} + \mathbf{W} \times \mathbf{h}_t$，其中 $\mathbf{W}$ 和 $\mathbf{b}$ 分别是维度适当的权重矩阵和偏置向量。

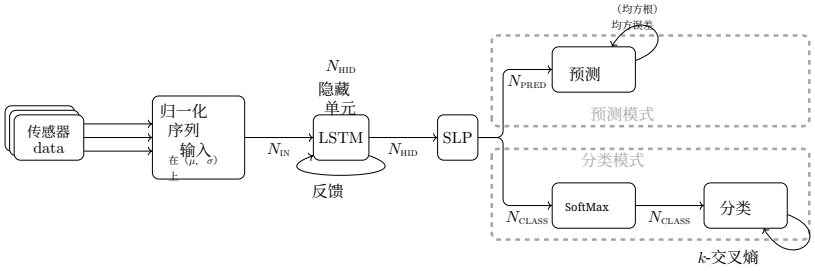


图 7：增强工具的流水线。传感器数据序列在输入长短期记忆网络之前，先按特征进行全局归一化，然后一次输入一个时间步。随后，长短期记忆网络的输出由多层感知机网络处理，该网络包含一个隐藏层或不包含隐藏层（实际相当于执行一次卷积步）。最后，根据是否为分类任务，在给出结果之前应用 SoftMax 层。

为了完成我们方法的流水线（摘要见图 7），我们还需要说明如何根据真实数据评估系统性能；也就是说，我们需要选择一个误差函数，以评估我们的估计值与真实值之间的差距（行话称为损失函数）。

- 预测任务：由于预测输出是一个实数，而且我们并不希望对高估和低估赋予不同权重，因此采用标准的（均方误差）均方根误差。分别用$\mathbf{y}_t$ 和$\mathbf{o}_t$表示输入$\mathbf{x}_t$对应的真实值和网络输出，回顾一下（均方误差）均方根误差的定义：

$$\mathrm{MSE} = \frac{1}{|\mathbf{X}|} \frac{1}{N_{\mathrm{PRED}}} \sum_{t=1}^{|\mathbf{X}|} \|\mathbf{o}_t - \mathbf{y}_t\|^2 , \qquad \mathrm{RMSE} = \sqrt{\mathrm{MSE}} . \tag{5}$$

- 分类任务：布尔分类损失的一种常见选择是 k交叉熵（带有 $k = 2$）。具体而言，由于我们的数据集中的正样本和负样本极不均衡，我们决定对正样本比例 $0 < p_r < 1$进行加权来计算交叉熵损失，而不是用合成数据扩充数据集，因为合成数据往往难以生成。回顾一下，应用于二分类任务的 softmax 会返回一个含两个值的向量 $\hat{\mathbf{o}}_t = (1 - \hat{o}_t, \hat{o}_t)$，其中$\hat{o}_t$ 表示选择正类的概率。因此，若 y_t 是与输入$\mathbf{x}_t$ 对应的真实标签位——也就是说，我们的目标$\mathbf{y}_t$ 为(1,0)或(0,1)——那么加权二元交叉熵定义为：

$$\text{CrossEntropy} = -\frac{1}{|\mathbf{X}|}\sum_{t=1}^{|\mathbf{X}|}\left(\big((1-p_r)\cdot y_t\cdot\ln(\hat{o}_t)\big)+\big(p_r\cdot(1-y_t)\cdot\ln(1-\hat{o}_t)\big)\right). \tag{6}$$

4 拥堵检测与短期预测

本节中，我们构建了一个拥堵事件带标签数据集，并使用前文介绍的基于人工神经网络的方法构建两个分类器：第一个分类器 $\mathfrak{F}_c$ 用于实时检测拥堵事件，第二个分类器 $\mathfrak{F}_p$ 用于对同类拥堵事件进行短期预测。这两个分类器分别充当“拥堵警报”和“拥堵预警”的触发器。

4.1 构建数据集

该数据集使用 $3T$ 传感器采集的数据构建。由于每个传感器（即每条车道）都会检测到拥堵事件，我们决定汇总所有车辆类别的数据，并按车道分别处理。因此，特征为所有车辆（所有类别）的总流量 f 和平均速度 v 。（实际上，事后恢复每个类别的一条信息通常相当容易。事实上，如果在慢车道检测到拥堵事件，而在快车道没有检测到，那么很可能只有重型车辆发生了拥堵，因为它们无法使用快车道，见第 2 节。）

我们按天拆分数据，以便在每个给定时间仅使用随机抽取的数据日批次进行增量训练阶段。因此，我们得到的数据集样本由一个双特征、长度为 1440（24 小时 $\times$ 60 分钟）的序列 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{1440})$ 构成，其中 $\mathbf{x}_t = (f_t, v_t)$、$t = 1, \dots, 1440$。数据集应使用一个长度为 1440 的布尔数组 $\mathbf{y} = (y_1, \dots, y_{1440})$ 进行标注，用以报告相应分钟 t 是否被标记为拥堵事件。数组 $\mathbf{y}$ 通过对三个不同计算过程进行逻辑组合来计算，每个过程都会产生一个布尔标志。具体而言，我们有：

$$\mathbf{y} = \mathbf{b}^{3T} \vee \mathbf{b}^{f} \vee \mathbf{b}^{v},$$

其中 $\vee$ 表示逻辑或，并且：

- $\mathbf{b}^{3T}$ 是由 $3T$传感器直接传输的拥堵事件标记。平均而言，每个传感器每天报告的拥堵事件数量相当少（$< 0.1\%$），因此，为了让数据集更加均衡，我们选择了正标签比例至少达到 $\sim 1\%$的样本（即$\sum_{t=1}^{1440} (b_t^{3T} \geq 15)$）。我们注意到，$3T$ 检测过程相当“保守”，也就是说，只有当传感器下方存在稳定队列时，它才倾向于报告拥堵事件。此外，拥堵事件的确切定义，以及用于检测拥堵事件的物理和计算过程均未公开，作者也对此并不知情。基于这些原因，我们决定通过以下两种启发式方法来扩充真实样本。

- $\mathbf{b}^{f}$ 仅考虑流量数组$\mathbf{f} := (f_1, \ldots, f_{1440})$，以及通过与三角核函数进行卷积得到的10倍高斯正则化$\mathbf{f}^*$，参见图3；如果传感器在 t 时检测到以下条件，则报告发生了拥堵事件（白天，即上午5点至晚上8点）：

$$f_{t-1} < 2, \quad \wedge \quad f_t < 2, \quad \wedge \quad f_t^* < 2, \quad \wedge \quad \left(\frac{1}{60}\sum_{s=t-60}^{t-1} f_s\right) - f_t^* > 2\,, \tag{7}$$

  其中 $\wedge$ 表示逻辑与运算符。这种启发式方法的思路是：当前时间的低流量与流量突然下降（即前一时间段的平均通量较高）结合起来，是发生拥堵事件的充分条件。

- $\mathbf{b}^{v}$ 仅考虑速度数组$\mathbf{v} := (v_1, \ldots, v_{1440})$，以及正则化$\mathbf{v}^*$（按前述方法获得）。如果传感器在白天（上午 5 点至晚上 8 点）检测到以下情况，则报告在时间 t 发生了拥堵事件：

$$v_t < v_{t-1}, \quad \wedge \quad 0 < v_t < 65, \quad \wedge \quad \max\{v_s^* - v_t \mid s = t-1, \ldots, t-15\} \geq 40\,. \tag{8}$$

  这种启发式方法的思路是：当前时间的低速度与速度突然下降（即前一时间段的速度较高）结合起来，是发生拥堵事件的充分条件。

备注。这两种启发式方法是在与数据所有者合作的过程中开发的，并通过了一个多月的人工直接观测验证。

备注。需要指出的是，这两种启发式方法都使用了在时间 t 尚不可用的数据（由于正则化），因此无法用于实时应用。

分析通量数据的平均值和标准差后，我们发现部署在 A4高速公路上的大多数传感器的表现与

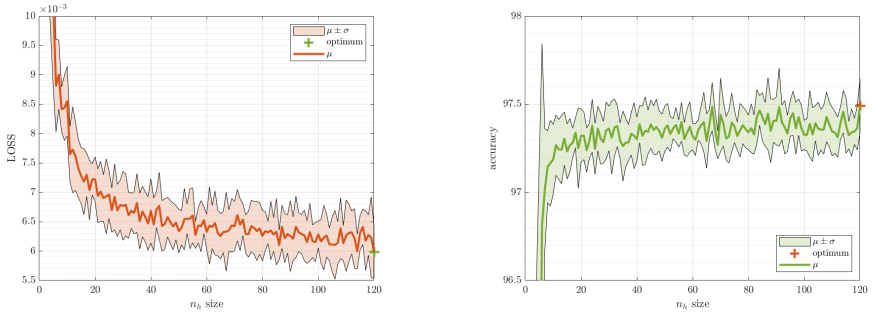


图 8： k-交叉熵（左侧为损失，右侧为准确率） $\mathfrak{F}_c$ 在不同 N_{HID}条件下训练所达到的结果。如图所示，较优的参数为 $N_{\text{HID}} = 120$ （即对应 2 小时），这表明使用更多数据或许能进一步改善结果。

其他传感器截然不同（$\mu \sim 1e{+}4$ 辆/天与 $2e{+}3$ 辆/天，以及 $\sigma \sim 1e{+}4$ 与 $5e{+}3$）。这并不令人意外，因为与 A23 和 A28 相比，A4高速公路连接的人口稠密地区更多。因此，我们将传感器分为两个互不重叠的集合，即高流量（HF）和低流量（LF）。LF 包括 A23、A28 上的所有传感器，以及 A4高速公路最左侧车道上的所有传感器（该路段有 3 条车道）。在本节余下部分及下一节中，我们只考虑 HF 传感器；针对 LF 传感器的流程和分析则与此类似。

我们将高流量119天数据集 $\{(\mathbf{X}^{(d)}, \mathbf{y}^{(d)})\}$， $d = 1, \ldots\ldots, 119$（171,360 个训练分钟）划分为 96 个训练日和 23 个验证日。增强后的平均样本阳性率为 $p_r = 4.2\%$，因此我们对k-交叉熵函数进行了适当加权。

4.2 模型训练

我们首先训练高流量分类器 $\mathfrak{F}_c$： 进行了多次训练，以估算长短期记忆网络的记忆单元合适大小 N_{HID} 。每次训练均使用 ADAM 优化器，对大小为 24 个样本的批次进行随机打乱，并训练 100 个周期；训练过程分为 9 次逐步调整学习率（每次调整训练 10 个周期）。

图 8 展示了各训练过程的损失和准确率结果。我们发现， $N_{\text{HID}} = 120$ 是最适合该模型的参数（其中 $N_{\text{IN}} = 2$、 $N_{\text{CLASS}} = 2$），准确率最高可达 97.70%（按 p_r 加权后为 99.75%），对应的损失值为 5.2×10^{-3}。我们提醒读者，$N_{\text{HID}} = 120$ 是我们的框架在实时应用中可用的最大数据量，见第 2 节。

其次，我们解决了训练分类器 $\mathfrak{F}_p$的问题。我们固定了模型的维度——即 $N_{\text{IN}} = 2$、 $N_{\text{HID}} = 120$ 和 $N_{\text{CLASS}} = 2$ ——并重点研究能够提前多少分钟预测 $\mathfrak{F}_c$ 的分类，同时又不至于使准确率下降过多。最简单的做法是生成一个带标签的

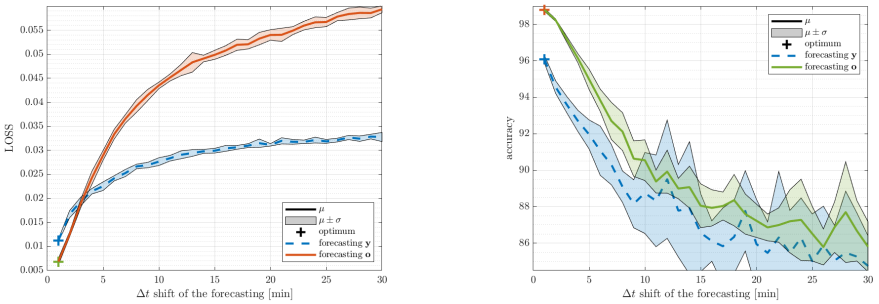


图 9：k-对于 $\mathfrak{F}_p$，随着预测 Δt 的时间窗口变化，训练过程所达到的交叉熵（损失，左侧）和准确率（右侧）（x轴）；预测 $\mathbf{y}^{(d)}$（蓝色虚线）和 $\mathbf{o}^{(d)}$（橙色/绿色）。正如预期，Δt越小，预测效果越好。

数据集，用于预测任务：将向量$\mathbf{y}^{(d)}$ 向前移动 Δt 分钟，以便将拥堵事件“提前”。然而，我们更倾向于预测 $\mathfrak{F}_c$ 本身的输出 $\mathbf{o}^{(d)}$（即概率向量的最大值索引），而不是 $\mathfrak{F}_c$的目标 $\mathbf{y}^{(d)}$。改变目标函数实际上有两个主要优点：（i）输出 $\mathbf{o}^{(d)}$ 在某种程度上是实际目标 $\mathbf{y}^{(d)}$的“正则化”版本，因此应该更容易预测；（ii）从理念上说，我们预期低流量警报之后会出现拥堵警报。不过，为了进一步修正数据集，我们还从新的目标向量中过滤掉了所有孤立的拥堵事件，实际上剔除了那些可能是 $\mathfrak{F}_c$假阳性的事件。

图 9 展示了使用所提出的两个数据集进行训练的结果，其中比较了不同 Δt 偏移下的损失和准确率。分析准确率可以发现，预测 $\mathbf{o}^{(d)}$ 通常更简单；相反，当目标设为 $\mathbf{y}^{(d)}$ 时，损失增长得更慢，这实际上表明，在 $\mathbf{o}^{(d)}$ 上训练对假阳性的判定不那么保守。我们发现，$\Delta t = 4$ 分钟是最合适的参数（尤其是在第二个数据集上，损失仍然相当），能够得到有意义的预测，同时平均准确率保持在 95% 以上（最高峰值为 96.50%，加权后为 98.82%），对应的平均损失为 0.24，因此与 $\mathfrak{F}_c$ 得到的结果相当。

4.3 实时分类器 $\mathfrak{F}_c$的性能评估

图10展示了低流量和高流量传感器对拥堵事件进行日长度分类的几个示例。我们报告了速度 v、流量 f以及分类的置信度（表示为经重新缩放至区间 $[-1,+1]$ 的 $\hat{}o$，即 $\hat{}o \cdot 2 - 1$）。可以看到，当下降仅在 v 中显著、而在 f 中不显著时，模型对事件分类略有不确定性（图10（左上））；然而，如果下降也涉及 f，那么即使.....模型也能够对拥堵进行分类

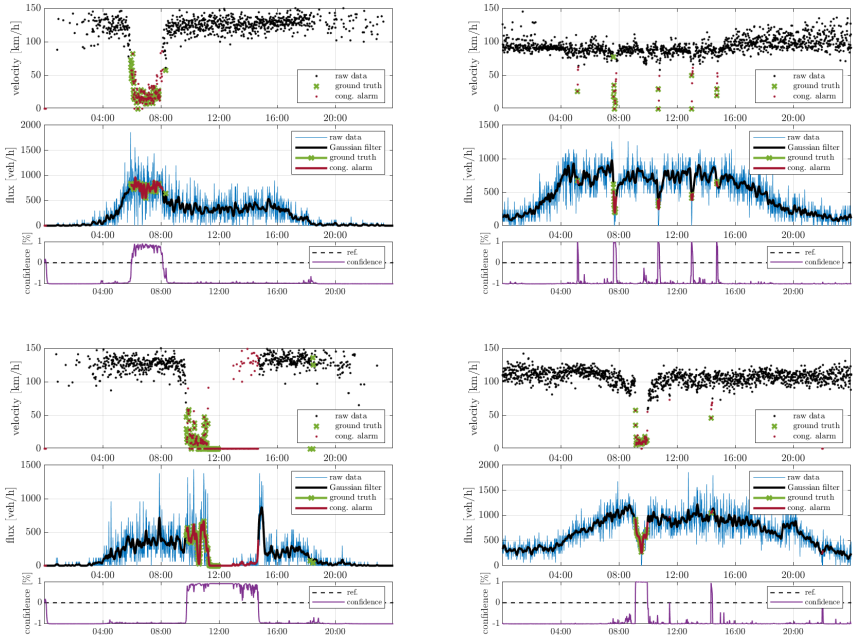


图10：拥堵事件的日长度分类。每幅图分别展示速度（上）、流量（中）以及模型分类的置信度（下）。左：低流量传感器；右：高流量传感器。

这些情况持续的时间很短（图10（右上））。在某些特定情况下，即使真实值不确定，模型也能够对关键事件进行分类，例如流量接近于零而速度较高时（图10（左下））。请注意，这是一种典型的夜间情形，此时没有拥堵；最后，当关键情况逐渐消失时，模型仍会保持拥堵警报开启：这是因为模型无法预知未来，与启发式方法所采用的正则化不同（图10（右下））。

4.4 短期预测分类器 $\mathfrak{F}_p$ 的性能评估

为说明预测拥堵事件的实际可能性，图11展示了我们常见的四种情形。图11（左上）展示了一种非常理想的情况：我们能够在拥堵实际形成前4分钟预测到拥堵。这是因为速度和流量都在下降之前稍早开始降低。图11（右上）展示了一种情形：持续时间较长的拥堵发生前，警报状态令人困惑，时开时关。在这种情况下，拥堵预警始终开启。图11（左下）展示了一种预测相当困难的情况，即使是受过训练的人也难以判断：拥堵伴随着速度骤降而突然开始，人工神经网络只能提前1分钟预测到。最后，图11（右下）展示了一种不确定的情况，它很可能会发展成排队（但至少在接下来的几分钟内并未如此）。预警处于开启状态；尽管从形式上看这是不正确的（拥堵预警之后并未出现拥堵警报），但这种不确定性充分说明了人工神经网络采取该行为的合理性。

5 预期交通流量的计算

本节将构建一个基于人工神经网络的预测器 $\mathfrak{P}$，以解决预期交通流量的中期预测（30 分钟）问题。需要注意的是，与上一节不同，本节将同一组内的传感器（按第 2 节中的定义）进行聚合，同时研究两种不同但相互关联的车辆类别（轻型和重型）。这一点很重要，因为这两种车辆类别的动力学差异很大，同时又紧密相连。

5.1 构建数据集

真实数据$\mathbf{y}$ 可以轻松地为每个传感器（组）生成，只需获取轻型车辆的实际流量$\mathbf{f}^{\mathrm{L}}$以及重型车辆的实际流量$\mathbf{f}^{\mathrm{H}}$，即T 未来几分钟内的流量。这些数据构成特征向量$\mathbf{x}$，即（$\mathbf{f}^{\mathrm{L}}$，$\mathbf{f}^{\mathrm{H}}$），以及对应的真实值$\mathbf{y}$：

$$y_t = \left(\frac{1}{\mathrm{T}} \sum_{s=t+1}^{t+\mathrm{T}} f_s^{\mathrm{L}},\ \frac{1}{\mathrm{T}} \sum_{s=t+1}^{t+\mathrm{T}} f_s^{\mathrm{H}} \right). \tag{9}$$

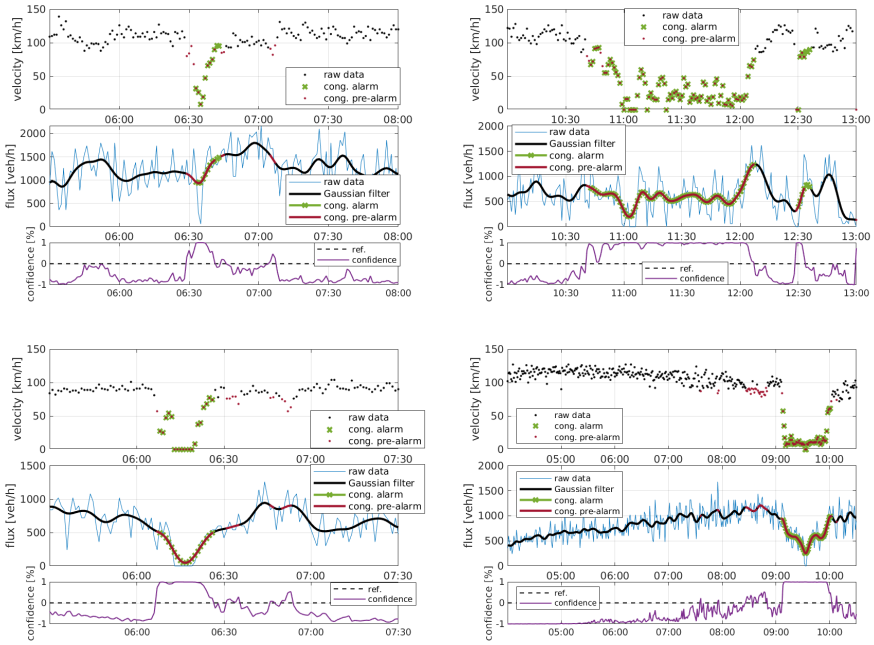


图11：左上：6:30 时，拥堵事件得到正确预警，提前了四分钟。右上：10:45 时，拥堵事件间歇性地开始出现，同时发出了明确的预警。左下：6:06 时，出现了难以预测的拥堵。右下：8:30 时，出现了状况不确定的情况，其特征是速度低于平均水平而流量正常；9:00 后形成了明显的拥堵（提前 1 分钟预警）。

关于数据集创建，我们选取了2021年全年采集的数据，对应348条按日记录（由于报告的读数问题，缺少17天的数据）。鉴于可用数据量很大，我们决定为每组传感器分别调整一个预测器。与先前的训练一样，我们将特征整理为日长度序列，并为每组 g构建了数据集 $D^{(g)} = \{\mathbf{x}^{(d)}; \mathbf{y}^{(d)}\}, 1 \leq d \leq 348$。我们从数据集中提取了大约一个月的样本作为测试集，其余部分则用作训练集。

5.2 模型训练

我们针对每组传感器，使用其对应的数据集 $D^{(g)}$训练了一个不同的模型 $\mathfrak{P}^{(g)}$ 。具体而言，我们重点关注代表所研究高速公路网络流入边界的四组传感器：A4高速公路上的威尼斯（g_1）和的里雅斯特（g_2）、A28高速公路上的科内利亚诺（g_3），以及A23高速公路上的乌迪内（g_4），见图 2。 g_1 是唯一部署在三车道路段上的一组传感器。

沿用图 7 中的命名方式，输入特征和输出特征的选择要求 $N_{\mathrm{IN}} = 2$ 和 $N_{\mathrm{PRED}} = 2$，而 N_{HID} 是一个自由参数。因此，我们进行了多次训练过程，以估算长短期记忆网络的记忆单元 N_{HID} 的合适大小。每次训练均使用 ADAM 优化器，并采用可变学习率；学习率在 5 个逐步递进的学习阶段内分段递减，每个阶段包含 50 个训练周期（总计 250 个训练周期）。

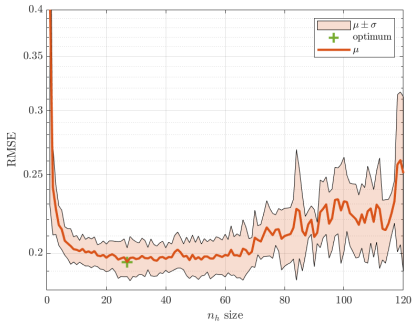


图 12：与 $\mathfrak{P}$训练对应的均方根误差值随规模 N_{HID}的变化。

图 12 展示了均方根误差随自由参数N_{HID}的变化。我们发现， $N_{\mathrm{HID}} = 27$ 是最能准确捕捉平均通量趋势的参数；由此得到的平均误差及相应统计数据见表 1。请注意，图 12 中报告的均方根误差是在归一化数据集上计算的，因此不能与表 1 中的结果进行比较。

	轻型车辆		重型车辆	
	双车道	三车道	双车道	三车道
误差均值的平均值	42.74	76.60	11.74	25.06
平均误差最大值	70.80	95.06	19.48	33.44
标准差均值	41.72	62.96	12.04	23.94
最大误差均值	422.64	744.40	164.46	251.24
平均流量均值	387.74	768.16	96.52	281.42
最大流量均值	5320.64	5864.48	1825.30	2047.10

表 1：多个数据集上多次训练所达到的误差统计（单位： [辆/小时]，使用 $N_{\mathrm{HID}} = 27$），涵盖四个数据集 $D^{(g_i)}$（对数据集中所有可用分钟取平均）。数值按车辆类别和车道数排序。平均流量和最大流量也作为相对误差统计的基准值列出。

由此产生一个问题：是否真的有必要针对每组传感器分别训练一个长短期记忆网络，还是一个人工神经网络就能适用于所有传感器。为回答这个问题，我们在

图13 $\mathfrak{P}^{(g_i)}$ 在数据集 $D^{(g_j)}$上测试得到的误差，其中 i,j =1,...,4。可以看出，强烈建议每组都训练尽可能多的人工神经网络

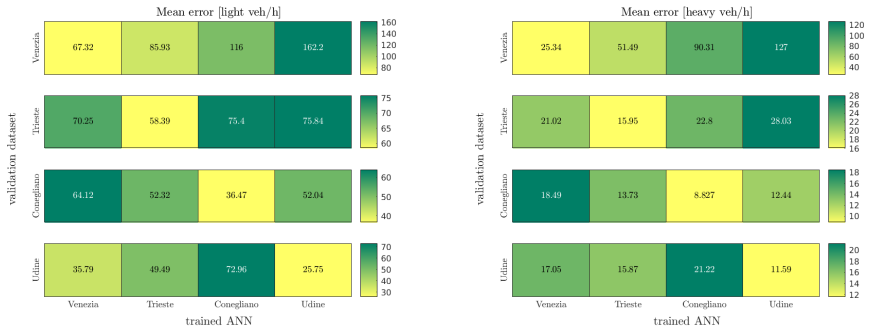


图13：将网络 $\mathfrak{P}^{(g_i)}$ $i \in \{1,2,3,4\}$ （x 轴）应用于相应的测试集 $D^{(g_j)}$ $j \in \{1,2,3,4\}$ （y 轴）得到的 [车辆/小时] 中的平均误差（$|y_t - o_t|$）（左侧为轻型车辆，右侧为重型车辆）。如图所示，每行最大值出现在主对角线上，这意味着将特定人工神经网络应用于不同的数据集并不合适。值得注意的是，行最大值不一定对应列最大值（例如，$i = 1, j = 4$）；这种现象是由某些数据集具有很强的规律性造成的。

考虑了传感器的数量。

5.3 性能评估

我们以传感器组 g_2 为例评估性能。图14展示了在测试集中每分钟进行预测所产生的误差直方图（单位：[辆/小时]）。图15

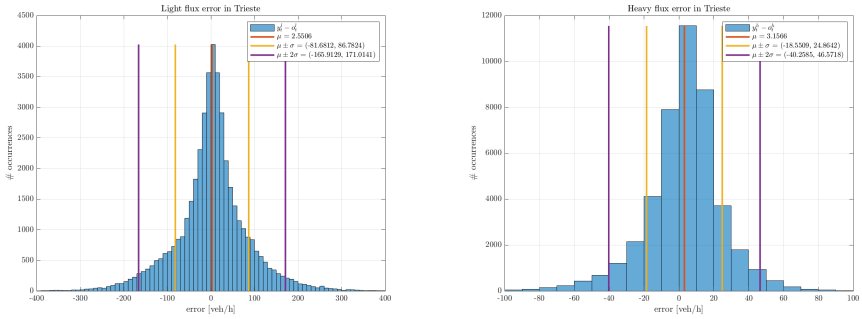


图14： $\mathfrak{P}^{g_2}$ 在其测试集上的预测误差直方图。轻型（左）和重型（右）车辆。误差关于 0 几乎对称。

则展示了 $\mathfrak{P}^{g_2}$ 一整天的结果。每分钟都将真实值（即未来30分钟内通过的车辆总数，单位：[辆/小时]）与人工神经网络预测值进行比较，并评估偏差。图中还给出了实际流量$\mathbf{f}$ 及其事后正则化结果，以供参考。再次可以看出，预测误差几乎对称；轻型车辆的误差通常小于200辆/小时，而重型车辆的误差则小得多。

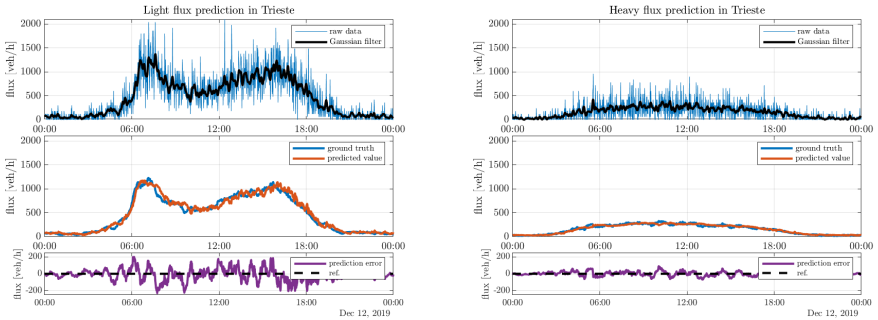


图15：对未来30分钟内每分钟的流量进行机器学习预测，并与实际测量的流量进行比较，轻型车辆（左）和重型车辆（右）。

最后，图16展示了一个单分钟预测示例。

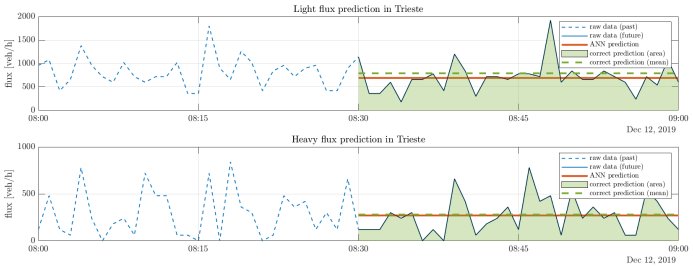


图16：对图15所示数据进行单分钟预测的放大图。在时间 t=8:30进行预测，此时虚线所示的数据已可用。预测值由橙色线下方的面积表示，而预测目标则由绿色阴影区域的数值表示（为便于比较，也以绿色虚线表示）。

6 使用机器学习增强数据为交通流模型提供数据

前文介绍的工具本身就很有用，但也可以与其他更经典的交通估计工具结合使用

并进行预测。本节尝试利用从数据中推断出的信息，以提高宏观微分模型的准确率。

6.1 临近预报

本节探讨在利用宏观微分型模型估计交通密度当前状态时，反转基本图可能带来的优势。如第 1 节所述，我们将整条道路 $[x_{\mathrm{MIN}}, x_{\mathrm{MAX}}]$ 划分为由固定传感器界定的连续路段。我们的目标是估计当前时刻 t_0的交通密度，因此从较早的时刻 $t_0 - \Delta t_{\mathrm{PAST}}$开始模拟，其中 $\Delta t_{\mathrm{PAST}} > 0$，并假设该时刻道路为空。设 N_S 表示道路路段的数量（传感器为 $N_S + 1$）。每个道路路段 $S^k := (s^k,\ s^{k+1})$，$k = 1,\ \ldots\ldots,\ N_S$，均由两个固定传感器界定，分别位于 $x = s^k$ 和 $x = s^{k+1}$。

我们先从简化情形入手，然后再转向实际情形。

6.1.1 简化情形

我们在每个道路路段 S_k上考虑经典的单车道单类别 LWR 模型

$$\begin{cases} \partial_t \rho^k(x,t) + \partial_x f(\rho^k(x,t)) = 0, & x \in S^k, \quad t \in (t_0 - \Delta t_{\mathrm{PAST}}, t_0) \\ \rho^k(x, t_0) \equiv 0, & x \in S^k \\ \rho^k(s_k, t) = \rho^k_{\mathrm{IN}}(t), & t \in [t_0 - \Delta t_{\mathrm{PAST}}, t_0] \\ \rho^k(s_{k+1}, t) = \rho^k_{\mathrm{OUT}}(t), & t \in [t_0 - \Delta t_{\mathrm{PAST}}, t_0] \end{cases} \tag{10}$$

其中，$\rho^k \in [0, \rho_{\mathrm{MAX}}]$ 表示某个最大密度为 $\rho^k_{\mathrm{MAX}} > 0$时的车辆密度，$\rho \mapsto f(\rho)$ 表示基本图。按惯例，假设$\rho \mapsto f(\rho)$是凹函数，并将 σ记为 f的最大值索引，即 $f(\sigma) = \max_\rho f(\rho)$（见图1）。方程（10）在每个路段上独立定义。还需指出，关系式（1）成立。

方程（10）通常通过数值近似求解。在定义域 $S^k \times [t_0 - \Delta t_{\mathrm{PAST}}, t_0]$上引入网格，空间步长为 Δx，时间步长为 Δt。时间区间划分为 N_t 个区间，而每个路段 S^k划分为 N^k_x 个长度为 Δx 的单元；时刻 n时，单元 C^k_j、$j = 1,\ \ldots\ldots,\ N^k_x$中的近似平均密度记为 $\rho^{k,\ n}_{\ j}$。方程（10）的任一守恒数值格式 [28] 具有如下形式

$$\rho^{k,n+1}_j = \rho^{k,n}_j - \frac{\Delta t}{\Delta x}\Big(F(\rho^{k,n}_j, \rho^{k,n}_{j+1}) - F(\rho^{k,n}_{j-1}, \rho^{k,n}_j)\Big), \qquad j = 1, \ldots, N^k_x, \tag{11}$$

其中 F是数值通量 （即对两个相邻单元界面处流量 f的近似）。例如，对于 Godunov 格式,

我们有

$$F(\rho_-,\rho_+) := \begin{cases} \min\{f(\rho_-), f(\rho_+)\} & \text{if } \rho_- \le \rho_+ \\ f(\rho_-) & \text{if } \rho_- > \rho_+ \text{ and } \rho_- < \sigma \\ f(\sigma) & \text{if } \rho_- > \rho_+ \text{ and } \rho_- \ge \sigma \ge \rho_+ \\ f(\rho_+) & \text{if } \rho_- > \rho_+ \text{ and } \rho_+ > \sigma \end{cases} . \quad (12)$$

例如，考虑给定路段 S^k 的右边界条件，它对应于下一个路段 S^{k+1} 的左边界条件。为便于阅读，我们也省略 ρ 中的索引 k：如果 $j = N_x^k$，可以采用一种

- 基于通量的方法：将传感器测得的通量数据 f_{s^k} 直接注入数值格式，作为界面 $S^k|S^{k+1}$ 处的数值出流通量 $F(\rho_{N_x}^n,\ \rho_{\text{OUT}}^n)$，而无需估计密度 ρ_{OUT}；- 基于密度的方法：通过估计密度 ρ_{OUT}，计算数值通量 $F(\rho_{N_x}^n,\ \rho_{\text{OUT}}^n)$。

原则上，这两种方法都正确，第一种似乎更实用，因为传感器只提供通量数据（也就是说，根本无法获取密度）。另一方面，传感器提供的通量数据 f_s 并不总是与数值格式所描述的解相容。事实上，任何处于可容许值集合之外的通量f_s

$$\left\{F(\rho_{N_x}^n, \rho) : \rho \in [0, \rho_{\text{MAX}}]\right\} \quad (13)$$

与（10）不兼容，并会导致负密度，因为流出的质量超过了道路上现有的质量，见图17左图。由此产生的问题是，如何确保传感器数据兼容：一种自然的解决方案是将传感器数据投影到可容许通量数据集（13）中。

对于基于密度的方法，其思路是利用第4节介绍的算法，帮助反演凹形基本图，即通过适当区分自由流状态和拥堵状态，从通量数据推导出密度数据。更具体地说，给定传感器流量数据 $f_s < f_{\text{MAX}}$，根据是否存在拥堵，在 ρ 和 ρ' 之间进行选择，其中 $\rho' \neq \rho$、 f（ρ） $= f$（ρ'） $= f_s$，见图 17 右图。

为了讨论这两种方法的差异，我们设计了一项简单的数值测试。考虑一条无限长的单车道道路。按照数学文献中的常见做法，我们将密度和速度都归一化到区间 $[0,1]$，并选择基本图为 f（ρ） $= \rho(1-\rho)$。在这种情况下，最大流量为 0.25，并在 $\rho = \sigma := 0.5$ 处达到。两个传感器分别位于 $x = s^k := 0.45$ 和 $x = s^{k+1} := 0.8$。

在初始时刻 $t = t_0 - \Delta t_{\text{PAST}}$，道路上的密度恒定为 $\rho = 0.45$，持续 $x < 0.52$，而道路其余部分为空（$\rho = 0$）。初始时刻过后不久，$x = b := 0.6$ 处发生事故，两个传感器之间形成了一个瓶颈。

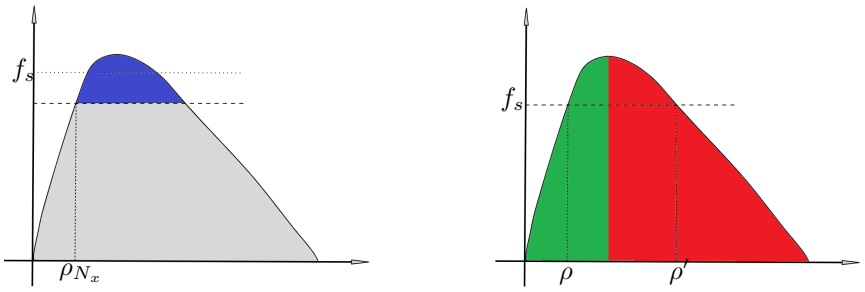


图17：基本图 $f = f(\rho)$。（左）数值通量 12 的一个不可采纳传感器流量数据示例：不存在与任何边界条件 ρ_{OUT} 相容的流量 $f_s > f(\rho_{N_x})$。（右）给定传感器流量数据 f_s 后，根据是否存在拥堵，选择相应的密度 ρ 或 ρ' 。

图18展示了参考模拟，也就是我们假设道路上实际发生的情况，也是理想情况下我们希望利用现有数据重现的情景：在初始时刻， x =0.52处立即出现稀疏波扇，道路右侧开始有车辆汇入。当足够多的车辆到达 $x = b$处的瓶颈时，便形成排队，并开始向上游回传。排队延伸至 s^k 处的传感器后，继续向上游回传。从瓶颈处开始，车辆以最大流量（$\rho = \sigma$）驶离，并正常行驶。

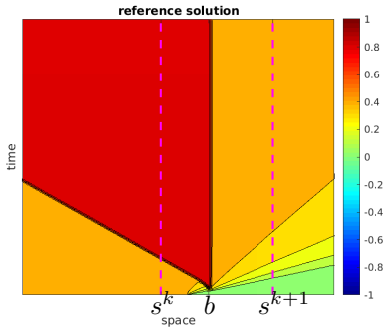


图18：学术测试。参考解

显然，如果只在传感器处观测交通流，我们就无法察觉事故及瓶颈在其实际位置的形成。只有当排队到达位置 $x = s^k$ 处的传感器时，事故的影响才会显现。

图19左图展示了采用基于密度的 方法所获得的结果。当在 $x = s^k$感知到排队时，记录到的流量$f_{s^k} = 0$ 经基本图反演后被正确转换为最大密度 $\rho = 1$ 。因此，排队继续向上游传播；而对于 $x > s^k$，交通流则以

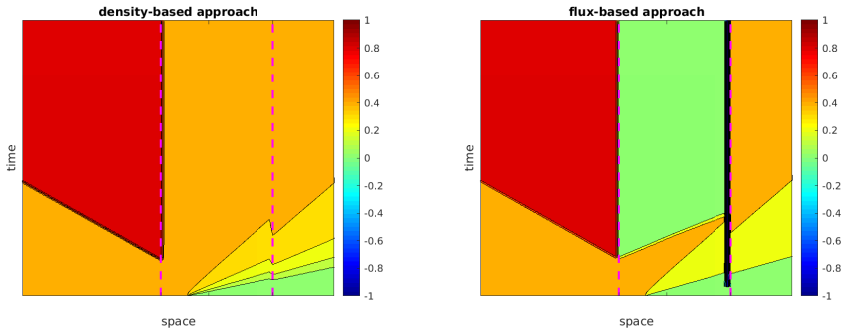


图19：学术测试。左：基于密度的方法。右：基于通量的方法

最大流量重新启动。在 $x = s^{k+1}$，无论采用哪种方法，得到的解都相同，交通流都以相同的动力学继续运行。

图19右侧展示了基于通量的方法所得到的结果。与之前一样，当排队到达 $x = s^k$ 时，传感器会在该处测得零流量，随后队列开始向上游回溯。由于通过传感器的流量为零，车辆无法继续驶离传感器，传感器之间的空间很快就会变空。此时，右侧传感器的通量数据与交通状况不相符（道路上空无一车，但传感器却检测到有车辆在移动）。因此，在 $x = s^{k+1}$处出现了负密度。对于 $x > s^{k+1}$，交通流动力学会利用测量流量正确恢复。

总之，在这种情况下，基于密度的方法更可取，因为它能更准确地捕捉真实情景。这是因为基于密度的方法实际上在模拟中加入了系统的其他信息，而不仅仅是原始通量数据，即对畅通／拥堵情景的区分。

6.1.2 实际场景

本节将考虑 Autovie Venete 提供的真实传感器数据，参见第 2 节。与前一次测试相比，主要区别在于，现在研究的是一条三车道高速公路和两类车辆（轻型和重型），其动力学相互耦合。通过一个偏微分方程组，将类似 LWR 的模型推广到这一情形。

$$\begin{cases} \partial_t \rho_{\mathrm{L}}^k + \partial_x f_{\mathrm{L}}(\rho_{\mathrm{L}}^k, \rho_{\mathrm{H}}^k) = 0, \quad x \in S_k, \quad t \in (t_0 - \Delta t_{\mathrm{PAST}}, t_0) \\ \partial_t \rho_{\mathrm{H}}^k + \partial_x f_{\mathrm{H}}(\rho_{\mathrm{L}}^k, \rho_{\mathrm{H}}^k) = 0, \quad x \in S_k, \quad t \in (t_0 - \Delta t_{\mathrm{PAST}}, t_0) \end{cases} \tag{14}$$

其中，ρ_{L}、f_{L} 和 ρ_{H}、f_{H} 分别表示轻型车辆和重型车辆的密度与流量。方程（14）补充了与（10）中相同的初始条件和边界条件。此外，两类车辆在道路上的行驶方式并不相同，因为重型车辆不得驶入最左侧车道。耦合

动力学及不均匀空间占用已在 [6]中推导，我们请读者参阅该论文，了解数学和数值方面的详细内容。这里仅回顾一下：由于系统存在两种状态，我们考虑一种相变 （参见 [11, 12, 15]），见图20：

- 当处于部分耦合相时，重型车辆会影响轻型车辆的动力学，反之则不会。此时，轻型车辆主要行驶在快车道上，而重型车辆不受它们影响。在这种情况下，系统 (14) 中的两个方程部分耦合，即 f_{H} 仅取决于ρ_{H},

$$\begin{cases} \partial_t \rho_{\mathrm{L}} + \partial_x f_{\mathrm{L}}(\rho_{\mathrm{L}}, \rho_{\mathrm{H}}) = 0, \\ \partial_t \rho_{\mathrm{H}} + \partial_x f_{\mathrm{H}}(\rho_{\mathrm{H}}) = 0. \end{cases}$$

- 当处于完全耦合相时，轻型车辆数量过多，无法仅容纳在快车道上，因此会进入慢车道，从而影响重型车辆的动力学。在这种情况下，两个方程完全耦合，符合系统 (14) 的一般形式。

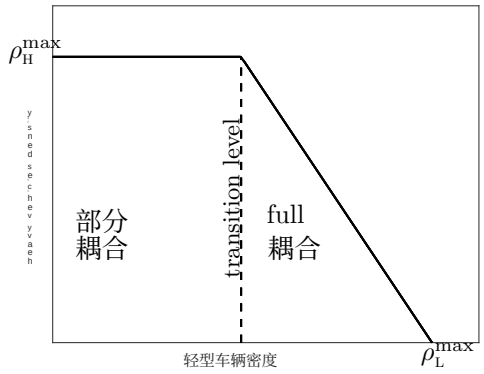


图20：模型（14）的两个阶段：轻型车辆的密度决定动力学是部分耦合还是完全耦合。

图21展示了基本图族，用于考虑在给定另一类别车辆密度的情况下，各车辆类别的流量-密度关系；更多详情请参见 [6]。

现在我们可以描述这项测试了。考虑一段长32 km、横跨两个路段的道路。在$x = 20.5$ km处的界面设有一个传感器。在 $x = 29.5$ km处，由车道数从3条减少到2条形成一个永久性瓶颈。

在真实场景中，Autovie Venete 的工作人员通过直接观察证实，瓶颈只会导致重型车辆拥堵；拥堵会向后传播，进而减慢轻型车辆的速度。与学术测试一样，只有当拥堵到达传感器时，模型才会感知到拥堵。

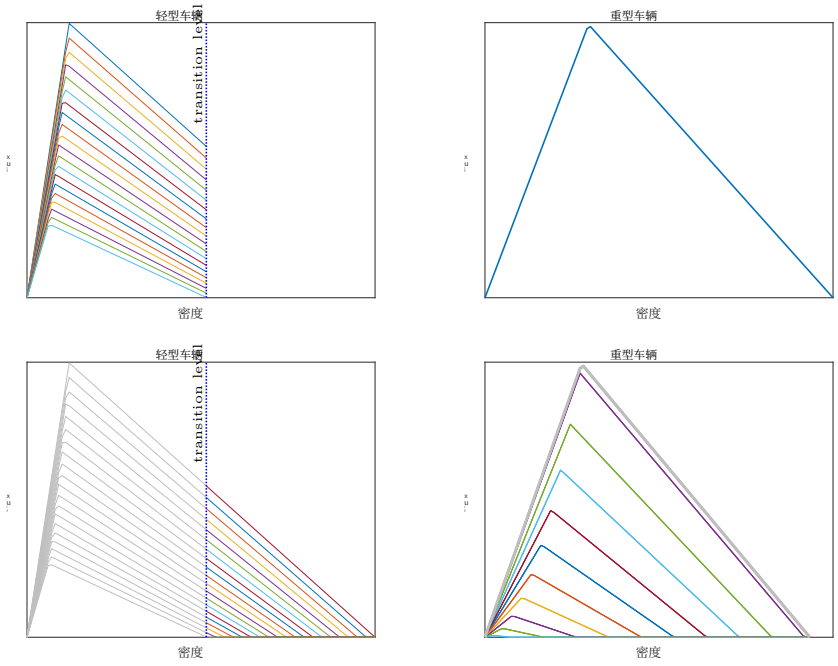


图21：两个阶段的基本图族。左：在部分耦合（上）和完全耦合（下）情况下，取 ρ_{H} 的若干值时的函数$\rho_{\mathrm{L}} \to f_{\mathrm{L}}(\rho_{\mathrm{L}},\ \rho_{\mathrm{H}})$。右：在部分耦合（上）和完全耦合（下）情况下，取 ρ_{L} 的若干值时的函数 $\rho_{\mathrm{H}} \to f_{\mathrm{H}}(\rho_{\mathrm{L}},\ \rho_{\mathrm{H}})$。

图22展示了采用基于通量和密度的方法时，轻型车辆和重型车辆的密度与速度。若采用基于通量的方法，重型车辆的拥堵路段会向传感器上游传播，并持续减慢轻型车辆的速度（其速度约为60公里/小时），而下游交通流畅通。若采用基于密度的方法，动力学则更为复杂，因为针对每个类别的修正 $\rho \to \rho'$ 都会改变另一类别的基本图，见图21：在下游，两类车辆都出现减速（重型车辆的速度约为20公里/小时，轻型车辆约为100公里/小时）；在上游，两类车辆都完全拥堵（速度为零）。同样，基于密度的方法与真实场景更加吻合，尤其是在传感器与瓶颈之间。

6.2 预测

在本节中，我们探讨估算到达交通流量对交通预测可能带来的优势。其思路是基于模型 (14)，从时间 t_0 到时间 $t_0 + \Delta t_{\mathrm{FUT}}$ 再次运行模拟器。与前一种情况不同，这里我们的目标是预测整条道路 $[x_{\mathrm{MIN}}, x_{\mathrm{MAX}}]$上的交通分布

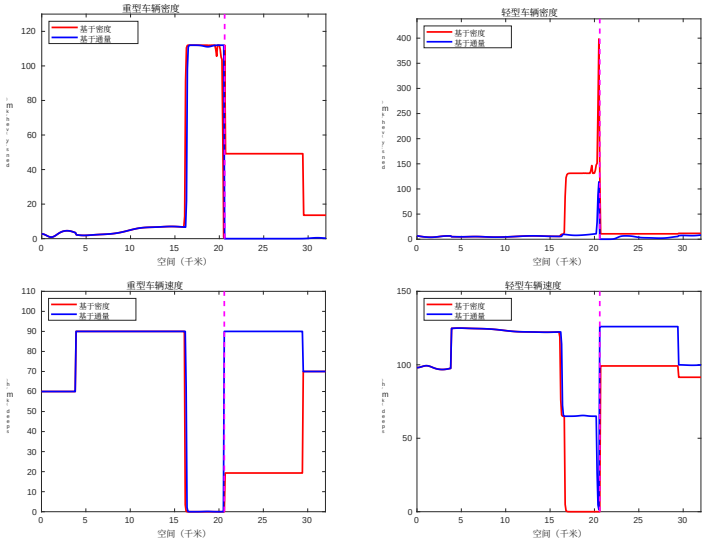


图22：使用真实数据进行测试。对于基于通量和密度的方法，在时间 $t = t_0$ 下轻型和重型车辆的密度和速度。

（传感器处没有中断），并以临近预报程序的结果作为密度的初始条件（ρ_{L}, ρ_{H}）。显然，这里无法再使用传感器数据，因为这些数据尚不可用，因此问题在于应采用哪些边界条件。理想情况下，应估算每个时间步 Δt的正确未来流入和流出，但考虑到交通流动力学的高度可变性，这极其困难。另一方面，最简单的方案是假设 $\rho_{\mathrm{IN}} = 0$ ，即假设没有车辆驶入道路，并假设 $\rho_{\mathrm{OUT}} = 0$ ，以确保流出量最大，但这会导致车辆从流入边界开始逐渐驶离道路。一种折中方案是将 $\rho_{\mathrm{OUT}} = 0$ 设为流出量最大，并保持流入量恒定，取最后可用的数据、最近几分钟的平均值，或由单独的程序预测出的某个值。这里我们采用第5节中构建的人工神经网络的结果，该网络可预测未来 30 分钟的交通流量，并将通量数据直接代入数值格式。

为了衡量模拟的误差，我们将任意未来时刻 $t_0 + \delta$,$\delta \in [0, \Delta t_{\mathrm{FUT}}]$ 的预测交通密度分布 $x \to \rho^F_{\{L,H\}}(x,\ t_0 + \delta)$ 与数据一旦可用（即时间 $t_0 + \delta$）便计算出的估计的临近预报 $x \to \rho^N_{\{L,H\}}(x,\ t_0 + \delta)$ 进行比较。交通密度之间的相对 L^1-距离按通常方式计算

$$E_{\{L,H\}}(t) := \frac{\|\rho^F_{\{L,H\}} - \rho^N_{\{L,H\}}\|_{L^1}}{\|\rho^N_{\{L,H\}}\|_{L^1}} \qquad \text{where} \qquad \|\rho\|_{L^1} := \int_{x_{\mathrm{MIN}}}^{x_{\mathrm{MAX}}} |\rho(x,t)|\, dx.$$

图23展示了轻型车辆和重型车辆的误差随时间变化的情况

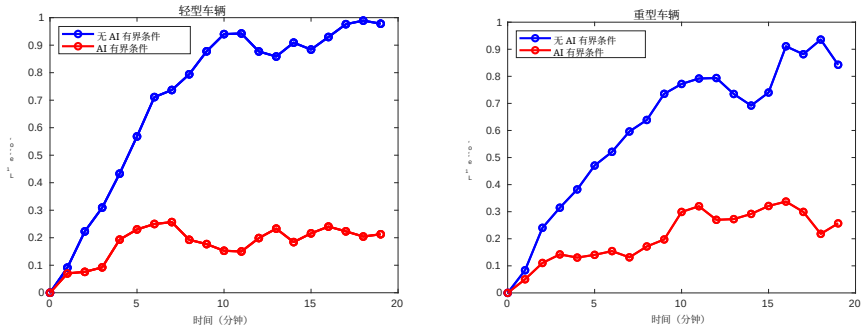


图23：人工神经网络生成的流入边界条件（红线）和零流入（蓝线）下，轻型车辆（左）和重型车辆（右）的归一化误差随时间的变化。

分别针对一段长16公里、设有两条车道的道路。我们比较了采用预测的入流通量值与简单假设零流入（最简单的选择）所产生的误差。值得注意的是，在人工神经网络辅助的交通预测中，误差起初有所增加，随后对轻型车辆和重型车辆均保持在约30%以下。这意味着，尽管交通分布相对于实际情况可能发生水平偏移，但总量（即车辆数量）并未出现过大的差异。相反，采用零流入时，误差会迅速增至100%，然后趋于稳定，这与预期一致。

7 结论

本文提出了一种混合模型驱动/数据驱动方法，用于重建和预测城郊公路及高速公路上的交通分布。与 [41, 23], 的思路类似，该方法使用人工神经网络作为真实数据与数学模型之间的连接环节，避免在人工神经网络的训练阶段使用数学模型。该方法能够利用机器学习的优势，从实时数据和历史数据中推断信息，再将处理后的信息传递给模型，使其能够立即加以利用。我们还发现，基于固定传感器测量数据的数据驱动方法很难推断任何类型的空间信息，也就是说，难以获取传感器之间交通流空间分布的信息。这正是我们认为数学模型对于交通状态估计至关重要的原因，因为它能够捕捉交通流在空间和时间上的正确因果关系。

致谢

作者谨向 Andrea Appella 和 Giulia Tatafiore 致以诚挚谢意，感谢他们在数据分析与理解方面提供的支持，以及为提升带标签数据集的质量所作的贡献。作者还要感谢 Paolo Ranut，他代表高速公路公司 Autovie Venete S.p.A. 为这项研究提供了支持。

资金支持

本研究是在“SMARTOUR：旅游智能平台”项目（编号：SCN 00166）框架内开展的，该项目由大学与研究部通过欧盟区域发展基金资助（PON研究与竞争力计划（2007—2013年））。

E.C.和M.B.还感谢意大利教育、大学与研究部（MIUR）为本研究提供支持，资助经费来自2017年PRIN项目（编号：2017KKJP4X，项目名称为“演化偏微分方程及其应用的创新数值方法”）。

本研究也部分由Autovie Venete S.p.A.资助。
所有作者均为意大利国家高等数学研究所（INdAM）数值计算科学组（GNCS）的成员。

参考文献

[1] A. Aw 和 M. Rascle. 交通流“二阶”模型的复兴。《*SIAM*应用数学杂志》, 60(3):916–938, 2000.[2] M. Barreau、M. Aguiar、J. Liu 和 K. H. Johansson. 用于交通密度识别与状态重构的物理信息驱动学习。载于*2021*年第*60*届*IEEE*决策与控制会议（*CDC*）, 第2653–2658页。IEEE, 2021.[3] N. Bellomo 和 C. Dogbe. 交通与人群建模：模型、推测与展望综述。《*SIAM*评论》, 53(3):409–463, 2011.[4] G. Bretti、M. Briani 和 E. Cristiani. 一种易于使用的网络交通流模拟算法：数值实验。《离散与连续动力系统系列*S*》, 7:379–394, 2014.[5] M. Briani 和 E. Cristiani. 一种易于使用的网络交通流模拟算法：理论研究。《网络与异质介质》, 9:519–552, 2014.[6] M. Briani、E. Cristiani 和 P. Ranut. 针对空间占用不均匀的多类别车辆动力学的宏观与多尺度模型：案例研究。《公理》, 10(2):102, 2021.

[7] 联邦公路研究院（BAST），贝吉施格拉德巴赫。公路路段检测站技术交付条件。技术报告，联邦交通、建筑与城市发展部（BMDV），2012年8月。[PDF],bast.de/tls.[8] G. Carlier 和 F. Santambrogio. 交通拥堵与Wardrop均衡的连续理论。《数学科学杂志》, 181(6):792–804, 2012.[9] W. Chen、F. Guo 和 F.-Y. Wang. 交通数据可视化综述。《*IEEE*智能交通系统汇刊》，16(6):2970–2984, 2015.[10] A. H. Chow、Y. Li 和 K. Gkiotsalitis. 动态交通建模中基本图的规范。《交通工程学报》, 141(9):04015015, 2015.[11] R. M. Colombo. 交通流中的双曲相变。《*SIAM J. Appl. Math.*》, 63(2):708–721, 2002.[12] R. M. Colombo、P. Goatin 和 B. Piccoli. 具有相变的道路网络。《*J. Hyperbolic Differ. Eq.*》, 7(1):85–106, 2010.[13] E. Cristiani 和 F. S. Priuli. 一种用于模拟网络交通流中Wardrop均衡的目的地保持模型。《网络与异质介质》, 10(4):857, 2015.[14] C. F. Daganzo. 元胞传输模型：一种符合流体动力学理论的高速公路交通动态表示方法。《交通研究 *B*：方法论》，28(4):269–287, 1994.[15] M. L. Delle Monache、K. Chi、Y. Chen、P. Goatin、K. Han、J. Qiu 和 B. Piccoli. 基于三维交通数据的三相基本图。《公理》, 10:17, 2021.[16] W. Du、Q. Zhang、Y. Chen 和 Z. Ye. 一种基于小波神经网络与改进鲸鱼优化算法的城市短期交通流预测模型。《可持续城市与社会》，69:102858, 2021.[17] S. Fan 和 D. B. Work. 一种包含缓行现象的异质多类别交通流模型。《*SIAM J. Appl. Math.*》, 75(2):813–835, 2015.[18] W. Fang、W. Zhuo、J. Yan、Y. Song、D. Jiang 和 T. Zhou. 注意力机制与长短期记忆网络的结合：一种用于交通流预测的深度学习网络。《物理学 *A*：统计力学及其应用》, 587:126485, 2022.[19] A. Ferrara、S. Sacone 和 S. Siri.《高速公路交通建模与控制》。施普林格，2018。

[20] M. Garavello 和 B. Piccoli。《网络交通流》。美国数学科学研究所，2006年。[21] Y. Gu、W. Lu、L. Qin、M. Li 和 Z. Shao。车道级交通速度短期预测：一种融合深度学习模型。《交通研究 C：新兴技术》，106:1–16，2019年。[22] D. Helbing。交通流及相关的自驱动多粒子系统。《现代物理评论》，73(4):1067，2001年。[23] M. Herty 和 N. Kolbe。交叉口交通流的数据驱动模型。预印本 arxiv:2212.08912，2022年。[24] Y.-L. Hsueh 和 Y.-R. Yang。一种基于长短期记忆网络的短期交通速度预测模型。《International Journal of Intelligent Transportation Systems Research》，19(3):510–524，2021年。[25] A. J. Huang 和 S. Agarwal。用于交通状态估计的物理信息深度学习：LWR 和 CTM模型示例。《IEEE智能交通系统开放期刊》，3:503–518，2022年。[26] F. Kessels。《交通流建模》。施普林格，2019年。[27] K. Lee、M. Eo、E. Jung、Y. Yoon 和 W. Rhee。基于深度神经网络的短期交通预测：综述。《IEEE Access》，9:54739–54756，2021年。[28] R. J. LeVeque。《守恒律数值方法》。Birkh¨auser，1992年。[29] M. J. Lighthill 和 G. B. Whitham。论运动学波 II：关于长距离拥堵道路交通流的理论。《伦敦皇家学会会刊A辑》，229:317–345，1955年。[30] S. Maerivoet 和 B. De Moor。道路交通的元胞自动机模型。《物理学报告》，419(1):1–64，2005年。[31] E. L. Manibardo、I. La˜na 和 J. Del Ser。道路交通预测的深度学习：它是否有所不同？《IEEE智能交通系统汇刊》，2021年。[32] A. Messner 和 M. Papageorgiou。METANET：高速公路网络宏观仿真程序。《交通工程与控制》，31(8- 9):466–470，1990年。[33] Z. Mo、R. Shi 和 X. Di。跟驰模型的物理信息深度学习范式。《交通研究 C：新兴技术》，130:103240，2021年。

[34] S. Modi、J. Bhattacharya 和 P. Basak。多步交通速度预测：一种采用潜在空间映射并考虑时空依赖关系的深度学习方法。《Expert Systems with Applications》，189:116140，2022年。[35] D. Ni、H. K. Hsieh 和 T. Jiang。将相图建模为随机过程及其在车辆交通流中的应用。《Appl. Math. Model.》，53:106–117，2018年。[36] H. J. Payne。高速公路交通与控制模型。《Mathematical Model of Public System》，第51–61页，1971年。[37] P. I. Richards。高速公路上的激波。《Oper. Res.》，4(1):42–51，1956年。[38] Z. (Sean) Qian、J. Li、X. Li、M. Zhang 和 H. Wang。异质交通流建模：一种务实的方法。Transp. Res. Part B，99:183–204，2017年。[39] T. Seo、A. M. Bayen、T. Kusakabe 和 Y. Asakura。高速公路交通状态估计：全面综述。Annual Reviews in Control，43:128–151，2017年。[40] R. Shi、Z. Mo 和 X. Di。用于交通状态估计的物理信息深度学习：由二阶交通模型提供信息的混合范式。载于*AAAI*人工智能会议论文集，第35卷，第540–547页，2021年。[41] R. Shi、Z. Mo、K. Huang、X. Di 和 Q. Du。用于交通状态与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》，23(8):11688–11698，2022年。[42] D. A. Tedjopurnomo、Z. Bao、B. Zheng、F. Choudhury 和 A. K. Qin。现代交通预测深度神经网络综述：趋势、方法与挑战。《IEEE Transactions on Knowledge and Data Engineering》，2020年。[43] B. T. Thodi、Z. S. Khan、S. E. Jabari 和 M. Men´endez。将运动波理论融入高分辨率交通速度估计的深度学习方法。《IEEE智能交通系统汇刊》，2022年。[44] F. van Wageningen-Kessels。评估多类别连续介质交通流模型的框架。Transp. Res. Rec.，2553(1):150–160，2016年。[45] F. van Wageningen-Kessels、H. Van Lint、K. Vuik 和 S. Hoogendoorn。交通流模型谱系。EURO Journal on Transportation and Logistics，4(4):445–473，2015年。

[46] H. Wang, D. Ni, Q.-Y. Chen, and J. Li. 平衡速度–密度关系的随机建模。《先进交通》，47:126–150，2013年。[47] K. Wang, C. Ma, Y. Qiao, X. Lu, W. Hao, and S. Dong. 一种基于1DCNN-LSTM-Attention 网络的混合深度学习模型，用于短期交通流预测。《物理学 *A*：统计力学及其应用》，583:126293，2021年。[48] S. Wang, J. Cao, and P. Yu. 时空数据挖掘的深度学习：综述。*IEEE*知识与数据工程汇刊，34(8):3681–3700，2022年。[49] Y. Wang, M. Zhao, X. Yu, Y. Hu, P. Zheng, W. Hua, L. Zhang, S. Hu, and J. Guo. 利用固定传感器和网联车辆对高速公路交通状态及模型参数进行实时联合估计：最新进展综述、方法与案例研究。《交通研究 *C*：新兴技术》，134:103444，2022年。[50] J. G. Wardrop. 道路交通研究的若干理论问题。土木工程师学会论文集，1(3):325–362，1952年。[51] G. B. Whitham. 线性波与非线性波。John Wiley & Sons，2011年。[52] W. Xiangxue, X. Lunhui, and C. Kaixun. 基于数据处理和长短期记忆网络-循环神经网络的城市道路网络交通数据驱动短期预测。《阿拉伯科学与工程杂志》，44(4):3043–3060，2019年。[53] J. Xing, W. Wu, Q. Cheng, and R. Liu. 多源数据融合下的城市道路网络交通状态估计：综述与新见解。《物理学 *A*：统计力学及其应用》，第127079页，2022年。[54] H. Yan, L. Fu, Y. Qi, D.-J. Yu, and Q. Ye. 用于短期交通流预测的鲁棒集成方法。《未来一代计算机系统》，133:395–410，2022年。[55] Y. Yuan, Z. Zhang, and X. T. Yang. 基于物理正则化高斯过程的宏观交通流建模：广义公式，预印本。https://arxiv.org/abs/2007.07762，2022年。[56] Y. Yuan, Z. Zhang, X. T. Yang, and S. Zhe. 基于物理正则化高斯过程的宏观交通流建模：对机器学习在交通领域应用的新见解。《交通研究 *B*：方法论》，146:88–110，2021年。[57] H. M. Zhang. 一种不具有类气体行为的非平衡交通模型。《交通研究 *B*：方法论》，36(3):275–290，2002年。[58] Z. Zhang, Y. Yuan, and X. Yang. 一种用于高速公路交通速度估计的混合机器学习方法。《交通研究记录》，2674(10):68–78，2020年。

[59] Z. Zhao, W. Chen, X. Wu, P. C. Chen, and J. Liu. 长短期记忆网络：一种用于短期交通预测的深度学习方法。*IET Intelligent TransportSystems*, 11(2):68–75，2017年。[60] G. Zheng, W. K. Chai, and V. Katos. 一种用于大规模道路网络交通速度预测的动态时空深度学习框架。*Expert Systems with Applications*，195:116585，2022年。