SocialVAE：人类轨迹预测利用时序潜在变量

1. SocialVAE 概述

由于人类决策过程的复杂性和不确定性，预测人类行为面临巨大挑战。现有方法主要面临以下局限：

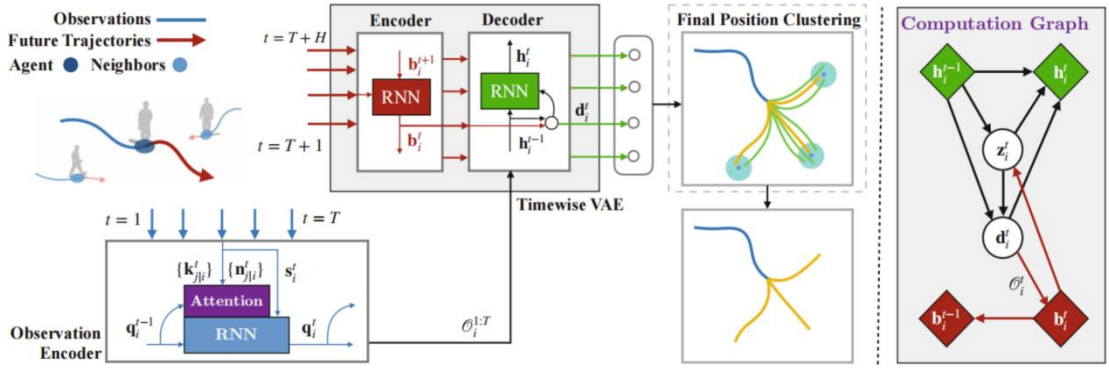
1） 确定性模型的不足：传统方法基于固定规则或单一模态预测，无法刻画同一情境下的多种可能轨迹。

2） 生成模型的缺陷：基于GAN或VAE的生成模型虽能建模多模态分布，但存在两大问题，第一个是静态潜在变量，现有VAE方法仅基于历史观测生成全局潜在变量，忽略了人类决策的动态时变性；第二个是社会交互建模不足，传统注意力机制依赖邻居的RNN隐藏状态，要求持续跟踪，无法处理动态变化的稀疏交互。

3） 小样本下的采样偏差：当从预测分布中抽取少量样本时，现有方法易产生偏差（如样本集中于高密度区域或落入低概率区域），导致预测多样性不足。

SocialVAE是一种用于人类轨迹预测的框架，通过结合时间变分自动编码器架构与社会交互建模来预测个体未来轨迹。实验结果表明，SocialVAE在ETH/UCY、SDD、SportVU NBA等多类基准数据集上均优于现有方法，实现了最先进性能，有效捕捉了人类导航的多模态特性与社会交互影响。

2. 技术细节



图表 1 SocialVAE 模型架构图

该框架包含多个关键模块：核心模块为时间变分自动编码器（Timewise VAE），其通过随机循环神经网络（RNN）在每个时间步引入潜在变量，对人类决策的动态随机性进行建模，生成模型利用 RNN 状态更新和潜在变量生成轨迹位移序列，

推理模型则通过后向 RNN（Backward RNN）基于完整真实轨迹近似后验分布；另有观测编码模块，其采用社会注意力机制（Social Attention），基于邻居的相对位置和相对速度及其社交特征计算注意力权重，动态融合邻居信息；此外，提出可选择后处理技术 Final Position Clustering（FPC），通过对预测轨迹的最终位置进行 K-均值聚类，保留各簇中最具代表性的样本，以减少低概率样本干扰，提升预测多样性与准确性。

2.1 观测编码

此模块意义在于获得 1-T 时间内智能体对场景的局部观测，包括自身观测信息和所有邻居合成信息。其中输入的信息有:（1）s_i^t: 智能体 i 的自身状态，包含由位置位移表示的速度和加速度信息（2）$n_{j|i}^t$: 邻居智能体 j 相对于 i 的局部状态，包含相对位置和相对速度（3）q_i^t: RNN 状态变量（4）$k_{j|i}^t$: 邻居的社交特征,该特征由三种几何特征定义: i 与 j 的欧氏距离; i 到 j 的方位角余弦值; 给定时间范围内 i 到 j 的最小预测距离。

$e_{j|i}^t$定义为为邻居节点 j 到目标节点 i 的边权重，结合$k_{j|i}^t$和q_i^{t-1}使用余弦相似度和神经网络计算得到:

$$e_{j|i}^t = \text{LeakyReLU}(f_q(q_i^{t-1}) \cdot f_k(k_{j|i}^t))$$

其中 fq 和 fk 为神经网络。

得到了每个邻居节点 j 到目标节点 i 的边权重，用各个邻居节点的边权重除以总权重，就可以得到该邻居节点的注意力权重$w_{j|i}^t$。

结合s_i^t，$w_{j|i}^t$，$n_{j|i}^t$，组合可学习特征提取神经网络可得到目标结果，即 1-T 时间内智能体对场景的局部观测:

$$O_i^t := [f_s(s_i^t), \sum_j w_{j|i}^t f_n(n_{j|i}^t)]$$

2.2 时间变分自动编码器

该模块包含 Encoder 和 Decoder 两部分，其中前者在预测时不参与计算，可

使模型能够从轨迹终点反推 “可能的决策路径”，帮助模型在早期时间步生成更分散的潜在变量，从而预测多模态轨迹。后者使用前向 RNN 和时间潜在变量进行轨迹预测。

1) 生成模型部分将 1-T 时间内智能体对场景的局部观测作为条件输入，得到位移序列的条件目标概率分布，将z_i^t设为时刻 t 引入的潜在变量，h_i^t设为 RNN 状态变量。将前面得到的位移序列的条件目标概率分布和z_i^t结合，可以得到生成模型:

$$p(d_i^{T+1:T+H}|O_i^{1:T}) = \prod_{t=T+1}^{T+H} \int_{z_i^t} p(d_i^t|d_i^{T:t-1}, O_i^{1:T}, z_i^t) p(z_i^t|d_i^{T:t-1}, O_i^{1:T}) dz_i^t$$

其中z_i^t为被积分项。生成模型两个积分项都可以表示成z_i^t和h_i^t的表达形式，即位移序列的条件目标概率分布可以由 RNN 状态变量推导得到。

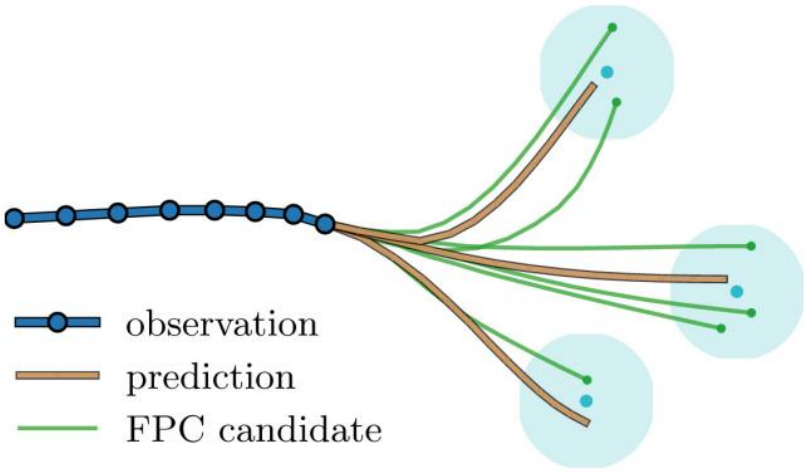
2) 推理模型将反向 RNN 状态变量设为b_i^t，设潜在变量的后验分布为 q。利用完整未来真实轨迹可以得到b_i^t的递归关系式。由反向 RNN 状态变量b_i^t和正向 RNN 状态变量h_i^{t-1}以及潜在变量后验分布 q 可以得到潜在变量采样式:

$$z_i^t \sim q_\emptyset(\cdot|b_i^t, h_i^{t-1})$$

其中∅为将b_i^t和h_i^{t-1}映射到后验参数的网络参数

这种设计使潜在变量同时捕获历史动态（前向）和未来目标（反向）的双向信息。

2.3 终点位置聚类



图表 2 FPC 模型图

首先以高于目标数量 K 的速率采样，然后对样本的终点位置进行 K-means 聚类，每个聚类仅保留终点最接近簇均值的样本，最终生成 K 个预测结果。

3. 模型表现

3.1 定量评估

实验的评价指标均为使用 20 次预测中的最优结果计算平均位移误差（ADE）和终点位移误差（FDE）。

在 ETH/UCY 实验中：相比条件 VAE 基线，SocialVAE 的 ADE 和 FDE 无论是否使用 FPC 均表现更优相比需要后处理的方法，无 FPC 时，SocialVAE 较 SGNet-ED 的 ADE 和 FDE 均提升约 12%，有 FPC 时，ADE 和 FDE 分别提升 21%和 30%。相比需要后处理的方法，SocialVAE+FPC 较 AgentFormer 的 ADE 和 FDE 分别提升约 9%和 13%，较 MemoNet 的 FDE 提升 5%，且推理速度更快、无需额外内存存储；

在 SDD 实验中：SocialVAE 的轨迹分布较其他 VAE 基线更准；SocialVAE+FPC 在 FDE 指标上显著优于现有基线；

在 NBA 数据集实验中：SocialVAE 在两个 NBA 数据集上同样表现出当前最优水平。

3.2 定性评估

ETH/UCY:（第 1 个实验）前三个场景中，热图很好覆盖了真实轨迹；第四个场景中，智能体在 8 帧观测中持续直线行走，但实际右转，尽管真实轨迹偏离平均预测，SocialVAE 的分布仍部分覆盖该轨迹；（第 2 个实验）第一帧中模型在关注右侧三个邻居的同时，对底部似乎正向目标行人移动的绿色邻居给予更多注意力，第 20 帧中模型忽略已改变方向的绿色邻居，将注意力转移至附近三个邻居（红、紫、黄），其中对朝向目标行人的黄色邻居关注度更高，对后方邻居关注度较低。两个场景中，左上角远处邻居和静止邻居（黄色点）均被忽略。

NBA 数据集案例分析：模型预测球员急剧转向、摆脱防守者接球的意图，预测分布给出多个符合该意图的方向，很好覆盖了球员实际选择的方向。其他一些场景中球员分别快速移动创造传球路线和冲击篮板，模型均更关注影响目标球员决策的队友和对手，预测轨迹清晰展现多模态特性，反映相同场景下球员可能的

多种反应行为。

消融实验：SocialVAE 包含四个关键组件：时间生成潜在变量（TL）、反向后验近似（BP）、基于社交特征的邻域注意力（ATT）和可选的终点聚类（FPC）。从没有任何一个组件慢慢增加组件个数，使 ADE/FDE 逐渐降低，可见，四个组件的协同作用显著降低 ADE/FDE 并实现最先进性能。

3.3 实验总结

SocialVAE 通过时间动态建模、双向特征提取、可观测社交特征驱动的注意力机制及后处理优化，展现出显著的优越性和创新性，在多模态预测准确性、动态交互建模和样本效率上实现突破，成为轨迹预测领域的新标杆。