



GMDNet: A Graph-Based Mixture Density Network for Estimating Packages' Multimodal Travel Time Distribution

一种面向物流网络中包裹的多峰行程时间分布预测的
基于图的混合密度网络

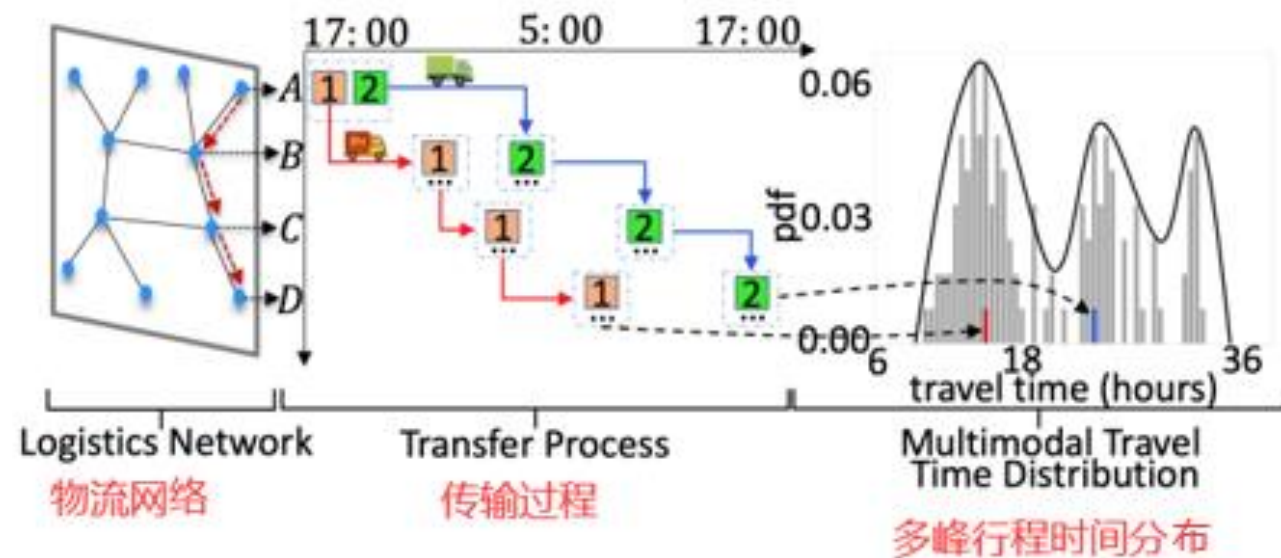
2023.6 《AAAI人工智能会议论文集》 DOI: 10.1609/aaai.v37i4.25578

汇报人：庞媛媛



1 研究背景：物流网络复杂VS包裹行程多样

- **物流平台关键任务：**估算包裹从起始节点到目的节点的路径行程时间分布TTD；
- **重要性：**消费者安排取货时间，减轻等待焦虑；物流平台提前制定调度计划；
- **面临挑战：**物流网络中复杂的空间关联和影响因素；包裹的行程时间分布是**多峰**的，即一条给定的路线可能不止一个输出。



给定物流网络路线的包裹MTTD



如何考虑物流网络的空间关联+影响因素+多峰行程时间分布？

2 研究内容

之前研究

- ❑ 基于链路的方法：不能很好地处理图结构数据和物流网络中的复杂相关性
- ❑ 公路网络中行程时间估算（TTE）：虽考虑到图结构数据但不能描述包裹的MTTD
- 拓展基于图的TTE方法的功能：突破输出仅限期望值或单一分布的局限；
- 提出基于图的混合密度网络GMDNet用于预测包裹的MTTD：集成图神经网络和混合密度网络优点+考虑物流网络的空间关联；训练采用期望最大化EM框架保证局部收敛，梯度下降更稳定；
- 两个真实物流数据集上实验，证明GMDNet性能优越；

3 相关研究

- **行程时间估算TTE**: 基于路段的方法: 效率高但无法纳入路线的上下文信息 (例: 交叉路口延误);
基于路线的方法
基于RNN的方法: 利用各种特征学习行程时间
基于图神经网络的方法: 捕捉网络中的时空依赖性

拓展基于图的TTE
方法的功能

上述方法的缺点: 输出仅限预期值或单峰分布

- **行程时间分布TTD**: 基于高斯马尔科夫随机场: 计算路径的行程时间分布
学习行程时间均值的模型: 建立方差与均值之间的关系推导行程时间分布
深度生成模型: 通过实时交通状况学习行程时间分布
基于深度图学习和GAN: 以分布形式估计行程时间

基于图的混合密度网络
+利用空间关联

上述方法的缺点: 没有探讨MTTD

- **多峰行程时间分布MTTD**: 基于高斯混合模型和马尔科夫链模型估计路网中MTTD
基于GAN估计路网中MTTD

上述方法的缺点: 通过链路对行程时间建模估计MTTD而不考虑整个网络的拓扑结构。

4 相关定义

定义1 物流网络

本质为有向图, $G=(\nu, \varepsilon, X, A, E)$ $\nu=\{v_1, \dots, v_N\}$ 每个节点对应一个物流实体（如商店、运转中心）， $\varepsilon=\{e_{ij} | v_i, v_j \in \nu\}$ 是边的集合, $X \in \mathbb{R}^{N \times d_v}$ 和 $E \in \mathbb{R}^{N \times N \times d_e}$ 分别是节点和边的特征, $A \in \mathbb{R}^{N \times N}$ 为相邻矩阵。

定义2 路径

元组 $R=(r, f)$ 表示, $r=[e_1, \dots, e_l]$ 是路径中的边序列, $f \in \mathbb{R}^{d_f}$ 是 r 中与路径相关的特征。

问题陈述

给定物流网络 G_t 和包裹在请求时间 t 的路线 R_t ，我们的目标是预测包裹的行程时间分布 $P(y_s | s)$ ， $s=(G_t, R_t)$ ， y_s 给定 s 的实际行程时间（即标签）。

5 GMDNet模型：根据输入路线和物流网络学习包裹的MTTD

□ 总体思路：MLE假设

为使模型具备多峰输出功能，利用混合密度网络 MDN 的来学习条件分布 $P(y_s | s)$ 。具体来说，通过**求解最大似然估计（MLE）来估计混合分量、权重**，将 K 个混合分量与混合权重 $\pi_s \in [0,1]^K$ 相结合，以产生混合高斯分布。给定一个假设空间 H ，我们寻求可以最大化似然 $\prod_{s \in D} P(y_s | s)$ 的最优 MLE 估计 h^* ：

$$\begin{aligned} h^* &= \arg \max_{h \in H} L(h | D) = \arg \max_{h \in H} \prod_{s \in D} P(y_s | s) \\ &= \arg \max_{h \in H} \prod_{s \in D} \sum_{k=1}^K P(y_s | \pi_s^k, s) P(\pi_s^k | s) \end{aligned} \quad (1)$$

混合密度网络MDN：一种用于建模复杂概率分布的神经网络架构，特别是涉及多模态（多个峰值）分布的情况。其输出是一个混合高斯分布，由多个高斯分量组成，每个高斯分量有3个参数：均值、方差（混合分量）和混合权重，混合权重表示表示每个高斯分量在混合分布中的比例，和为1。混合分量为每个高斯分量的特征。

最大似然估计MLE：从已知的数据中找到最有可能生成这些数据的参数值。

潜变量 π_s ：不能直接观测到但有用的变量（交通流量和拥堵情况、天气和环境因素、道路质量和施工情况等）。

边缘化引入潜变量：去除一个或多个变量的影响从而得到另一个变量的分布。

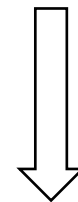
补充：
班级的学生成绩分布的模型

一个重要因素：学生的智力水平（无法观察）

引入潜变量I代表智力水平

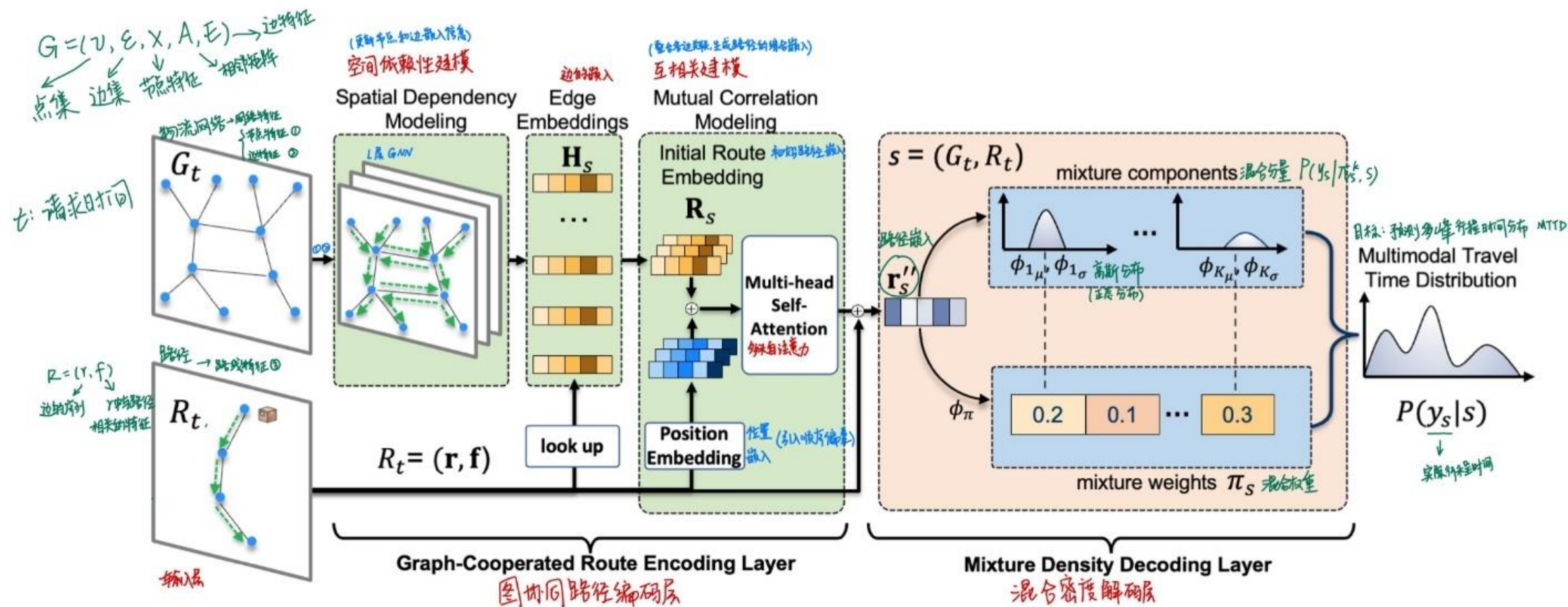
边缘化掉智力水平

$P(\text{成绩})$



$P(\text{成绩} | \text{智力水平})P(\text{智力水平})$

5 GMDNet模型：根据输入路线和物流网络学习包裹的MTTD



- 输入层：网络特征+路线特征
- 图协同路径编码层：空间依赖性建模+互相关建模 {输入s；输出路径嵌入 r''_s }
- 混合密度解码层 {输入路径嵌入；输出混合权重 π_s 和混合分量 y_s }
- 通过EM框架进行模型训练：训练+预测

5 GMDNet模型：根据输入路线和物流网络学习包裹的MTTD

□ 输入层：请求时间为 t ，输入物流网络和路线

①网络特征 Network Features

令 a_{ij} 为相邻矩阵 $A \in \mathbb{R}^{N \times N}$ 的第 (i, j) 项。如果包裹可以从节点 i 传输到节点 j ($i \neq j$)，则 $a_{ij} = 1$ ；如果 $i = j$ ，则 $a_{ij} = -1$ 。否则 $a_{ij} = 0$ 。

给定节点 $v_i \in V$ ，**节点特征向量** X_i 的公式为： $X_i = (x_i^{in}, x_i^{out}, weekday, hour)$ 。 x_i^{in} 是所有上游节点传入的包裹数量， x_i^{out} 是发送到所有下游节点的包裹数量。 $weekday$ 和 $hour$ 分别是星期几和一天中的几时。

给定时间 t 的一条边 $(i, j) \in \mathcal{E}$ ，则**边特征向量**为 $e_{ij} = (e_{ij}^{in}, e_{ij}^{out}, e_{ij}^{stay}, e_{ij}^{trans}, a_{ij}, weekday, hour)$ 。 $e_{ij}^{in} / e_{ij}^{out}$ 分别是**从节点 i 向节点 j 送入/送出的包裹数量**。 e_{ij}^{stay} 是从节点 i 到节点 j 的包裹在节点 i 的平均停留时间。 e_{ij}^{trans} 是从节点 i 到节点 j 的平均传输时间。 a_{ij} 是节点 i 和 j 之间的邻近度。

②路线特征 Route Features

路线 R_i 中的边序列记为 r 。**路线 R_i 的相关特征** f ： $f = (f^{out}, f^{trans}, weekday, hour)$ 。 f^{out} 是从起始节点发送到目的节点的包裹数量。 f^{trans} 是从起始节点到目的节点的平均行程时间。

上述特征中， $x_i^{in}, x_i^{out}, e_{ij}^{in}, e_{ij}^{out}, e_{ij}^{stay}, e_{ij}^{trans}, f^{out}, f^{trans}$ 是在 t 之前的给定时间窗口（本文为一天）内计算的。

5 GMDNet模型：根据输入路线和物流网络学习包裹的MTTD

□ 图协同路径编码层：对物流网络中的空间依赖性进行建模（得到点的嵌入和边的嵌入），并整合了路径中边之间的相互关系，以生成路径的综合表示（得到路径嵌入）。

■ 空间依赖性建模：得到各个节点和边的嵌入

step1 对给定节点和边特征进行线性变换得到节点和边的嵌入作为输入

step2 通过图神经网络对节点和边之间的空间依赖关系建模，每层更新节点和边的嵌入

节点嵌入和边的嵌入更新公式：
$$u_i^{l+1} = f(u_i^l, \text{Agg}\{u_i^l, u_{ij}^l : j \in N_i\}) \quad (2)$$

$$h_{ij}^{l+1} = g(h_{ij}^l, \text{Agg}\{h_{ij}^l, u_i^l, u_j^l\}) \quad (3)$$

$$u_i^{l+1} = u_i^l + \sigma_1(BN(W_1^l u_i^l + \sum_{j \in N_i} \sigma_2(h_{ij}^l) \odot W_2^l u_j^l)) \quad (4)$$

$$h_{ij}^{l+1} = h_{ij}^l + \sigma_1(BN(W_3^l h_{ij}^l + W_4^l u_i^l + W_5^l u_j^l)) \quad (5)$$

step3 得到空间相关性编码输出： U_s, H_s 节点的嵌入和边的嵌入

5 GMDNet模型：根据输入路线和物流网络学习包裹的MTTD

- 图协同路径编码层：对物流网络中的空间依赖性进行建模（得到点的嵌入和边的嵌入），并整合了路径中边之间的相互关系，以生成路径的综合表示（得到路径嵌入）。
- 互相关建模：整合路径中各边之间的相互关联，生成路线的综合嵌入。

step1 由边的嵌入堆叠得到初始路线的嵌入以便将顺序偏差纳入模型（路径的前后顺序关系）

step2 采用多头自注意力机制整合路线中各边的相互信息，得到更新的路径嵌入

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V}, \quad (6)$$

$$\text{MHSelfAttention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \oplus(\text{head}_1, \dots, \text{head}_h)\mathbf{W}^O, \quad (7)$$

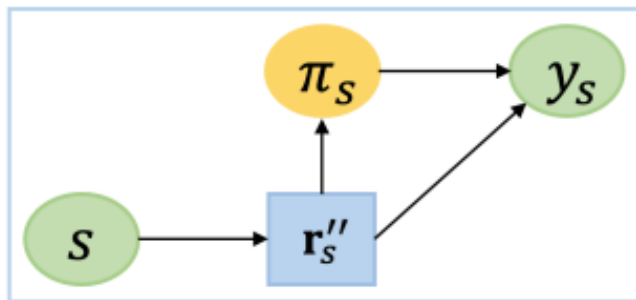
$$\text{head}_j = \text{Attention}(\mathbf{Q}\mathbf{W}_j^Q, \mathbf{K}\mathbf{W}_j^K, \mathbf{V}\mathbf{W}_j^V), \quad (8)$$

step3 与路径相关特征相连得到最终的路径嵌入

5 GMDNet模型：根据输入路线和物流网络学习包裹的MTTD

□ 混合密度解码层：根据路径嵌入对混合权重和混合分量进行建模

建模过程用叶贝斯网络表示



叶贝斯网络：建模随机变量之间条件依赖关系的图形化表示方法。

节点：随机变量；有向边：因果关系/条件概率关系。

绿色节点是观测到的随机变量，黄色节点是未观测到的节点，蓝色节点是确定性路径嵌入

■ 混合权重建模为具有K种可能状态的分类分布 $\sum_{k=1}^K P(\pi_s^k) = 1$

条件分布由确定性路径嵌入计算 $P(\pi_s | s) = \phi_{\pi}(\mathbf{r}_s'') = \sigma(f_{\pi}(\mathbf{r}_s'')), \quad (9)$

■ 混合分量的条件分布来自高斯分布族 $p(y_s | \pi_s^k, s) = \frac{1}{(2\pi)^{1/2} \sigma_k(s)} \exp\left\{-\frac{(y_s - \mu_k(s))^2}{2\sigma_k^2(s)}\right\}. \quad (10)$

■ 采用 $\phi_{k_{\mu}}, \phi_{k_{\sigma}}$ 网络根据路径嵌入输出混合高斯分布的参数 $\mu_k(s) = \phi_{k_{\mu}}(\mathbf{r}_s'') = f_{k_{\mu}}(\mathbf{r}_s''), \quad (11)$

$\sigma_k(s) = \phi_{k_{\sigma}}(\mathbf{r}_s'') = \exp(f_{k_{\sigma}}(\mathbf{r}_s'')), \quad (12)$

5 GMDNet模型：根据输入路线和物流网络学习包裹的MTTD

□ 通过EM框架进行模型训练预测：

■ 训练：采用期望最大化方法最大化似然估计中似然函数的对数 $\log L(h|\mathcal{D}) = \sum_s \log \sum_{k=1}^K P(y_s|\pi_s^k, s) P(\pi_s^k|s). \quad (13)$

step1 引入指示变量后通过Jensen不等式编写对数似然的下界 $\log L(h|\mathcal{D}) \geq \mathbb{E}_{\mathcal{Z}|\mathcal{D}}[\log L(h|\mathcal{D})]. \quad (14)$

step2 最大化对数似然的下界 $\mathbb{E}_{\mathcal{Z}|\mathcal{D}}[\log L(h|\mathcal{D})] = \sum_{s \in \mathcal{D}} \sum_{k=1}^K E[z_s^k|\mathcal{D}] P(y_s|\pi_s^k, s) P(\pi_s^k|s). \quad (15)$

step3 通过边缘化推导标准化项并计算指示变量的后验概率来执行EM算法的E-step

$$E[z_s^k|\mathcal{D}] = P(z_s^k = 1|s) = \frac{1}{Q} P(y_s|\pi_s^k, s) P(\pi_s^k|s), \quad (16) \quad Q = \sum_{k=1}^K P(y_s|\pi_s^k, s) P(\pi_s^k|s). \quad (17)$$

step4 避免单峰分布输出,在 $P(\pi_s|s)$ 上应用可选的狄利克雷正则化向 λ 得目标函数 $\mathcal{L}_{\text{obj}} = \mathbb{E}_{\mathcal{Z}|\mathcal{D}}[\log L(h|\mathcal{D})] + \sum_{s \in \mathcal{D}} \log \lambda(\pi_s|\alpha). \quad (18)$

step5 使用梯度下降实现M-step使方程18最大化，反复进行直至收敛获得最优的最大似然估计MLE假设

EM算法步骤： 初始化：选择初试的模型参数值

→E-step：计算隐函数数据的期望

→M-step：更新模型参数以最大化期望的对数似然

→重复迭代直至模型参数收敛或达到预设值

→终止

5 GMDNet模型：根据输入路线和物流网络学习包裹的MTTD

□ 通过EM框架进行模型训练预测：

■ 预测

Algorithm 1: Prediction by GMDNet. 通过GMDNet预测

Input: The graph and route of a package $s = (G_t, R_t)$ at t . 包裹 s 在 t 时刻的物流网络和路径

Output: Travel time distribution $P(y_s|s)$. 行程时间分布

- 1: // Graph-Cooperated Route Encoding Layer //图协同路径编码器
- 2: **for** $l = 1, \dots, L$ **do**
- 3: Update node embeddings by Equation 4; 通过方程4更新节点嵌入
- 4: Update edge embeddings by Equation 5; 通过方程5更新变的嵌入
- 5: **end for**
- 6: Obtain r_s'' according to Equation 6 - Equation 8; 通过方程6-8获得路径嵌入 r_s''
- 7: // Mixture Density Decoding Layer //混合密度解码层
- 8: Output mixture weights $P(\pi_s|s)$ by Equation 9; 通过方程9输出混合权重
- 9: Output parameters of mixture components $\mu_k(s), \sigma_k(s)$ by Equation 11 and Equation 12; 通过方程11、12输出混合分量的参数
- 10: $P(y_s|s) = \sum_{k=1}^K P(\pi_s^k|s) \times \mathcal{N}(\mu_k(s), \sigma_k(s))$; 得行程分布时间
- 11: **return** $P(y_s|s)$;

6 实验

□ **数据集：** 菜鸟网络中两个地区的真实物流数据集，2022.2.6~2022.3.8物流网络中包裹的行程信息。

Dataset		\mathcal{D}_1	\mathcal{D}_2
平均行程时间	Average travel time (hours)	15.8	16.1
	Average number of edges in a route	3	3
路径的平均边数	#training samples 训练样本数	358,822	492,831
	#validation samples 验证样本数	59,529	77,572
	#test samples 测试样本数	54,257	76,003

□ **基准：**

- History Average (HA)
- LightGBM
- Wide-Deep-Recurrent (WDR)
- 混合密度网络 (MDN)
- 核密度估计 (KDE)
- 宽-深-循环网络 (GCGTTE)
- 梯度下降训练的GMDNet (GMDNet-GD)

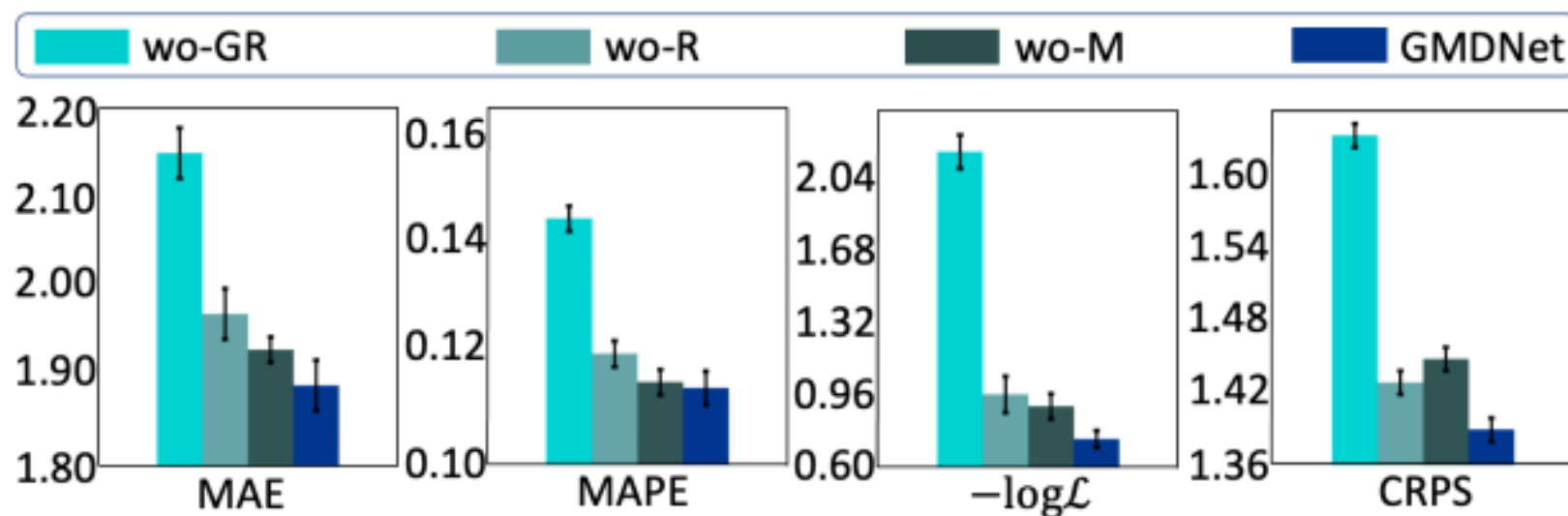
6 实验

- 评估指标：较大的 $\log L$ 和较小的MAE/MAPE/CRPS模型性能更好
- 平均绝对误差MAE和平均绝对百分比误差MAPE：衡量预测与标签之间的偏差程度
- 对数似然 $\log L$ 和连续排名概率分数CRPS：衡量预测与观测结果的匹配程度
- 实验结果：GMDNet在两个数据集上表现优于其他方法

Method	\mathcal{D}_1				\mathcal{D}_2			
	MAE(h)	MAPE(%)	$\log \mathcal{L}$	CRPS	MAE(h)	MAPE(%)	$\log \mathcal{L}$	CRPS
HA	3.75	27.2	-	-	3.57	25.5	-	-
LightGBM	2.06	13.6	-	-	1.81	12.1	-	-
WDR	2.10 ± 0.07	13.9 ± 0.56	-	-	1.85 ± 0.06	12.6 ± 0.83	-	-
MDN	2.15 ± 0.03	14.2 ± 0.20	-2.19 ± 0.08	1.64 ± 0.01	1.90 ± 0.04	13.3 ± 0.32	-1.61 ± 0.06	1.37 ± 0.01
KDE	2.04	12.6	-2.02	1.61	1.83	11.7	-1.41	1.35
GCGTTE	2.56 ± 0.13	16.2 ± 0.76	-4.27 ± 0.98	1.95 ± 0.18	2.47 ± 0.14	15.7 ± 0.81	-4.11 ± 0.85	1.65 ± 0.14
GMDNet-GD	1.92 ± 0.08	11.5 ± 0.75	-0.81 ± 0.16	1.43 ± 0.04	1.66 ± 0.08	11.0 ± 0.70	-0.99 ± 0.12	1.23 ± 0.03
GMDNet	1.89 ± 0.03	11.3 ± 0.28	-0.73 ± 0.02	1.39 ± 0.01	1.63 ± 0.02	10.8 ± 0.37	-0.94 ± 0.04	1.21 ± 0.01

6 实验

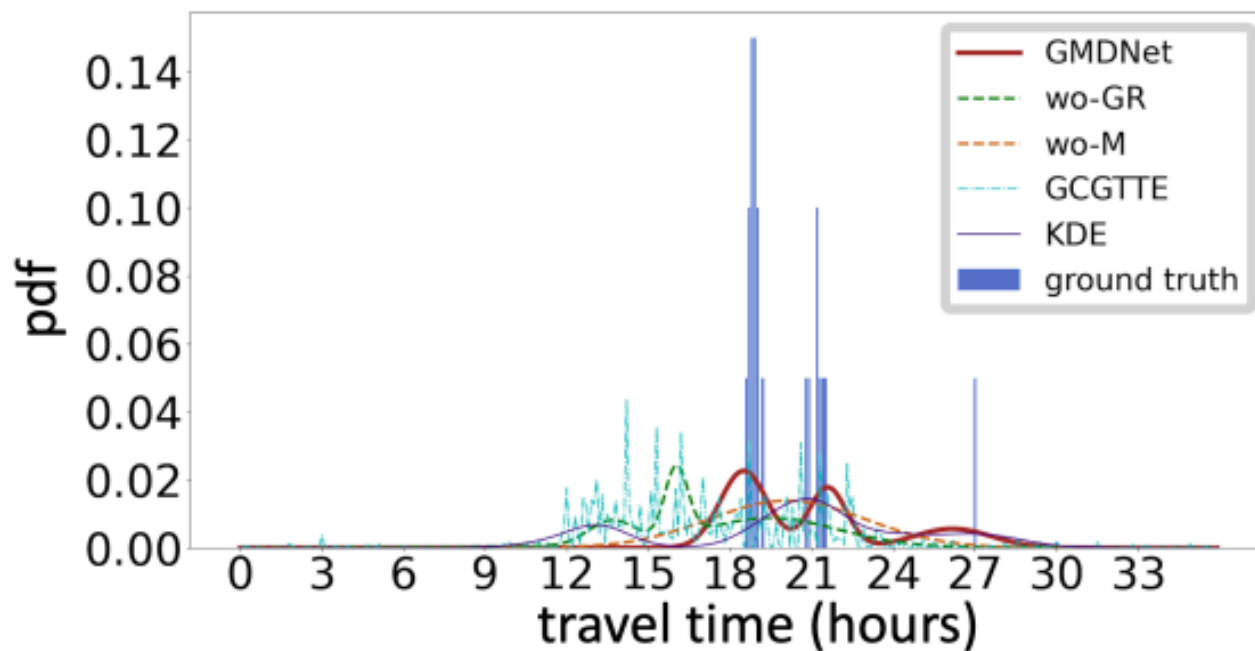
□ 成分分析：分析比较GMDNet的三个变体



- **Wo-GD:** 多层感知机代替图协同路径编码层→证明考虑物流网络中空间依赖性和相互关联性的必要
- **Wo-R:** 去除互相关建模模块→证明路径中整合边之间相互信息的重要性
- **Wo-M:** 混合分量数设置为1→证明对多峰行程时间分布的建模更近似真实分布情况

6 实验

□ 案例分析：绘制wo-GR/wo-M/GCGTTE/KDE/GMDNet/实际行程时间分布



- **Wo-GD和KDE:** 无法很好地近似包裹行程时间分布，原因：不了解图结构信息，无法建模空间依赖性；
- **Wo-M:** 产生单峰分布，占据实际分布的大部分，但无法贴近实际；
- **GCGTTE:** 未能贴近实际，原因：未考虑空间依赖性&GAN训练困难且不稳定；
- **GMDNet:** 能够学习物流网络中空间相关性和影响因素，贴近实际的多峰分布



7 结论

通过研究物流网络中包裹的MTTD

- 为研究以图结构数据为条件的MTTD学习提供极具吸引力的研究机会
- 提出一种基于图的混合密度网络GMDNet用于准确预测包裹的行程时间分布

其中，图协同路径编码层模拟物流网络中复杂空间的依赖性和路径中各边之间的互关联性

混合密度解码层输出多峰分布以扩展图的TTE功能

训练过程采用EM框架保证局部收敛和结果稳定

- 在真实数据集上证明了GMDNet有效性

谢谢！

