

基于时态知识图谱的车站客流组织模型*

孟 歌

(中国铁道科学研究院集团有限公司电子计算技术研究所, 100081, 北京)

摘 要 [目的]城市轨道交通车站存在客流压力逐渐增大、智能化手段不足等问题。为提升车站的客流组织效率和智能化水平,需采用更先进的信息化方法和手段。[方法]阐述了 TKG(时态知识图谱)的发展情况及特征。引入车站客流的时态信息,构建了基于 TKG 的车站客流组织模型。将该模型应用于北京地铁 2 号线某车站的客流组织中,建立了该站的 TKG 模型,通过 ArangoDB 数据库软件进行图形化展示,以研究站内客流组织的动态演化过程。[结果及结论]该模型可为车站客流组织提前预判、快速响应、有效实施、智能预警等提供信息化手段及技术支撑。

关键词 城市轨道交通;车站;客流组织;时态知识图谱
中图分类号 U293.1

DOI:10.16037/j.1007-869x.2024.09.017

Research on Station Passenger Flow Organization Model Based on Temporal Knowledge Graph

MENG Ge

(Institute of Computing Technology, China Academy of Railway Sciences Co., Ltd., 100081, Beijing, China)

Abstract [Objective] Urban rail transit stations are facing the challenges of increasing passenger flow pressure and insufficient intelligent means. To improve the efficiency and intelligent level of passenger flow organization in station, it is necessary to adopt more advanced informational method and means. [Method] The development and characteristics of temporal knowledge graph (TKG) are described. By introducing the temporal information of the station passenger flow, a passenger flow organization model based on TKG is established. The model is applied to a station passenger flow organization on Beijing Metro Line 2, and the TKG model for this station is established. A graphical display is performed using ArangoDB database software to study the dynamic evolution process of the station passenger flow organization. [Result & Conclusion] The model can provide informational means and technical support for the advanced prediction, rapid response, effective implementation and intelligent early warning in the station passen-

ger flow organization.

Key words urban rail transit; station; passenger flow organization; TKG

KG(知识图谱)具有信息处理速度快、关联查询效率高、语义处理能力强等优点,是许多智能应用的核心,适用于客流组织文本等非结构化数据的处理和展示。目前,KG 已经被广泛应用于风险控制^[1]、智能问答^[2]、运维知识管理^[3]等领域。然而,传统 KG 常常被用作静态知识的表达,较少考虑知识的时间因素。

城市轨道交通车站客流的变化速度快,具有极强的时效性,只有将站内客流组织措施与时效状态下的客流紧密结合,才能达到快速、有效组织客流的目标。因此,传统 KG 难以对站内客流进行描述,需要引入客流的时态信息,构建基于 TKG(时态知识图谱)的车站客流组织模型。本文将该模型应用于北京地铁 2 号线某车站的客流组织中,通过 ArangoDB 数据库软件进行图像可视化,以研究站内客流组织的动态演化过程。

1 TKG 概述

KG 在 2012 年被 Google 提出^[4]。从 2013 年开始,KG 在科研、工业等领域得以应用。KG 具有强大的语义处理和互联组织能力,是通过图的方式表达实体、概念及两者关系的一种方法。相较于静态的 KG,TKG 具有更强大的动态演化能力,尤其适用于时间要素与应用场景紧密耦合的情况。在城市轨道交通车站客流组织中,时效性尤为重要,可应用 TKG 进行站内客流组织。

相较于 KG 的传统三元组,TKG 增加了时间单元,为四元组形式: $G = \{T, E, R, F\}$ 。其中: T 代表时间区间集合; E 代表实体集合; R 代表关系类别集合; F 代表知识集合。设 TKG 中的知识为时态知识

* 中国铁道科学研究院集团有限公司科研项目(2022YJ079);中国国家铁路集团有限公司科技研究开发计划项目(P2021X009)

f , 则有:

$$\text{s. t. } \begin{cases} f(t, m, r, n) \in F \\ t \in T \\ r \in R \\ m, n \in E \end{cases} \quad (1)$$

式中:

t ——知识的时间内容;

r ——知识的关系类别;

m, n ——知识的头、尾实体。

设 t_i 为第 i 个三元组生效时间点, q 为三元组的数量, 则有:

$$t = \{t_i \mid t_1 \leq t_i \leq t_q\} \quad (2)$$

在获取四元组的基础上, 为了保证 TKG 的可靠性和完整性, 需要利用知识推理对未知的内容进行补全和预测。补全是通过挖掘既有知识 $f(t, m, r, n)$ 中隐含的规律, 对其中的未知实体 f 或未知关系 r 进行推理, 以补全图谱中缺失的内容。预测是根据车站客流组织的历史知识, 提取事件演化特征, 预测客流组织知识在未来一段时间内的有效性, 并将知识转化为对时间单元的预测。三元组 (m, r, n) 在时间点 t_i 发生概率 P 的计算式为:

$$P(m, r, n) \Rightarrow t_i \quad (3)$$

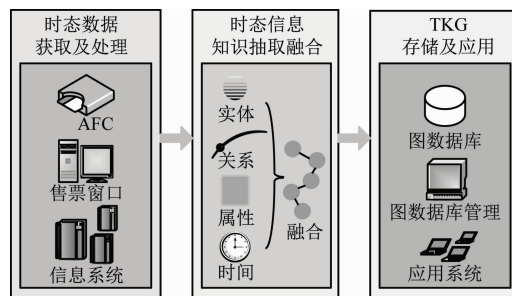
2 车站客流组织 TKG

文本是知识表达的载体之一。客流组织文本是城市轨道交通车站构建客流组织 TKG 的主要数据源, 需要进行自然语言处理。此外, 城市轨道交通车站的时态信息往往是以结构化数据形式存在的, 主要包括进站量、出站量、站内客流等实时数据, 这些数据仅需要确认四元组关系。

车站客流组织 TKG 构建的主要步骤包括: 时态数据获取及处理、时态信息知识抽取及融合、TKG 存储及应用。城市轨道交通车站客流组织 TKG 框架示意图如图 1 所示。

2.1 时态数据获取及处理

围绕车站客流组织自然语言文本及实时客流数据开展时态数据获取工作, 以得到知识元组的标准化结构。车站客流组织文本通常为结构化或半结构化的数据, 以业务管理部门收集为主, 业务管理系统收集为辅。车站客流数据通常通过 AFC、窗口售票等途径获取, 具有时效性强、非线性波动等特点。通过数据降噪、减维、清洗等手段, 对原始数据进行处理, 以提高数据整体质量。



注: AFC—自动售检票。

图 1 车站客流组织 TKG 框架示意图

Fig. 1 TKG framework of station passenger flow organization

2.2 时态信息知识抽取融合

时态信息知识抽取主要包括“时态、实体、关系、属性”4 个核心内容, 实体信息的抽取是其中最重要的环节。基于领域知识库采用自动化语料标注方法, 可自动、快速地对实体信息进行抽取。实体信息的自动化标注克服了以往人工字典定义标注的局限性及时间限制。

在实际应用环境中, 车站客流组织文本对时态的描述常具有歧义, 即: 同一个时态实体可能对应多个时刻, 同一个时刻也可能指代多个时态实体。针对这种情况, 需要在知识库中进行语义消歧, 从而达到知识融合的目的。此外, 车站客流组织文本中的四元组往往存在部分信息缺失的情况, 需要通过知识推理, 补全缺失内容, 进而达到提升知识可用性和完整性的目的。

2.3 TKG 存储及应用

KG 存储通常采用基于 RDF(资源描述框架) 及图数据库的存储方式。ArangoDB 数据库是一个兼容键/值对、图、文档 3 个类别模型的数据库, 具有统一的数据库查询语言, 可支持这 3 个类别模型的组合使用。其中, 图数据库以图的逻辑和联系来存储、读取数据, 实体和关系信息都具有属性, 数据的获取效率较高。

将 TKG 应用于车站客流组织中, 可以克服传统客流组织方法时效性低、响应速度慢等缺点。TKG 在车站智能决策、客流组织分析、智能预警等场景中具有较广阔的应用前景。

3 构建车站客流组织 TKG

3.1 车站客流组织 TKG 的架构

TKG 从构建方向上可以分为“由顶向下”和“由底向上”两种。在具有城市轨道交通领域知识

图谱的前提下,车站客流组织 TKG 的模式层宜采用“由顶向下”的方法构建,即逐层细化,由顶层向下依次定义概念实体、属性、层级关系与时态等,进而形成客流组织概念基本架构。数据层宜采取“由底向上”的构建方法,即采用从具体到抽象的技术路线。

由于客流组织和客流数据信息有多个数据源(如数据库、管理文档等),应先进行数据的结构化,抽取实体信息,并建立各实体信息间的关联。然后对从各渠道获取到的知识进行消歧和融合,将客流组织措施等要素进行分解。在客流组织实体与客流组织概念间建立映射关系,进而获得模式层与数据层的映射关系,完成客流组织 TKG 的构建。最后,通过 ArangoDB 数据库实现客流组织 TKG 的图形化展示。车站客流组织 TKG 构建流程如图 2 所示。

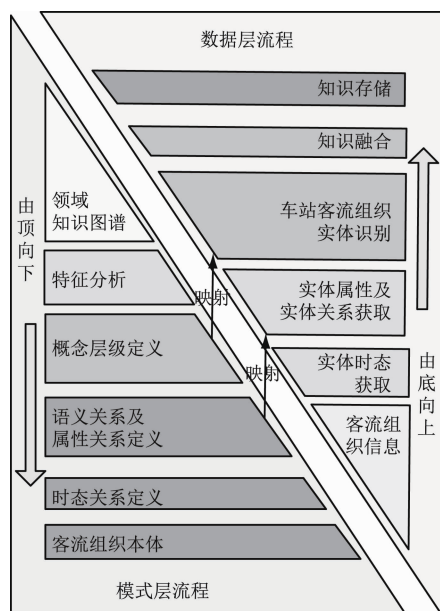


图2 车站客流组织 TKG 构建流程

Fig.2 TKG construction process of the station passenger flow organization

在构建车站客流组织 TKG 的过程中,TKG 的本体是客流组织的集合,本体模型的主要元素包括概念定义、实体属性、实体间关系、时态关系 4 个部分。

1) 概念定义指城市轨道交通车站客流组织领域中各类实体的集合,包括部门、人员、地点、事件、措施等。

2) 实体属性指各种属性的集合,包括部门属

性、人员属性、事件属性、措施属性等。

3) 实体间关系指所有关系的集合,如隶属关系、“概念-实体/属性”关系等。

4) 时态关系指包含时间因素的实体关系,如“概念-时间-实体/属性”等随着时间变化的关系。

城市轨道交通车站客流组织 TKG 共计包含 7 类实体、5 类实体属性、10 类实体关系及 6 类时态关系,其部分信息如表 1 所示。

表1 车站客流组织 TKG 的部分信息

Tab.1 Part of TKG information for the station passenger flow organization

项目	内容	备注
概念定义	部门	描述车站部门类型
	人员	描述车站人员类型
	事件	描述客流组织事件类型
	措施	描述客流组织措施类型
实体属性	部门属性	描述车站部门属性情况,如部门名称、部门任务等
	人员属性	描述人员属性情况,如姓名、性别、职务等
	事件属性	描述事件属性情况,如事件名称、事件级别等
	措施属性	描述措施属性情况,如措施名称、措施内容等
实体关系	接到	描述人员与部门的关系
	处理	描述人员与事件的关系,
	请求	描述部门与部门的关系
	位于	描述事件与地点的关系
	立即	描述时态瞬时生效
	5 min 内	描述时态从即时起的 5 min 内生效
时态关系	24 h	描述时态从即时起的 24 h 内生效
	在某个时刻	描述时态在某个时刻生效

采用 ALBERT (轻量双向编码器表征)-BiLSTM(双向长短期记忆网络)-Attention (注意力)-CRF(条件随机场)架构,构建车站客流组织 TKG。其中:上下文信息的结合采用 ALBERT 模型,该方法是轻量级的 BERT(双向编码器表征)模型,更适用于城市轨道交通车站客流组织场景,能够将对象以向量形式进行计算,并获得更全面的客流组织文本信息;数据训练采用 BiLSTM 方法,该方

法可以结合上下文输出,解决了梯度问题;Attention 机制可通过降噪来提高特征抽取的准确性;CRF 模型可以解决相邻标签问题,并根据概率排序,进而提高整个框架对实体及关系的抽取性能。

车站客流组织实体及关系抽取结构示意图如图 3 所示。在对客流组织信息进行收集后,将文本转化为向量序列 C (C 为文本序列)输入 ALBERT 中,则可获得向量序列 L (L 为降维序列); L 经过特征提取后获得特征向量 M ,将 M 输入 BiLSTM 中,

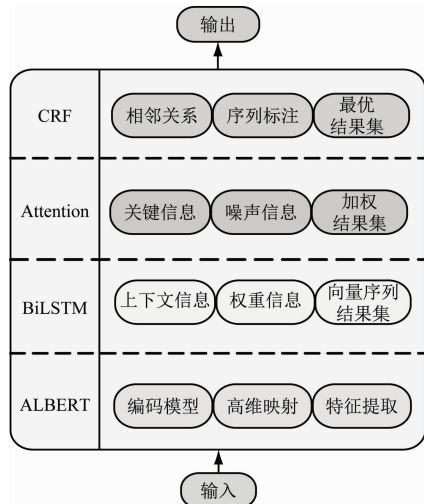


图 3 车站客流组织实体及关系抽取结构示意图

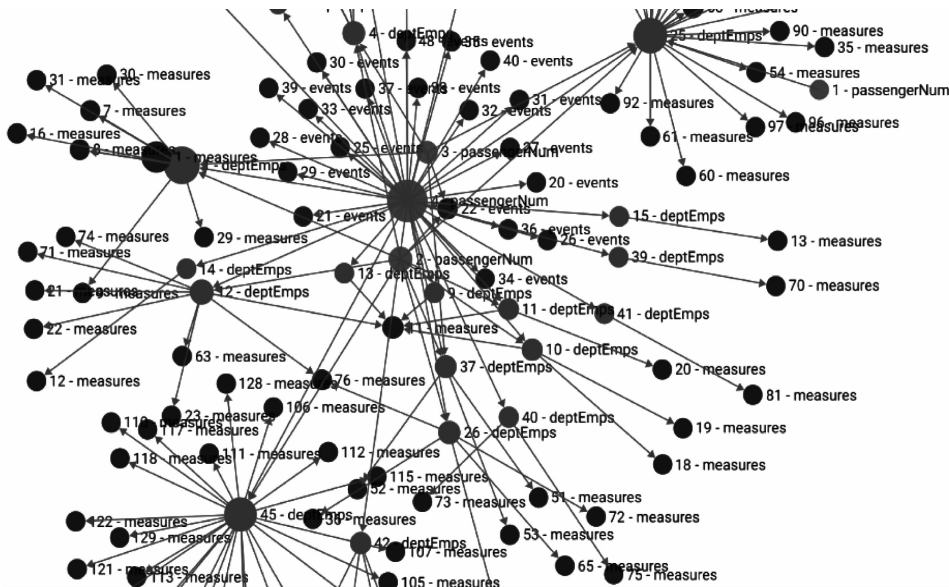
Fig. 3 Schematic diagram of entity and relationship extraction structure for the station passenger flow organization

分析上下文关系,得到新的向量序列 U (U 为关系序列);引入 Attention 机制,对向量序列 U 赋予权重,再通过 CRF 模型综合分析相邻关系依赖情况;最终输出最佳结果集,即车站客流组织实体及关系的抽取结果。

3.2 车站客流组织 TKG 应用

城市轨道交通车站客流组织情况复杂,涉及线网中多个车站,且车站与车站,车站与 OCC(运营控制中心)间均具有联动关系。考虑到各站的客流组织措施有所不同,需要由车站向上层层推进,在构建完成各站知识图谱的基础上,逐步形成覆盖整个线网、可联动的 TKG。本文以北京地铁 2 号线某车站为例展开研究。在该站 TKG 的构建过程中,引入客流数据,通过进站、出站闸机获取实时的进出站人数及站内人数,并通过客流预测规律对未来的客流数据进行预测,以期在 TKG 的基础上提出客流组织措施,辅助车站管理人员进行决策。

车站客流组织 TKG 数据库图形化展示截图如图 4 所示。由图 4 可以看出,通过对地铁客流、事件信息等数据进行处理,可获得相应的客流组织策略。引入时态元组后,可获取与现场时效相关的信息,挖掘车站客流的历史演变规律并进行知识推理,进而提前预判或实时辅助车站进行客流组织决策。



注:30-measures 指编号为 30 的措施;14-deptEmps 指编号为 14 的部门;20-events 指编号为 20 的事件;2-passengerNum 指编号为 2 的乘客。余类同。

图 4 车站客流组织 TKG 数据库图形化展示截图

Fig. 4 Screenshot of the visual display of TKG database of the station passenger flow organization

在图4的基础上,车站根据客流组织现场信息设置分级预警策略,形成基于TKG的车站客运组织智能预警系统,其系统架构如图5所示。该系统可以根据车站客流波动、预警策略等进行实时监测和智能预警,为车站客流组织提前预判、快速响应提供信息化手段和数据支撑。

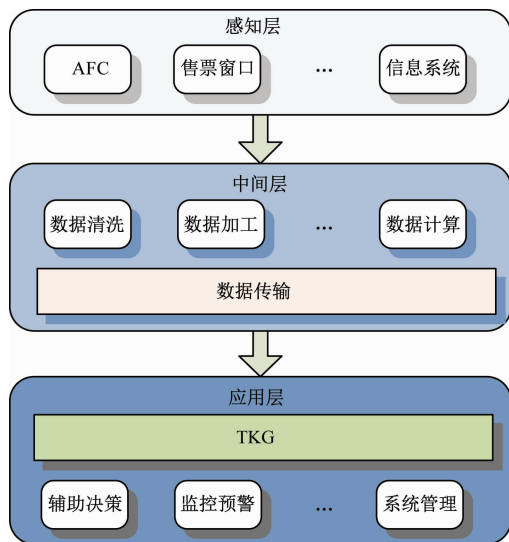


图5 基于TKG的车站客运组织智能预警系统架构示意图
Fig.5 Schematic diagram of the architecture of intelligent early warning system for the station passenger transport organization based on TKG

4 结语

针对城市轨道交通车站客流组织智能化手段不足、客流压力持续增大等问题,本文结合车站客流组织的实际情况,提出了基于TKG的车站客流组织模型。通过引入时态信息,结合车站客流数据及客流组织文本,利用ALBERT-BiLSTM-Attention-CRF架构,构建了车站客流组织TKG。该模型可为车站客流组织提前预判、快速响应、有效实施、智能

预警等提供信息化手段及技术支持。在未来的研究中,应进一步研究TKG中的知识推理算法,以提高城市轨道交通车站客流组织效率。

参考文献

- [1] 苗文文,王伟,邵志国. 基于CiteSpace软件的地铁运营安全风险研究文献计量分析[J]. 城市轨道交通研究, 2021, 24(9): 142.
MIAO Wenwen, WANG Wei, SHAO Zhiguo. Bibliometric analysis of research on metro operation safety risk based on CiteSpace[J]. Urban Mass Transit, 2021, 24(9): 142.
- [2] 马自力,王淑莹,张海柱,等. 基于知识图谱的智能问答意图识别联合模型[J]. 计算机工程与应用, 2023, 59(6): 171.
MA Zili, WANG Shuying, ZHANG Haizhu, et al. Joint model of intelligent Q & A intent recognition based on knowledge graph[J]. Computer Engineering and Applications, 2023, 59(6): 171.
- [3] 郭恒,黎荣,张海柱,等. 多域融合的高速列车维修性设计知识图谱构建[J]. 中国机械工程, 2022, 33(24): 3015.
GUO Heng, LI Rong, ZHANG Haizhu, et al. Construction of knowledge graph of maintainability design based on multi-domain fusion of high-speed trains[J]. China Mechanical Engineering, 2022, 33(24): 3015.
- [4] 唐进君,庾昊南,刘佑,等. 基于BERT-Bi-LSTM-CRF模型的自主式交通系统参与主体识别方法[J]. 交通信息与安全, 2022, 40(5): 80.
TANG Jinjun, TUO Haonan, LIU You, et al. A method for identifying the participants of autonomous transportation system based on a BERT-Bi-LSTM-CRF model[J]. Journal of Transport Information and Safety, 2022, 40(5): 80.

· 收稿日期:2022-06-07 修回日期:2022-08-05 出版日期:2024-09-10
Received:2022-06-07 Revised:2022-08-05 Published:2024-09-10
· 通信作者:孟歌,助理研究员,mg.clx@163.com
· ©《城市轨道交通研究》杂志社,开放获取CC BY-NC-ND协议
© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license

敬请关注《城市轨道交通研究》微信视频号

《城市轨道交通研究》微信视频号聚焦轨道交通行业内的热点问题、焦点问题,以及新技术、新成果,邀请相关专业领域内的专家学者及高级管理人员以视频方式解读和评述,是您及时获知行业资讯、深度了解轨道交通各专业领域的最佳平台。您还可以通过该平台查阅往期论文、查询稿件进度、开具论文录用通知书。敬请关注。

