**4．主要研究内容**

（五）（李弟诚）未来城市基层社会海量数据智能分析及大模型构建技术（见《1-科技报告-20230421（西南交大李天瑞）》《2-自评价报告-20230421（西南交大李天瑞）》）

1. 城市知识语义协同挖掘技术

城市知识语义协同挖掘技术整体架构如图4-5-1所示。本项目在智慧城市知识库的基础上，对知识库中的知识三元组（微观）和整个知识库（宏观）进行知识粒化分析和知识多语义表示。

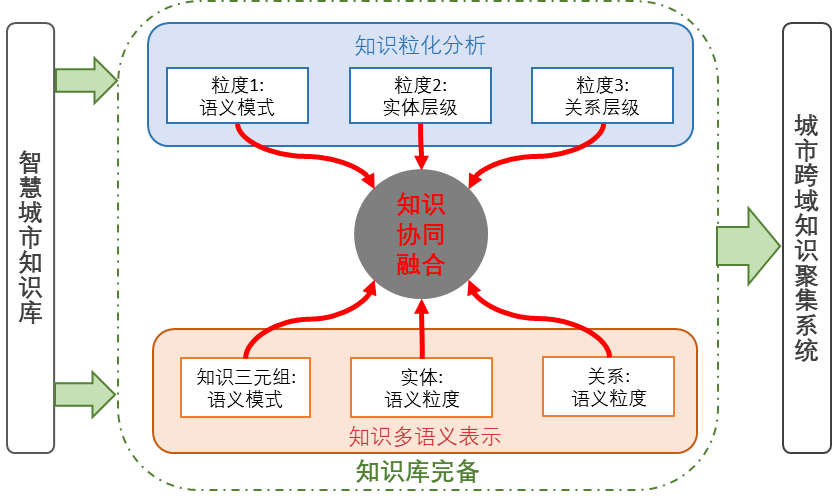


图4-5-1 城市知识语义协同挖掘总体架构图

架构的核心在于解决城市大数据处理过程中存在的模态缺失、时空关联复杂、信息融合困难等问题。通过结合关联计算方法、特征学习、粒计算的粒化分析以及知识构建，旨在实现知识语义的协同挖掘。具体目标是通过知识图谱嵌入进行表示学习，从而完成知识图谱的补全任务，包括实体对齐、图谱融合和图谱补全等。在时空数据处理的挑战方面，本项目通过采用知识融合和协同表征技术，致力于实现时空语义的协同挖掘。

1.1 基于时空数据的知识融合挖掘技术

为了从城市多源异构时空大数据中挖掘知识，本项目提出了一种时空知识融合表征模型（RFP-KMN），该模型能够有效融合时空数据与知识图谱。图4-5-2展示了RFP-KMN的整体结构。该模型包含两个主要模块：第一个模块负责提取关系和实体，包括城市大数据预处理、关系提取（R建模）和实体提取（V建模）；第二个模块则负责挖掘知识三元组，主要涉及数据融合和知识三元组挖掘。

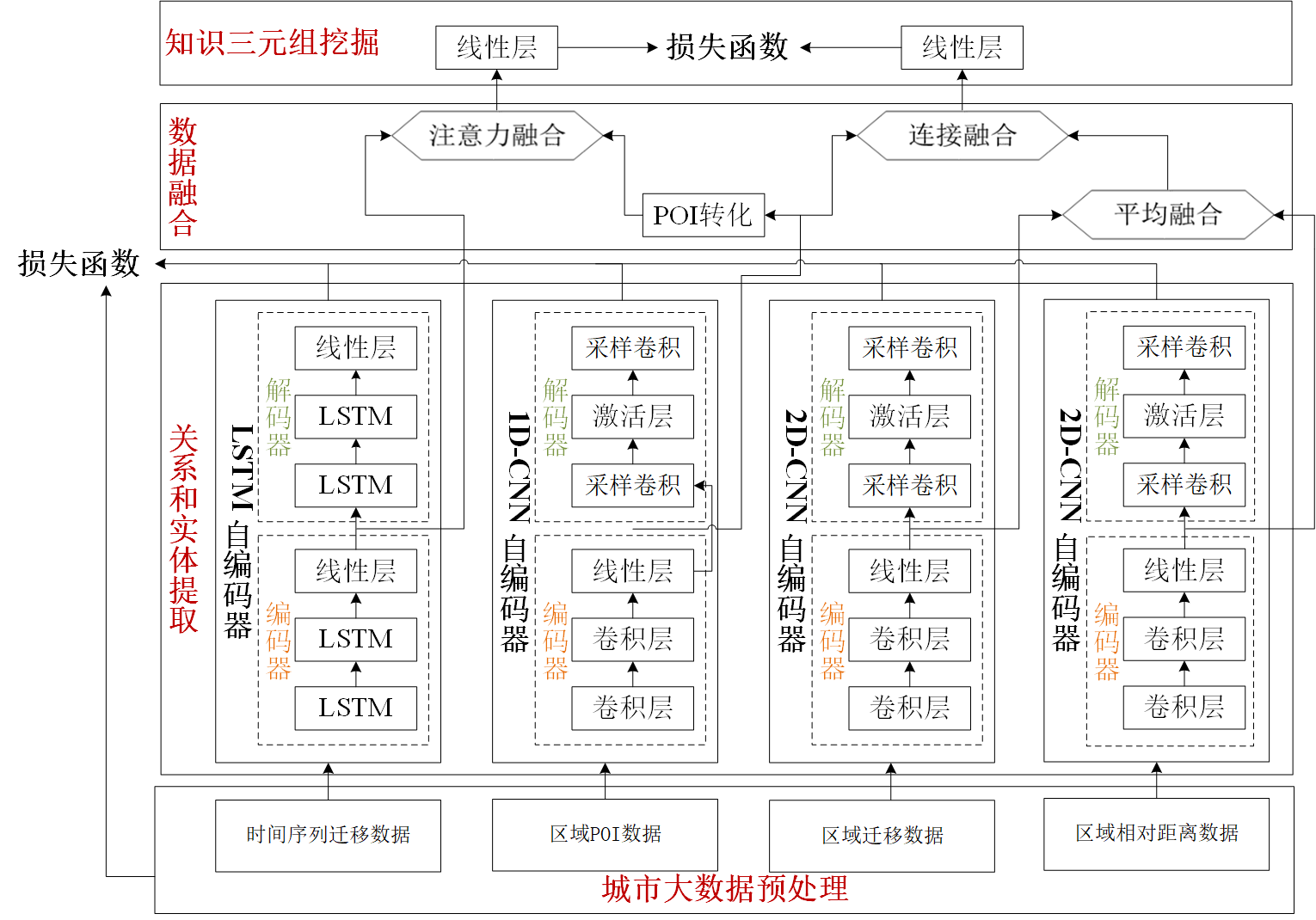


图4-5-2 时空知识融合表征模型

具体而言，时间序列迁移数据通过LSTM自编码器进行建模（R建模），区域特征则通过2D-CNN自编码器进行建模（V建模）。区域POI语义通过1D-CNN自编码器建模。数据融合在此过程中增强了R和V的表示学习能力。而知识三元组的挖掘则基于平移-距离建模方法进行构建，尤其是使用Vec层对前一层输出的维度进行调整。

1.2 时空知识协同挖掘技术

在时空知识融合表征模型的基础上，本文提出了时空知识协同表征模型，如图4-5-3所示。该模型包括嵌入的切割、嵌入的交互和嵌入的卷积三个步骤。

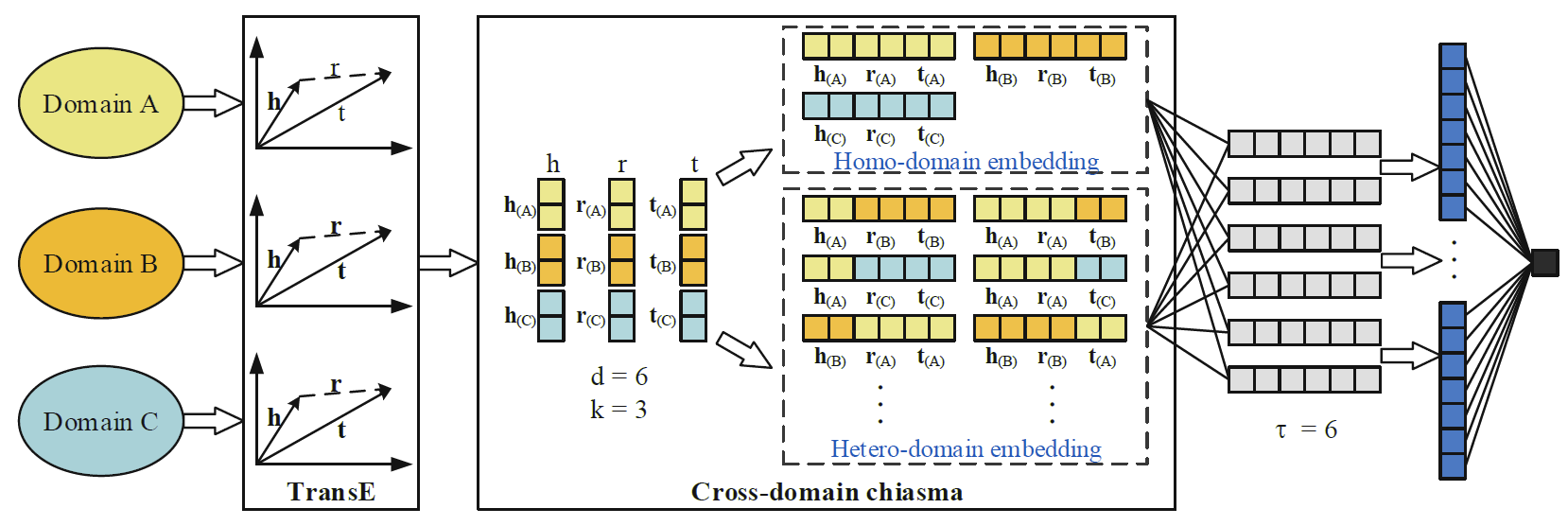


图4-5-3 时空知识协同表征模型

在嵌入的切割模块中，采用经典的知识图谱嵌入技术TransE对多个知识图谱进行表示学习，从而实现嵌入的初始化。接着，将实体和关系平均切分成k个部分（k为知识图谱的数量），每个子实体/子关系代表了实体/关系的一个语义。因此，通过将实体/关系切分成多个部分，可以实现实体/关系的多语义表示。

在嵌入的交互模块中，设计了绑定规则来实现实体/关系的多语义交互，包括头实体-关系绑定和关系-尾实体绑定两种规则。然后，对不同部分进行两两交互，分为同域交互和异域交互，如图4-5-3中间部分所示。同域交互可以获取实体/关系在同一域中的语义信息，而异域交互则获取实体/关系在其他域中的语义信息。

在嵌入的卷积模块中，使用τ个卷积对交互后的实体/关系信息进行特征提取，并通过全连接层计算最终的得分。

1. 城市知识动态演化发现技术

城市知识库中知识的动态演化总体技术框架如图4-5-4所示。该框架基于智慧城市知识库，围绕多时间粒度的信息，研究了基于时间粒度的知识动态演化模型和联邦多粒度动态知识图谱嵌入模型。具体成果包括：（1）构建了基于多时间粒度的知识动态演化模型，并研发了预测未来时刻实体间关系的技术，形成了面向城市知识库迭代更新的知识生长模型；（2）在知识动态演化的基础上，提出了联邦多粒度动态知识图谱嵌入模型，结合联邦学习和动态知识图谱嵌入方法，从多粒度角度实现了各方无需共享知识图谱的情况下，进行动态知识图谱嵌入的交流与优化。

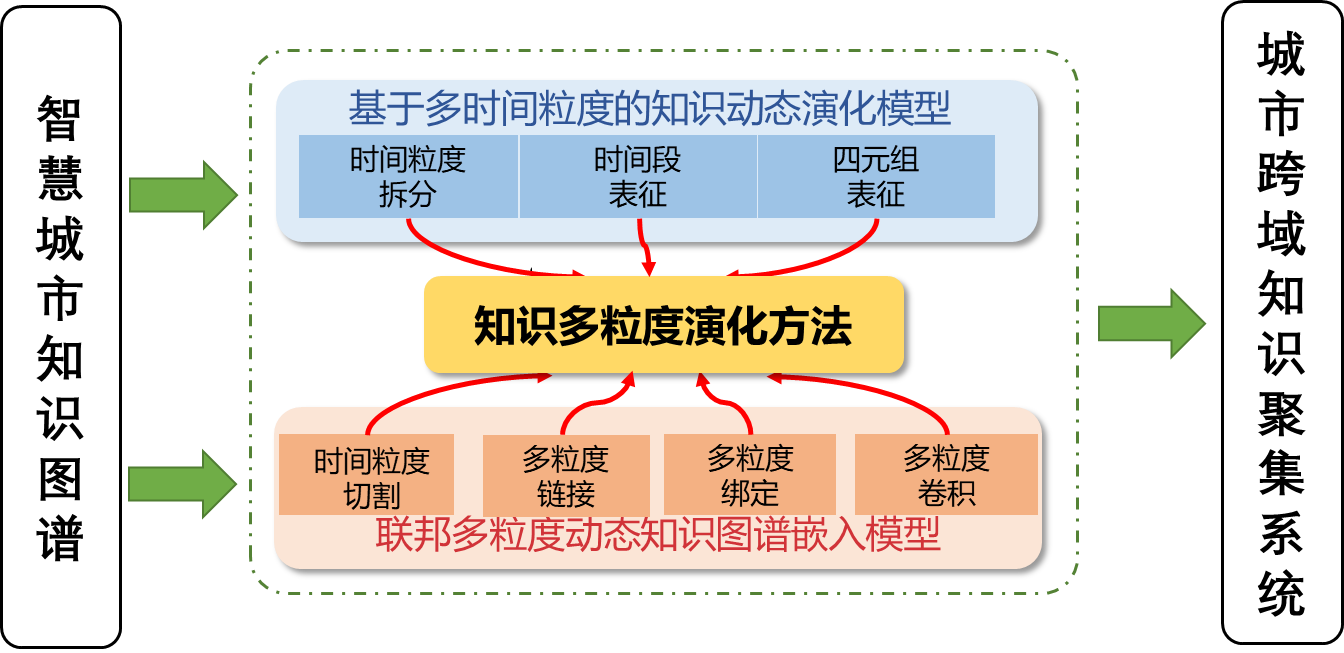


图4‑5-4 城市知识多维度动态演化架构图

2.1 多时间粒度的嵌入框架

多时间粒度的嵌入框架包含多个步骤。首先，该框架将知识四元组拆分为知识三元组和时间。接着，为三元组中的头实体、关系和尾实体在嵌入空间中初始化嵌入向量，同时为每个时间单位分别初始化三个嵌入向量，这些向量用于正弦函数和余弦函数的振幅、频率和初相。随后，通过组合正弦和余弦函数来构建时间的周期性表示，用以表征知识图谱中的时间信息。框架进一步将三元组的嵌入向量与时间的表示拼接，形成知识四元组的最终嵌入表示，并采用DistMult方法对这些表示进行学习。多时间粒度的知识图嵌入框架如图4-5-5所示。

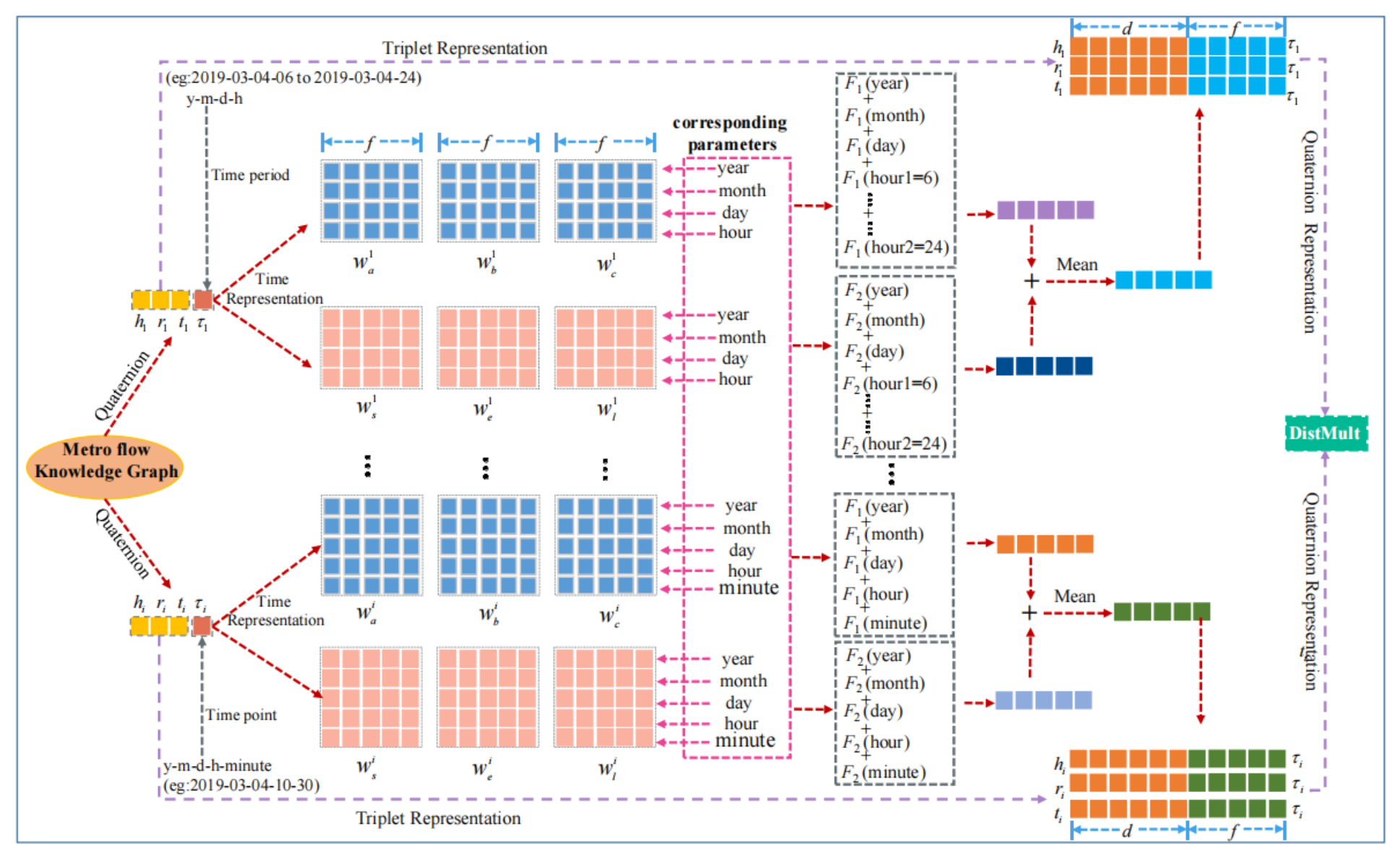


图4‑5-5 多时间粒度的知识嵌入框架

此外，框架会从动态知识图谱中随机选择不同粒度的知识四元组，对每个时间单位的时间表示采用特定的可学习参数向量进行建模，并通过正弦波的组合来表示时间点或时间段。四元组的嵌入向量通过将时间表示与实体和关系的嵌入向量相结合，利用DistMult方法进行学习。模型将时间知识图中的事实四元组分为训练集、验证集和测试集，利用交叉熵构建损失函数，最终通过最小化损失来优化四元组的嵌入表示。

2.2 联邦多粒度动态知识图谱嵌入技术

该模型基于时空知识的动态演化，提出了一种联邦多粒度动态知识图谱嵌入框架，如图4-5-6所示。其每一轮迭代遵循联邦学习的三个阶段：首先，服务器初始化全局嵌入模型，并将其分发至参与训练的客户端；随后，客户端在本地基于各自的动态知识图谱数据集训练多粒度动态知识图谱嵌入模型（MDKGE），并上传训练后的模型参数至服务器；最终，服务器根据多粒度聚合规则对接收到的参数进行更新。MDKGE模型包括三个主要步骤：多粒度切割、多粒度链接和多粒度卷积。



图4-5-6 联邦多粒度动态知识图谱嵌入框架

在动态知识图谱中，每个三元组包含相应的时间信息，显示该事实在特定时间有效，不同粒度可包括年、月和日。多粒度动态知识图谱嵌入的目标是更精确地表征不同时间粒度下的实体和关系。模型通过多粒度切割，将实体向量、关系向量及时间向量根据年、月、日等粒度划分，并通过链接操作将切割后的向量按粒度结合，以实现同一粒度下的信息交互。接着，模型应用多粒度卷积方法，将多种时间粒度的信息嵌入知识图谱中，构成动态嵌入的四元组表示。

此外，在联邦训练过程中，客户端将MDKGE模型参数发送至服务器，服务器根据聚合规则更新全局模型参数。针对多粒度嵌入需求，提出了两种不同的聚合更新规则：（1）多粒度聚合规则，允许客户端根据特定时间粒度选择上传部分参数，以减少通讯量并增强嵌入信息的针对性；（2）经典联邦聚合规则，通过将客户端模型的所有参数平均聚合。这些规则不仅确保了模型的隐私性和安全性，同时也使得FedMDKGE能够在客户端之间实现有效的数据隐私保护和信息交互。

**5. 创新性**

创新点5（李弟诚）

（1）城市知识语义协同挖掘技术

1）多层次特征与粒化分析融合

该技术创新性地结合了深度学习的多层次特征学习和粒计算的多层次粒化分析，针对城市大数据中模态缺失和信息融合的挑战，通过多层知识粒度表示加强了对知识图谱的表征。这种多层次协同挖掘方式提升了知识图谱的语义丰富性和准确性。

2）跨时空数据的知识图谱融合表征

本项目提出了一种时空融合表征模型（RFP-KMN），该模型整合了时空数据与知识图谱嵌入技术，通过LSTM和2D-CNN自编码器提取时间序列和区域特征，形成时空特征的复合表征，并借助TransE模型构建知识三元组。此技术在时空数据挖掘方面有效增强了实体关系的语义表达力。

3）基于注意力机制的特征融合

本项目通过引入注意力机制实现区域特征与POI语义的深度融合。该机制在跨区域交通流量分析中通过动态关注特定区域的POI语义特征，增强了交通模式的嵌入表达，这种基于注意力的融合方法进一步提升了时空知识挖掘的精度与适应性。

4）知识图谱多语义表示与嵌入交互

在嵌入的切割模块中，本项目将实体和关系切分为多个子语义表示，通过绑定规则实现跨领域的多语义交互，以提取不同语义层次的实体关系。此创新的多语义交互机制有效解决了实体间语义丰富度不足的问题，提升了语义表示的细腻度和跨领域协同挖掘能力。

（2）城市知识动态演化发现技术

1）基于多时间粒度的知识动态演化模型

该模型在时间维度上引入了多粒度时间切片，使知识库能够在不同时间尺度上刻画城市数据动态。研究团队通过构建包含头实体、关系、尾实体和时间的知识四元组，并基于地铁流量数据进行建模，使用正弦波函数组合方式对时间信息进行多粒度表征，增强了未来流量预测的准确性。此项创新的核心在于，模型整合了多时间粒度，动态地关联和预测时空数据中的实体关系，有效提升了动态演化知识图谱的时空覆盖度。

2）联邦多粒度动态知识图谱嵌入模型

本项目在动态知识图谱嵌入中结合了联邦学习框架，实现了数据隐私保护和知识共享。该模型通过联邦学习的多轮训练过程，使不同客户端能够在各自的本地数据上进行训练，然后将模型参数汇总，服务器通过多粒度聚合规则更新全局模型。此项创新点在于，该模型在知识嵌入时融合了多粒度时间信息，并通过联邦学习的多粒度卷积方式，保障了跨领域知识的动态、分布式更新，提高了数据安全性和隐私性。

3）多粒度切割、链接与卷积的嵌入表征

本项目提出了一种多粒度切割方法，将知识图谱中的实体、关系和时间向量根据时间粒度进行切割，并通过链接操作实现不同粒度的交互，再利用多粒度卷积操作捕捉不同时间尺度下的信息特征。这项创新在于，通过多粒度嵌入切割和卷积方法，能够将知识四元组的时间维度转换为三行矩阵，生成多种特征向量，从而更精细地表征动态实体关系。

4）多粒度聚合与经典联邦聚合规则的结合

本项目设计了多粒度聚合规则，将多个卷积核参数在各个粒度上进行融合操作，从而保留各粒度的特征信息，避免信息丢失，增强知识四元组的表征能力。通过将该规则与经典联邦聚合规则结合，提升了多客户端联合模型的泛化性和预测效果。