



(12) 发明专利

(10) 授权公告号 CN 119988570 B

(45) 授权公告日 2025. 07. 11

(21) 申请号 202510462490.5

(22) 申请日 2025.04.14

(65) 同一申请的已公布的文献号

申请公布号 CN 119988570 A

(43) 申请公布日 2025.05.13

(73) 专利权人 集美大学

地址 361000 福建省厦门市集美区银江路
185号

(72) 发明人 许伟坚 方煌杰 杨长进 高志斌
赖联有 孙养龙

(74) 专利代理机构 厦门原创专利事务所(普通
合伙) 35101

专利代理师 黄巧香

(51) Int. Cl.

G06F 16/3329 (2025.01)

G06F 16/334 (2025.01)

G06F 16/36 (2019.01)

G06N 5/022 (2023.01)

G06N 5/04 (2023.01)

G06F 21/62 (2013.01)

G06F 21/60 (2013.01)

H04L 9/06 (2006.01)

H04L 67/12 (2022.01)

(56) 对比文件

CN 119520009 A, 2025.02.25

WO 2024240011 A1, 2024.11.28

审查员 卢明英

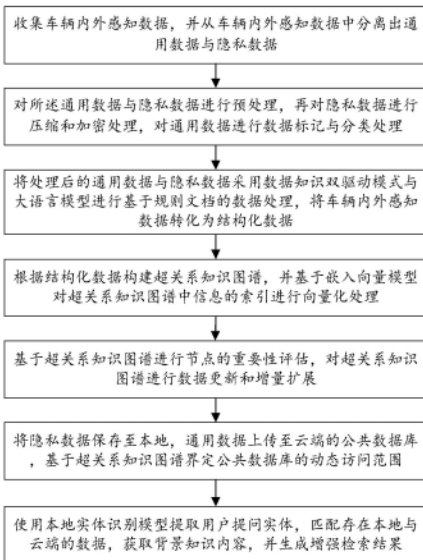
权利要求书4页 说明书11页 附图2页

(54) 发明名称

基于知识检索增强的车联网时空动态数据的
决策支持方法

(57) 摘要

本发明公开了一种基于知识检索增强的车联网时空动态数据的决策支持方法,包括:收集车辆内外感知数据,分离出通用数据与隐私数据;对数据进行预处理,对隐私数据进行压缩和加密处理;采用数据知识双驱动模式与大语言模型进行数据处理,转化为结构化数据;构建超关系知识图谱,基于嵌入向量模型进行向量化处理;基于超关系知识图谱进行节点的重要性评估,进行数据更新和增量;将隐私数据保存至本地,通用数据上传至云端的公共数据库,界定公共数据库的动态访问范围;使用本地实体识别模型提取用户提问实体,匹配存在本地与云端的数据,获取背景知识内容,并生成增强检索结果。本发明解决了车联网环境下决策支持信息的获取和数据的使用安全问题。



1.一种基于知识检索增强的车联网时空动态数据的决策支持方法,其特征在于,包括如下步骤:

步骤1、收集车辆内外感知数据,并从车辆内外感知数据中分离出通用数据与隐私数据;

步骤2、对所述通用数据与隐私数据进行预处理,再对隐私数据进行压缩和加密处理,对通用数据进行数据标记与分类处理;

步骤3、将处理后的通用数据与隐私数据采用数据知识双驱动模式与大语言模型进行基于规则文档的数据处理,将车辆内外感知数据转化为结构化数据;

步骤4、根据结构化数据构建超关系知识图谱,并基于嵌入向量模型对超关系知识图谱中信息的索引进行向量化处理;具体包括:

步骤41、基于结构化数据构建超关系知识图谱,将结构化数据中的实体和关系分别表示为超关系知识图谱中的节点和边,每个节点包含实体信息,边表示实体之间的关系;超关系由一个主三元组和若干辅助属性值描述组成,所述主三元组表示为(主体,关系,客体),辅助属性值描述以键值对的形式表示;所述超关系知识图谱表示为:

$$G = (V, E) = \sum (source(n), class(n), target(n), class(n), relation(e), Level(e))$$

其中, V 是节点集合,每个节点 $node(n) \in V$ 包含属性 $class(n)$, $class(n)$ 表示节点键值类别,以列表形式保存多对键值,其中, $class(n)$ 包括键 Key 以及值 $Value$;

E 是边集合,定义为 $E \subseteq V \times V$,即边是节点集合 V 中两个节点的有序对;边 $e \in E$ 连接节点 $source(n)$ 和节点 $target(n)$,边 e 包含属性 $relation(e)$ 和属性 $Level(e)$, $relation(e)$ 描述的是边的关系, $Level(e)$ 描述的是关系强度评级;

步骤42、对每个节点和边添加描述信息,使用键值对保存节点的补充信息;

步骤43、对于具有多个关系的节点,支持多边描述,并标记和区分相同节点与边的信息,评估不同信息的重要程度,并对信息的重要程度进行标记;

步骤44、通过超关系知识图谱捕捉上下文信息,增强超关系知识图谱的上下文理解能力,识别和整合实体间的多维关系,以及实体与关系的属性信息,最终得到一份具有丰富语义的超关系知识图谱;

步骤45、使用嵌入向量模型将超关系知识图谱内的键、节点与关系进行嵌入式向量编码,并将向量信息写入到各个键、节点与关系的列表内,完成索引的向量化;

步骤5、基于超关系知识图谱进行节点的重要性评估,对超关系知识图谱进行数据更新和增量扩展;具体包括:

步骤51、使用EasyGraph技术对超关系知识图谱进行分析;

步骤52、采用基于层间邻域信息熵的时序网络节点重要性评估方法,确定各个节点与边的重要程度,并将其写入 $Level(e)$ 中,结合用户的权重倾向用于最后信息匹配选取时的重要度评估;

步骤53、在已构建好的超关系知识图上进行增量扩展,采用基于规则文档进行再一轮数据处理,构成原超关系知识图谱和新超关系知识图谱,原超关系知识图谱作为本体图谱,新超关系知识图谱作为增量图谱;

步骤54、根据时序网络节点重要性评估方法界定实体对齐的范围,并根据本体图谱的规模进行分块大小的设置,根据实体对齐的范围和分块大小将本体图谱进行分块,形成多个节点块;

步骤55、以 $Level(e)$ 中标识的重要节点为中心,将本体图谱的节点块与增量图谱的节点进行匹配,获取相似度最高的节点块;

步骤56、在结构化数据中,使用大语言模型将 $class(n)$ 、 $relation(e)$ 和 $Level(e)$ 进行节点信息的智能融合, $source(n)$ 与 $target(n)$ 进行相同实体的知识融合,将增量图谱与本体图谱不匹配的部分作为新的结构化数据扩展到本体图谱上,完成关于扩展内容在本体图谱的增量生成;

步骤6、将隐私数据保存至本地,通用数据上传至云端的公共数据库,基于超关系知识图谱界定公共数据库的动态访问范围;

步骤7、使用本地实体识别模型提取用户提问实体,匹配存在本地与云端的数据,获取背景知识内容,并生成增强检索结果。

2. 如权利要求1所述的基于知识检索增强的车联网时空动态数据的决策支持方法,其特征在于,所述步骤1具体包括:

步骤11、收集车辆内外感知数据,所述车辆内外感知数据包括车辆的环境信息、车辆的状态信息、交通流量信息、车联网服务平台信息、车辆的位置信息、车辆归属地信息、车联网用户社区群组信息、停车场的布局信息、用户个人信息、车辆行驶轨迹信息以及车载音频交互信息;

步骤12、将所述车辆内外感知数据转换为用于预训练语言模型识别的文本数据,并使用结构化格式进行保存;

步骤13、使用基于深度学习的预训练语言模型进行预训练,并针对车联网中的隐私数据类型,通过冻结大模型修复小模型的方法冻结预训练语言模型的部分参数,只对关键层进行微调,实现预训练语言模型具备对数据分离脱敏的功能;

步骤14、使用微调后的预训练语言模型对车辆内外感知数据进行识别,并分离出通用数据和隐私数据,所述通用数据包括车辆的环境信息、车辆的状态信息、交通流量信息、车联网服务平台信息、车辆的位置信息、车辆归属地信息、车联网用户社区群组信息以及停车场的布局信息;所述隐私数据包括用户个人信息、车辆行驶轨迹信息以及车载音频交互信息。

3. 如权利要求1所述的基于知识检索增强的车联网时空动态数据的决策支持方法,其特征在于,所述步骤2具体包括:

步骤21、对所述通用数据与隐私数据进行清洗、缺失值处理和异常值处理;

步骤22、基于隐私数据构建语义知识库,在语义知识库中分析隐私数据的语义含义,并根据不同用户和场景的语义特征以及字符出现的频率,构建一个哈夫曼树,为每个字符分配一个编码进行压缩,并保存压缩后的隐私数据和编码表;

步骤23、通过差分隐私技术在隐私数据中加入噪声来掩盖原始数据的真实值;

步骤24、使用AES加密算法对隐私数据进行加密处理,并在加密处保留原文类别的加密标识;

步骤25、对通用数据进行数据标记与分类,以匹配不同情景下本地与云端的数据交互。

4. 如权利要求1所述的基于知识检索增强的车联网时空动态数据的决策支持方法,其特征在于,所述步骤3具体包括:

步骤31、在数据驱动部分,利用车辆内外感知数据来训练和优化大语言模型,实现大语言模型自动识别和提取车辆内外感知数据中的实体和关系;在知识驱动部分,引入专家的知识,设置规则文档,包括制作规则文档、限制大语言模型的输出结构、提供实体与关系的提取案例、关系重要性评级的评估案例以及类别属性判别的评估案例,以指导大语言模型的数据处理和关系评估,然后将数据驱动和知识驱动结合起来,形成数据知识双驱动模式;

步骤32、选取大语言模型,将大语言模型部署在云端,本地车辆使用API从云端进行大语言模型的调用;或者在车联网区块内部署大语言模型,使用私有内网通过端口转接方式进行大语言模型的调用;

步骤33、采用数据知识双驱动模式和大语言模型进行车辆内外感知数据的评估从实际的车辆内外感知数据中学习模式和规律,提供规则文档限制模型交互的输出内容与格式,制定文档分块的策略,利用先验知识来指导大语言模型的输出;

步骤34、根据车辆内外感知数据的信息量进行分块大小的决策,信息量大于预设值的车辆内外感知数据则使用小分块,信息量不大于预设值的车辆内外感知数据则使用大分块;

步骤35、通过分块数据进行基于规则文档的大语言模型交互后获得实体信息,提取出构建超关系知识图谱所需的实体信息,并转化为结构化数据。

5. 如权利要求4所述的基于知识检索增强的车联网时空动态数据的决策支持方法,其特征在于,所述步骤35具体包括:

步骤351、在使用大语言模型进行第一次交互时,记录信息来源的数据分块位置;

步骤352、根据数据的知识密度与分块大小进行知识的提取,完成第一次的大语言模型交互;

步骤353、经由第一次的大语言模型交互后,将提取的数据传送给大语言模型进行二次判决,评估超关系知识图谱内的节点与边的关系,以关系强度评级进行评级结果输出;

步骤354、对于关系强度评级中的边描述关系评级有所变化的数据进行第三次判决,提取第一次交互时记录的信息来源的数据分块位置与大语言模型进行再一次交互,获得实体信息,并转换为结构化数据,所述结构化数据包括实体、关系及其属性信息。

6. 如权利要求1所述的基于知识检索增强的车联网时空动态数据的决策支持方法,其特征在于,所述步骤52的时序网络节点重要性评估方法具体包括:

步骤521、根据邻居拓扑重叠系数 $c_i(t-1, t)$ 计算节点与其邻居持续出现的连边的层间相似性:

$$c_i(t-1, t) = \frac{\sum_j a_{ij}(t-1) a_{ij}(t)}{\sum_j a_{ij}(t-1)} \cdot \sum_j a_{ij}(t)$$

其中, i 表示第 i 个节点, j 表示第 j 个节点, t 表示第 t 个时间切片, $a_{ij}(t)$ 表示在第 t 个时间切片的邻接矩阵 $A^{(t)}$ 中节点 i 和节点 j 之间的边;

步骤522、根据不同时间切片之间的影响构建时序超邻接矩阵 $TSAM$;

$$TSAM = \begin{bmatrix} A^{(1)} & I^{(1,2)} & \dots & I^{(1,t)} \\ 0 & A^{(2)} & \dots & I^{(2,t)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A^{(t)} \end{bmatrix}$$

其中, $A^{(t)}$ 表示第 t 个时间切片的邻接矩阵, 且 t 为正整数, s 表示第 s 个时间切片, s 为正整数且 $s < t$, $I^{(s,t)}$ 表示第 s 个时间切片和第 t 个时间切片之间的影响矩阵, $TSAM$ 是 $N \times N$ 的时序超邻接矩阵, 用于描述在第 s 个时间切片和第 t 个时间切片之间各节点的关系或影响;

步骤523、基于所构建的时序超邻接矩阵 $TSAM$ 计算最大特征值对应的特征向量, 即各时间切片内各节点的特征向量中心性, 用于判断节点在图谱网络中的重要程度。

7. 如权利要求1所述的基于知识检索增强的车联网时空动态数据的决策支持方法, 其特征在于, 所述步骤6具体包括:

步骤61、将隐私数据作为本地超关系图谱文件保存至本地, 将通用数据上传至云端的公共数据库, 支持多车共同访问;

步骤62、通过对每个语义记录打上一个基于规则文档的权限许可范围, 对不同权限开放不同范围的语义库集合: 对于相近地域范围的邻居节点, 开放公开共享语义中关于本地地图上的交通地理信息; 对于同单位群组关系的邻居节点, 则开放共享单位园区内的地理标签、专属停车位以及邻居车辆的历史轨迹这些隐私数据以支持多车共同访问;

步骤63、通过构建的超关系知识图谱, 依据节点键值类别的各类键值对识别车辆间的聚类关系, 将具有相似特征或行为的车辆划分为同一个群体, 使得车辆在隐私安全范围内免授权互访, 且互访的数据都是与各自时空强相关的有效数据;

步骤64、管理员结合用户群组内的基础设施对云端数据进行载入和更新, 实现通用数据的共享和检索。

8. 如权利要求1所述的基于知识检索增强的车联网时空动态数据的决策支持方法, 其特征在于, 所述步骤7具体包括:

步骤71、基于本地实体识别模型提取问题的实体与表述;

步骤72、通过嵌入向量模型对问题的实体与表述进行向量化, 将文本信息转换为向量信息;

步骤73、通过超关系知识图谱中的聚类信息来识别车辆间的聚类关系, 动态的确定访问范围;

步骤74、将向量信息与本地超关系图谱文件和权限内云端的公共数据库中的索引向量进行相似性匹配, 得到背景知识内容;

步骤75、将匹配得到的背景知识内容与用户提问内容结合, 形成提示内容并与大语言模型进行交互;

步骤76、将大语言模型交互后的答复返回给本地, 并基于步骤2的加密流程反向解密, 得到最终的增强检索结果。

基于知识检索增强的车联网时空动态数据的决策支持方法

技术领域

[0001] 本发明涉及数据处理技术领域,尤其涉及一种基于知识检索增强的车联网时空动态数据的决策支持方法。

背景技术

[0002] 在数字化时代,车联网技术通过将车辆与互联网连接,实现了车辆间以及车辆与基础设施之间的信息交换,这不仅提高了驾驶安全性和效率,同时也带来了数据安全和隐私保护的挑战。在车联网技术中,数据安全问题尤为突出,在行驶过程中车辆会生成和交换大量的实时数据,包括车辆状态信息、环境感知数据以及用户私人数据。这些数据的隐私性要求我们必须采取有效的保护措施,以确保用户在享受智能决策支持的同时,其隐私和数据安全得到充分保障。

[0003] 在基于图谱的检索增强系统下,实现车辆信息的实时更新也是一个挑战,现有的检索增强系统大多不支持知识基于时序的增量生成,这意味着无法有效地处理和更新车辆数据。为了应对这一挑战,以便能够持续地、动态地更新超关系知识,确保数据库中车辆信息的准确性和有效性,需要提出一种基于知识检索增强的车联网时空动态数据的决策支持方法。

发明内容

[0004] 有鉴于此,本发明的目的在于提出一种基于知识检索增强的车联网时空动态数据的决策支持方法。

[0005] 为了实现上述的技术目的,本发明所采用的技术方案为:

[0006] 本发明提供了一种基于知识检索增强的车联网时空动态数据的决策支持方法,包括如下步骤:

[0007] 步骤1、收集车辆内外感知数据,并从车辆内外感知数据中分离出通用数据与隐私数据;

[0008] 步骤2、对所述通用数据与隐私数据进行预处理,再对隐私数据进行压缩和加密处理,对通用数据进行数据标记与分类处理;

[0009] 步骤3、将处理后的通用数据与隐私数据采用数据知识双驱动模式与大语言模型进行基于规则文档的数据处理,将车辆内外感知数据转化为结构化数据;

[0010] 步骤4、根据结构化数据构建超关系知识图谱,并基于嵌入向量模型对超关系知识图谱中信息的索引进行向量化处理;

[0011] 步骤5、基于超关系知识图谱进行节点的重要性评估,对超关系知识图谱进行数据更新和增量扩展;

[0012] 步骤6、将隐私数据保存至本地,通用数据上传至云端的公共数据库,基于超关系知识图谱界定公共数据库的动态访问范围;

[0013] 步骤7、使用本地实体识别模型提取用户提问实体,匹配存在本地与云端的数据,

获取背景知识内容,并生成增强检索结果。

[0014] 进一步的,所述步骤1具体包括:

[0015] 步骤11、收集车辆内外感知数据,所述车辆内外感知数据包括车辆的环境信息、车辆的状态信息、交通流量信息、车联网服务平台信息、车辆的位置信息、车辆归属地信息、车联网用户社区群组信息、停车场的布局信息、用户个人信息、车辆行驶轨迹信息以及车载音频交互信息;

[0016] 步骤12、将所述车辆内外感知数据转换为用于预训练语言模型识别的文本数据,并使用结构化格式进行保存;

[0017] 步骤13、使用基于深度学习的预训练语言模型进行预训练,并针对车联网中的隐私数据类型,通过冻结大模型修复小模型的方法冻结预训练语言模型的部分参数,只对关键层进行微调,实现预训练语言模型具备对数据分离脱敏的功能;

[0018] 步骤14、使用微调后的预训练语言模型对车辆内外感知数据进行识别,并分离出通用数据和隐私数据,所述通用数据包括车辆的环境信息、车辆的状态信息、交通流量信息、车联网服务平台信息、车辆的位置信息、车辆归属地信息、车联网用户社区群组信息以及停车场的布局信息;所述隐私数据包括用户个人信息、车辆行驶轨迹信息以及车载音频交互信息。

[0019] 进一步的,所述步骤2具体包括:

[0020] 步骤21、对所述通用数据与隐私数据进行清洗、缺失值处理和异常值处理;

[0021] 步骤22、基于隐私数据构建语义知识库,在语义知识库中分析隐私数据的语义含义,并根据不同用户和场景的语义特征以及字符出现的频率,构建一个哈夫曼树,为每个字符分配一个编码进行压缩,并保存压缩后的隐私数据和编码表;

[0022] 步骤23、通过差分隐私技术在隐私数据中加入噪声来掩盖原始数据的真实值;

[0023] 步骤24、使用AES加密算法对隐私数据进行加密处理,并在加密处保留原文类别的加密标识;

[0024] 步骤25、对通用数据进行数据标记与分类,以匹配不同情景下本地与云端的数据交互。

[0025] 进一步的,所述步骤3具体包括:

[0026] 步骤31、在数据驱动部分,利用车辆内外感知数据来训练和优化大语言模型,实现大语言模型自动识别和提取车辆内外感知数据中的实体和关系;在知识驱动部分,引入专家的知识,设置规则文档,包括制作规则文档、限制大语言模型的输出结构、提供实体与关系的提取案例、关系重要性评级的评估案例以及类别属性判别的评估案例,以指导大语言模型的数据处理和关系评估,然后将数据驱动和知识驱动结合起来,形成数据知识双驱动模式;

[0027] 步骤32、选取大语言模型,将大语言模型部署在云端,本地车辆使用API从云端进行大语言模型的调用;或者在车联网区块内部署大语言模型,使用私有内网通过端口转接方式进行大语言模型的调用;

[0028] 步骤33、采用数据知识双驱动模式和大语言模型进行车辆内外感知数据的评估从实际的车辆内外感知数据中学习模式和规律,提供规则文档限制模型交互的输出内容与格式,制定文档分块的策略,利用先验知识来指导大语言模型的输出;

[0029] 步骤34、根据车辆内外感知数据的信息量进行分块大小的决策,信息量大于预设值的车辆内外感知数据则使用小分块,信息量不大于预设值的车辆内外感知数据则使用大分块;

[0030] 步骤35、通过分块数据进行基于规则文档的大语言模型交互后获得实体信息,提取出构建超关系知识图谱所需的实体信息,并转化为结构化数据。

[0031] 进一步的,其特征就在于,所述步骤35具体包括:

[0032] 步骤351、在使用大语言模型进行第一次交互时,记录信息来源的数据分块位置;

[0033] 步骤352、根据数据的知识密度与分块大小进行知识的提取,完成第一次的大语言模型交互;

[0034] 步骤353、经由第一次的大语言模型交互后,将提取的数据传送给大语言模型进行二次判决,评估超关系知识图谱内的节点与边的关系,以关系强度评级进行评级结果输出;

[0035] 步骤354、对于关系强度评级中的边描述关系评级有所变化的数据进行第三次判决,提取第一次交互时记录的信息来源的数据分块位置与大语言模型进行再一次交互,获得实体信息,并转换为结构化数据,所述结构化数据包括实体、关系及其属性信息。

[0036] 进一步的,所述步骤4具体包括:

[0037] 步骤41、基于结构化数据构建超关系知识图谱,将结构化数据中的实体和关系分别表示为超关系知识图谱中的节点和边,每个节点包含实体信息,边表示实体之间的关系;超关系由一个主三元组和若干辅助属性值描述组成,所述主三元组表示为(主体,关系,客体),辅助属性值描述以键值对的形式表示;所述超关系知识图谱表示为:

[0038] $G = (V, E) = \sum (source(n), class(n), target(n), class(n), relation(e), Level(e))$

[0039] 其中, V 是节点集合,每个节点 $node(n) \in V$ 包含属性 $class(n)$, $class(n)$ 表示节点键值类别,以列表形式保存多对键值,其中, $class(n)$ 包括键 Key 以及值 $Value$;

[0040] E 是边集合,定义为 $E \subseteq V \times V$,即边是节点集合 V 中两个节点的有序对;边 $e \in E$ 连接节点 $source(n)$ 和节点 $target(n)$,边 e 包含属性 $relation(e)$ 和属性 $Level(e)$, $relation(e)$ 描述的是边的关系, $Level(e)$ 描述的是关系强度评级;

[0041] 步骤42、对每个节点和边添加描述信息,使用键值对保存节点的补充信息;

[0042] 步骤43、对于具有多个关系的节点,支持多边描述,并标记和区分相同节点与边的信息,评估不同信息的重要程度,并对信息的重要程度进行标记;

[0043] 步骤44、通过超关系知识图谱捕捉上下文信息,增强超关系知识图谱的上下文理解能力,识别和整合实体间的多维关系,以及实体与关系的属性信息,最终得到一份具有丰富语义的超关系知识图谱;

[0044] 步骤45、使用嵌入向量模型将超关系知识图谱内的键、节点与关系进行嵌入式向量编码,并将向量信息写入到各个键、节点与关系的列表内,完成索引的向量化。

[0045] 进一步的,所述步骤5具体包括:

[0046] 步骤51、使用EasyGraph技术对超关系知识图谱进行分析;

[0047] 步骤52、采用基于层间邻域信息熵的时序网络节点重要性评估方法,确定各个节点与边的重要程度,并将其写入 $Level(e)$ 中,结合用户的权重倾向用于最后信息匹配选取时

的重要度评估；

[0048] 步骤53、在已构建好的超关系知识图上进行增量扩展,采用基于规则文档进行再一轮数据处理,构成原超关系知识图谱和新超关系知识图谱,原超关系知识图谱作为本体图谱,新超关系知识图谱作为增量图谱；

[0049] 步骤54、根据时序网络节点重要性评估方法界定实体对齐的范围,并根据本体图谱的规模进行分块大小的设置,根据实体对齐的范围和分块大小将本体图谱进行分块,形成多个节点块；

[0050] 步骤55、以 $Level(e)$ 中标识的重要节点为中心,将本体图谱的节点块与增量图谱的节点进行匹配,获取相似度最高的节点块；

[0051] 步骤56、在结构化数据中,使用大语言模型将 $class(n)$ 、 $relation(e)$ 和 $Level(e)$ 进行节点信息的智能融合, $source(n)$ 与 $target(n)$ 进行相同实体的知识融合,将增量图谱与本体图谱不匹配的部分作为新的结构化数据扩展到本体图谱上,完成关于扩展内容在本体图谱的增量生成。

[0052] 进一步的,所述步骤52的时序网络节点重要性评估方法具体包括：

[0053] 步骤521、根据邻居拓扑重叠系数 $c_i(t-1,t)$ 计算节点与其邻居持续出现的连边的层间相似性：

$$[0054] \quad c_i(t-1,t) = \frac{\sum_j a_{ij}(t-1)a_{ij}(t)}{\sum_j a_{ij}(t-1)} \cdot \sum_j a_{ij}(t)$$

[0055] 其中, i 表示第 i 个节点, j 表示第 j 个节点, t 表示第 t 个时间切片, $a_{ij}(t)$ 表示在第 t 个时间切片的邻接矩阵 $A^{(t)}$ 中节点 i 和节点 j 之间的边；

[0056] 步骤522、根据不同时间切片之间的影响构建时序超邻接矩阵 $TSAM$ ：

$$[0057] \quad TSAM = \begin{bmatrix} A^{(1)} & I^{(1,2)} & \dots & I^{(1,t)} \\ 0 & A^{(2)} & \dots & I^{(2,t)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A^{(t)} \end{bmatrix}$$

[0058] 其中, $A^{(t)}$ 表示第 t 个时间切片的邻接矩阵,且 t 为正整数, s 表示第 s 个时间切片, s 为正整数且 $s < t$, $I^{(s,t)}$ 表示第 s 个时间切片和第 t 个时间切片之间的影响矩阵, $TSAM$ 是 $N \times N$ 的时序超邻接矩阵,用于描述在第 s 个时间切片和第 t 个时间切片之间各节点的关系或影响；

[0059] 步骤523、基于所构建的时序超邻接矩阵 $TSAM$ 计算最大特征值对应的特征向量,即各时间切片内各节点的特征向量中心性,用于判断节点在图谱网络中的重要程度。

[0060] 进一步的,所述步骤6具体包括：

[0061] 步骤61、将隐私数据作为本地超关系图谱文件保存至本地,将通用数据上传至云端的公共数据库,支持多车共同访问；

[0062] 步骤62、通过对每个语义记录打上一个基于规则文档的权限许可范围,对不同权限开放不同范围的语义库集合:对于相近地域范围的邻居节点,开放公开共享语义中关于本地地图上的交通地理信息;对于同单位群组关系的邻居节点,则开放共享单位园区内的地理标签、专属停车位以及邻居车辆的历史轨迹这些隐私数据以支持多车共同访问；

[0063] 步骤63、通过构建的超关系知识图谱,依据节点键值类别的各类键值对识别车辆间的聚类关系,将具有相似特征或行为的车辆划分为同一个群体,使得车辆在隐私安全范围内免授权互访,且互访的数据都是与各自时空强相关的有效数据;

[0064] 步骤64、管理员结合用户群组内的基础设施对云端数据进行载入和更新,实现通用数据的共享和检索。

[0065] 进一步的,所述步骤7具体包括:

[0066] 步骤71、基于本地实体识别模型提取问题的实体与表述;

[0067] 步骤72、通过嵌入向量模型对问题的实体与表述进行向量化,将文本信息转换为向量信息;

[0068] 步骤73、通过超关系知识图谱中的聚类信息来识别车辆间的聚类关系,动态的确定访问范围;

[0069] 步骤74、将向量信息与本地超关系图谱文件和权限内云端的公共数据库中的索引向量进行相似性匹配,得到背景知识内容;

[0070] 步骤75、将匹配得到的背景知识内容与用户提问内容结合,形成提示内容并与大语言模型进行交互;

[0071] 步骤76、将大语言模型交互后的答复返回给本地,并基于步骤2的加密流程反向解密,得到最终的增强检索结果。

[0072] 采用上述的技术方案,本发明与现有技术相比,其具有的有益效果为:

[0073] 本发明利用基于深度学习的预训练语言模型进行预训练,以捕捉车联网环境中数据的复杂性、多样性和时空相关性,针对隐私数据类型,使用标注数据集对模型进行微调,提高隐私数据的识别和抽取性能。对初步处理的隐私数据和通用数据进行保护,包括缺失值处理、噪声数据处理,确保数据的安全性和准确性,构建语义知识库,实现数据的压缩和隐私保护。结合数据处理规则和数据知识双驱动模式的大语言模型,将车辆数据转化为结构化数据,构建超关系知识图谱,其中包含节点和边的属性信息。构建基于BERT的嵌入向量模型,对车辆知识和通用知识进行编码,利用超关系知识图谱挖掘车联网中邻居节点间的时空相关性,引入时序网络节点重要性评估方法确定节点间的关系强度,完成知识增量生成。将通用数据上传到云端的公共数据库,实现多车共同访问和数据共享,通过这一系列的技术步骤,构建一个既能保护车辆和用户隐私,又能高效处理和利用车联网数据的智能系统。

[0074] 综上所述,本发明提出的基于知识检索增强的车联网时空动态数据的决策支持方法,将数据知识双驱动、增量的检索增强生成与知识图谱、语义知识库、超关系、节点重要性和数据的动态注意力机制与车联网需求结合,不仅有效解决了车联网中数据实时高效处理与安全性的问题,还为用户智能决策提供了新的思路和技术手段,对于推动车联网的智能化和安全性提升具有重要意义。

附图说明

[0075] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以

根据这些附图获得其他的附图。

[0076] 图1是本发明实施例提供的一种基于知识检索增强的车联网时空动态数据的决策支持方法的执行流程图。

[0077] 图2是本发明实施例提供的知识图谱的示意图。

[0078] 图3是本发明实施例提供的结构框架。

[0079] 图4是本发明实施例提供的增强检索示意图。

具体实施方式

[0080] 下面结合附图和实施例,对本发明作进一步的详细描述。特别指出的是,以下实施例仅用于说明本发明,但不对本发明的范围进行限定。同样的,以下实施例仅为本发明的部分实施例而非全部实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0081] 请参见图1-图4,本发明的一种基于知识检索增强的车联网时空动态数据的决策支持方法,包括如下步骤:

[0082] 步骤1、收集车辆内外感知数据,并从车辆内外感知数据中分离出通用数据与隐私数据;

[0083] 在本实施例中,所述步骤1具体包括:

[0084] 步骤11、收集车辆内外感知数据,所述车辆内外感知数据包括车辆的环境信息、车辆的状态信息、交通流量信息、车联网服务平台信息、车辆的位置信息、车辆归属地信息、车联网用户社区群组信息、停车场的布局信息、用户个人信息、车辆行驶轨迹信息以及车载音频交互信息;

[0085] 步骤12、将所述车辆内外感知数据转换为用于预训练语言模型识别的文本数据,并使用结构化格式进行保存;

[0086] 步骤13、使用基于深度学习的预训练语言模型进行预训练,以学习车联网时空动态复杂环境中数据的复杂性、多样性和时空相关性,并针对车联网中的隐私数据类型,通过冻结大模型修复小模型的方法冻结预训练语言模型的部分参数,只对关键层进行微调,在减少计算量的同时保持模型性能,实现预训练语言模型具备对数据分离脱敏的功能;使用标注数据集对模型进行微调,提高模型在隐私数据的识别和抽取性能;

[0087] 步骤14、使用微调后的预训练语言模型对车辆内外感知数据进行识别,并分离出通用数据和隐私数据,所述通用数据包括车辆的环境信息、车辆的状态信息、交通流量信息、车联网服务平台信息、车辆的位置信息、车辆归属地信息、车联网用户社区群组信息以及停车场的布局信息;所述隐私数据包括用户个人信息、车辆行驶轨迹信息以及车载音频交互信息。

[0088] 步骤2、对所述通用数据与隐私数据进行预处理,再对隐私数据进行压缩和加密处理,对通用数据进行数据标记与分类处理;

[0089] 在本实施例中,所述步骤2具体包括:

[0090] 步骤21、对所述通用数据与隐私数据进行清洗、缺失值处理和异常值处理,以提高数据质量,确保数据的安全性和准确性;

[0091] 数据清洗: 去除停用词、拼写校正、更正拼写错误,确保文本的准确性、去除特殊

字符和数字、清理文本中的非文本元素和不必要的数字、词干提取与词形还原。

[0092] 缺失值处理方法为众数插值,众数插值:对于分类标签或短语,如果存在缺失值,可以使用众数插值。例如,一个数据集中的标签是“A”、“B”和“C”,而某个标签缺失,使用出现频率最高的标签来填补。

[0093] 异常值检测方法使用IQR法。 $IQR = Q3 - Q1$,其中, $Q1$ 是第一四分位数, $Q3$ 是第三四分位数。如果数据点低于 $Q1 - 1.5 \times IQR$ 或高于 $Q3 + 1.5 \times IQR$,则认为是异常值。

[0094] 步骤22、基于隐私数据构建语义知识库,在语义知识库中分析隐私数据的语义含义,并根据不同用户和场景的语义特征以及字符出现的频率,构建一个哈夫曼树,为每个字符分配一个编码进行压缩,并保存压缩后的隐私数据和编码表,以便解压缩时使用;

[0095] 步骤23、通过差分隐私技术在隐私数据中加入噪声来掩盖原始数据的真实值;这样即使数据被泄露,也无法直接识别出原始数据,保护数据的原始性质不被直接揭露,确保数据传输和存储的安全性,实现一种个性化的隐私保护策略。

[0096] 步骤24、使用AES加密算法对隐私数据进行加密处理,并在加密处保留原文类别的加密标识;确保数据在共享和使用过程中的安全性和隐私性;

[0097] 加密算法(AES): $C = E_k(P)$ 其中, C 是密文, E_k 是使用密钥 k 的加密函数, P 是明文。进行数据的加密和清洗,用于后续的本地与外部加密索引,防止隐私数据的泄露。

[0098] 步骤25、对通用数据进行数据标记与分类,以匹配不同情景下本地与云端的数据交互。

[0099] 步骤3、将处理后的通用数据与隐私数据采用数据知识双驱动模式与大语言模型进行基于规则文档的数据处理,将车辆内外感知数据转化为结构化数据;

[0100] 在本实施例中,所述步骤3具体包括:

[0101] 步骤31、在数据驱动部分,利用车辆内外感知数据来训练和优化大语言模型,实现大语言模型自动识别和提取车辆内外感知数据中的实体和关系;在知识驱动部分,引入专家的知识,设置规则文档,包括制作规则文档、限制大语言模型的输出结构、提供实体与关系的提取案例、关系重要性评级的评估案例以及类别属性判别的评估案例,以指导大语言模型的数据处理和关系评估,然后将数据驱动和知识驱动结合起来,形成数据知识双驱动模式;

[0102] 步骤32、选取大语言模型,将大语言模型部署在云端,本地车辆使用API从云端进行大语言模型的调用;或者在车联网区块内部署大语言模型,使用私有内网通过端口转接方式进行大语言模型的调用;

[0103] 步骤33、采用数据知识双驱动模式和大语言模型进行车辆内外感知数据的评估从实际的车辆内外感知数据中学习模式和规律,提供规则文档限制模型交互的输出内容与格式,制定文档分块的策略,利用先验知识来指导大语言模型的输出;

[0104] 步骤34、根据车辆内外感知数据的信息量进行分块大小的决策,信息量大于预设值的车辆内外感知数据则使用小分块,信息量不大于预设值的车辆内外感知数据则使用大分块,提升大语言模型的信息抽取效率与可信度;

[0105] 步骤35、通过分块数据进行基于规则文档的大语言模型交互后获得实体信息,提

取出构建超关系知识图谱所需的实体信息,并转化为结构化数据。

[0106] 在本实施例中,所述步骤35具体包括:

[0107] 步骤351、在使用大语言模型进行第一次交互时,记录信息来源的数据分块位置;

[0108] 步骤352、根据数据的知识密度与分块大小进行知识的提取,完成第一次的大语言模型交互;

[0109] 步骤353、经由第一次的大语言模型交互后,将提取的数据传送给大语言模型进行二次判决,评估超关系知识图谱内的节点与边的关系,以关系强度评级进行评级结果输出;

[0110] 步骤354、对于关系强度评级中的边描述关系评级有所变化的数据进行第三次判决,提取第一次交互时记录的信息来源的数据分块位置与大语言模型进行再一次交互,获得实体信息,并转换为结构化数据,所述结构化数据包括实体、关系及其属性信息。

[0111] 步骤4、根据结构化数据构建超关系知识图谱,利用超关系知识图谱实现车联网中邻居节点间数据的时空相关性的超关系挖掘,并基于嵌入向量模型对超关系知识图谱中信息的索引进行向量化处理;

[0112] 在本实施例中,所述步骤4具体包括:

[0113] 步骤41、基于结构化数据构建超关系知识图谱,将结构化数据中的实体和关系分别表示为超关系知识图谱中的节点和边,每个节点包含实体信息,边表示实体之间的关系;所述超关系知识图谱如图2所示,图2中的圆圈表示节点,线条表示边;所述超关系知识图谱表示为:

[0114] $G = (V, E) = \sum (source(n), class(n), target(n), class(n), relation(e), Level(e))$

[0115] 其中, V 是节点集合,每个节点 $node(n) \in V$ 包含属性 $class(n)$, $class(n)$ 表示节点键值类别,以列表形式保存多对键值,其中, $class(n)$ 包括键 Key 以及值 $Value$;

[0116] E 是边集合,定义为 $E \subseteq V \times V$,即边是节点集合 V 中两个节点的有序对;边 $e \in E$ 连接节点 $source(n)$ 和节点 $target(n)$,边 e 包含属性 $relation(e)$ 和属性 $Level(e)$, $relation(e)$ 描述的是边的关系, $Level(e)$ 描述的是关系强度评级;

[0117] 其中,超关系由一个主三元组和若干辅助属性值描述组成,所述主三元组表示为(主体,关系,客体),分别由步骤4的 $(source(n), relation(e), target(n))$ 表示,而辅助属性值描述以 $class(n)$ 键值对的形式表示。通过引入超关系的概念,将实体之间的关系扩展到可以包含多个实体的复杂关系,这涉及到识别和整合实体间的多维关系,以及实体与关系的属性信息。

[0118] 其中,超关系格式如下:如果有一条知识为:”张三参加了位于北京国际会议中心2024年11月11日的会议,参加这次会议的原因是年度总结。”获取超关系内容如下表:

[0119]

source(n)	class(n)	target(n)	class(n)	relation(e)
张三	键: “时间”, 值: “2024年11月11日” 键: “地点”, 值: “北京国际会议中心” 键: “原因”, 值: “年度总结”	北京国际会议中心	键: “时间”, 值: “2024年11月11日” 键: “人员”, 值: “张三”	张三参加了会议

[0120] 步骤42、对每个节点和边添加描述信息,使用键值对保存节点的补充信息;

[0121] 步骤43、对于具有多个关系的节点,支持多边描述,并标记和区分相同节点与边的信息,评估不同信息的重要程度,并对信息的重要程度进行标记;

[0122] 步骤44、通过超关系知识图谱捕捉上下文信息,如源文档、时间限定词和置信分数,为系统提供更细致的知识表示,这些信息将被用于增强超关系知识图谱的上下文理解能力,识别和整合实体间的多维关系,以及实体与关系的属性信息,最终得到一份具有丰富语义的超关系知识图谱;实现车联网中邻居节点之间数据的时空相关性的超关系挖掘;

[0123] 步骤45、使用嵌入向量模型(基于BERT的嵌入向量模型)将超关系知识图谱内的键、节点与关系进行嵌入式向量编码,并将向量信息写入到各个键、节点与关系的列表内,完成索引的向量化。结合超关系知识图谱中的结构化数据

source(n)、target(n)、class(n)、relation(e)、Level(e),使用基于BERT的嵌入向量模型构建一个更全面的向量表示,不仅包含了文本的语义信息,还融入了实体间的关系和键值对等属性信息,为构建索引和知识图谱的节点描述列表提供了更丰富的结构化数据。

[0124] 步骤5、基于超关系知识图谱进行节点的重要性评估,对超关系知识图谱进行数据更新和增量扩展;

[0125] 在本实施例中,所述步骤5具体包括:

[0126] 步骤51、使用EasyGraph技术对超关系知识图谱进行分析,利用其混合编程的特点实现对图谱处理与图谱分析的加速,有效地管理和查询图谱内车辆的实时状态、行驶路径、环境交互等信息,结合用户决策需求,提高决策支持的响应速度和决策效率;

[0127] 步骤52、采用基于层间邻域信息熵的时序网络节点重要性评估方法,确定各个节点与边的重要程度,并将其写入Level(e)中,结合用户的权重倾向用于最后信息匹配选取时的重要度评估;准确地衡量节点在网络中的影响力;

[0128] 步骤53、在已构建好的超关系知识图上进行增量扩展,采用基于规则文档进行再一轮数据处理,构成原超关系知识图谱和新超关系知识图谱,原超关系知识图谱作为本体图谱,新超关系知识图谱作为增量图谱;

[0129] 步骤54、根据时序网络节点重要性评估方法界定实体对齐的范围,并根据本体图谱的规模进行分块大小的设置,根据实体对齐的范围和分块大小将本体图谱进行分块,形成多个节点块;

[0130] 步骤55、以 $Level(e)$ 中标识的重要节点为中心,将本体图谱的节点块与增量图谱的节点进行匹配,获取相似度最高的节点块;

[0131] 步骤56、在结构化数据中,使用大语言模型将 $class(n)$ 、 $relation(e)$ 和 $Level(e)$ 进行节点信息的智能融合, $source(n)$ 与 $target(n)$ 进行相同实体的知识融合,将增量图谱与本体图谱不匹配的部分作为新的结构化数据扩展到本体图谱上,完成关于扩展内容在本体图谱的增量生成。

[0132] 在本实施例中,所述步骤52的时序网络节点重要性评估方法具体包括:

[0133] 步骤521、根据邻居拓扑重叠系数 $c_i(t-1,t)$ 计算节点与其邻居持续出现的连边的层间相似性:

$$[0134] \quad c_i(t-1,t) = \frac{\sum_j a_{ij}(t-1)a_{ij}(t)}{\sum_j a_{ij}(t-1)} \cdot \sum_j a_{ij}(t)$$

[0135] 其中, i 表示第 i 个节点, j 表示第 j 个节点, t 表示第 t 个时间切片, $a_{ij}(t)$ 表示在第 t 个时间切片的邻接矩阵 $A^{(t)}$ 中节点 i 和节点 j 之间的边;

[0136] 步骤522、根据不同时间切片之间的影响构建时序超邻接矩阵 $TSAM$:

$$[0137] \quad TSAM = \begin{bmatrix} A^{(1)} & I^{(1,2)} & \dots & I^{(1,t)} \\ 0 & A^{(2)} & \dots & I^{(2,t)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A^{(t)} \end{bmatrix}$$

[0138] 其中, $A^{(t)}$ 表示第 t 个时间切片的邻接矩阵,且 t 为正整数, s 表示第 s 个时间切片, s 为正整数且 $s < t$, $I(s,t)$ 表示第 s 个时间切片和第 t 个时间切片之间的影响矩阵, $TSAM$ 是 $N \times N$ 的时序超邻接矩阵,用于描述在第 s 个时间切片和第 t 个时间切片之间各节点的关系或影响;

[0139] 步骤523、基于所构建的时序超邻接矩阵 $TSAM$ 计算最大特征值对应的特征向量,即各时间切片内各节点的特征向量中心性,用于判断节点在图谱网络中的重要程度。

[0140] 步骤6、将隐私数据保存至本地,通用数据上传至云端的公共数据库,基于超关系知识图谱界定公共数据库的动态访问范围;

[0141] 在本实施例中,所述步骤6具体包括:

[0142] 步骤61、将隐私数据作为本地超关系图谱文件保存至本地,将通用数据上传至云端的公共数据库,支持多车共同访问;

[0143] 步骤62、通过对每个语义记录打上一个基于规则文档的权限许可范围,对不同权限开放不同范围的语义库集合:对于相近地域范围的邻居节点,开放公开共享语义中关于本地地图上的交通地理信息;对于同单位群组关系的邻居节点,则开放共享单位园区内的地理标签、专属停车位以及邻居车辆的历史轨迹这些隐私数据以支持多车共同访问;

[0144] 步骤63、通过构建的超关系知识图谱,依据节点键值类别的各类键值对识别车辆间的聚类关系,将具有相似特征或行为的车辆划分为同一个群体,使得车辆在隐私安全范围内免授权互访,且互访的数据都是与各自时空强相关的有效数据,以更精准地控制数据的共享范围,更高效地实现有用数据的快速检索;

[0145] 步骤64、管理员结合用户群组内的基础设施对云端数据进行载入和更新,实现通

用数据的共享和检索。这样既保护车辆的隐私,又实现通用知识的共享和检索,并极大地提升了系统的二次知识构建速率。

[0146] 步骤7、使用本地实体识别模型提取用户提问实体,匹配存在本地与云端的数据,获取背景知识内容,并生成增强检索结果。

[0147] 在本实施例中,所述步骤7具体包括:

[0148] 步骤71、基于本地实体识别模型提取问题的实体与表述;

[0149] 步骤72、通过嵌入向量模型对问题的实体与表述进行向量化,将文本信息转换为向量信息;

[0150] 步骤73、通过超关系知识图谱中的聚类信息来识别车辆间的聚类关系,动态的确定访问范围;

[0151] 步骤74、将向量信息与本地超关系图谱文件和权限内云端的公共数据库中的索引向量进行相似性匹配,得到背景知识内容;这一过程考虑了节点的重要性和上下文信息,以调取最相关的背景知识。提高决策背景支持的精确度;

[0152] 步骤75、将匹配得到的背景知识内容与用户提问内容结合,形成提示内容并与大语言模型进行交互;

[0153] 步骤76、将大语言模型交互后的答复返回给本地,并基于步骤2的加密流程反向解密,得到最终的增强检索结果。完成数据使用闭环,满足用户在车联网环境下的数据使用安全以及决策支持信息的智能获取需求。

[0154] 以上所述仅为本发明的部分实施例,并非因此限制本发明的保护范围,凡是利用本发明说明书及附图内容所作的等效装置或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本发明的专利保护范围内。

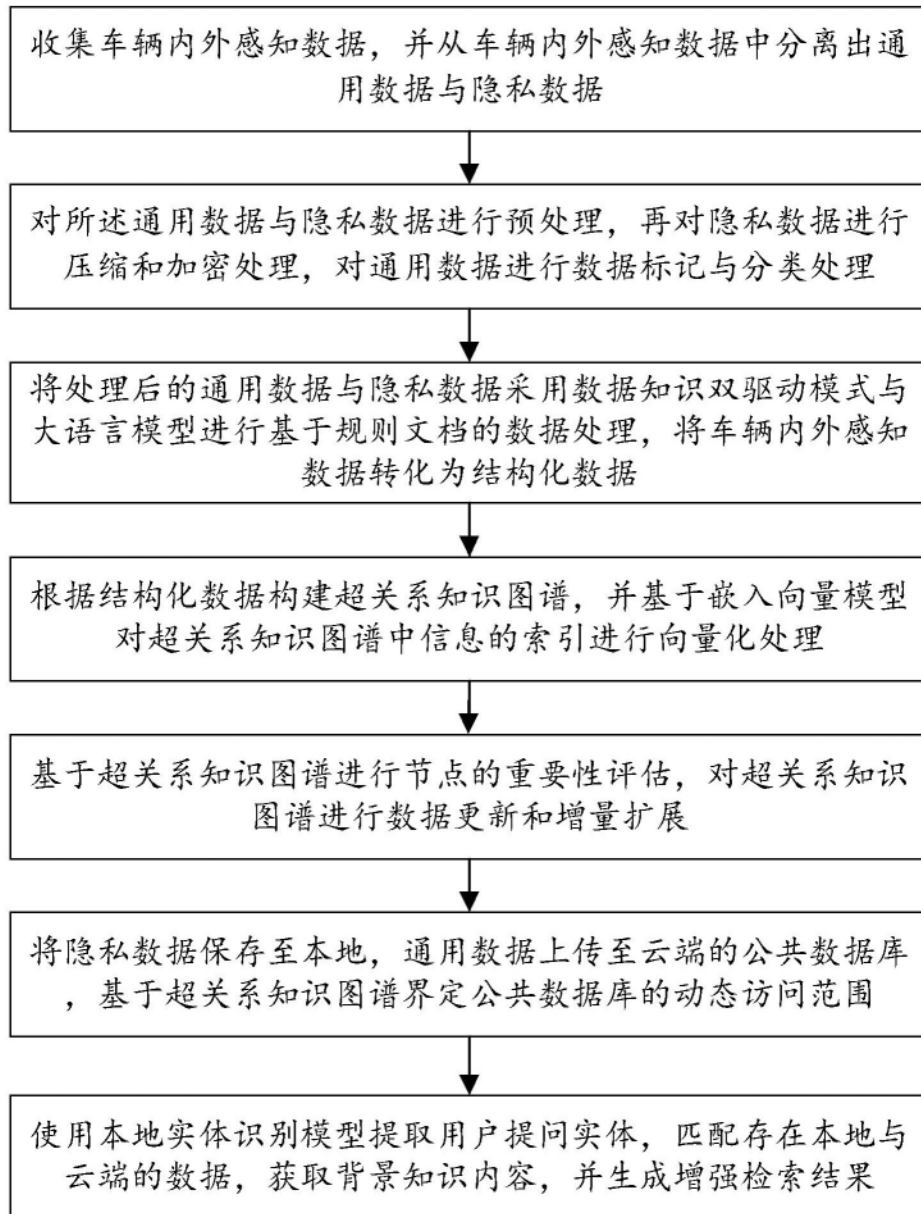


图1

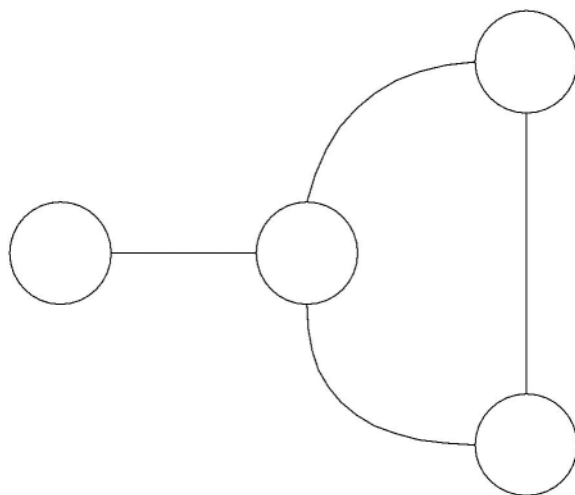


图2

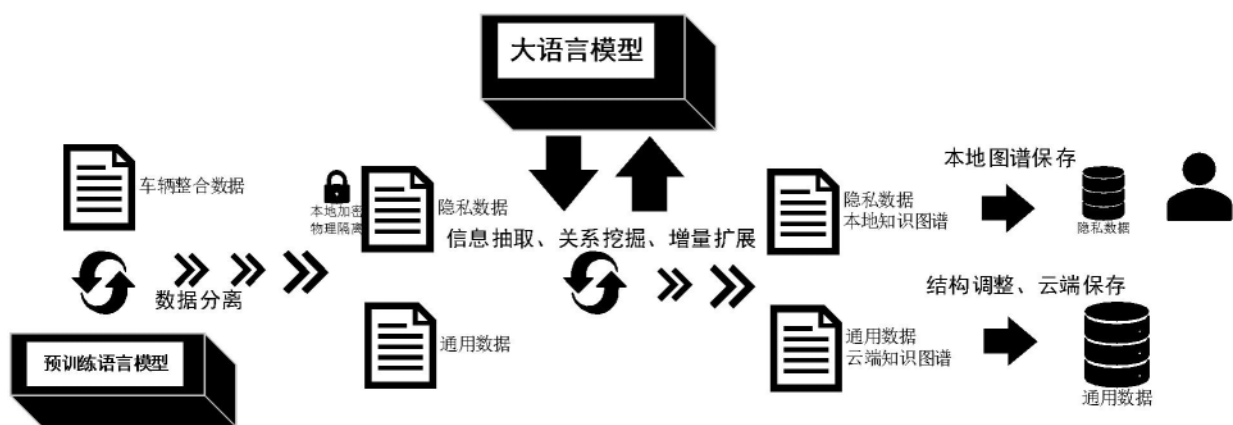


图3

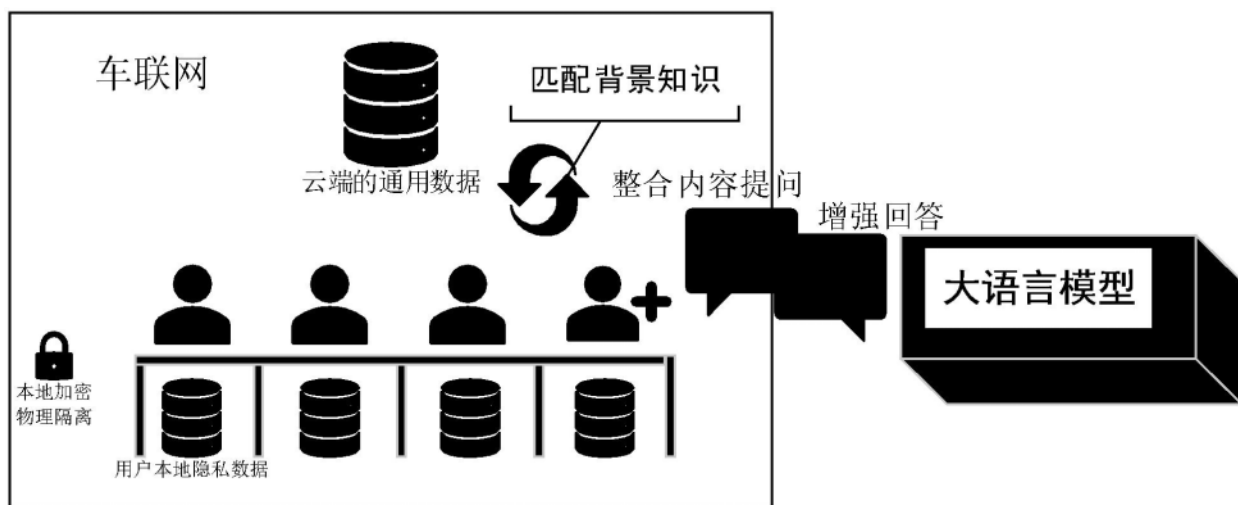


图4