**北京航空航天大学计算机学院**

**硕士学位论文开题报告**

**论文题目**：基于图神经网络的轨迹预测~~及多阶段意图分析~~算法研究与实现

**专 业**：计算机科学与技术

**研究方向**：智能交通

**研 究 生**：杨思逊

**学 号**：SY2006306

**指导教师**：翁敬农 研究员

**北京航空航天大学计算机学院**

2021年12月16日

目 录

[1. 论文选题的背景与意义 1](#_Toc90057571)

[2. 国内外研究现状 2](#_Toc90057572)

[2.1. 意图预测 2](#_Toc90057573)

[2.1.1. 意图在认知体系中的位置 2](#_Toc90057574)

[2.1.2. 意图分类 3](#_Toc90057575)

[2.1.3. 意图推测方式 4](#_Toc90057576)

[2.2. 轨迹预测 6](#_Toc90057577)

[2.2.1. RNN 7](#_Toc90057578)

[2.2.2. GNN 8](#_Toc90057579)

[2.2.3. GAN 8](#_Toc90057580)

[2.2.4. Others 9](#_Toc90057581)

[2.3. 结合轨迹预测和轨迹预测 9](#_Toc90057582)

[2.3.1. 预测潜在目的地 10](#_Toc90057583)

[2.3.2. 预测操作驾驶意图 10](#_Toc90057584)

[3. 论文的研究内容及拟采取的技术方案 11](#_Toc90057585)

[3.1. 研究目标 11](#_Toc90057586)

[3.2. 研究内容 11](#_Toc90057587)

[3.2.1. 对于车辆行驶的历史轨迹数据的建模 11](#_Toc90057588)

[3.2.2. 对于车辆之间交互方式的建模 12](#_Toc90057589)

[3.2.3. 对于交通语义信息的建模 12](#_Toc90057590)

[3.2.4. 对于意图的建模 13](#_Toc90057591)

[3.3. 研究方案 13](#_Toc90057592)

[3.3.1. 车辆道路交互编码器 13](#_Toc90057593)

[3.3.2. 历史轨迹数据建模模块 13](#_Toc90057594)

[3.3.3. 意图解码器 13](#_Toc90057595)

[3.3.4. 轨迹解码器 14](#_Toc90057596)

[4. 关键技术及难点 14](#_Toc90057597)

[4.1. mode collapse问题 14](#_Toc90057598)

[4.2. 模型准确率和推理时间的平衡 14](#_Toc90057599)

[4.3. 交互建模问题 15](#_Toc90057600)

[5. 论文研究计划 15](#_Toc90057601)

[6. 主要参考文献 15](#_Toc90057602)

# 论文选题的背景与意义

车联网系统是传统汽车产业、交通领域与信息通信行业加速融合发展的产物，在提升车辆行驶安全、提高交通效率、提供出行信息服务和支持自动驾驶等方面具有重要的意义。根据《车联网（智能网联汽车）产业发展行动计划》等文件，国家对高速公路产业的发展指明了方向。高速公路产业需要完成数字化的转型，实践新技术在高速公路产业的应用，推动智慧高速试点工程，完善高速公路车路协同体系的建设标准，探索车路协同体系的产业模式。

在现代城市中，交通场景的复杂性与多样性决定了车联网预警机制在这些场景下的重要作用，预警系统需要即时准确地预测当前场景下车辆的未来轨迹，为其未来可能发生的冲突及时向车辆进行预警，具体应用场景如下：

1. 高速公路的分合流区域是交通流发生变化的区域。在分合流区域，由于主线和匝道的车速差异大、变道行为多、变道窗口期短、驾驶员视野受限等原因，容易产生违法驾驶行为，从而容易发生交通事故、降低通行效率。通过建设分合流安全预警及诱导系统，对分合流区域过往车辆进行预警提醒，尽量避免主线车辆与匝道车辆的碰撞事故发生。

2. 隧道出入洞口段为光线突变段，行驶通过该段时，驾驶员的视觉生理反应需要消耗时间，产生“白洞效应”和“黑洞效应”，导致出入口附近的交通事故频发；另一方面，由于隧道内视距受限，易引发二次事故。此处隧道通行场景是指车辆在通行隧道前、隧道中、隧道后的全过程安全预警及诱导场景，重点减少由于隧道黑白洞效应引发的交通事故，同时实现洞内事故提前告知功能。

3. 团雾等恶劣气象、积雪凝冰等路面异常等问题均会对车辆通行安全及效率产生重大影响，导致高速公路无法实现全天候通行。车路协同预警、诱导服务，能够实现特定恶劣气象条件下车辆的安全通行。

在这些关键交通场景中，预警系统除了根据【静态交通信息】，如一般安全提醒、危险道路结构、限速预警、车内标牌等进行例行提示；根据【长效动态交通信息】如道路拥堵预警，合流区域车流量、事故发生提示之外，还需要感知预测【即时动态交通信息】如前方车辆慢行、停驶、逆行、急刹车等驾驶意图以及其他异常行为，对驾驶人员做出及时必要的提示。其中最后一项信息获取难度最大：需要道路边缘自治节点或者车辆依据特定算法及时检出判断，且误报漏报的安全隐患也更大。

因此，在关键场景驾驶预警系统中，【预测即时动态交通状态的算法】是关键核心，需要进行研究与初步实现，它应尝试解决如下问题：

1. 关键车辆的定位：关键车辆指的是其正常驾驶操作会对本车驾驶决策产生影响的车辆，限于目前算力的局限性，预警系统无法无限制地追踪当前道路上的所有车辆，因此需要构建关系图，其节点包含与【ego-car】在同一语义背景下的车辆，追踪它们的状态变化并进行推理。

2. 及时、准确地识别关键车辆的意图：在确定了关键车辆的驾驶意图之后，一方面能够通过多任务学习知识，使之提高对关键车辆轨迹预测的精度；另一方面能够及时对驾驶员发出明确预警。

3. 建立合理的轨迹预测模型：交通场景背景的复杂度、快速动态的交通场景变化，车辆行为的不确定性，以及出于安全考虑的预测精度共同决定了模型需要正确建模道路拓扑、车辆之间的关系网和车-路交互矩阵。合理的模型不仅应有较高的精确度，也应具备泛化能力。

# 国内外研究现状

## 轨迹预测

轨迹预测的方式有很多，此处我们将其分为三类：

1. **基于物理模型的方法**：通过建立牛顿运动方程来对物体进行预测。遵循的是感知-预测的流程，即通过运动物体过去的位置与速度直接求解计算得出未来的具体位置。而基于模型的复杂程度，这些方法还能够被继续分为单模型方法：只使用单一模型进行估计，以及多模型方法：首先定义一个模型方法集，在估计过程中选择或者综合利用其中的方法。其有点在于简单、快捷、无需数据进行训练，但是无法掌握时空上的高级语义信息，适用于简单以及对精度要求不高的场景。
2. **基于学习运动模式的方法**：实际上是通过轨迹数据学习出运动物体的运动模式，它能够考虑的信息不仅有物理定律，还具有感知周围环境信息以及建模交互的能力，遵循的是感知-学习-预测的流程。据不完全统计，使用基于学习运动模式方法进行轨迹预测的研究占所有轨迹预测研究比例最高，2019年达到了80%以上。由于深度学习能够处理多模态数据、具有多种成熟网络结构的优势，它承担了大量轨迹预测的研究工作。因此我们将基于学习模式的方法分为基于经典模型和基于神经网络的方法。这类方法适合场景运动物体较多，交互复杂的情况，但是需要大量专门用于训练的数据，场景单一又使得模型泛化成为一个问题。当地图结构改变，是否还能够保证预测质量？相比于基于物理的方法，学习运动模式的方法缺乏可解释性，导致前者已经能够实用而后者只能用于对安全性要求不高、场景结构不变化或变化较少的场景。限于本文研究内容以及近年来研究趋势，在此仅介绍基于深度学习方法的轨迹预测方式。
3. **基于规划的方法**：为了弥补基于学习运动模式方法缺乏可解释性的不足，基于规划的方法提出要明确指出运动物体的目标或实际策略。通过判断模型是直接给出判断运动是否合适的标准，可以将基于规划的方法分为正向规划方法与逆向规划方法，后者一般使用逆强化学习进行标准的建模。如果地图信息完全，该模型相比基于物理的方法更准确，相比学习运动模型方法泛化性更强。但是随着场景复杂度的增加（人数），预测时间长度的增加，运行时间也会指数倍增加。

传统基于数据挖掘的方法在时空数据预测上表现差主要由两个原因：一，时

空数据本质上是连续的，而传统方法更适合离散数据；二，时空数据之间的相关

性相较传统数据更加复杂，传统方法难以捕捉。而深度神经网络能解决这两个问题。

### RNN

如同其他序列预测任务一样，循环神经网络是轨迹预测任务的首选架构，很多研究都使用了长短期记忆网络(LSTM)来对历史信息进行建模：[[[1]](#endnote-2)]使用的Social-LSTM网络最为经典，它为每一位行人的轨迹都使用独立的LSTM网络进行建模，并且使用了池化机制建模行人之间的交互行为；[[[2]](#endnote-3)]在Social-LSTM的基础上做了延伸：加入了能够理解语义的池化层，因此加入了对场景静态障碍物的考虑；[[[3]](#endnote-4)]在数周观察数据的基础上使用LSTM对特定环境、时间段的行人运动进行建模，表征人运动模式的向量被扩充，向其中加入了时间、场景信息。[[[4]](#endnote-5)]提出了新的池化机制：池化时考虑周围行人的相对重要性；在预测多种交通参与者轨迹时，[[[5]](#endnote-6)]在池化时考虑了组别信息，并且对预测自行车轨迹的LSTM结构进行了修改。

下面的工作使用LSTM不仅对历史轨迹进行建模，还读取了其他信息进行：[[[6]](#endnote-7)]使用了两条LSTM序列，一条的输入是自机(ego-vehicle)的转向、速度的历史数据、场景特征，输出自机的未来轨迹，另一条的输入是其他交通参与者的历史位置，联合第一条序列的输出进行其他目标轨迹的预测；[[[7]](#endnote-8)]在上述基础上进一步加入了预测意图的LSTM序列，输出关于预测目标是否会横跨街道的信息，这条序列的输入为行人的运动信息、图像信息以及周围场景语义；[[[8]](#endnote-9)]首先将交通参与者的形状、速度信息，以及他们基于周围区域的相对位置嵌入进向量，以此作为LSTM的输入，之后再使用多级CNN提取当前场景的总体运动模式。

但是[[[9]](#endnote-10)]指出，上述基于RNN，存在记忆的网络使用单一的状态向量代表一个运动物体的所有时空信息，但是当涉及某一方面具体的特征时，无法将其从整体的状态向量中提取出来，而且由于该向量表征的状态是非结构化且全局的，RNN在一步步更新向量之后，最终无法在长时间序列上得到较好的结果，因此需要外部的信息增强模型的预测能力，即外部记忆。受记忆增强网络的启发，[9]将其应用到了轨迹预测任务当中，使用AE作为骨架，加入记忆增强模块，取得了较好的结果。

### GNN

由于图神经网络的发展以及交通场景天然的非欧结构，越来越多的研究将图结构引入模型中，并且占据着越来越重要的地位。

Social-BiGAT[[[10]](#endnote-11)]使用LSTM对每条轨迹进行建模，使用图神经网络对行人之间的交互进行建模；Social-STGCNN[[[11]](#endnote-12)]直接使用图卷积网络，而行人所代表的节点之间的权重则由它们之间的距离决定，之后使用时间卷积网络(TCN)进行轨迹预测；SGCN[[[12]](#endnote-13)]在Social-STGCNN思想的基础上加入了图简化机制，使用非对称稀疏图建模交互并且引入了非对称卷积操作，在前述工作的性能上得到了提升；STAR[[[13]](#endnote-14)]模型在图结构的基础上使用了Transformer模型，对空间、时间的相关性分别进行建模；AgentFormer[[[14]](#endnote-15)]不同于STAR，使用Transformer模型直接对所有预测目标的时空数据进行建模，并设计了独特的嵌入算法与双重注意力机制对时间数据的有序性、行人数据的排列不变性进行了建模；RSGB[[[15]](#endnote-16)]注意到远处的行人有时候也存在较强的相互作用，因此在模型中加入社会心理学的先验知识，使得模型能够对行人进行正确分组。

上述介绍的基于图的网络都使用的是静态图，一旦观察到场景中的所有人之后，网络构造图结构之后，该结构便不会再改变，而下述工作中的图结构则会因行人的运动状态不同而发生改变：[[[16]](#endnote-17)]提出的EvolveGraph较好地对多样化的数据进行了整合，并且提出图演化机制来对场景中行人的整体运动进行建模，以此输出预测轨迹；DGAN[[[17]](#endnote-18)]的思想有所不同，它结合了动态注意力区域机制与图注意力机制，它们对司机的意图以及在不同交通场景下的驾驶习惯进行了建模。

### GAN

此外，使用生成式对抗网络进行轨迹预测的研究也有很多，根据[[[18]](#endnote-19)]，上述的网络结构如CNN、RNN、GCN、AE都被应用到了轨迹预测中，然而它们仍然存在不足：首先，基于这些架构的预测网络本质上还是判别式模型，难以真正对运动轨迹中的随机性进行建模；而且，这些结构难以对抽象数据进行推理，因此无法准确对影响人车轨迹运动的抽象因素进行建模。而GAN的性质可以解决这两个问题：它能学习到的轨迹分布相较之前使用的高斯模型更加灵活，天然适合多模态输出。

在轨迹预测领域比较经典基于GAN的模型有[[[19]](#endnote-20)]提出的SocialGAN，所有预测目标的轨迹被联合预测，使得预测结果中不会出现碰撞。在对网络进行训练时，它们加入了新的损失函数，使得生成分布的空间确实包含所有可能路径，并且同时也考虑到历史轨迹。在交互建模方面，引入了新的池化机制，使用一个全局向量表示当前场景下共有特征；[[[20]](#endnote-21)]提出的GD-GAN所习得的特征向量能够进行轨迹预测和不同社交小组的识别；[[[21]](#endnote-22)]在SocialGAN的基础上加入了Info-GAN[[[22]](#endnote-23)]的结构。

最近较为出名的网络则是SoPhie[[[23]](#endnote-24)]，它既考虑到了场景语义，也为行人交互建立了模型。在场景语义提取与交互建模模块里都使用了注意力机制；Social-BiGAT10基于Bicycle-GAN[[[24]](#endnote-25)]的架构，建立了生成器输入与输出的双射函数，此外还引入了VGG进行场景信息捕捉，使用了图注意力网络建模行人交互；CoL-GAN[[[25]](#endnote-26)]使用CNN网络作为判别器，不同于上述GAN的判别器，该网络的判别器能够判断轨迹每一段是否真实。更近一些，[[[26]](#endnote-27)]结合自动编码器与GAN，联合预测行人轨迹；APOIR[[[27]](#endnote-28)]进行了两阶段预测：首先预测行人潜在偏好的分布，然后再进行轨迹预测，再使用判别器进行辨别。

### Others

除此之外，还有另外一些使用深度学习网络，但是核心重点不在基础网络结构的研究：[[[28]](#endnote-29)]是第一个使用多任务学习进行轨迹姿态预测的网络，此外，面对多模态的数据，该研究使用了注意力机制决定了各数据源的重要性；[[[29]](#endnote-30)]使用多阶段预测方法，首先基于历史轨迹预测行人的潜在目标，然后使用CVAE在潜在目标的基础上生成轨迹；类似PECNet，[[[30]](#endnote-31)]仍然重视预测行人目标：它使用两个自动编码器实现目标点预测的由粗到精；Social NCE[[[31]](#endnote-32)]不同于以上网络的训练思想，它使用对比学习减轻了轨迹预测时的covariate shift问题，然后通过增加负样例使得模型在避免碰撞的测试中效果更好。

## 结合意图预测和轨迹预测

目前许多研究工作将汽车行驶过程中意图推断与轨迹预测的任务结合起来，主要是为了提升轨迹预测的准确率。而在这些工作中，意图主要分两个方面：当前时间范围内汽车的潜在目的地与具体操作驾驶意图。

### 预测潜在目的地

[[[32]](#endnote-33)]认为汽车在运动过程中受到道路的限制较大，大部分工作使用的栅格化地图并未能让深度学习模型准确显式地捕捉到道路信息，因此他们的工作额外对道路进行建模，最终使用注意力机制得到车辆最终可能位于的道路节上，基于此生成未来轨迹。

[[[33]](#endnote-34)]为了解决汽车运动多模态特征方面的“mode collapse”问题时，提出了使用热力图进行轨迹预测。文章首先基于运动车辆的历史信息以及高清地图所包含的多种语义信息得到车辆目标的可能热力图，再此基础上，设计了两种采样算法来减小minFDE与MR，在此基础上生成的可能轨迹既覆盖了所有可能的运动交互模式，也较为准确。截至论文发布时，该研究在Argoverse Motion Forecasting Benchmark上取得了第一名。之后该团队对前述方法进行了改进，提出了GO-HOME[[[34]](#endnote-35)]：它们效仿VectorNet[[[35]](#endnote-36)]，将道路作为结构化数据输入，建立道路关系图，基于道路具体几何形状拓扑结构生成热力图，达到了更好的效果。

谷歌和waymo的联合研究[[[36]](#endnote-37)]中，实现了一个三步走的预测方式：首先根据车辆的运动信息进行可能终点集合的预测，然后基于这些终点生成轨迹，最后对这些轨迹进行可信性判别，选出k条最有可能的路线。在处理车与道路的关系方面，研究者首先使用同样是谷歌提出的VectorNet[35]处理结构化的道路信息，再结合预测的可能终点集合让下游模块进行处理。该方法在Argoverse、INTERACTION、SDD数据集上都取得了SOTA的结果。无独有偶，在2021年的DenseTNT: Waymo Open Dataset Motion Prediction Challenge中，清华大学团队提出了DenseTNT[[[37]](#endnote-38)]，对比之前的TNT方法，无需额外的启发式算法对目的地进行初步生成，而是直接对可能目的地区域进行密集采样，再采取基于注意力机制对每个终点进行打分。注意到之前前述工作用以生成目的地的NMS算法导致了【suboptimal】结果，团队重新提出了新的采样算法。这项工作在前述竞赛中取得了第一名的成绩。

### 预测操作驾驶意图

[[[38]](#endnote-39)]的工作不仅预测汽车的未来轨迹，还对车辆的简单行为：刹车、加速、匀速进行分类。作者首先建立了交通动态图来模拟当前场景下所有车辆的交互，随着时间变化，车辆空间位置和相对位置的变化也导致了动态图的变化，而这些动态图邻接矩阵的特征向量被自编码器处理之后则进行了谱聚类，这一切的则是为了对车辆的驾驶风格进行分类，且也对模型的长时段预测友谊，文章最终证明了该方法预测结果的误差上限。

[[[39]](#endnote-40)]工作的思想较为新颖：作者提出假设：道路上其他车辆的未来运动受到自车策略的影响，因此模型在对所有车辆的轨迹进行建模之后，首先规划出自车的可行路线，既作为对驾驶意图的预测，也是对当前行驶的规划，基于此，再预测出其他车辆的轨迹。

[[[40]](#endnote-41)]作者基于历史轨迹建模使用双头softmax函数分别对车辆再在纵向和横向的操作进行了分类，在得到车辆N×N的运动状态，即车辆驾驶意图之后，再对之后的轨迹进行预测。

# 论文的研究内容及拟采取的技术方案

本节主要讲述本课题的研究目标、研究内容与拟采取的技术方案。

## 研究目标

随着交通智能化，车路互联越受到人们的重视，在关键场景如分合流区域、隧道区域，【即时动态交通信息的感知与预测算法】对于汽车驾驶的辅助功能显得尤为重要。它需要做到基于视觉的交通场景理解，即时对关键车辆的意图进行识别以及轨迹预测，针对这个需求，提出如下研究目标：

1. 设计基于图神经网络的道路车辆轨迹预测以及意图识别模型，在公开数据集上进行训练，使之具有较高的轨迹预测和意图识别准确度，针对意图，还应有较低的假阳性识别率。
2. 将所设计训练模型在所收集的现实数据集上进行测试，证明所设计的轨迹预测与意图识别框架具有可移植性。

## 研究内容

基于以往的工作，轨迹预测任务需要观察目标车辆历史轨迹，进而给出其看可能的未来轨迹分布。此外，由于轨迹预测模型需要具有在现实场景下的应用能力，一般也要将交通信息：如道路结构、车辆相对位置、~~路权~~等语义信息考虑进去，本节将一一进行介绍。

### 对于车辆行驶的历史运动数据的建模

轨迹预测使用的车辆行驶数据包括汽车位置、航向角、加速度等物理信息。历史运动信息是典型的时间序列信息，因此传统机器学习方法捕捉数据间【长时段关系】的能力不如深度学习方法，主流方法一般使用不易受梯度消失问题影响的长短期记忆模型（LSTM）来处理历史运动信息。然而LSTM模型因其需要多次读入数据，运行时耗费较多时间，难以满足预测的实时性要求，因此近年来又有工作使用时间卷积网络或者在自然语言处理领域大火的Transformer结构来替代LSTM。此外，对于历史轨迹数据的稀疏采样也被证明可以缓解长时段数据关系难以捕捉的问题。

因此，考虑到实际情况中对于模型实时性与正确性的要求，使用何种结构模型达到较好的运动数据建模效果是本文研究内容之一，本研究内容将从时间卷积网络入手，对比其与注意力机制、LSTM等主流方法的表现差异。

### 对于车辆之间交互方式的建模

在车辆密度较高的实际交通场景中，驾驶员操控汽车时需要对周围车辆的意图有基本的判断，会注意与自己有发生冲突可能的车辆。而深度学习模型在模仿驾驶员考虑这个问题时，一般会采取数据融合的方式，让每辆车对应的实体之间进行信息传递，这就是大部分方法使用的池化机制。

而随着对于图神经网络的研究深入，目前也有越来越多的工作基于场景建立动态图，其中一个节点对应一个车辆实体，边则是车辆之间关系的抽象表示。如何利用图论来对复杂动态的车辆关系进行描述也是一个较为复杂的问题，本文拟从有向图开始，探究如何正确建立图的邻接矩阵，以及从图中如何提取出关键有效信息。

### 对于交通语义信息的建模

传统的机器学习方法限于其表示能力，无法处理高维场景信息，而基于深度学习的方法，借助计算机视觉研究成果，能够将交通语义信息如“可行驶区域”、“车道中心线”、“道路节点”等场景信息以二值图或者重叠起来以伪图片（pesudo-picture）的格式，将交通全景以俯视图的形式，交由卷积网络进行处理。

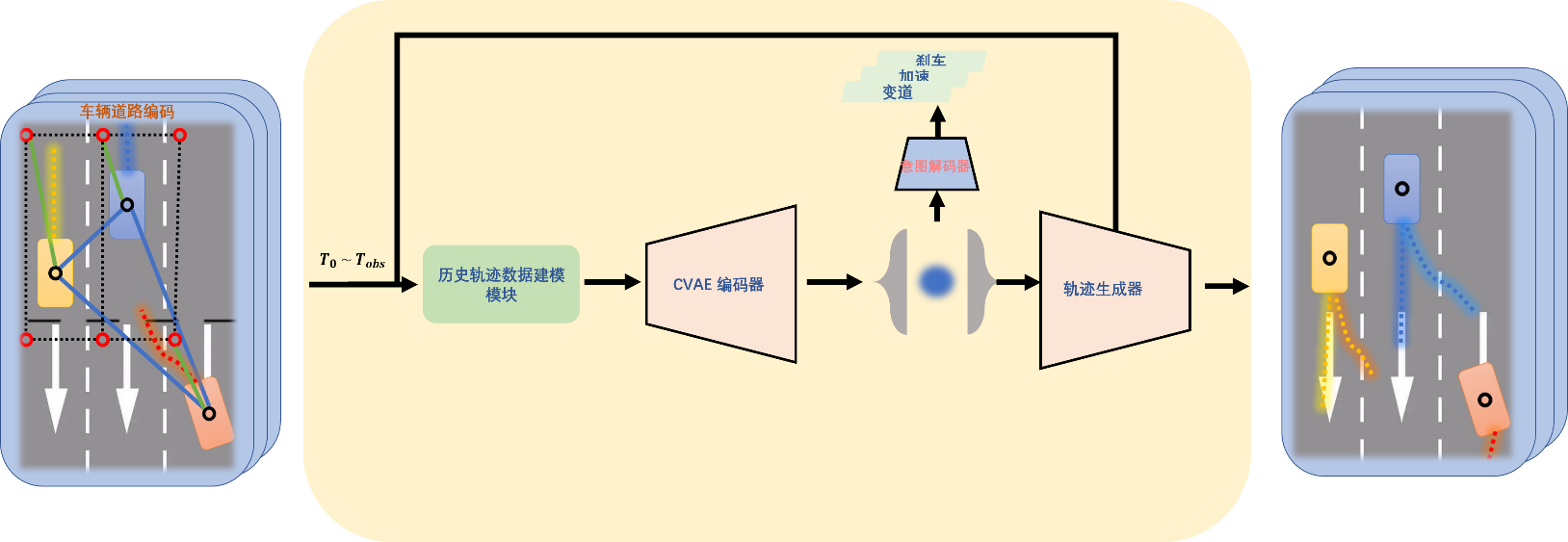
而由谷歌所提出将道路场景结构化的VectorNet35较为独特：它将道路结构视作向量而不是图片来进行处理，显式表示了道路结构的同时也减少了计算量。

该部分待解决的问题是选择哪些交通语义信息能在预测准确率和计算效率之间找到平衡，以及如何结构化交通场景。

### 对于意图的建模

意图预测可以作为一个独立的任务，也可以作为轨迹预测的辅助任务进行。本研究的计划是后者，因此意图预测在预测模型结构中的位置成为重点。本文拟将多阶段预测车辆的潜在目的地、换道意图、以及轨迹，因此如何设计不同阶段意图所对应的计算模块顺序是本文研究内容之一。同时验证意图预测在本任务中是否能够提升预测准确度也是烧灼实验所会关注的。

## 研究方案

下图为本研究所提出模型的示意图，大致分4为个模块：车辆道路交互编码器、历史轨迹数据建模模块、意图生成模块以及轨迹生成模块。本文将按顺序对这四个模块进行介绍。

**图3-3 模型方案示意图**

### 车辆道路交互编码器

本文拟采用构建图的方式来对车辆交互进行建模：将车辆实体作为图中的节点，而节点属性则为车辆历史数据，每一个时刻对应一张图。而道路节作为交通场景实体也作为节点引入前述图中，节点属性为道路节点的绝对位置。仿效[[[41]](#endnote-42)]基于此异构图，建立车辆-车辆、车辆-道路、道路-道路三个邻接矩阵。而临界矩阵的计算则是基于一定规则：可以基于距离、也可以基于图注意力机制、也可以使用多层感知机来自适应学习规则。

一张图建立之后，使用图神经网络中的图卷积机制，或者生成式图神经网络来对每个节点的状态进行迭代，最终达到所有节点互相交换信息、准确对交通场景下的交互进行建模。

### 历史轨迹数据建模模块

首先，选择时间卷积网络作为处理节点历史数据的基本框架，在此基础上尝试添加多尺度采样来提升模型捕获数据间【长效关联】的性能。在历史数据长度方面，除了在与其他已发表方法对比时使用共同设置：读入1.2s数据，预测未来0.8s情形外，需观察延长读入历史信息的长度对预测精度带来的变化。此方式可对历史数据建模方式的有效程度进行定量分析，基于此方法，可以更换建模方式为其他方法：LSTM或者注意力机制，通过对比择优。

### 意图解码器

条件变分自编码器(CVAE)作为能够引入中间变量的深度学习结构，适用于本文的意图生成。因此本模型将使用CVAE的典型方法进行训练和测试。

第一步：在训练时，将完整的历史轨迹信息和未来历史轨迹信息作为输入，训练意图解码器-1，通过损失函数的设置，让意图解码器-2的结果与1号尽量相近。

第二步：模型训练好之后，在测试阶段中，仅使用历史轨迹信息作为输入，将此时意图解码器-2的输出结果作为轨迹解码器的输入。

在本模块中，由于需要输出两种意图：潜在目的地与转向换道意图，因此解码器架构也是研究重点，本文拟从双头CVAE的结构出发，在此基础上探究结构对于多级意图推断的影响。

### 轨迹解码器

通过意图解码器的结果以及前述模块关于场景信息、车辆交互信息、车辆运动历史信息的建模，轨迹解码器将综合利用这些数据，进行轨迹解码。轨迹解码器拟使用多层感知机，以二维高斯分布的形式对轨迹进行输出。

# 关键技术及难点

## mode collapse问题

轨迹预测问题本质上是一个多模态问题，在实际情况中，由于无法获知驾驶员的未来真实预定目的地与意图，因此需要在每一次预测中进行多种可能性的预测。但是由于训练数据标签仅存在一种可能，在训练时容易出现相似场景不同标签的数据的同时存在使得模型无法正确捕获两种正确模式的情况，我们称之为mode collapse问题。

目前该问题存在多种解决方案，如训练采样器，修改损失函数迫使模型生成相互正交独立的结果。但是要真正解决这个问题，需要模型在语义层面上理解当前交通场景。

## 模型准确率和推理时间的平衡

理论上，使用越多的交通场景数据，便能更精确地对当前场景建模，这个需求在目前提供多样数据的车辆数据集如waymo、argoverse等不难实现，使用模拟数据集更是能够生成尽可能多的图像视频数据。然而在许多实际场景中，作为预警辅助系统的关键算法环节，使用大量甚至冗余图像视频数据的模型难以部署，它们会消耗大量计算资源，并且无法保证实时性。因此使用哪些交通场景数据，如何设计模型使其计算时间满足实际应用需求，是一个较难且必须进行取舍的问题。

## 交互建模问题

实际交通场景中的车辆交互方式非常复杂且多变，如何利用图论以及图神经网络的研究成果来对实际交互关系进行抽象描述是一个较为重要且存在难度的问题。在图卷积方面，存在着一般图卷积、非对称卷积、基于道路的特殊卷积等方式，在邻接矩阵计算方面，存在着基于距离、神经网络、综合以上二者的方式。使用上述理论、方法的工作目前较多且成为当下基于此领域的主流，而在这方面的正确性证明还未多见。

# 论文研究计划

**表5-1 论文研究计划表**

|  |  |  |
| --- | --- | --- |
| 序号 | 起止时间 | 工作内容 |
| 1 | 2021年12月 | 提交开题报告 |
| 2 | 2021年1月——2021年2月 | 收集查询相关文献 |
| 3 | 2021年3月——2021年4月 | 研究设计轨迹预测、意图推断算法框架 |
| 4 | 2021年5月——2021年6月 | 递交中期报告，准备中期答辩 |
| 5 | 2021年7月——2021年8月 | 实现整体算法流程 |
| 6 | 2021年9月——2021年10月 | 进行实验验证 |
| 7 | 2021年11月——2021年12月 | 撰写毕业论文，准备答辩 |

# 主要参考文献

1. []Alahi A, Goel K, Ramanathan V, Robicquet A, Fei-Fei L and Savarese S (2016) Social LSTM: Human trajectory prediction Prepared using sagej.cls Rudenko et al. 29 in crowded spaces. In: Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR). pp. 961–971 [↑](#endnote-ref-2)
2. [] Bartoli F, Lisanti G, Ballan L and Bimbo AD (2018) Context-aware trajectory prediction. In: Proc. of the IEEE Int. Conf. on Pattern Recognition. pp. 1941–1946. [↑](#endnote-ref-3)
3. [] Sun L, Yan Z, Mellado SM, Hanheide M and Duckett T (2018) 3DOF pedestrian trajectory prediction learned from long-term autonomous mobile robot deployment data. In: Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA). pp. 1–7. [↑](#endnote-ref-4)
4. [] Hasan I, Setti F, Tsesmelis T, Del Bue A, Galasso F and Cristani M (2018) MX-LSTM: mixing tracklets and vislets to jointly forecast trajectories and head poses. In: Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR). pp. 6067–6076. [↑](#endnote-ref-5)
5. [] Saleh K, Hossny M and Nahavandi S (2018a) Cyclist trajectory prediction using bidirectional recurrent neural networks. In: Australasian Joint Conference on Artificial Intelligence. Springer, pp. 284–295. [↑](#endnote-ref-6)
6. [] A. Bhattacharyya, M. Fritz, and B. Schiele, “Long-term on-board prediction of people in traffic scenes under uncertainty,” in CVPR, 2018. [↑](#endnote-ref-7)
7. [] A. Rasouli, I. Kotseruba, T. Kunic, and J. K. Tsotsos, “Pie: A largescale dataset and models for pedestrian intention estimation and trajectory prediction,” in ICCV, 2019. [↑](#endnote-ref-8)
8. [] R. Chandra, U. Bhattacharya, A. Bera, and D. Manocha, “Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions,” in CVPR, 2019. [↑](#endnote-ref-9)
9. [] Marchetti, Francesco, et al. "Mantra: Memory augmented networks for multiple trajectory prediction." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. [↑](#endnote-ref-10)
10. [] Vineet Kosaraju, Amir Sadeghian, Roberto Mart´ın-Mart´ın, Ian Reid, Hamid Rezatofighi, and Silvio Savarese. Socialbigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks. In NeurIPS, pages 137–146, 2019. [↑](#endnote-ref-11)
11. [] Abduallah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In CVPR, pages 14424–14432, 2020 [↑](#endnote-ref-12)
12. [] Shi, Liushuai, et al. "SGCN: Sparse Graph Convolution Network for Pedestrian Trajectory Prediction." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. [↑](#endnote-ref-13)
13. [] Cunjun Yu, Xiao Ma, Jiawei Ren, Haiyu Zhao, and Shuai Yi. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In ECCV, pages 507–523, 2020. [↑](#endnote-ref-14)
14. [] Yuan, Ye, et al. "AgentFormer: Agent-Aware Transformers for Socio-Temporal Multi-Agent Forecasting." arXiv preprint arXiv:2103.14023 (2021) [↑](#endnote-ref-15)
15. [] Jianhua Sun, Qinhong Jiang, and Cewu Lu. Recursive social behavior graph for trajectory prediction. In CVPR, pages 660–669, 2020. [↑](#endnote-ref-16)
16. [] Li, Jiachen, et al. "Evolvegraph: Multi-agent trajectory prediction with dynamic relational reasoning." arXiv preprint arXiv:2003.13924 (2020). [↑](#endnote-ref-17)
17. [] Dong, Bo, et al. "Multi-modal Trajectory Prediction for Autonomous Driving with Semantic Map and Dynamic Graph Attention Network." arXiv preprint arXiv:2103.16273 (2021). [↑](#endnote-ref-18)
18. [] Gao, Nan, et al. "Generative adversarial networks for spatio-temporal data: A survey." arXiv preprint arXiv:2008.08903 (2020). [↑](#endnote-ref-19)
19. [] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. 2018. Social gan: Socially acceptable trajectories with generative adversarial networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2255–2264. [↑](#endnote-ref-20)
20. [] Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes. 2018. Gd-gan: Generative adversarial networks for trajectory prediction and group detection in crowds. In Asian Conference on Computer Vision. Springer, 314–330. [↑](#endnote-ref-21)
21. [] Javad Amirian, Jean-Bernard Hayet, and Julien Pettré. 2019. Social ways: Learning multi-modal distributions of pedestrian trajectories with GANs. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. [↑](#endnote-ref-22)
22. [] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. 2016. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In Advances in Neural Information Processing Systems. 2172–2180. [↑](#endnote-ref-23)
23. [] Amir Sadeghian, Vineet Kosaraju, Ali Sadeghian, Noriaki Hirose, Hamid Rezatofighi, and Silvio Savarese. 2019. SoPhie: An attentive gan for predicting paths compliant to social and physical constraints. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 1349–1358. [↑](#endnote-ref-24)
24. [] Jun-Yan Zhu, Richard Zhang, Deepak Pathak, Trevor Darrell, Alexei A Efros, Oliver Wang, and Eli Shechtman. 2017. Toward multimodal image-to-image translation. In Advances in Neural Information Processing Systems. [↑](#endnote-ref-25)
25. [] Shaohua Liu, Haibo Liu, Huikun Bi, and Tianlu Mao. 2020. CoL-GAN: Plausible and collision-less trajectory prediction by attention-based GAN. IEEE Access (2020), 1–10. [↑](#endnote-ref-26)
26. [] Qiang Gao, Fengli Zhang, Fuming Yao, Ailing Li, Lin Mei, and Fan Zhou. 2020. Adversarial mobility learning for human trajectory classification. IEEE Access 8 (2020), 20563–20576. [↑](#endnote-ref-27)
27. [] Fan Zhou, Ruiyang Yin, Kunpeng Zhang, Goce Trajcevski, Ting Zhong, and Jin Wu. 2019. Adversarial point-of-interest recommendation. In The World Wide Web Conference. 3462–34618. [↑](#endnote-ref-28)
28. [] Liang, Junwei, et al. "Peeking into the future: Predicting future person activities and locations in videos." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. [↑](#endnote-ref-29)
29. [] Mangalam, Karttikeya, et al. "It is not the journey but the destination: Endpoint conditioned trajectory prediction." European Conference on Computer Vision. Springer, Cham, 2020. [↑](#endnote-ref-30)
30. [] Liang, Junwei, et al. "The garden of forking paths: Towards multi-future trajectory prediction." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. [↑](#endnote-ref-31)
31. [] Liu, Yuejiang, Qi Yan, and Alexandre Alahi. "Social nce: Contrastive learning of socially-aware motion representations." arXiv preprint arXiv:2012.11717 (2020). [↑](#endnote-ref-32)
32. [] LUO, Chenxu, et al. Probabilistic multi-modal trajectory prediction with lane attention for autonomous vehicles. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020. S. 2370-2376. [↑](#endnote-ref-33)
33. [] GILLES, Thomas, et al. HOME: Heatmap Output for future Motion Estimation. arXiv preprint arXiv:2105.10968, 2021. [↑](#endnote-ref-34)
34. [] GILLES, Thomas, et al. Gohome: Graph-oriented heatmap output forfuture motion estimation. arXiv e-prints, 2021, S. arXiv: 2109.01827. [↑](#endnote-ref-35)
35. [] GAO, Jiyang, et al. Vectornet: Encoding hd maps and agent dynamics from vectorized representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. S. 11525-11533. [↑](#endnote-ref-36)
36. [] ZHAO, Hang, et al. Tnt: Target-driven trajectory prediction. arXiv preprint arXiv:2008.08294, 2020. [↑](#endnote-ref-37)
37. [] GU, Junru; SUN, Chen; ZHAO, Hang. Densetnt: End-to-end trajectory prediction from dense goal sets. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. S. 15303-15312. [↑](#endnote-ref-38)
38. [] CHANDRA, Rohan, et al. Forecasting trajectory and behavior of road-agents using spectral clustering in graph-lstms. IEEE Robotics and Automation Letters, 2020, 5. Jg., Nr. 3, S. 4882-4890. [↑](#endnote-ref-39)
39. [] SONG, Haoran, et al. Pip: Planning-informed trajectory prediction for autonomous driving. In: European Conference on Computer Vision. Springer, Cham, 2020. S. 598-614. [↑](#endnote-ref-40)
40. [] HASAN, Mohamed, et al. Maneuver-Aware Pooling for Vehicle Trajectory Prediction. arXiv preprint arXiv:2104.14079, 2021. [↑](#endnote-ref-41)
41. [] ZENG, Wenyuan, et al. LaneRCNN: Distributed Representations for Graph-Centric Motion Forecasting. arXiv preprint arXiv:2101.06653, 2021. [↑](#endnote-ref-42)