Proceedings of the 29th ACM SIGKDD Conference on
Knowledge Discovery and Data Mining. 2023

汇报人：刘泽华

Uncertainty-Aware Probabilistic Travel Time Prediction for On-Demand Ride-Hailing at DiDi

Hao Liu, Wenzhao Jiang, Shui Liu, Xi Chen

The Hong Kong University of Science and Technology (Guangzhou)
The Hong Kong University of Science and Technology (Guangzhou)
Didichuxing Co. Ltd
Didichuxing Co. Ltd

研究问题

➢问题描述

- ➢出行时间估计（Travel Time Estimation，TTE）
- ➢从源地址（origin）到目的地址（destination）的预期出行时间

➢重要性

- ➢DiDi每天处理非常多的TTE请求，响应质量直接影响顾客的用户体验和司机-乘客匹配的效率
- ➢路径规划、订单派送、价格动态调整

➢问题定义

- ➢路网-有向加权图，$G=(V,E)$, $v_i\in V, e_{ij}\in E$, v_i表示交叉路口（road intersection），e_{ij}表示v_i和v_j之间的路段（road segment）
- ➢路径：$r=\{e_1,\ldots,e_i,\ldots,e_k\}$
- ➢查询：四元组$q=(l_o,l_d,t,r)\rightarrow y_i$
- ➢定义：给定一次查询q，目的是预测$\hat{y}=F(q,G)$，其中F是从历史查询$Q=\{q_1,q_2,\ldots,\}$学到的函数

研究现状

➢现有方法

- ➢WDR[32]利用RNN提取时空特征的高阶信息
- ➢Google Maps[4]利用GNN建模路网的空间依赖性
- ➢ConSTGAT[6]提出一种3D图注意力机制捕捉空间和时间依赖性
- ➢……

➢存在问题

- ➢主要关注于提升预测精度，但忽视出行时间的**不确定性**
- ➢交通状况动态变化：突发事故引发的交通拥堵、临时交通管制
- ➢出行时间服从长尾分布（long-tail）
- ➢路段在不同交通状况下的出行时间可能差异很大

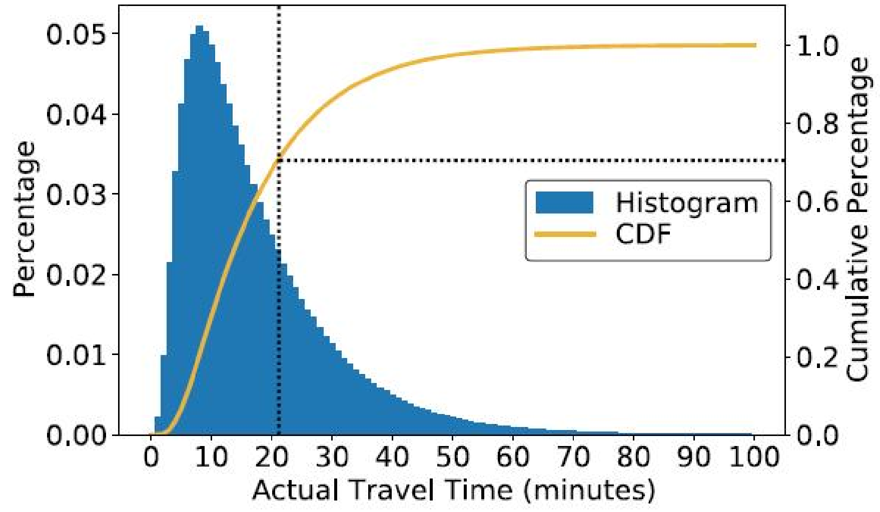


Figure 1: The long-tail travel time distribution of real-world historical trips in Beijing from May 2021 to June 2021 on DiDi's ride-hailing platform.

将TTE任务视为**确定性的端到端回归（deterministic end-to-end regression）**任务，即预测单标签值

解决思路

➢现有方法

- ➢WDR[32]利用RNN提取时空特征的高阶信息
- ➢Google Maps[4]利用GNN建模路网的空间依赖性
- ➢ConSTGAT[6]提出一种3D图注意力机制捕捉空间和时间依赖性
- ➢……

➢存在问题

- ➢主要关注于提升预测精度，但忽视出行时间的不确定性
- ➢交通状况动态变化：突发事故引发的交通拥堵、临时交通管制
- ➢出行时间服从长尾分布（long-tail）
- ➢路段在不同交通状况下的出行时间可能差异很大

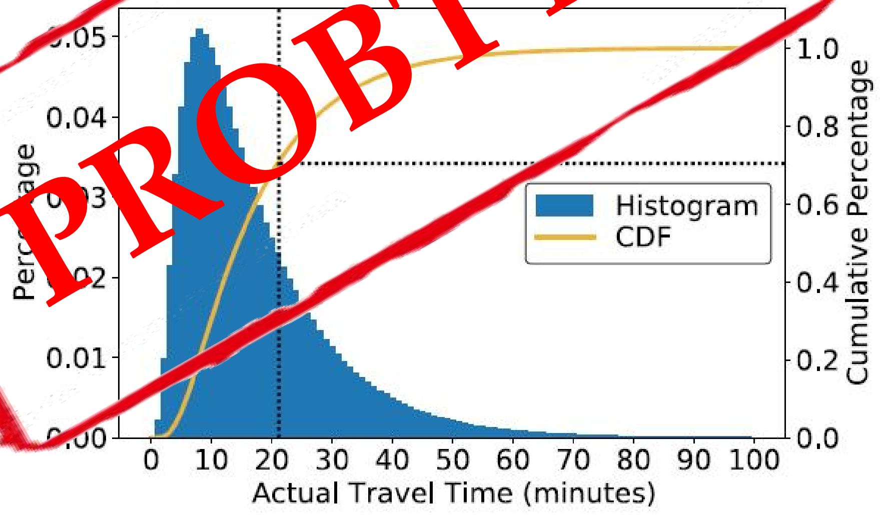


Figure 1: The long-tail travel time distribution of real-world historical trips in Beijing from May 2021 to June 2021 on DiDi's ride-hailing platform.

提出一种概率框架（probabilistic framework），感知TTE任务中的不确定性（uncertainty-aware）

面临挑战

➢未知的交通不确定性

- ➢Unobservable travel uncertainty：给定路段的交通时间高度依赖它的未来情况（比如交通堵塞，天气情况，突发事故）->全部未知
- ➢利用其他模型预测这些未知的未来交通状态->误差累积

➢随路段变化的分布

- ➢Route-varying distributions：不同路段的出行时间分布不同
- ➢高速路相比市中心道路的出行时间分布更加平稳（交叉路口和红绿灯更少）->怎样建模route-wise的出行时间分布

➢服务效率

- ➢Service utility：DiDi现有的TTE服务涉及复杂的基础基础设施流水线，从零开发概率TTE服务代价太大->考虑当前架构经济高效地实现升级

通过传感器收集到的实际交通时间是一个标量->单标签预测分布

Overview

➤损失函数（Joint Optimization）

- ➤$L = L_{reg} + \lambda_1 L_{cla} + \lambda_2 L_{exp}$, $L_{reg} = \frac{1}{|Q|}\sum_{i=1}^{|Q|} |y_i - \hat{y}_i|$

➤其余三个主要模块

- ➤ALLS, DTTE, RWPR

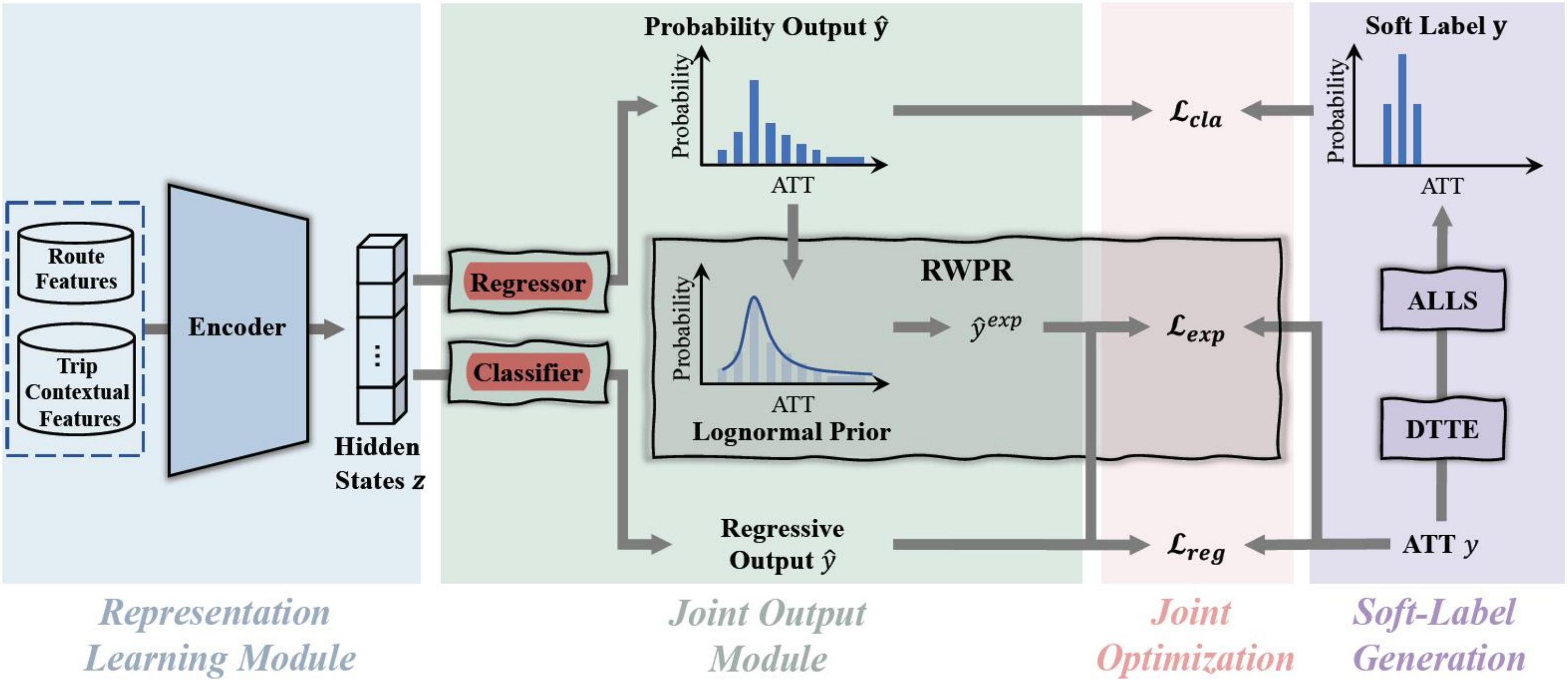


Figure 2: Overall framework of ProbTTE.

Overview

- ➢将单标签的回归预测问题转换为分类问题
  - ➢Distributional Travel Time Estimation, DTTE
  - ➢将连续的旅行时间离散化为各个类别
- ➢保留不同类别间的原始关系
  - ➢Adaptive Local Label Smoothing, ALLS
  - ➢在没有显式的分布信号下自动学习类间关系
- ➢结合先验知识
  - ➢Route-Wise Prior Regularization, RWPR
  - ➢从大规模历史数据得到先验分布知识进而指导优化

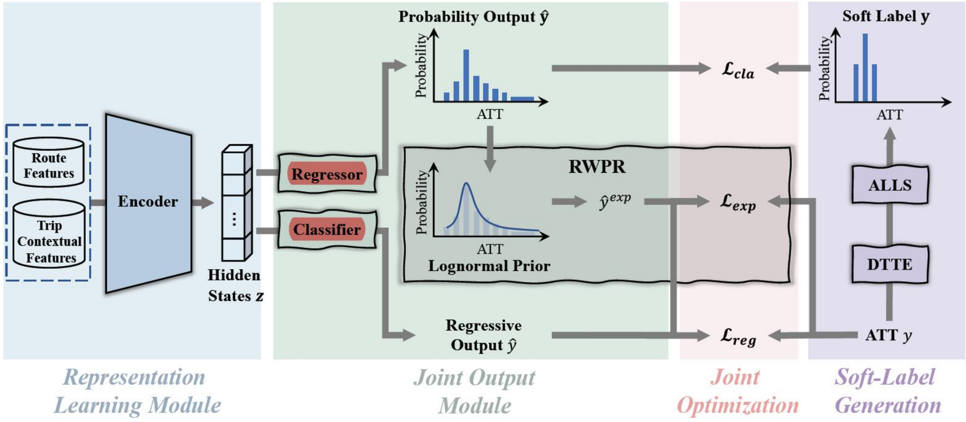


Figure 2: Overall framework of ProbTTE.

DTTE (Distributional Travel Time Encoding)

➢数据分析

- ➢数据呈长尾分布
- ➢同样的预测误差在短途旅行中比在长途旅行中用户体验更严重
- ➢10s in 10min vs. 10h

➢独热编码

- ➢根据实值的旅行时间分段
- ➢划分区间时不是采用等间隔，而是设置两种时间间隔（$\Delta t \ll \Delta T$）
- ➢区间$[b_j, b_{j+1}), 0 \le j \le M + N$

$$b_j = \begin{cases} j \cdot \Delta t, & \text{if } 0 \le j \le M \\ M \cdot \Delta t + j \cdot \Delta T, & \text{if } M+1 \le j \le N \\ +\infty, & \text{if } j = M+N+1 \end{cases}$$

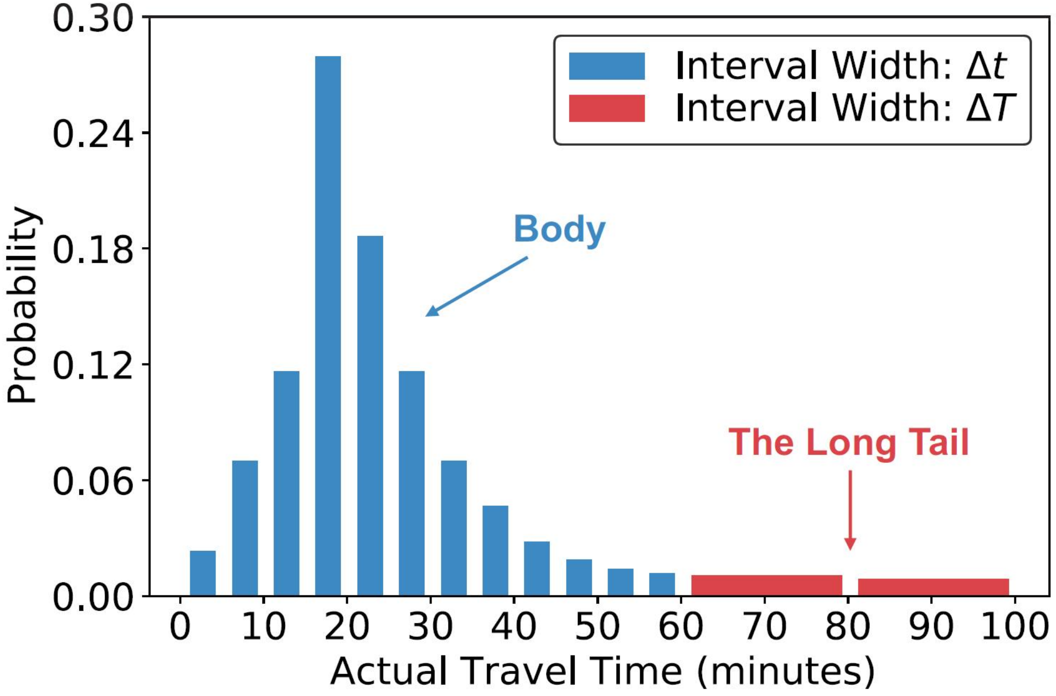
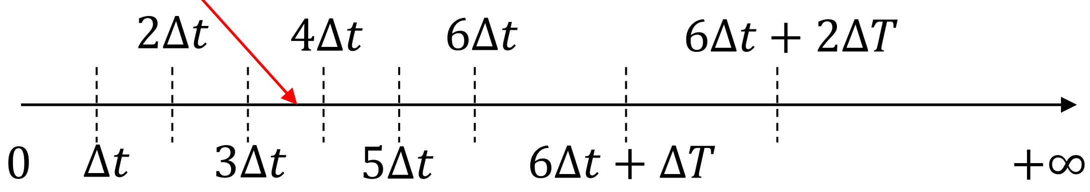


Figure 3: An illustrative example of travel time label discretization. DTTE encodes long-tail classes in more coarse-grained intervals, i.e., $\Delta t \ll \Delta T$.



ALLS, Adaptive Local Label Smoothing

- 独热编码?
  - Over-confident on the targeted class
- Label Smoothing
  - $1 \rightarrow 1-\epsilon, 0 \rightarrow \frac{\epsilon}{|C|-1}$
  - 将所有非target类视作同等重要性
  - DTTE类数量可能非常多，收敛难
- Adaptive Local Label Smoothing
  - 类关系有序 $\boldsymbol{p}, \boldsymbol{\tau}$

$$
\mathrm{y}[j]=\begin{cases} p, & \text{if } |j-c|=0, \\ \frac{1-p}{2\tau}, & \text{if } 1 \leq |j-c| \leq \tau, \\ 0, & \text{else}, \end{cases}
$$

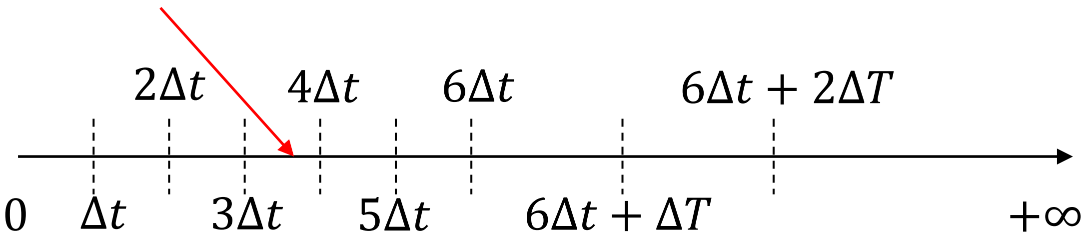
(0, 0, 0, 1, 0, 0, 0, 0, 0)

↓

(0.01, 0.01, 0.01, 0.92, 0.01, 0.01, 0.01, 0.01, 0.01)

↓

(0, 0.02, 0.02, 0.92, 0.02, 0.02, 0, 0, 0)



RWPR, Route-Wise Prior Regularization

➢对数高斯分布

- ➢ $y_i \sim p_i(y) = \frac{1}{y\sqrt{2\pi\sigma_i^2}} \exp(-\frac{(\log y - \mu_i)^2}{2\sigma_i^2})$
- ➢ $p_i[j] = p_i\left(\frac{b_j + b_{j+1}}{2}\right) \cdot (b_{j+1} - b_j), 0 \leq j \leq |C| - 2$
- ➢ $p_i[j] = p_i\left(\frac{2b_j + \Delta T}{2}\right) \cdot \Delta T, j = |C| - 1$
- ➢ $p_i \in R^{|C|} = (p_i[0], p_i[1], \ldots, p_i[|C| - 1])$

➢求解预测结果

- ➢ $\min_{\mu_i,\sigma_i} L_{prior} = -\hat{y}_i \cdot \log p_i$
- ➢ $\mu_i = \hat{y}_i \cdot \log b\,, \sigma_i^2 = \hat{y}_i \cdot \left(\log b - \mu_i \cdot 1_{|C|}\right)^2$
- ➢ $\hat{y}_i^{max} = \operatorname{argmax}_y p_i(y) = \exp(\mu_i - \sigma_i^2)$
- ➢ $\hat{y}_i^{exp} = E[y_i] = \exp(\mu_i + \frac{\sigma_i^2}{2})$

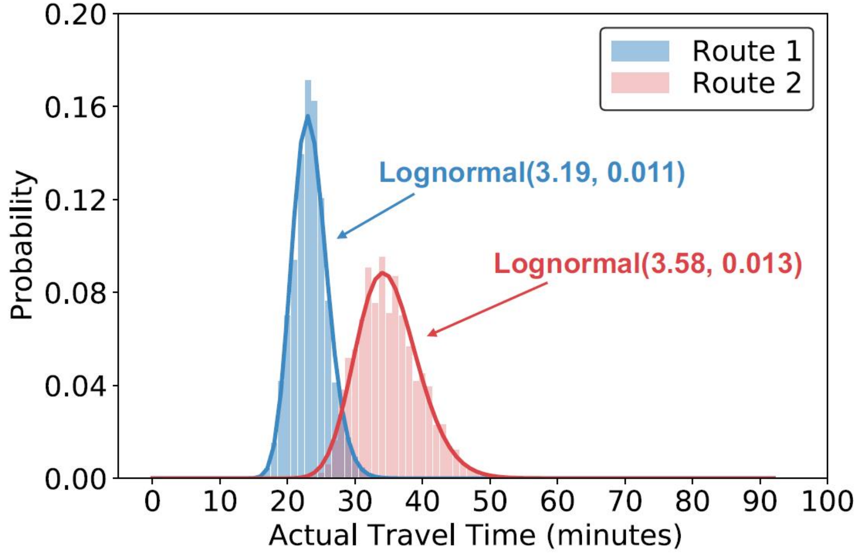
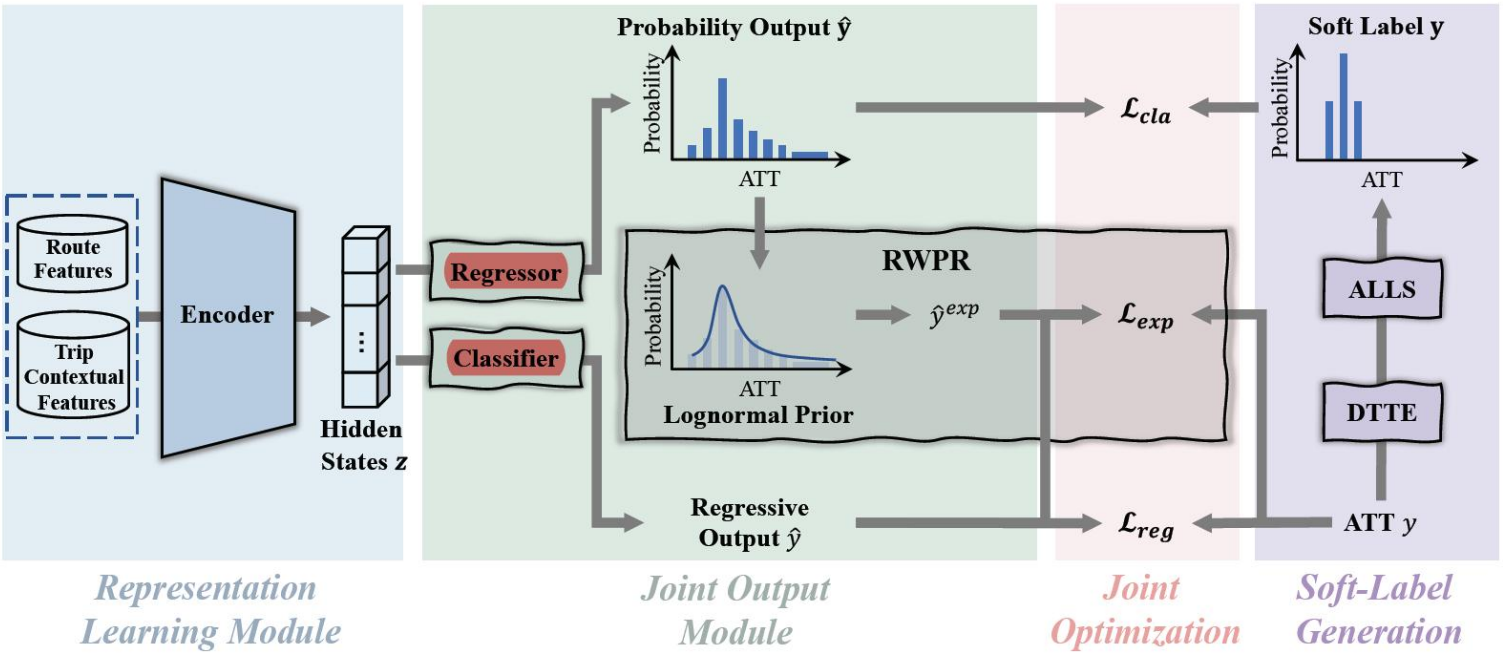


Figure 4: The probabilistic distribution of two different routes, which can be well fitted by two different log-normal prior distributions.

Overview

➢损失函数（Joint Optimization）

- ➢$L = L_{reg} + \lambda_1 L_{cla} + \lambda_2 L_{exp} \quad (\lambda_1 = 4 \times 10^4, \lambda_2 = 1)$
- ➢$L_{reg} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} |y_i - \hat{y}_i|$
- ➢$L_{cla} = -\frac{1}{|Q|} \sum_{i=1}^{|Q|} y_i \log \hat{y}_i$
- ➢$L_{exp} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} |\hat{y}_i^{exp} - y_i|$



实验验证

➢出行时间预测

➢Travel time prediction

➢$y_i = \lambda\hat{y}_i + (1-\lambda)\hat{y}_i^{exp}$

➢订单调度

➢Trip distribution score for order dispatching

➢利用PROBTTE反映出行不确定性信息

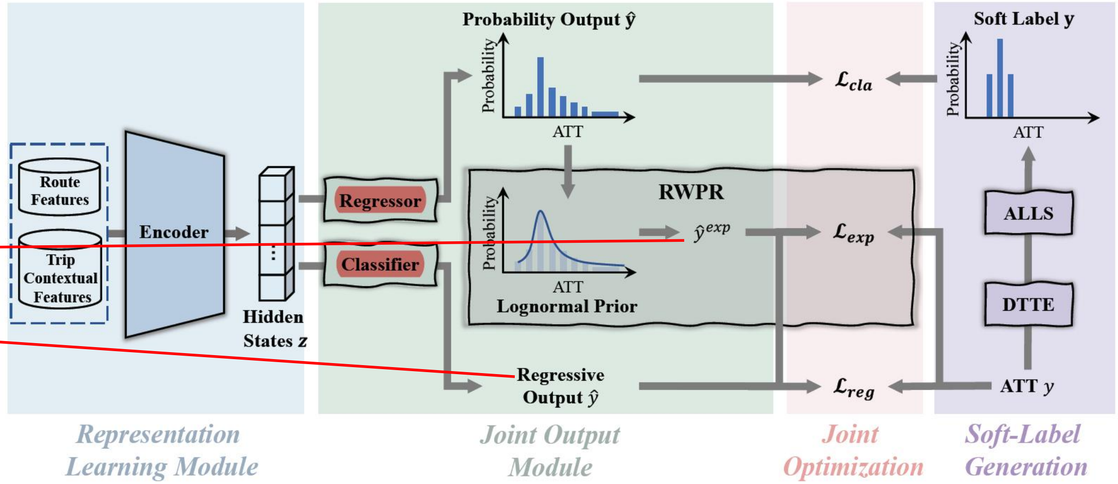
➢直观上，分布更集中在峰值附近，不确定性越小，分数越高

➢过滤出不确定性较高的潜在司机

➢Distribution score

➢给定概率向量$\hat{y} \in \mathrm{R}^{|C|}$和最大概率类m

➢$s(\hat{y}) = \sum_{0 \le i < j \le m} (\hat{y}[j] - \hat{y}[i])(j-i) + \sum_{m \le i < j \le |C|-1} (\hat{y}[i] - \hat{y}[j])(j-i)$



➢实验设置

➢数据集：北京，上海（约2个月）

➢指标：MAE, RMSE, MAPE, SR(Satisfaction Rate)；

➢BCR (Bad Case Rate)：实际出行时间超过估计时间限度的比例

➢Baseline：RouteETA, GBRT, WDR, HierETA

实验结果

➢出行时间预测

- ➢online和offline结果接近，因此只汇报offline结果
- ➢对于所有的基线模型均获得提升，且提升明显
- ➢PROBTTE对现有的WDR神经网络架构没有更改
- ➢GBRT优于RouteETA但弱于深度模型
- ➢上海的MAE和RMSE优于北京但MAPE和SR差于北京，北京的平均出行时间更长而上海的出行环境更复杂

Table 2: Overall performance of travel time prediction on Beijing and Shanghai datasets.

Model	Beijing				Shanghai			
	MAE (sec)↓	RMSE (sec)↓	MAPE↓	SR(%)↑	MAE (sec)↓	RMSE (sec)↓	MAPE↓	SR(%)↑
RouteETA	153.29	250.06	0.1512	43.07	137.52	203.88	0.1728	38.64
GBDT	140.23	230.70	0.1401	46.22	125.02	189.19	0.1557	42.13
WDR	113.52	191.92	0.1258	55.82	97.35	146.79	0.1284	52.16
HierETA	115.59	194.62	0.1146	55.14	98.92	151.27	0.1579	51.76
ProbTTE	**111.12**	**187.01**	**0.1111**	**56.48**	**96.04**	**145.09**	**0.1208**	**52.68**

实验结果

➢联合优化的有效性

- ➢所有的三个模块对最终性能都有贡献
- ➢PROBTTE-WoSR：移除ALLS和RWPR模块
- ➢PROBTTE-WLS：将ALLS模块替换为普通的平滑策略
- ➢PROBTTE-WoR：移除RWPR模块
- ➢PROBTTE-WND：将对数高斯分布替换为高斯分布

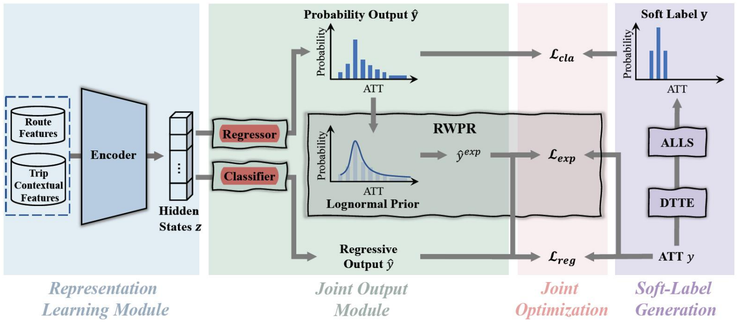


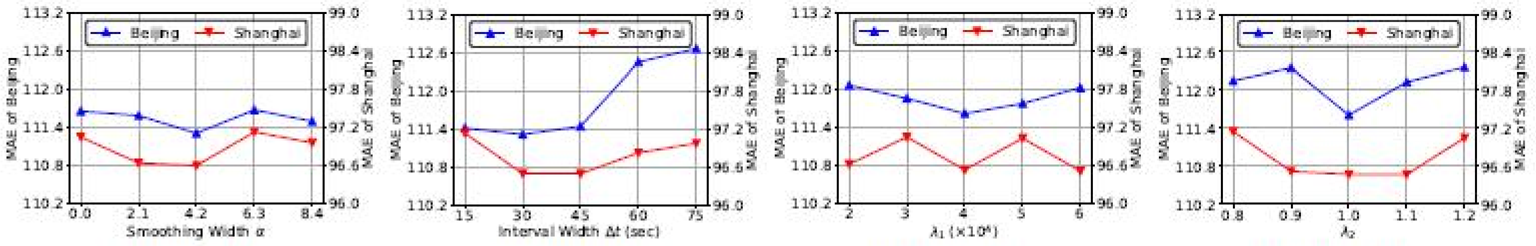
Table 3: Ablation study on Beijing and Shanghai datasets.

Model variants	Beijing				Shanghai			
	MAE (sec)↓	RMSE (sec)↓	MAPE↓	SR(%)↑	MAE (sec)↓	RMSE (sec)↓	MAPE↓	SR(%)↑
ProbTTE -*WoSR*	112.46	190.67	0.1367	56.24	96.44	146.6	0.1453	52.49
ProbTTE -*WLS*	112.16	189.76	0.1372	56.19	97.07	147.64	0.1482	52.23
ProbTTE -*WoR*	112.42	190.56	0.1367	56.20	96.75	147.59	0.1461	52.40
ProbTTE -*WND*	111.68	189.03	0.1369	56.32	96.33	146.42	0.1461	52.55
ProbTTE	**111.12**	**187.01**	**0.1111**	**56.48**	**96.04**	**145.09**	**0.1208**	**52.68**

实验结果

➢参数敏感性实验

➢调整超参后的效果变化在合理范围内



(a) Effect of α. (b) Effect of Δt. (c) Effect of λ_1. (d) Effect of λ_2.

Figure 5: Parameter sensitivity analysis on two datasets.

➢订单调度结果

➢WoDS: current order-dispatching without distribution score

➢WDS: optimized order-dispatching incorporating distribution score

➢在线A/B test，BCR下降，统计学意义

Table 4: Improvement on order dispatching services.

	Service I	Service II
WoDS	3.04%	4.08%
WDS	2.96%	3.7%
Improvement	2.8%	9.16%
P-value	0.03	0.01

Contributions

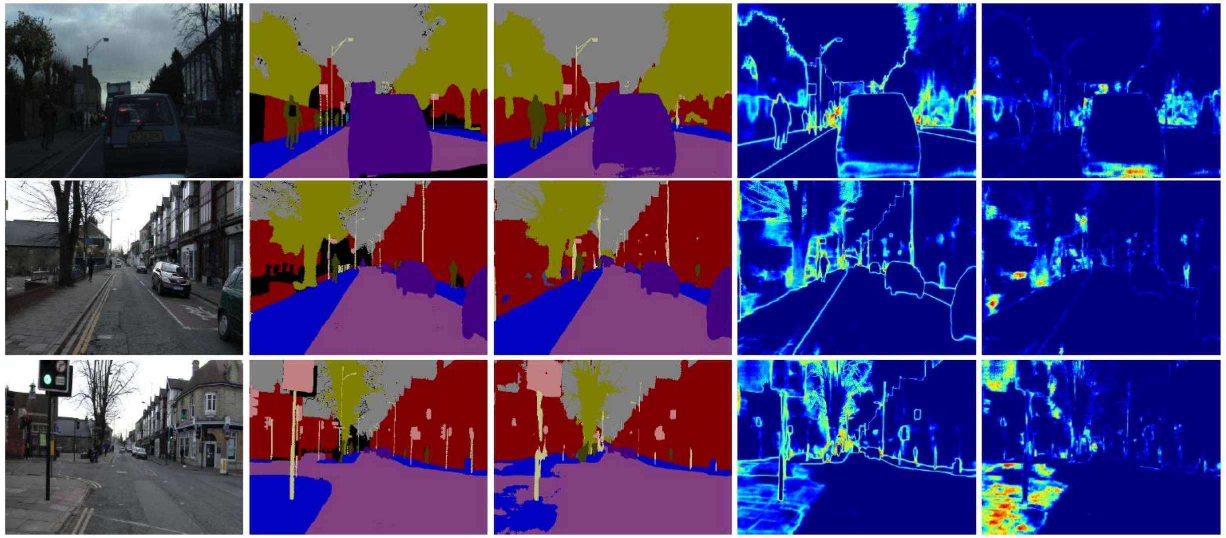
- 首次研究TTE的概率方法
  - The first attempt to investigate the probabilistic travel time estimation for the ride-hailing platform.
- 提出PROBTTE
  - We propose PROBTTE, a probabilistic framework to improve the accuracy and utility of the TTE service. PROBTTE is agnostic to the prediction model and can be easily integrated with existing TTE alternatives.
- 在DiDi平台部署PROBTTE
  - We deploy PROBTTE at DiDi's ride-hailing platform, which successfully improves the travel time prediction accuracy and order-dispatching quality.
- 实验结果证明方法有效性
  - Extensive experiments on the real-world TTE datasets and two order-patching services demonstrate the effectiveness of our proposed framework against five baselines and the previous order-dispatching strategy.

思考：贝叶斯神经网络+不确定性?

不确定性的分类

- **Aleatoric Uncertainty (AU)**
  - 随机不确定性
  - 捕捉数据固有的噪声
- **Epistemic Uncertainty (EU)**
  - 认知不确定性
  - 衡量模型参数的不确定性
- **Bayesian Framework**
  - 将两种不确定性统一

语义分割任务



(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

Figure 1: **Illustrating the difference between aleatoric and epistemic uncertainty** for semantic segmentation on the CamVid dataset [8]. *Aleatoric* uncertainty captures noise inherent in the observations. In (d) our model exhibits increased aleatoric uncertainty on object boundaries and for objects far from the camera. *Epistemic* uncertainty accounts for our ignorance about which model generated our collected data. This is a notably different measure of uncertainty and in (e) our model exhibits increased epistemic uncertainty for semantically and visually challenging pixels. The bottom row shows a failure case of the segmentation model when the model fails to segment the footpath due to increased epistemic uncertainty, but not aleatoric uncertainty.

[1] Kendall, Alex, and Yarin Gal. "What uncertainties do we need in Bayesian deep learning for computer vision?." Advances in neural information processing systems 30 (2017).

不确定性->交通任务 思路1：回归->分类 思路2：贝叶斯框架->分布

参考文献

[4] Austin Derrow-Pinion, Jennifer She, David Wong, Oliver Lange, Todd Hester, Luis Perez, Marc Nunkesser, Seongjae Lee, Xueying Guo, Brett Wiltshire, et al. 2021. Eta prediction with graph neural networks in google maps. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 3767–3776.

[6] Xiaomin Fang, Jizhou Huang, FanWang, Lingke Zeng, Haijin Liang, and Haifeng Wang. 2020. Constgat: Contextual spatial-temporal graph attention network for travel time estimation at baidu maps. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2697–2705.

[32] Zheng Wang, Kun Fu, and Jieping Ye. 2018. Learning to estimate the travel time. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 858–866.

Proceedings of the 29th ACM SIGKDD Conference on
Knowledge Discovery and Data Mining. 2023

汇报人：刘泽华

谢谢大家！

Hao Liu, Wenzhao Jiang, Shui Liu, Xi Chen

The Hong Kong University of Science and Technology (Guangzhou)
The Hong Kong University of Science and Technology (Guangzhou)
Didichuxing Co. Ltd
Didichuxing Co. Ltd