

# 论文技术方案

---

## DeepSD: Supply-Demand Prediction for Online Car-hailing Services using Deep Neural Networks

我们的目标是预测未来几分钟内某个地区的打车供需缺口。根据预测，我们可以通过提前调度司机来平衡供需。我们提出了一个端到端的框架称为深度供需(DeepSD)使用一种新的深度神经网络结构。我们的方法可以从打车服务数据中自动发现复杂的供求模式，而只需要最少量的手工制作的特征。此外，我们的框架具有高度的灵活性和可扩展性。

### 问题定义

问题假设当前日期是第 $d$ 天，当前时间段是 $t$ ，给定过去的订单数据和过去的环境数据，我们的目标是预测每个区域 $a$ 的供需缺口 $gap_a^{d,t}$ ，即未来10分钟的供需缺口。

### 已有解决方案

#### 运单数据的处理

基于手动设计的特征提取（比如特征工程），将结果输入到逻辑回归或者random forest。

存在的问题：特征工程通常需要大量人力，相关的原理解释较少

部分训练只基于某些属性（时间戳，起始位置），训练简化的同时造成信息缺失以及预测准确率降低

### 本文模型

我们提出了一个基于深度学习方法的端到端框架。我们的方法可以自动学习跨不同时空属性(例如地理位置、时间间隔)的模式，这允许我们在一个统一的模型中处理所有数据，而不是手动将其分成子模型。与其他现成的方法(如随机森林)相比，我们的模型需要最少量的特征工程(即手工制作的特征)，但产生更准确的预测结果。

我们设计了一种新颖的神经网络架构，允许人们将诸如天气和交通数据的“环境因素”数据非常容易地结合到我们的模型中。另一方面，我们可以轻松地利用订单数据中包含的多个属性，而不会丢失太多信息。

我们利用嵌入方法将高维特征映射到一个更小的子空间。在实验中，我们发现嵌入方法显著提高了预测精度。此外，通过嵌入，我们的模型还可以自动发现不同地区和时间段的供需模式之间的相似性。

我们进一步研究了模型的可扩展性。在实际应用中，将新的额外属性或数据源合并到已经训练好的模型中是非常常见的。通常，我们必须从头开始重新训练模型。然而，我们模型的剩余学习组件可以通过简单的微调策略来利用这些已经训练好的参数。在实验中，我们发现微调可以显著加快模型的收敛速度。

#### 前提

#### 嵌入方法

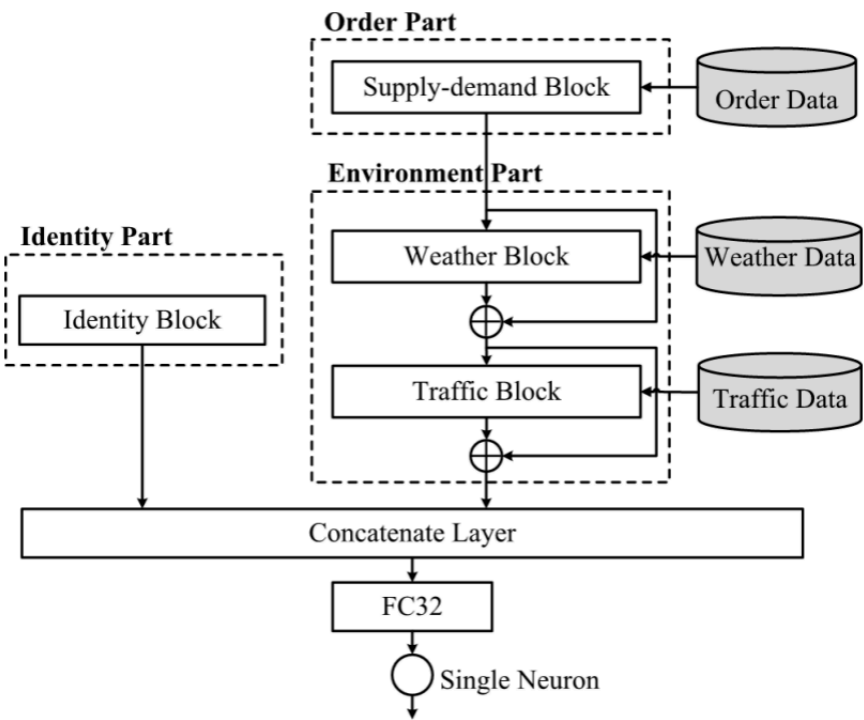
嵌入方法通过将每个分类值映射到低维空间。此外，嵌入方法的一个重要性质是语义相似的范畴值在嵌入空间中通常非常接近。例如，在我们的问题中，我们发现如果两个不同的区域共享相似的供需模式，那么它们的区域id在嵌入空间中是接近的。

残差网络

残差网络解决了训练非常深的模型在反向传播过程中梯度消失/爆炸问题

残差连接容易学习到恒等映射

基础版DeepSD



我们首先处理身份部分中的“身份特征”(区域ID、时隙、星期几)。

处理订单数据的订单部分。订单部分是我们模型中最重要的部分。

考虑[t-L, t]时间戳的订单集表示为 $S_{d,t}$ ,并聚合成一个实时供需向量

环境部分。环境部分处理天气数据和交通数据。

最后，我们说明了如何连接不同的块。

我们使用三个嵌入层分别嵌入ArealID、TimeID和WeekID。然后，我们通过一个连接层连接三个嵌入层的输出。

连接层将向量列表作为输入，并简单地输出向量的连接。

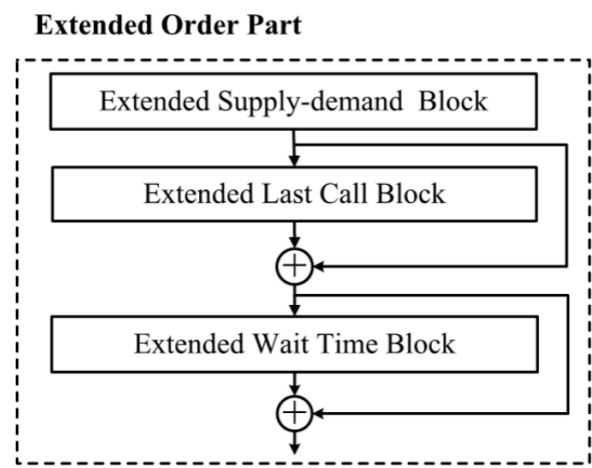
进阶版DeepSD

扩展订单部分代替了基础版的订单部分，扩展订单部分由三个块组成：第一块扩展供需块以设计良好的结构扩展了原有的供需块。后两块分别为最后呼叫块和等待时间块，它们与扩展供应-需求块具有相同的结构。与只使用订单数量的基本版本相比，新的块还包含乘客信息。

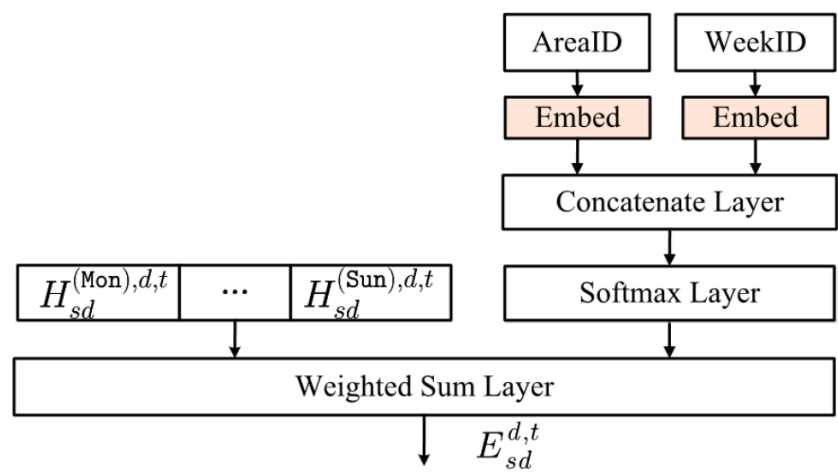
供需块分两个阶段。在第一阶段，我们在第d天获得时间区间[t-L, t]的经验供求向量。这种经验供求向量是一种

基于历史订单数据的估计。在第二阶段，我们使用实时供需向量和经验供需向量来构建我们的扩展供需区块。

扩展订单部分结构如图



历史供需向量结构如图



## 大宗货运场景的适用性

我们的模型中不包括weather block以及traffic block。

在我们的场景里，订单部分不需要包含乘客信息，也就没有最后呼叫和等待时间的概念。但是我们货车装载的货物有一个额外的装卸货等待时间需要考虑。可以替代进阶版本的等待时间。

## Deep Multi-View Spatial-Temporal Network for Taxi Demand Prediction

传统的需求预测方法大多依赖于时间序列预测技术，无法模拟复杂的非线性时空关系。深度学习的最新进展通过从大规模数据中学习复杂的特征和相关性，在传统的挑战性任务(如图像分类)上表现出优异的性能。这一突破启发了研究人员在交通预测问题上探索深度学习技术。然而，现有的交通预测方法仅仅独立地考虑了空间关系(如使用CNN)或时间关系(如使用LSTM)。我们提出了一个深度多视图时空网络(DMVST-Net)框架来建模空间和时间关系。具体来说，我们提出的模型由三个视图组成:时间视图(通过LSTM对未来需求值与最近时间点之间的相关性进行建模)、空间视图(通过本地CNN对本地空间相关性进行建模)和语义视图(对共享相似时间模式的区域之间的相关性进行建模)。在大规模真实出租车需求数据上的实验证明了该方法的有效性。

## 问题定义

出租车需求预测问题：如何通过历史出租车数据来预测一个地区在未来时间戳的出租车请求数量

## 已有解决方案

### 时间序列方法预测区域流量

利用ARIMA进行

在时间序列预测中考虑空间联系、外部语义信息（天气、规律性事件、地点）

共性问题：虽然考虑到不同因素的影响，但是未能捕捉复杂的非线性时空相关性

### 深度学习方法

可捕捉复杂的非线性时空相关性

#### LSTM

模拟序列相关性

复杂顺序交互作用的建模 解决RNN的梯度消失和梯度爆炸

#### CV

将城市交通看成图, 某时段流量为像素值

#### CNN

捕捉空间相关性，将整个城市的交通视为图像

对空间相关性建模 获取相邻空间信息

#### 神经网络

[Wanget al.\(2017\)](#)利用多种来源的上下文数据设计了一个神经网络框架，并预测出租车供需之间的缺口。

共性问题：没有同时考虑空间关系以及时间序列关系

与本文方法最大的不同

我们在联合深度学习模型中同时考虑了空间关系和时间序列关系。

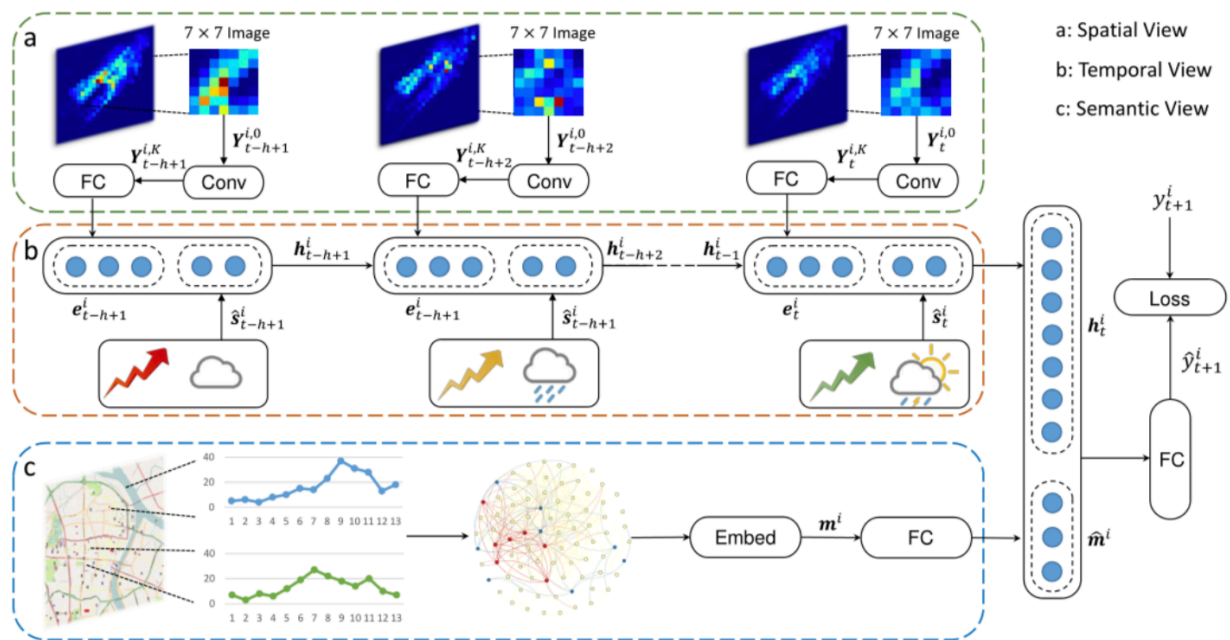
## 本文模型

我们提出了一个统一的多视图模型，该模型综合考虑了空间、时间和语义关系。联合利用CNN和LSTM来捕捉空间和时间的复杂非线性关系

我们提出了一个局部CNN模型，该模型捕捉区域相对于其邻居的局部特征，过滤弱相关的偏远地区。

我们基于需求模式的相似性构建了一个区域图，以便对相关但在空间上相距遥远的区域进行建模。通过图形嵌入捕捉区域的潜在语义。

模型框架



(一)空间视图

使用本地CNN来捕捉附近区域之间的空间依赖性。本地CNN包括几个卷积层。最后使用完全连接的层来进行低维表示。

(二)时间视图

采用LSTM模型，该模型从空间视图中提取表示，并将它们与相应时间的上下文特征连接起来。

(三)语义视图

首先构建区域的加权图(用权重表示功能相似性)。将节点编码成向量并使用全连接层进行联合训练。最后，使用全连接神经网络进行预测。

训练流程

---

**Algorithm 1:** Training Pipeline of DMVST-Net

---

**Input:** Historical observations:  $\mathcal{Y}_{1,\dots,t}^L$ ; Context features:  $\mathcal{E}_{t-h,\dots,t}^L$ ; Region structure graph  $G = (V, E, D)$ ; Length of the time period  $h$ ;  
**Output:** Learned DMVST-Net model

- 1 Initialization;
- 2 **for**  $\forall i \in L$  **do**
- 3     Use LINE on  $G$  and get the embedding result  $\mathbf{m}^i$ ;
- 4     **for**  $\forall t \in [h, T]$  **do**
- 5          $\mathcal{S}_{spa} = [\mathbf{Y}_{t-h+1}^i, \mathbf{Y}_{t-h+2}^i, \dots, \mathbf{Y}_t^i]$ ;
- 6          $\mathcal{S}_{cox} = [\mathbf{e}_{t-h+1}^i, \mathbf{e}_{t-h+2}^i, \dots, \mathbf{e}_t^i]$ ;
- 7         Append  $\langle \{\mathcal{S}_{spa}, \mathcal{S}_{cox}, \mathbf{m}^i\}, y_{t+1}^i \rangle$  to  $\Omega_{bt}$ ;
- 8     **end**
- 9 **end**
- 10 Initialize all learnable parameters  $\theta$  in DMVST-Net;
- 11 **repeat**
- 12     Randomly select a batch of instance  $\Omega_{bt}$  from  $\Omega$ ;
- 13     Optimize  $\theta$  by minimizing the loss function Eq. (9) with  $\Omega_{bt}$
- 14 **until** *stopping criteria is met*;

---

## 大宗货运场景的应用

与本文模型不同，我们的数据中不包含语义信息。我们需要综合考虑时间、空间维度。

### 潜在运力计算

基于订单数据得到历史运输过该流向的司机

基于流向空间数据得到运输过相关流向的司机（建立图结构，在其上得到不同司机流向的相似性）

### 任务完成时间预测

基于轨迹数据提取司机偏好（运输习惯、停留时间、停留次数等）以及客户偏好（收货时间（基于货车停留时间，找出常用的卸货时间））

## Learning to Estimate the Travel Time

本文提出了一种基于浮动车数据预测车辆行驶时间的新型机器学习解决方案。首先，我们基于大量有效特征将ETA表示为纯粹的时空回归问题。其次，我们调整不同的现有机器学习模型来解决回归问题。此外，我们提出了一种宽-深度递归（WDR）学习模型，以准确预测在给定出发时间沿给定路线的旅行时间。然后，我们联合训练宽线性模型，深度神经网络和递归神经网络，以充分利用这三种模型的优势。

## 问题定义

ETA：一对起点和目的地位置之间的旅行时间估计

准确的ETA将提高运输系统的效率，降低用户的旅行成本，节省能源消耗并减少机动车污染。

# 旅行时间估计的经典解决方案

## 基于路线的解决方案：

给定路线的总旅行时间被表示为通过每个路段的旅行时间和每个交叉路口的延迟时间的总和。

### 缺点：

- 1) 尽管可用数据量和数据源的多样性显著增加，但可用数据的空间 - 时间覆盖仍然非常稀疏。不足以监控整个道路网络的实时交通模式；
- 2) 运输系统是一个动态系统。很难以明确的形式对未来的移动模式进行物理建模。例如，当车辆在未来时间到达时，很难预测特定路段中的交通状况是什么以及特定交叉路口处的交通灯是什么。
- 3) 模型计算中将整个旅行时间分为几个部分。这可能导致估计误差的累积，并且整体估计精度下降；
- 4) 忽略了不同的驾驶员和骑车者的个性化特征对旅行时间的影响。

## 数据驱动的解决方案：

- 基于路线的解决方案中使用机器学习来预测每个路段的交通速度和行驶时间
- 基于其历史旅行直接预测未来时间段的整个路线的行驶时间

### 缺点：

- 1) 仍然存在不充分的数据覆盖问题。这种方法主要是在高速路上进行调查，路段很少，交通状况更稳定，数据覆盖更好；
- 2) 旅行时间预测限于几条固定路线。可扩展性受限；
- 3) 忽略了许多关键信息，例如交通信息和个性化信息，这使得它们无法获得高预测精度。

## 两种方法缺点总结

数据覆盖不足，泛化能力弱和信息使用不足的根本缺点限制了现有方法的有效性。

# 本文模型

## 特征提取

为基于位置的数据建立了丰富有效的特征系统，包括浮动车数据，道路网数据和用户行为信息。我们将这些特征概括为几种类型：空间信息，时间信息，交通信息，个性化信息和增强信息。

空间信息：首先将车辆轨迹映射到基础道路网络，并获得路段和交叉口的顺序。然后我们提取构成路线的所有构建块的特征，例如路段，交叉路口和交通灯信息。此外，在路线经过的区域内提取POI信息。

时间信息：交通网络中的交通状况对旅行时间有直接影响。

个性化信息：不同人有不同驾驶偏好。引入个性化信息，包括驾驶员档案，车手档案和车辆档案等。

增强信息：所有其他可用信息用作增强功能，包括天气信息和交通限制等。

## 制定回归问题

重新定义问题将 $y = [y_1, y_2, \dots, y_N]$ 表示为每个样本的地面实况标签，其中， $y = e_i - s_i \in \mathbb{R}^+$ 是已知的行程时间，其计算为到达时间 $e_i$ 和出发时间 $s_i$ 之间的时间间隔。

## 目标：最小化MAPE

$$\min_f \sum_{i=1}^N \frac{|y_i - f(x_i)|}{y_i} + \Omega(f)$$

其中 $f(x_i)$ 是路径 $x_i$ 的ETA，函数 $f$ 是回归模型。为了保证对看不见的数据的良好预测性能，引入额外的正则化术语来控制过度拟合来约束模型的复杂性。

## 解决方案

### 1. 梯度提升决策树 (GBDT)

$$\min_f \sum_{i=1}^N \frac{|y_i - f(x_i)|}{y_i} + \sum_t \left( \gamma L_t + \frac{1}{2} \|\mathbf{w}_t\|_2^2 \right)$$

这是一个凸优化问题。然而，目标是不平滑的，可以使用Huber损失来近似MAPE函数或采用次梯度方法来解决优化问题

### 2. 分解机器 (FM)

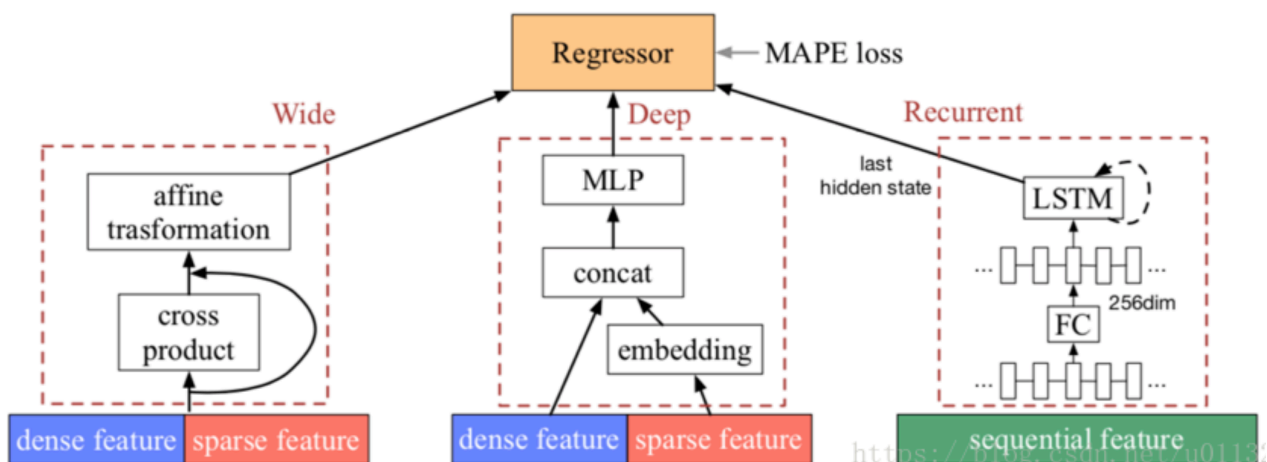
$$\min_{\mathbf{w}, \mathbf{V}} \sum_{i=1}^N \frac{|y_i - f(x_i)|}{y_i} + \lambda_1 \|\mathbf{w}\|_2 + \lambda_2 |\mathbf{w}| + \lambda_3 \|\mathbf{V}\|_F$$

通过梯度方法解决。我们遵循MF的在线优化算法之一，使用自适应次梯度法 (AdaGrad) 更新 $\mathbf{V}$ ，并使用FTRL更新 $\mathbf{w}$

### 3. 广泛深度复发学习 Wide-Deep-Recurrent Learning

GBDT很难适应大型功能集。FM的性能高度依赖于特征的表示，并且模型表示能力是有限的。

我们联合训练宽线性模型，深度神经网络和递归神经网络，以结合它们的好处。它平衡了一个模型中的记忆，泛化和表示能力，有效地缓解了现有方法的局限性。



宽模型首先将输入要素投影到高维特征空间中。通过计算输入要素的交叉积来实现。之后应用仿射变换  $y = \mathbf{w} \cdot \mathbf{x} + b$

深度模型首先使用要素嵌入层将稀疏输入要素转换为密集要素。该嵌入是通过紧凑的特征向量（在out设置中大小= 20）来表示高维稀疏特征的每个类别。然后，密集输入特征与嵌入特征连接，并馈入前馈神经网络。



LSTM可以捕获每个段的本地信息以及序列中的长期依赖性。因此，我们引入它来模拟我们问题中的路段序列。

顶层的回归量结合了宽模型和深模型以及LSTM的输出，以提供最终预测。

#### 4.GBDT，FM和WDR网络解决时空回归问题比较

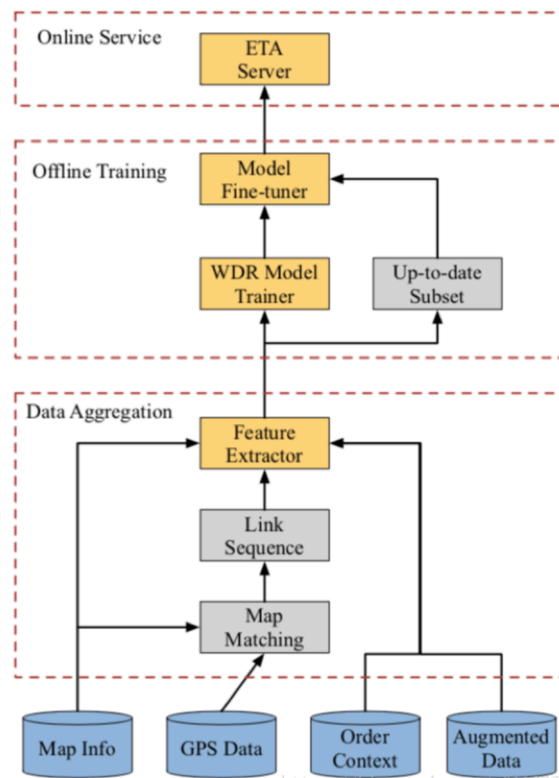
GBDT基于决策树将非线性变换引入到模型表示中。它可以被视为一个简单的深层模型。当可用的训练数据和特征集都受到限制时，GBDT是一个合理的选择。

FM可以看作是两个部分的模型组合：一个是线性模型部分，它是一个宽线性模型;另一部分代表非线性特征交互，它相当于一个两层MLP，也可以看作是一个简单的深层模型。因此，FM是一种简化的宽和深模型。当训练数据和特征集设置的较大时，可以使用它。

在WDR学习中，深度和宽度模型部分具有比FM更高的复杂度，如果可以使用足够的训练数据训练模型，则具有更好的表示能力。循环模型部分引入了更多表示每个路段的详细信息的能力。通过利用更多信息，WDR学习获得了更好的预测性能。此外，WDR模型还可应用于更一般的序列学习问题。

#### 解决方案整体架构

在数据聚合模块中，我们首先将GPS轨迹与道路网络匹配，以获得相应的路段序列。然后，特征提取器聚合道路网络信息，轨迹信息，订单上下文信息和增强信息，以产生用于训练模块的输入数据。在特征提取之后，我们基于大规模历史数据开始离线训练。然后，我们使用一组最新生成的数据进一步微调模型，以确保模型适应最新的数据分布。一旦完成微调，模型就会被推送到在线服务器。



模型输入：道路网络信息，轨迹信息，订单上下文信息和增强信息

模型输出：ETA预测结果

## 大宗货运场景下模型的适用性

该模型是基于滴滴出行的数据，floating-car data可以被分成两类，driver-pickup data和trip data。driver-pickup data是网约车司机接单之后去接乘客的那一段,而trip是真正从起始点送到目的地的那一段。

模型可以调整后适用于我们的场景。

在大宗货运场景中，我们计算的到厂时间为去程和返程的总时间，不需要考虑接单时间，不过需要考虑运输达到目的地之后的卸货时间，故而模型的输入数据需要调整。此外我们数据不包括天气信息还有交通限制这种数据，不存在增强数据这一块。

## 实验

### 评价指标

我们在实验中使用多个评估指标。对于离线实验，我们采用三个经典度量，包括平均绝对百分比误差（MAPE），平均绝对误差（MAE）和均方误差（MSE）来评估竞争方法。对于在线比较，我们使用四个评估指标，

包括MAPE，APE20，bad case rate和underestimate rate。最后三个指标在工业中广泛用于评估实时ETA服务的性能。它们的详细定义如下：

- APE20：绝对百分比误差（APE）小于20%的预测百分比（越高越好）。
- Bad case rate：预测值的百分比超过50%或绝对误差（AE）大于180秒，测量极端不良情况的百分比（越低越好）。
- underestimate rate：低估预测的百分比（越低越好）。

## Multi-task Representation Learning for Travel Time Estimation

根据起终点及出发时刻来预估到达时间（ETA）是智能交通系统中的一项关键的任务。当前大多数估算方法均假设行程的具体路线已经给定。但在实际应用中此假设往往不能成立，因为具体的路线常因交通状况、用户偏好等因素而动态地改变。此外，依靠起终点来推断行程的路线通常耗时较长且易于出错。因此，在实际应用中我们需要一种能够无需路线信息来预测到达时间的方法。该问题有着多方面的挑战性，比如非常有限的可用信息以及其复杂的时空依赖性。在本文中，我们提出了一种用于预估到达时间的多任务表征学习模型（MURAT）。该模型产生的特征能够保留现实世界中行程的各种属性，同时也利用了道路网络及时空先验知识。此外，我们进一步提出了通过多任务学习在训练阶段利用历史行程的路径信息来提高模型性能的方法。我们在两个大规模真实数据集上进行了实验，结果表明本文提出的方法较现有技术有着显著地提高。

## 已有解决方案

### 旅行时间估计问题

#### 基于路径

需要路线信息来生成预测

法一：估计单个链路上的行程时间，然后对给定路径上所有链路行程时间求和。用环路检测器数据或者浮动车数据来估计每条链路的行驶时间

缺点：未考虑路段之间的过渡时间

法二：考虑花在交叉口的时间，并连接子路径已实现更精确传播时间估计

首先挖掘频繁模式来进一步改进这种基于子路径方法，然后找到连接频繁路径的最佳方式，平衡长度和支持度。

主要缺点：需要估计路径，既耗时又容易出错。

## 基于起止

在没有路线信息的情况下预测行驶时间进行起点-终点时间估计

在OD旅行时间估计问题中存在几个主要挑战。首先，由于缺少路径信息，只有有限的原始输入特征，即起点、终点和出发时间。此外，这些原始特征通常难以用于模型，例如，基于两条道路的纬度/经度来测量两条道路之间的相似性或道路网络距离并非无关紧要。因此，为了处理由原始特征提供的信息的不足，需要学习一种利用来自基础道路网络结构和问题的时空属性的丰富信息的特征表示。

现有的工作都是直接使用原始特征来构建模型，对旅行时间估计，关注非常有限。

法一：基于最近邻的方法，该方法通过对具有相似起点和终点的所有历史行程的缩放行程时间进行平均来估计行程持续时间。然而，这种非参数方法很难推广到没有邻居可用或邻居数量非常有限的情况。

法二：ST-NN，一种用于行程时间估计的多层前馈神经网络。ST-NN首先将出发地和目的地的离散纬度和经度作为输入来预测旅行距离，然后将该预测与时间信息相结合来估计旅行持续时间。

共同的局限性：基本的道路网络结构和时空属性在很大程度上被忽略

## 问题描述

问题1 (OD 行程时间估计)。给定出发地、目的地和出发时间，我们的目标是使用历史出行数据集 $X$ 以及基础道路网络 $g$ 来估计持续时间。

挑战：

- 1) 基础道路网络中复杂的时空依赖性
- 2) 进行在线预测时可用的信息量有限。

## 本文模型

我们提出了一个新的表示学习框架MURAT来解决起点-终点旅行时间估计问题。利用拓扑信息以及时空先验知识。

在处理输入信息不足之前，我们提出了利用基础道路网络结构以及时空平滑度的方法。我们提出了一种多任务学习方法，在训练阶段利用路径信息来学习更有意义的表示，从而提高学习性能。

## 先验知识

空间和时间的平滑度。连续的或空间上邻近的观测值往往与相似的值相关联。本文模型在表示链接和时空因子之前，使用无监督的图嵌入和图拉普拉斯正则化。

跨任务的共享因素。许多感兴趣的任务是由与其他任务共享的因素来解释的。在本文中，我们通过联合学习行驶时间估计以及许多其他相关的现实世界任务来利用这一先验知识，这些任务可能具有共同的因素，例如，预测行驶距离和路径中路段的数量。

## 道路网络的表示学习

### 将基础道路网络结构纳入模型的方法

出行时间受到基础道路网络结构的强烈影响。例如，高速公路相反方向上的两个邻近位置对于同一目的地可能具有显著不同的行驶时间。这是因为基础道路网络施加的限制，阻止了车辆在欧几里得空间中自由移动。因此，希望将道路网络结构结合到学习过程中。

法一：识别位置的相应链接。直觉上，同一链路上的邻近位置对于同一目的地将具有相似的行程时间。

一种表示链接信息的潜在方式是将链接标识符作为数字特征或one-hot编码分类特征来提供。然而，这些方法不能捕捉网络结构，因为每个链接是独立定义和学习的。

一种替代方法是使用无监督的图形嵌入方法，例如拉普拉斯特征图，DeepWalk，以学习每个链接的表示。虽然这些表示保留了图上定义的相似性，但是它们很难利用任务特定的监督信号。

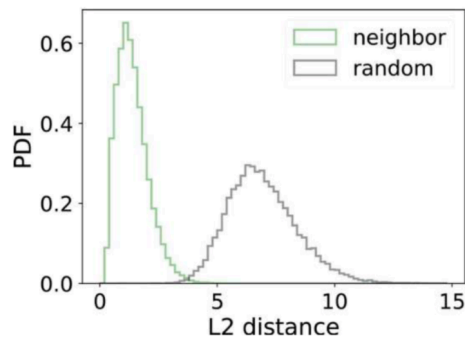
### 图拉普拉斯正则化

使用监督损失和非监督损失的组合作为目标。添加额外的图拉普拉斯算子作为无监督损失。

### 无监督的预训练

具体来说，我们首先学习基于无监督图嵌入技术的每个链接的表示，然后我们使用学习到的表示作为链接嵌入的初始化，稍后将基于监督信号对其进行微调。

该方法会有帮助的原因可以从正则化效应和优化效应的角度解释。



如图显示了基于DeepWalk嵌入计算的分布。我们观察到相邻链接嵌入之间的L2距离比两个随机链接之间的距离小得多。我们假设这种初始化具有与图拉普拉斯正则化子相似的效果。

## 时空表示学习

### 将时空先验知识整合到表示学习中

除了道路网络结构之外，学习表示的另一个重要方面是我们对空间和时间域的先验知识，例如时空平滑度、交通的周期性和重现性。我们通过在嵌入空间中构造空间图和时间图来整合这些先验知识。

在空间和时间图中，每个节点作为固定长度的向量嵌入，然后通过对这些嵌入向量应用图拉普拉斯正则化来增强时空平滑先验和周期性。

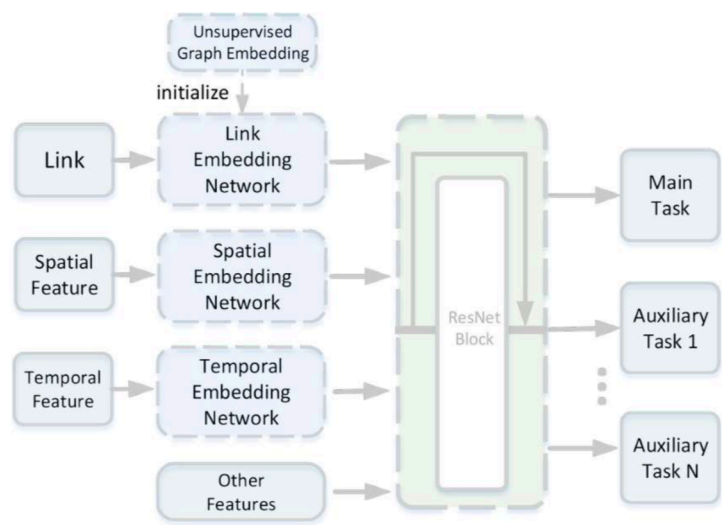
此外，我们使用空间和时间嵌入的分布式表示。对于空间表示，节点(即空间网格)的嵌入被表示为其纬度和经度的嵌入的连接，而对于时间表示，节点(即时间间隔)被表示为“一天中的时间”和“一周中的每一天”的嵌入的连接。

与分别表示每个位置/时间间隔相比，这种分布式表示具有以下主要优点:

- (1)参数数量少
- (2)通过共享嵌入，它隐含地强加了在同一纬度/经度的位置以及在一天中的同一时间或一周中的同一天的间隔应该具有相似表示的先验。

## 多任务表示学习

架构:



用于行程时间估计的多任务表示学习模型(MURAT)的系统架构。该模型首先将原始链接信息和时空信息嵌入到学习空间中。然后，将学习到的表示与其他数字特征一起输入到深度残差网络中，该网络使用来自多个相关任务的监督信号进行联合训练。

算法框架:

从路径中提取各种摘要，例如，行驶距离、行程中的链接数量、交通灯数量和转弯数量，并将它们用作待预测的辅助任务。具体地，提出了多任务学习框架，以联合学习主任务（预测旅行时）以及各种辅助任务（预测不同的路径摘要）。我们使用硬参数共享框架，其中不同的任务共享模型的大部分，除了每个任务有一个专用的输出层。

---

**Algorithm 1:** Multi-task Representation Learning (MURAT)

---

**Data:** Link graph  $\mathcal{G}_L$ , Spatial graph  $\mathcal{G}_S$ , temporal graph  $\mathcal{G}_T$

**Result:** Learned representations:  $E_L, E_S, E_T$

**begin**

$E_L \leftarrow \text{GraphEmbed}(\mathcal{G}_L)$  // Unsupervised pre-training

$E_S, E_T \leftarrow \mathcal{N}(0, 1)$

**for**  $i \leftarrow 1 \cdots N$  **do**

$\mathbb{L}^{(i)} \leftarrow []$

$\mathbf{x}^{(i)}, \mathbf{y}^{(i)} \leftarrow \text{sample}(\text{data}, i)$

$E_{L, \mathbf{x}^{(i)}} \leftarrow \text{Embed}(E_L, \mathbf{x}^{(i)})$  // Link embedding

$E_{S, \mathbf{x}^{(i)}} \leftarrow \text{Embed}(E_S, \mathbf{x}^{(i)})$  // Spatial embedding

$E_{T, \mathbf{x}^{(i)}} \leftarrow \text{Embed}(E_T, \mathbf{x}^{(i)})$  // Temporal embedding

$\hat{\mathbf{y}}^{(i)} \leftarrow \text{ResNet}(\Theta, [E_{L, \mathbf{x}^{(i)}}, E_{S, \mathbf{x}^{(i)}}, E_{T, \mathbf{x}^{(i)}}])$

        // Aggregate losses from multiple tasks

**for**  $k \leftarrow 1 \cdots K$  **do**

$\mathbb{L}^{(i)} \leftarrow [\mathbb{L}^{(i)}, \ell_k(\hat{\mathbf{y}}^{(i)}, \mathbf{y}^{(i)})]$

**end**

$\ell^{(i)} \leftarrow f_\ell(\mathbb{L}^{(i)}) + \lambda_S \ell_{\mathcal{G}_S}(E_S) + \lambda_T \ell_{\mathcal{G}_T}(E_T)$

$E_L, E_S, E_T, \Theta \leftarrow \text{AdamOpt}([E_L, E_S, E_T, \Theta], \ell^{(i)})$

**end**

**return**  $E_L, E_S, E_T$

**end**

---

## 大宗货运场景模型的适用性

本文模型主要应用于出租车场景。

每个接送行程包含来源、目的地、出发时间、旅行时间和路径。我们提取各种路径摘要作为辅助任务，包括行驶距离、道路数量、信号灯数量、转弯。对于道路网络，使用滴滴出行提供的北京商业地图，根据道路连接信息构建道路图。利用地图匹配获取每个位置的相应链接信息。

我们货车运单11月的数据中包含起点和流向以及轨迹信息。考虑用地图匹配技术获取实际运行的道路轨迹。

数据轨迹形式如下：

