武汉大学学报(信息科学版)
Geomatics and Information Science of Wuhan University
ISSN 1671-8860,CN 42-1676/TN

《武汉大学学报(信息科学版)》网络首发论文

题目：城市知识体系
作者：郑宇
DOI：10.13203/j.whugis20220366
收稿日期：2022-06-17
网络首发日期：2022-07-05
引用格式：郑宇．城市知识体系[J/OL]．武汉大学学报(信息科学版).
https://doi.org/10.13203/j.whugis20220366

网络首发时间：2022-07-05 17:14:34
网络首发地址：https://kns.cnki.net/kcms/detail/42.1676.TN.20220704.1741.001.html

DOI: 10.13203/j.whugis20220366

引用格式：ZHENG Yu. The Knowledge System for Intelligent Cities[J].Geomatics and Information Science of Wuhan University，2022,DOI:10.13203/J.whugis20220366（郑宇. 城市知识体系[J].武汉大学学报•信息科学版,2022,DOI:10.13203/J.whugis20220366）

城市知识体系

郑 宇[1,2,3]

1 京东城市（北京）数字科技有限公司，北京,100176
2 京东智能城市研究院，北京，100176
3 计算机与人工智能学院，西南交通大学，四川 成都，611756

摘要：智能城市的建设已经从过往的以业务为中心进入了以数据为中心的阶段。随着数据的不断积累以及智能算法的发展，智能城市的应用也开始从“简单直接地使用数据”向“挖掘数据蕴含的知识”来解决问题演进。类比以数据为中心的智能城市建设阶段需要城市数据管理体系，以实现数据的有效归集、管理和复用；在未来以知识为中心的时代，人们也需要建立相应的城市知识体系，以确保知识的有效组织、沉淀和复用。首次提出了智能城市知识体系，包括城市知识的内容定义、知识的表达形式、知识的产生和知识的应用四部分以及各部分中的规范和方法。城市知识体系可帮助从业者快速掌握行业知识、有效组织知识、高效构建以知识为中心的智能城市应用，发挥知识的价值，并让不同应用产生的知识不断沉淀、复用和融合，让智能城市可以更快、更好的发展。

关键词：城市计算；智能城市；知识体系；城市知识框架

收稿日期：2022-06-17
项目来源：国家重点研发计划项（2019YFB2101805）；国家自然科学基金（62076191）。
作者简介：郑宇，博士，京东集团副总裁，京东智能城市研究院院长，京东科技首席数据科学家，上海交通大学讲座教授，南京大学、香港科技大学等多所知名高校的客座教授和博士生导师。他是城市计算领域的先驱和奠基人，也是大数据和人工智能领域的领军人物和实践者，担任人工智能顶尖国际期刊 ACM TIST 的主编、IEEE 智能城市操作系统国际标准组主席、国家重点研发计划项目首席科学家、总负责人。他带领团队建设了雄安智能城市的数字底座、北京的政府运行一网协同体系以及南通等多个城市的城市治理一网统管平台。2013 年被 MIT 科技评论评为全球杰出青年创新者；2016 年评为美国计算机学会杰出科学家；2019 年，他作为大陆首位受邀学者在国际人工智能顶尖会议 AAAI 上发表主旨演讲。2020 年 11 月，他因在时空数据挖掘和城市计算领域的杰出贡献，被评为 IEEE Fellow。2021 年，因为他在智能城市领域做出的杰出贡献，被授予首都劳动奖章。2021 年 8 月，他当选 ACM 数据挖掘中国分会主席。msyuzheng@outlook.com

The Knowledge System for Intelligent Cities

ZHENG Yu [1,2,3]

1 JD Intelligent Cities Technology Co., Ltd., Beijing 100176, China
2 JD Intelligent Cities Research, BDA, Beijing 100176, China
3 School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China

Abstract: Data has been playing an increasingly important role in building intelligent Cities in the last decade. Recently, the advances in artificial technology has boosted a rising trend of using knowledge mined from data rather than raw data to tackle urban challenges. Just like designing data management frameworks for intelligent cities in the last decade, we need to define the knowledge management framework for the coming new era, in order for that knowledge generated by different applications can be constantly accumulated and consistently shared between each other. Otherwise, new knowledge islands will be created while we are spending a lot of efforts to break down data islands. To address this issue, we propose the knowledge system, consisting of the definition of the content of knowledge in intelligent cities and the framework of knowledge representation, generation and application. The knowledge system can help professionals to acquire domain knowledge quickly, mine knowledge from data efficiently, apply knowledge to solve problems effectively, and share knowledge among each other consistently, thus leading to smarter and greener cities.

Keywords: urban computing；intelligent cities；knowledge systems；the framework of urban knowledge

经过几十年的发展，智慧城市的建设已经从过往的以业务为中心迈向了以数据为中心的阶段，各级、各地政府都意识到了数据的重要性，加快开展数据的共享和汇聚工作。随之而来的是各种数据管理体系的设立、数据平台的建设和数据治理工作的开展。

近几年，随着数据的不断积累以及各种感知技术和智能算法的发展，智慧城市的应用也开始从“简单直接的使用数据”向“挖掘数据蕴含的知识”来解决问题演进。例如，从简单的居民查询个人的社保和公积金缴纳情况，向利用这些数据来分析计算居民的社会信用演进；从简单的企业在线缴纳水费和电费，到利用这些水电数据来分析企业的经营状况，甚至判断行业的发展态势。这些都是在用数据中蕴含的知识（而非数据本身）来分析和解决问题。此时，如何发掘、管理和应用好知识就变成了一个新的挑战和机遇[1-2]。

类比以数据为中心的智慧城市建设阶段需要城市数据管理体系，以实现数据的有效归集、管理和复用；在未来以知识为中心的时代，人们也需要建立相应的城市知识体系，以确保知识的有效组织、沉淀和复用。数据管理体系和城市知识体系的关系如下：

1)城市数据管理体系。包括城市数据类型的定义，以及数据的产生、接入、存储、查询和应用过程中的方法和规范。以支撑具体的智慧城市应用为目标，按照该数据管理体系来组织和处理数据的过程也被称为数据治理。

2）城市知识体系。包括知识体系内容的定义，以及知识的表达、产生和应用。因为知识来源于数据，城市知识体系跟城市数据管理体系并不是割裂的，知识体系必须要建立在数据体系之

上，并借助后者的某些方法和规范来落地。另一方面，在以知识为中心的智能城市应用中，数据治理过程也需要知识体系的指引。

但目前城市知识体系没有明确的定义和框架，导致各种智慧城市应用在挖掘数据和使用知识时无规范和方法论可循，数据挖掘和知识产生严重依赖个人能力和人工操作，知识的重复挖掘、模型的重复构建现象显现，知识无法沉淀、积累和复用，在人们不断破除数据孤岛的同时，一座座知识孤岛正在逐渐形成。

本文首次讨论智能城市知识体系，包括城市知识的内容和构建方法，及其表达形式、产生过程和应用方法。该知识体系既可以帮助从业者快速掌握智能城市的业务知识，从数据中挖掘知识，并利用知识解决问题，又可以让来源于不同项目的知识不断沉淀积累，并相互融通复用，大大提高智能城市应用的建设效率，降低开发成本。

1 城市知识体系概述

智能城市知识体系是指城市中的实体、实体间关系及其相关属性的知识化表达，以及用结构化、体系化的手段从中提炼出来的内涵和演变出来的外延。城市知识体系包括城市知识的内容定义、知识的表达形式、知识的产生和知识的应用四部分，以及其中的规范和方法。利用智能城市知识体系可以描述城市里发生的现象，洞察其本质，呈现其演进过程，掌控其发展态势，并支持和伴随城市中各类智能应用的发展和迭代，从而不断积累与升华。

图 1 展示了城市知识体系的框架。城市知识体系的最底层是城市知识内容的定义，即把城市中的实体抽象为人、地、事、物和组织五类，并定义好实体的属性、实体间的关系和关系的属性。这些实体、实体间关系及属性具有高度的概括性和普适性，是业务知识的高度凝炼和泛化表达，描述了城市中不断演变现象背后的本质规律，用有限、稳固的内容空间来衍生出无穷、变化的知识表达。城市知识体系的内容也会指引知识的产生和应用。

面向这些城市实体，利用城市感知技术，产生描述这些实体的原始数据，然后通过数据管理体系的数据接入将数据汇聚，并将数据分为结构化数据、非结构化数据和时空数据三大类分别存储起来。

根据上层应用的需求，开始治理数据，并通过对结构化数据的再组织、对非结构化数据建模和对时空数据的重构来对数据作知识转化，生成基础知识表达。之后，通过知识萃取和融合过程，利用统一的数据管理和构建专题库、AI（artificial intelligence）模型和知识图谱的能力，将基础知识转化为高级知识。

基础知识表达来自于单一种类的数据，其形态和产生过程不由上层特定业务和应用决定。高级知识表达可以是来自不同基础知识表达的组合，且跟上层业务和应用紧密相关。基础知识表达和高级知识表达构成了城市知识的表达形式，对应于图 1 中两块黄色的部分。知识的转化、知识的融合与萃取构成了知识的产生，对应于图 1 中的两块蓝色部分。

最后，这些高级知识向上支撑态势感知、分析研判和监测预警等基于知识的应用功能，上层应用通过可视化、查询检索和服务调用等接口，按照安全规范和权限设置使用这些应用功能和高级知识，解决应急响应、生态文明、城市治理等不同业务上的难题。

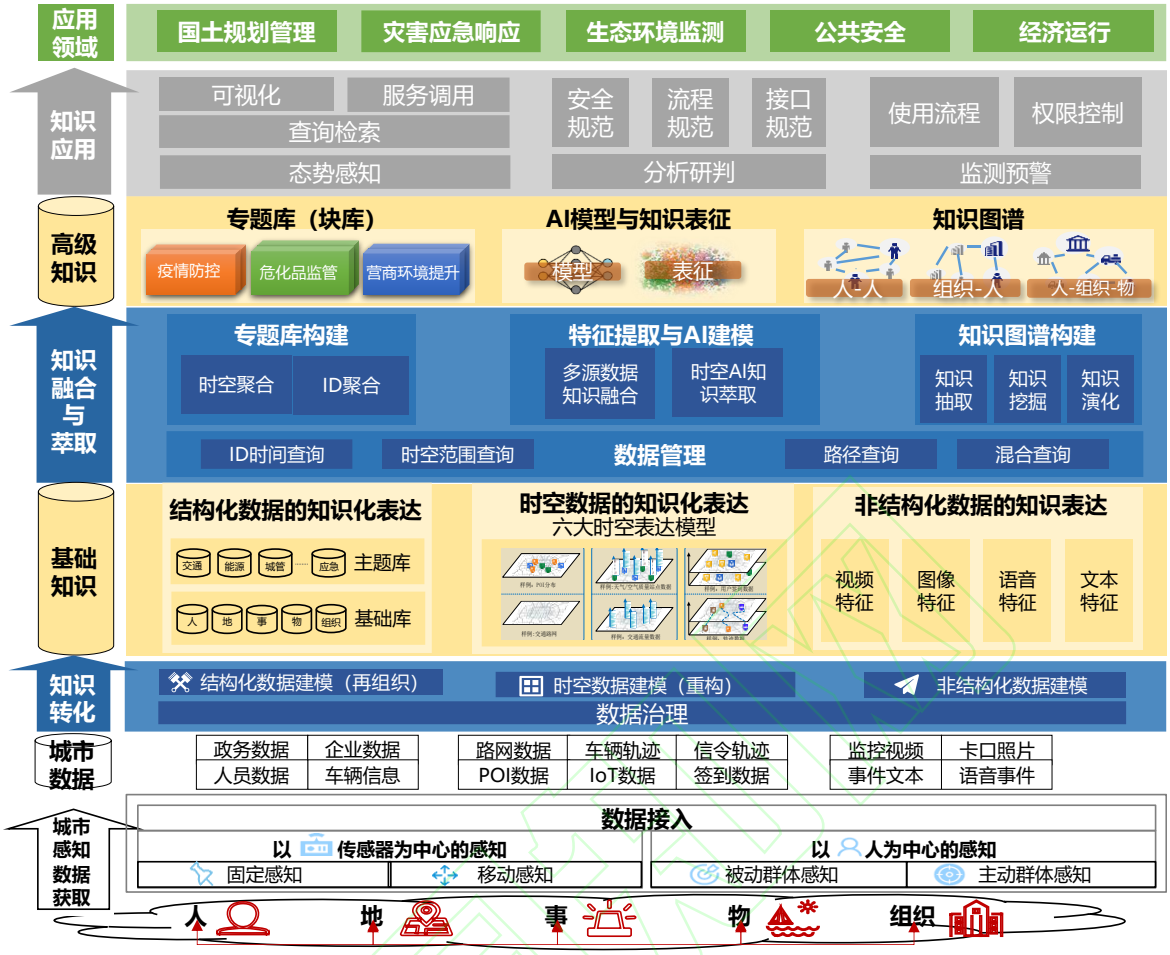

图 1 智能城市知识体系

Fig.1 Framework of Knowledge Systems for Intelligent Cities

2 城市知识体系的价值和意义

城市知识体系之所以变得如此重要，有以下三方面的原因和价值：

1）智能城市领域从直接使用数据的时代向利用知识来解决问题的时代演进。

在数据时代需要城市数据管理体系，在知识的时代就需要城市知识体系。过去需要面向应用来治理数据，以后也需要面向应用来构建和组织知识。这些都需要清晰的流程、统一普适的规范、高效的方法论的支撑。如此，从业者才能快速开展工作、少走弯路，利用自动化、半自动化的方法，高效建立基于知识的智能城市应用，发挥知识的价值。

2）不同智慧城市应用中知识的复用、沉淀和融合。

在使用知识的过程中，不同的应用会设计各自的模型、产生各自的知识，虽然很多模型完全一样，产生的知识也相同，但由于缺乏统一的标准和体系，各自的命名规则差异很大，相互之间无法识别，导致知识无法沉淀，模型无法复用，知识无法融合。这不仅会造成极大的资源浪费和管理混乱，也不利于智能城市的发展。

例如，在一个项目中，工作人员 1 定义了一个指标 T，其计算方法是 T=A+B-C；工作人员 2 定义了一个指标 M，其计算方法是 M=X+Y-Z。看上去两个指标完全不同，即便是让人来看也不一定能识别他们是一致的。但实际上 A 和 X 完全是一个意思，B 和 Y、C 和 Z 也是完全一样的东西，只是大家的命名规则不同。如果没有统一的知识体系，完全相同的东西也无法识别。有了城市知识体系之后，就很容易知道 A 和 X 是某类实体的某项具体属性，B 和 Y 是某类关系的具体属性等等，这些信息在设计这个指标时就需要指定好。基于此就可以自动地判断 T 和 M 是一回事，如果

T 在系统里已经有了，就无需再去设计和计算 M 了。

3）业务知识的沉淀和泛化表达。

很多人新进入智能城市领域时，感觉最缺乏的就是业务知识，如城市应急、交通、医疗和基层治理等领域的关键指标、行业痛点、业务关系等。过往需要长期的工作积累、阅读大量的文献、学习国家的政策方针和地方政府报告才能掌握相关行业知识。其实，人们在阅读和学习行业文献后，往往都会提炼出相应的知识点，并把它们组织成体系，业务知识的沉淀就是知识体系的构建过程。如果有一个事先构建好的知识体系，从业者可以通过直接装载这个知识体系来快速掌握行业知识，加速解决问题，避免了知识体系构建过程漫长、内容不全面的问题。

此外，过往的项目历程、方案、设计、报告等都属于碎片化的经验，在日后新的场景中被再次直接重复使用的概率很小。而知识体系中的内容源于大量案例、经验和文献的抽象和提炼，具有很强的泛化能力，具有更强的实用价值。

例如，在交通领域的知识体系中，乘客作为人这类实体跟交通工具作为物这类实体存在“乘坐”关系。如果知道乘客可以乘坐公交车，就能引申出乘客可以乘坐地铁，因为地铁和公交车都是交通工具，都属于物这一类实体。进一步，可以计算公交车的载客率，即用乘坐公交车的人数除以公交车的容量，这里乘坐公交车的人数就是乘坐关系这条边的属性之一。类比公交车载客率，可以进一步泛化出地铁的载客率，因为两者都是交通工具，因此该指标可以泛化到同类别的其它实体上。后者虽然也极具业务价值，但可能从未出现在任何的文献中，只有依靠知识体系来泛化产生。

3 城市知识的内容定义

城市知识体系中的知识内容包括人、地、事、物和组织五类实体，以及这些实体的属性、实体间的关系和关系的属性[3]。这五类实体在不同的领域会有更加细化和具体的实体分类，实体属性存在跨领域的共性属性和针对某个领域的特有属性。实体间关系在不同的领域也会有一定的差异和延展，因此，也存在跨领域的共性实体关系属性和领域特定属性。无论是跨领域还是面向特定领域，实体和实体关系的种类、实体属性和实体关系属性的数量都是有限空间、且高度精炼，但某个属性的具体取值空间可以随着业务的变化而增长、变化。通过对每个属性赋予具体值，实体和实体关系在应用中完成实例化。

3.1 实体和实体属性

在人、地、事、物和组织五类实体中，物和地的区别在于是否可移动（但并不一定要动），如汽车是物，可移动，停车场是地，不可移动；交通信号灯是物，道路是地。这五类实体在不同的领域会有更加细化和具体的实体分类，例如，在交通领域，人有司机、乘客、警察等更加具体的实体分类；地有公交车站、加油站、停车场等更加具体的实体分类。

实体的属性包括跨领域的共性属性和领域特有属性两部分。图 2 展示了这五类实体的共性属性，属性部分用蓝色线条描述。例如，无论是在哪个领域，人都会有姓名、出生日期、身高、体重等共性属性；地都会有名称、位置、功能和尺寸（长宽高）等共性属性；物都会有名称、功能、型号等共性属性。

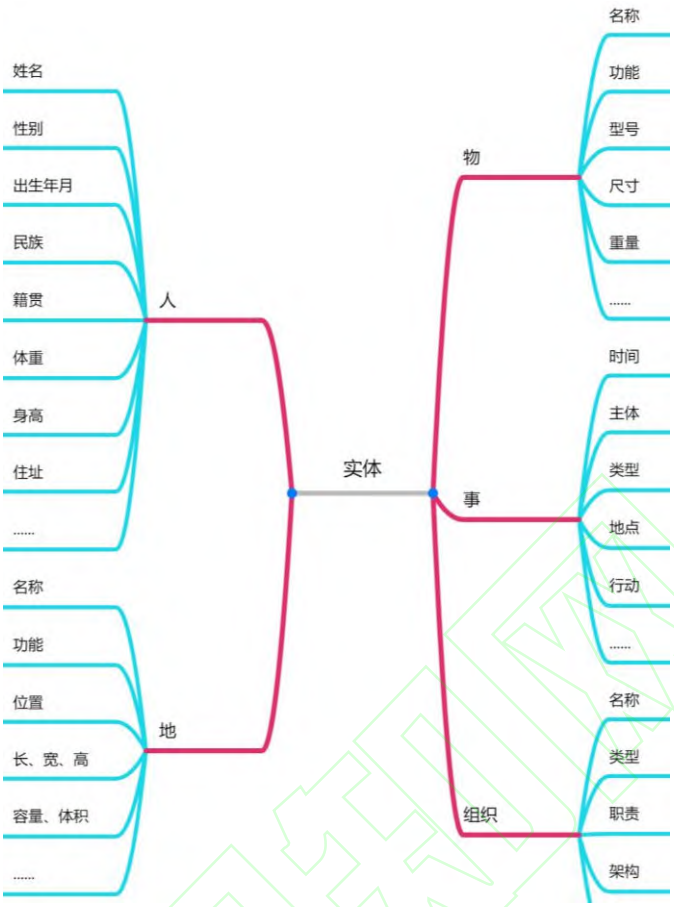

图 2 城市知识体系内容中的五类实体及其共性属性

Fig. 2 Five Types of Entities and Their Common Attributes of Urban Knowledge Systems

但如图 3 所示，在交通领域，人有司机、警察和乘客等更加具体的实体类别。除了人的共性属性外，司机这个实体还会有驾龄、驾照类型、违章扣分等面向领域的特有实体属性；警察会有警号、警种和警衔等特有实体属性。在医疗领域，人有医生、护士和病人等更加具体的实体类别。除了人的共性属性外，医生有等级、职称和专业方向等特有属性。人的共性属性加上这些特有属性构成了交通和医疗领域中不同类别人的实体属性。

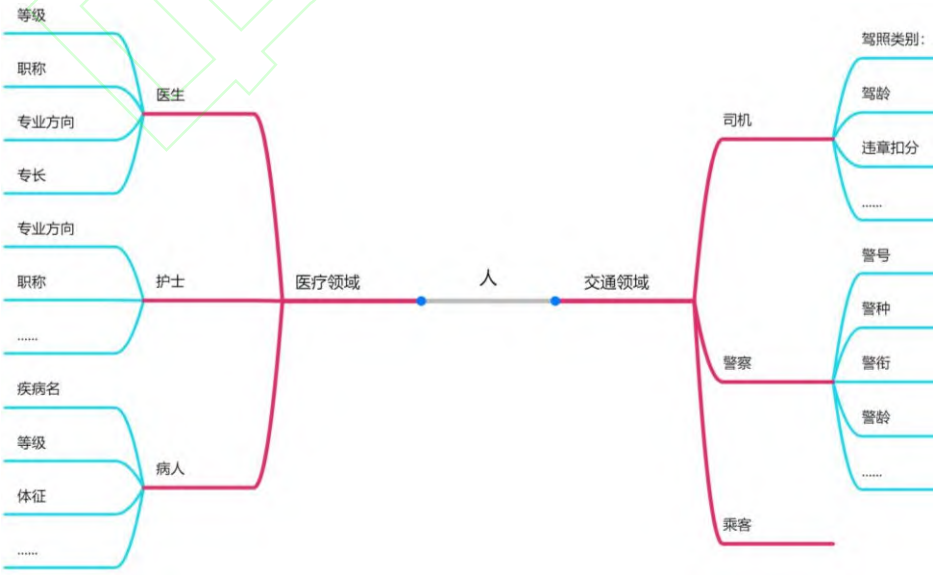

图 3 人实体在不同领域的细分类别和特有属性

Fig. 3 Subdivision Categories and Unique Attributes of Human Entities in Different Fields

面向业务领域定义的实体类别需要具有高度的概括性，具有相同属性的实体应被归为一类，避免将同类实体按照某个属性划分成细类。例如，警察应该作为一个实体类别，把警衔作为该实体的属性，把警监、警督、警司作为警衔属性的具体取值。而不应该设立警监、警督、警司 3 个实体类别，因为它们的属性一样。否则，同类实体可以按照每种属性的取值划分成无穷多的实体类别，失去了城市知识体系用有限空间的精炼表达来衍生出无穷变化的价值。反过来，小轿车和大货车虽然都笼统地称为车，但因为其功能有显著差异，因此它们的重要属性差别很大（如货车有货箱形式、货箱尺寸、驾驶室类型等），不能在同一类别中统一。而且，货车跟货物之间存在承载关系（小轿车没有），小轿车跟人存在承载关系（货车没有），即实体间关系也不一样。因此，需要将它们分成两类不同的实体。

3.2 实体之间的关系

3.2.1 实体关系的种类

人、地、事、物和组织五类实体之间存在同类关系、相互关系以及三者及以上关系。关系的类型也有跨领域的共性关系和领域特有的关系。同类关系和相互关系都是两个实体间的关系，无需借助第三类实体来产生，且实体双方有直接感知。图 4 展示了实体间的部分共性关系和部分关系的属性。

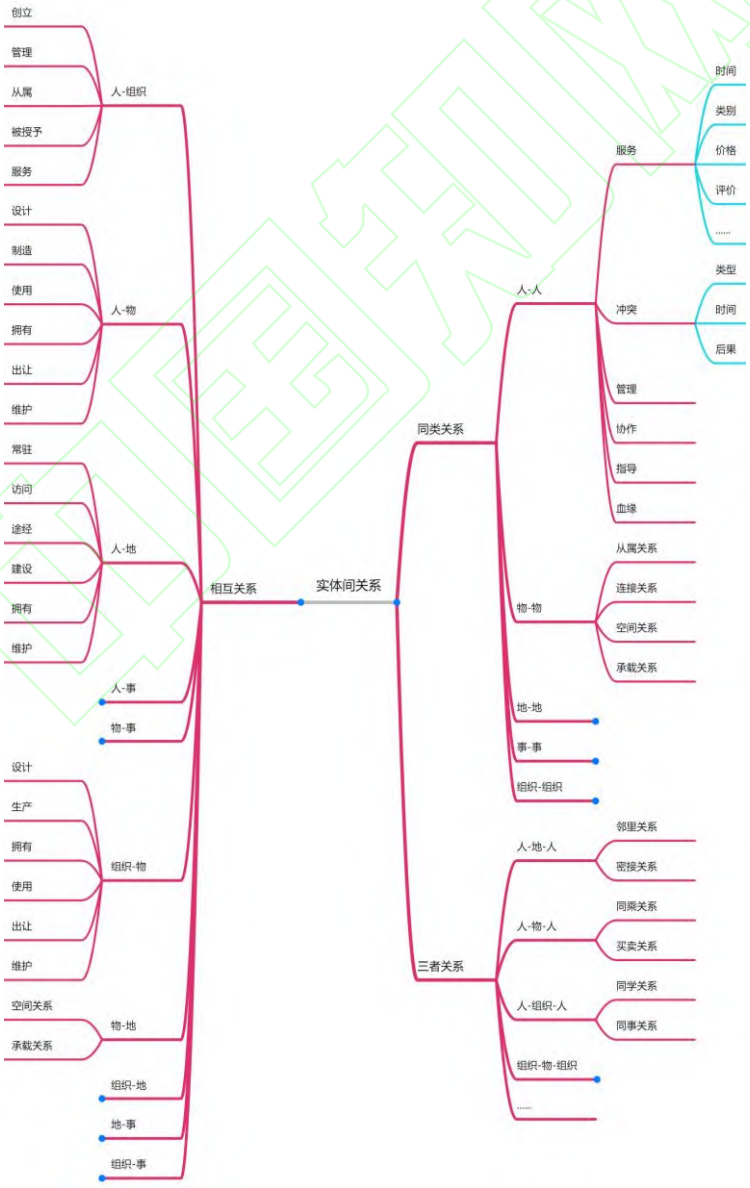

图 4 实体间的共性关系和关系的属性样例

Fig. 4 Examples of Common Relationships Between Entities and Attribute of Relationships

1)同类关系：这五类实体间存在的同类关系包括“人-人”“物-物”“事-事”“地-地”“组织-组织”。如图 4 右上部分所示，“人-人”关系具体包括“服务、冲突、协作、管理、指导和血缘”等跨领域的共性关系。在某些特定领域里可能还存在一些特殊的关系。“物-物”关系包括“从属、连接和空间”等跨领域共性关系。进一步，“人-人”的“服务”关系包含“时间、类别、价格、评价”等共性属性。

例如，一个司机在 2022 年 4 月 5 日 8 点为某车主提供了代驾服务，收取了 40 元的服务费，获得了车主的 5 星好评。在这个案例中，司机和车主两个具体的人实体产生了“服务”关系，服务关系的属性包括，时间：2022 年 4 月 5 日 8 点,类型：代驾,价格：40 元,评价：5 星好评。

2)相互关系：城市实体间的相互关系包括“人-地”“人-事”“人-物”“人-组织”“地-事”“地-物”“地-组织”“事-物”“事-组织”“物-组织”共十类。这里两个实体的关系是对称的，只是在交换了实体的顺序后，关系会有不同的表达方式。如图 4 的左半部分所示，例如“人-组织”关系中的“从属”属性，跟“组织-人”关系的“管辖”属性是一个意思，即一个人从属于某个组织，跟某个组织管辖某个人是一个意思，没有本质区别，在此不再区分，选择其中一种表达即可。

3)三者及以上关系：此类复杂关系其实可以通过两个实体间的关系组合得到，这些三者及以上关系很少用，但也要避免错误的把他们归类到两者关系中。如图 4 右下角的“人-地-人”关系中的邻里关系可以通过两个“人-地”的常驻关系推断出来，即两个人都住在同一小区，则可以推出他们是邻里关系。“人-物-人”关系中的同乘关系（同乘一趟火车或飞机）实际上是从交通领域中两个“人-物”的乘坐（专有）关系推断而出，即两个都乘坐了同一辆车。“人-物-人”中的买卖关系，也是一个“人-物”的出让关系跟另一个“人-物”的拥有关系的串联组合而成。

这里把有些关系定义为三者及以上关系而非两者关系有三点原因。以邻里关系为例子，把它定义为“人-地-人”三者之间的关系，而非“人-人”关系的原因如下：首先，邻里关系必须要以地作为条件，即“他是我在×××社区的邻居”。由于一个人在不同阶段可能会有不同的住所，或者同时拥有多套房产，在缺少地作为条件时，邻里关系有歧义，缺乏区分度。其次，如果将邻里关系直接定义为“人-人”关系，其中的两位当事人可能并没有感知，这种关系也不是他们主动建立前来的。因为大部分住在同一小区的业主并不一定相互认识，甚至并不知道有对方的存在。最后，如把任何两者以上组合产生的关系都安置到两者关系上，实体间关系的类别将无限膨胀，失去了作为标准体系去指引知识的产生和使用的意义。因此，在设计实体间关系的规范时，应尽可能让两者关系的数量精简，关系种类的名称有共性，可涵盖本质相同但有细枝末节差异的细分关系，且保持不同关系种类有显著的区分度。例如，在“人-物”关系中，出让关系可以涵盖卖出、捐赠和赠送等有细节差异的细分关系，这些细分的关系可以作为出让关系的类型属性的具体赋值。

3.2.2 实体关系的属性

在定义实体间关系的属性时，要区分哪些属性是实体的属性，哪些是（连接两个实体的）边上的属性，避免重复定义和错位定义。

实体上的属性不因其他实体而改变，可以独立决定。例如，人的姓名、年龄等属性跟他去过什么地方、设计过什么物品没有关系；一辆轿车的型号和排量等属性也跟谁来驾驶它无关。

边上的属性需要同时受两端实体节点的影响来决定。例如，司机为乘客提供的服务，属于人和人之间的服务关系，该服务的费用和评价等属性既跟司机提供的服务质量、车型标准有关，也跟乘客乘坐的距离、时间和体验有关，不能由一方单独决定，这些属性只能放在边上。人和组织的从属关系也一样，薪酬这个属性也只能放在边上，因为，既要考虑个人的能力和贡献，也要考虑组织的薪酬体系。

在实体上已经定义的属性，实体间边上将不再包含该属性。例如，人访问一个地点，形成了人与地的访问关系，这个关系上不需要再定义位置这个属性，因为该属性由地单独决定，在地类

实体已经定义了，无需在边上再次定义，更加不需要在人的实体上再定义位置属性。

4 城市知识的表达

城市知识的内容在产生和应用的过程中还需要有具体的表达形式。如图 1 黄色部分所示，城市知识的表达分为基础知识表达和高级知识表达两大类，每类又分别包括三小类，共六类表达方式，基础知识表达是产生高级知识表的基础。

4.1 基础知识表达

来自于单一种类的数据，不由上层具体的业务和应用决定，包括以下三种表达方式。

1）结构化数据知识表达。按照城市知识体系的内容中，对来自不同源头的结构化数据进行再组织，构建人、地、事、物、组织五类基础库和面向交通、能源、应急等不同领域的主题库。

基础库：既可以建立跨领域通用的基础库，也可以设立每类实体细分类别的基础库。例如，在政务数据中有人口、法人、宏观经济、自然资源与空间地理四大基础库；人口基础库是人这类实体的通用属性的一种知识表达，法人库实际上是组织这类实体的通用属性表达，因此，不同领域可共用这些基础库。同时，也可以建立司机、医生、孤寡老人等细分“人”类实体的基础库，以及政府部门、餐饮机构等细分组织类的基础库。除了向上支撑数据汇总和分析，基础库经常要向下承担跨部门数据共享的职责，因此对实时性和可扩展性要求较高。

主题库：按照某些主题，如交通、能源、应急、环保等领域，对多个基础库和政务数据来源进行知识的再组织，构建不同的主题库，如交通主题库、能源主题库等。主题库更多面向某领域的分析需求，强调数据的汇聚和归集。

专题库和主题库是城市知识体系中实体、实体属性、实体间关系及其属性的一种具体表达方式。基础库和主题库的数据都来自同一类型的结构化数据（如政务数据），且都具有普适性，不是为上层的某个具体应用而构建。再组织的过程中，也可能会涉及到简单的计算，例如，通过多个数据字段的加减乘除来产生一些指标。虽然结构化数据的基础知识表达不是为某一应用专门设定，但也可以被上层应用直接调用，例如，交通类的拥堵指数、道路平均时速、高速公路总里程、公交站点数量等共性指标可以在领导驾驶舱、指挥中心的大屏上和部门的业务系统中展现。

2）时空数据的知识表达

如图 5 所示，时空数据按照数据结构可分为点数据和网数据；按照时空信息是否动态变化可分为三类：时空静态、空间静态时间动态和时空动态，总共有 2×3=6 种数据。针对这六类数据，设计了六类知识表达模型：时空静态点模型、空间静态时间动态点模型、时空动态点模型、时空静态网模型、空间静态时间动态网模型、时空动态网模型[3]。把六类时空数据转化为相应的知识表达后，就可以开始后续的知识管理、萃取、融合和查询。

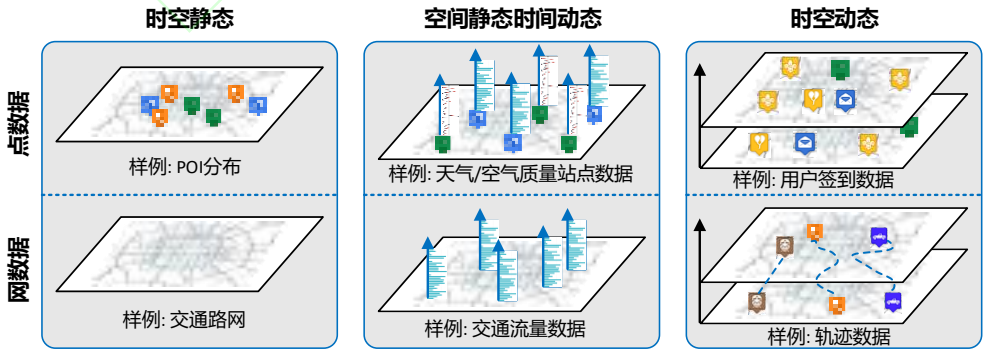

图 5 六类时空数据对应的基础知识表达

Fig. 5 Basic Knowledge Representation Corresponding to Six Types of Spatiotemporal Data

3)非结构化数据的知识表达。以视频、图像、语音和文本为代表的非结构化数据，需要经过分析处理变成结构化的知识表达才能被使用和查询。这些基础知识表达后续会作为产生高级知识

表达的输入。

例如，一个文本可以采用 One-hot、bag-of-words、n-gram、TF-IDF 等表达方式。以 TF-IDF 为例，一个文本可以表达为一个词向量，向量中的每一个元素对应词典中的一个词，该元素的具体值为对应词在该文档中出现的次数(TF)。对于包含多个文本的文本集合，可以计算每个文本中包含的词在其它文本中出现的频次(IDF)，TF 和 IDF 相乘就能得到每个文本中每个元素（词）的最终值。因此，一个文本就可以用这样一个向量来表达。当然这些基础的知识表达还面临很多挑战，如维度太高、数据稀疏、词语关联性无法体现等问题，因此，后续还需要借助更高阶段的知识萃取和融合技术来产生更好的高级知识表达。

图像可以由其包含的颜色、形状、纹理和空间关系等特征向量来表达。例如，每一副图片都可以生成颜色直方图，描述不同色彩在整幅图像中所占的比例。由于颜色直方图不关心每种色彩所处的空间位置，即无法描述图像中的对象或物体，因此,每一副图片还需要用一些关键点来表达，每一个关键点由一个固定长度的向量表示。有了关键点后，还可以进一步描述它们之间的相对位置关系。这些基础知识表达可以称为后续 AI 模型的输入，进而产生更高级别的知识表达。

4.2 高级知识表达

高级知识表达跟业务和应用紧密相关，可以通过对某类基础知识表达方式的再组织来产生，也可以是来自不同基础知识表达的融合。高级知识表达包括专题库、AI 模型和特征、知识图谱三个子类表达方式。

4.2.1 专题库

面向疫情防控、危化品管理或营商环境提升等专题应用，将不同领域、不同类别数据产生的知识进一步组织、融合，构建专题库，作为高级知识表达的一种形式。

例如，危化品全流程管理包括生产、储存、使用、经营、运输和处置六大环节，涵盖企业、个人、车辆、仓库和政府部门，揽括了人、地、事、物、组织五类实体，涉及到应急、交通运输、公共安全、生态环境等多个主题。因此，需要把不同的基础库、主题库中的知识再次组织，也需要进一步融合车辆运输轨迹等时空知识表达和从视频数据中提炼的车牌号、违章等非结构化数据的知识表达，才能形成一个有效的危化品全流程管理专题库。

4.2.2 AI 模型和特征

在利用知识解决问题的应用中，构建 AI 模型是一个重要环节，训练好的模型和提取的特征本身也是一种高级知识的表达。

AI 模型：其中蕴含的知识包括模型的结构、模型的参数、选择的特征以及训练的方法等。例如，一个训练好的决策树模型，其本身就蕴含了大量的分类规则，就是一种高级知识的表达。在一个场景中训练好的模型，可以用于类似场景，解决类似问题（条件是两个场景下数据是同分布的）。模型中的规则也能拿出来单独使用。

另外，针对一个文本集训练好的 LDA 模型也是一种高级知识，它将一篇文章表达为一系列主题的分布，将一个主题表示为一系列词的分布。单个文本的高级知识能支撑文本检索和相似度匹配等应用，整个集合的主题和词的分布也能让我们很快的理解一个文本集合的含义。

特征：选择的特征除了能被其它模型复用外，特征的具体值也可以作为重要指标在大屏、领导驾驶舱等很多业务场景直接使用。

例如，根据移动物体的运行轨迹计算的前进方向改变率，即在单位距离里改变前进方向的幅度，可用于判断移动物体当时采用的交通工具种类（如开车、骑自行车、步行等）。因为，步行时人的移动自由度最大，可随时转身或改变行进方向，人也很难走成直线，因此方向改变率一定比受道路约束的车辆更加频繁、幅度更大。这个特征还可以用于疲劳驾驶预警，如果自驾行为产生了蛇形前进的轨迹，导致方向改变率变大，则说明存在不正常情况，需要及时预警。

像公交车空载率、各地铁站最近半小时进站人数之类的特征，即可作为各交通系统中流量预测的特征，其具体的值也可以直接在大屏和领导驾驶舱中作为重要指标直接展示，帮助整个城市的态势感知。

4.2.3 知识图谱

知识图谱可能是最接近人脑组织知识方式的一种表达，它以实体为节点，以实体间的关系为边来构建图，节点和边都有相应的属性。这种表达形式可能是最为直观的体现城市知识体系内容的一种表达方式，但人脑到底是如何组织知识的，暂时还不得而知。因此，本文认为知识图谱是城市知识体系内容的一种高级表达形式，但不是唯一的表达形式，也不是直接的等价对应关系。

以知识图谱来表达高级知识，可以某个实体或某条边为关注对象，动态关联、组合多维数据，形成以该对象为中心的知识表达视角。同时，可以顺着图谱的脉络不断查看关联信息，挖掘更深层次的知识。知识图谱跟专题库有所不同，后者重点明确、固定，需要提前生成相应的数据库和页面来承载和展示相关知识。基于知识图谱的表达无需提前形成大量静态展示页面，且探索的对象可能存在变化和迁移。

例如，构建一个企业（组织）、危化品（物）、车辆（物）、司机（人）和仓库（地）之间的知识图谱。图 6 展示了 4 家企业、3 个仓库、2 种危化品、1 个司机和 1 台车辆之间的知识图谱关系。企业 1 生产了两种危化品，即跟危化品 A 和 B 有“生产”关系。企业 1“拥有”一座仓库 P，用于“承载”危化品 A 和 B。企业 1“出让”危化品 B，企业 3 通过买入从而“拥有”了危化品 B，三者构成了销售关系。在完成交易后，企业 1 委托企业 2 来完成对危化品的运输，即企业 1“使用”企业 2“拥有”的车辆 A 来“承载”危化品 B。更为具体的讲，企业 2 指派“从属”于其公司的司机 X“驾驶”企业 2“拥有”的车辆 A 前往企业 1“拥有”的仓库 P，通过在仓库 P 的“驻留”完成车辆 A 对危化品 B“承载”。之后司机 X“驾驶”车辆 A 前往企业 3“拥有”的仓库 Q，通过“驻留”完成卸货，实现仓库 Q 对危化品 B 的“承载”，让企业 3 最终“拥有”了危化品 B。

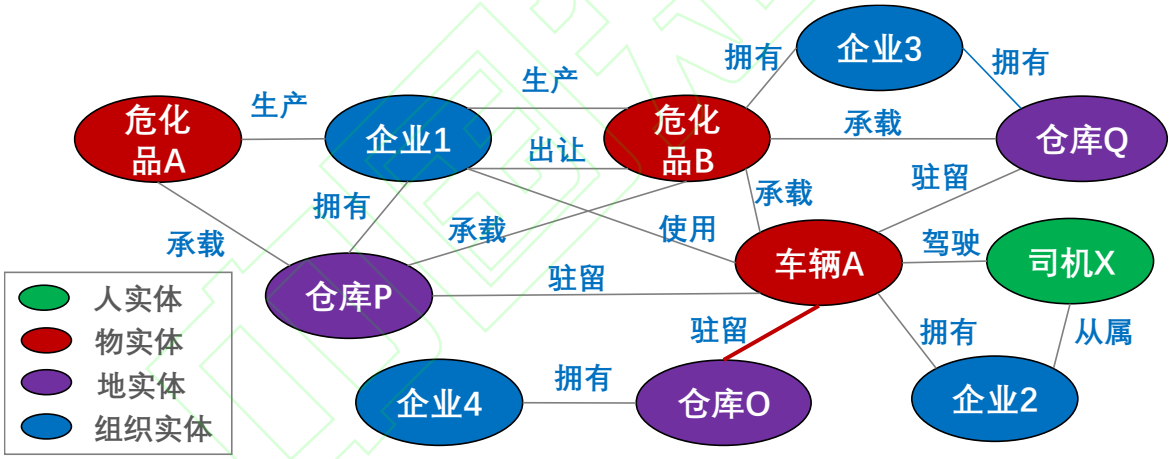

图 6 基于知识图谱的危化品全流程管理知识表达

Fig. 6 Knowledge Representation of the Closed-Loop Management of Hazardous Chemical Based on Knowledge Graph

有了以上的知识表达方式后，可以某个企业为中心，把其他企业、车辆、仓库和危化品的信息关联起来，形成以该企业为目标的知识描述形态。该表达方式也可灵活切换至任意实体（如车、司机等）作为展示中心。同时，该知识图谱还可以顺着某中心的外围节点继续探索更深层次的知识。

例如，通过分析车辆 A 的 GPS(global positioning system)轨迹，发现其曾经“驻留”于仓库 Q，而仓库 Q“从属”的企业 4 因不具备危化品生产条件已经被关停整改。此时，这个“驻留”关系就需要向相关管理部门预警。通过这条“驻留”边的属性（如起始时间、频次等）、仓库 Q 的属性（位置、等级等）及其承载的危化品种类等信息便可精准的分析出该非法运输行为的时间、地点和事项，车辆所属企业和可能驾驶该车辆的司机。如继续深挖，还能找到这些危化品被运送到哪些下游企业和仓库，把整个非法经营的交易链找出来。

5 城市知识的产生

这里的知识的产生指如何产生 6 种具体的知识表达形式。知识的产生包括如图 1 所示的两个蓝色阶段：知识的转化、知识的融合与萃取。知识的转化以原始数据为输入，产生基础知识表达；知识的融合与萃取以基础知识为输入，产生高级知识表达。知识的产生过程需要城市知识体系内容的指引。

5.1 知识的转化

1）数据治理。知识转化阶段面向接入的结构化、非结构化和时空数据三类原始数据，根据上层应用的需求，开始数据治理工作，对数据进行清洗、整理并建立相应的数据仓库存储数据，数据治理工作遵循数据管理体系的规范[4]。数据清洗过程中建立的数据仓库一般叫归集库，为了保证数据可溯源，每一个数据来源都会建立相应的归集库。归集库定位于原始数据的存储，不存在基于知识的再组织过程。在构建基于知识的智能城市应用时，数据治理过程需要根据城市知识的内容来设计元数据管理，并对数据接入作标准化处理。

2）结构化数据建模。通过对存储好的结构化数据的再组织来构建基础库和主题库，一个基础库中的知识可以来源于多个归集库，一个主体库的知识可以来源于多个基础库。在知识的再组过程中，依靠唯一实体标识（ID）来关联不同来源的数据。例如，根据个人的身份证号码来关联其职业、医疗保险、住址以及拥有的房产等信息。基础库、主题库的构建需要根据城市知识内容中定义的实体属性和实体间关系来聚合数据。

3）非结构化数据建模。通过对非结构化数据建模来提取颜色、形状和纹理等基础特征，形成非结构化数据的基础知识表达。

例如，文本文件为了得到前面介绍的 TF-IDF 基础知识表达，需要先对一个文本集合中所有文档的内容分词，形成文本集合对应的词典，并将每个文档转化为若干词语的集合，再确定每个文档包含的词语在本文档中出现的次数(即 TF)；同时建立词语与文档的反向对应关系(倒排表)，即一个词在哪些文档中出现过，计算 IDF。最后可以为每个文档计算出 TF 和 IDF 的乘积，得到一个 TF-IDF 向量。文本预处理的方法很多，不在此一一介绍。

针对图像数据，可以用成熟的统计公式来计算颜色直方图中的各种指标，也可以使用 SIFT（scale-invariant feature transform）算法来提取图片中的关键点信息。这些基础特征可以被很多不同的应用使用，产生的动机跟某个具体的业务关联性弱，因此属于基础知识表达的产生。

4）时空数据建模。通过对时空数据的重构生成六类时空知识表达模型，每一类表达模型后续都配有相应的管理、查询、挖掘算法和 AI 模型。将数据装入这些模型后，就能利用后续能力对这些数据进行流水线般的分析和处理，得到高级知识。

5.2 知识的融合与萃取

该阶段提供两个层次的能力，第一层是统一的数据管理能力，第二层是分别构建专题库、AI 模型和知识图谱三方面的能力。数据管理能力形成了对第二层能力的支撑。

1）数据管理。对基础库和主题库提供各种结构化数据仓库的查询和管理能力。对六类时空基础知识模型提供时空范围查询、ID 时间点查询、最近邻查询和可达范围查询等管理能力。对非结构化数据特征提供倒排表、相关性计算和相似性排序等查询和管理能力。

2）专题库构建。通过对不同基础知识表达的再组织和聚合，形成面向应用的专题库。在这个过程中有基于唯一标识和基于时空范围两种不同的聚合方式，两种方式都需要城市知识体系内容的指引.

唯一标识聚合法：是一种根据实体唯一标识将不同来源的知识对齐、聚合的方法。它根据城市知识内容中为每类实体定义的属性，把同一实体分散在不同地方的属性信息聚合起来，也可以根据定义的实体间关系把相关联实体属性的内容聚合起来。例如，可以身份证为唯一标识把同一个人的职业、收入、房产、交通违章记录等融合到一起；可以社会信用号为唯一标识把同一企业的产品、税收、产值等信息融合到一起。过往的专题库大都通过这种方法构建，利用结构化数仓

的管理能力就能够实现。

时空范围聚合法：在城市治理的很多场景中，很多应用需要研究探查的对象并不存在预先分配的一个 ID，需要依靠空间范围、外加时间区间来动态聚合不同源头的知识和数据，这点需要高效的时空知识管理能力来支撑，也需要知识内容中定义的实体间关系来指引[5-6]。

例如，在治理一个社区时，需要根据这个社区的空间范围来聚合在某个时间段（如最近两周内）出入该社区的人员和车辆信息，以及这个社区内产生的消费和视频等多源数据，然后才能精准的治理社区的安全隐患或周边的停车问题。

这里并没有一个天然的 ID 存在，车辆的轨迹数据里并没有它经过的社区名称这个字段，传感器、车辆和人流等数据都需要通过社区所占据的空间范围以及应用关注的时间段被聚合到一起，从而建立了跟该应用的关联。即便针对同一社区，不同的应用场景关注的时间跨度不一致（三天、三周或三个月），空间范围大小也会有变化（社区内、或社区周边道路、或一刻钟商圈），需要关联的数据和知识的种类也不一致，这些都不能提前通过预设 ID 来解决。

还有很多应用，如在户外举办的活动和集会，其空间范围根据活动地点临时选择，无法事先固定下来。像突发灾难这类应急事件，更加是无法预知时间和空间范围，但更加需要针对这个专题来快速融合不同维度的知识和数据。针对这样的场景，更加不可能找到一个特定的 ID 来关联不同源头的数据。

这里利用了人-地的“驻留关系”、物-地的“承载关系”和“空间关系”以及人-物的“驾驶关系”等实体间关系来建立不同信息的关联，从而完成不同知识的聚合。

3）特征提取与 AI 建模。特征提取是 AI 建模的前序工作和重要基础。首先，特征选取的有效性将很大程度上决定模型结果的精准度。其次，特征本身也是一种知识表达，好的特征能广为流传，在多个场景中被使用。最后，面对高实时性场景，特征提取的效率对 AI 模型的价值体现至关重要。面临海量数据，如果没有高效的数据管理能力，简单的特征提取会变得异常缓慢，使得 AI 模型不能快速产生智能分析的结果，失去其本身的价值。

例如，要评估一起交通事故带来的影响，需要以事故点为圆心，快速提取当前时间段以及历史对应时间段周边 500 m、1 km、2 km 的车辆数量和道路状态作为特征，输入预测模型。这需要对海量的手机信令或者车辆的 GPS 轨迹作快速的时空范围查询操作，此时如果通过遍历数据来筛查满足条件的数据，将花费数小时甚至更长的时间，不能满足应急场景下的高实时性、紧迫性需求，完全失去了模型分析判断的价值。

AI 建模包括模型种类的选择、模型结构的设计、模型特征的选择、模型参数的训练和模型的发布。首先，针对实际问题，分析其属于哪类数据科学问题，如分类、聚类、回归、异常检测等，由此选定模型的类别。进一步根据面临的数据类型和同一类别中不同模型的特性来选定具体的 AI 模型。

以文本数据为例，在之前产生的 One-Hot、TF-IDF 等基础知识表达之上，可以利用 SVD(singular value decomposition)、LDA（latent Drichlet allocation）、pLSA（probabilistic latent semantic analysis）等主题模型、或 Word2vec 等基于词向量固定表征、或者 BERT 等基于词向量动态表征等方法来产生更加高级的知识表达。

选定具体模型后，需要根据实际问题来设计模型结构。例如，一个神经网络模型需要多少层级、输入层有多少变量、输出层级有多少变量等。模型结构的设计通常也会跟特征选择结合起来思考，尤其是在输入层的设计上。之后就是利用已有数据对模型进行训练，不断调整其参数，使得结果达到最优。期间，AI 模型的设计者也会根据训练结果来反向调整模型的结构和特征选择。

一方面，建立好的 AI 模型本身就是一种高级知识的表达方式。另一方面，城市知识体系也能帮助 AI 建模的过程。例如，在设计贝叶斯网络的模型结构时，对实体间关系的提前认知将帮助构建模型中不同节点的边；对于实体专业属性的认知将帮助模型设计有差异性的特征。

4）知识图谱构建。以知识体系内容中定义的实体、实体属性和实体间关系及属性作为指引，结合实际应用的需要作适当筛选，利用接入的数据和转化的基础知识来实例化相应实体的属性和

实体间的连接边，从而快速构建出相应的知识图谱。构建知识图谱的知识既可以来源与同一种基础知识表达，也可来源于不同的基础知识表达的融合。

如图 6 中的知识图谱，根据城市知识体系内容中定义的人、地、物、组织四类实体、实体的重要属性和实体间的关系，针对危化品全流程管理的实际需求，选择危化品企业（组织）、危化品运输车辆（物）、司机（人）和仓库（地）作为主要实体，重点关注“司机-车”“司机-企业”“企业-仓库”“企业-危化品”“车-危化品”“危化品-仓库”和“车-仓库”等主要实体间关系，将这些实体和实体关系作为知识图谱的主体框架。

然后，利用基础知识的专题库、主题库和六类时空数据知识表达，对以上实体的属性、实体间关系及关系的属性实例化（即建立实际边、对属性赋值）。比如，危化品企业、运输车辆、仓库、司机的知识来自于应急主题库。通过车辆表单中所属企业的字段，可以建立具体企业和某车辆之间的边；通过企业的员工名录，可以建立其企业和下属的司机之间的关系；通过企业的经营记录中的危化品订单，可以建立起司机驾驶过某运输车辆。通过将危化品车辆 GPS 轨迹转化为时空动态网知识模型，对其进行驻留点分析，再跟危化品企业的空间位置匹配，可以得到某辆车在某个仓库驻留过，从而建立起该车辆跟仓库的驻留关系。

6 城市知识的应用

在智能城市的不同业务领域中，态势感知、分析研判和监测预警是最为常见的基于知识的应用功能，这些功能通常需要设定一些有重要意义的关注指标，通过获取相关数据，动态计算、更新这些指标，并将相关信息推送到指挥中心的大屏和领导驾驶舱，帮助政府掌控城市状态。有时也需要设计较为复杂的智能模型来预测这些指标，或者将这些指标作为重要特征输入某些模型，以实现对重要事项的研判和预警，辅助政府决策[6]。在以上应用中，以产生知识为导向的数据治理、指标或特征的设计、智能模型的构建是最为重要的三件事情，也是城市知识体系发挥重要作用的地方。

6.1 指导数据治理的过程

为支撑基于知识的应用而开展的数据治理，需要根据城市知识体系来设计元数据管理，对数据接入作标准化处理，并对不同来源的数据作有效聚合。

指导元数据管理设计：在支撑基于知识的应用时，需要根据城市知识体系的内容来设计整个数据管理体系的元数据，根据知识体系的结构来设计元数据的树形结构，根据知识体系中实体和实体关系的类别做好表的标签化和命名，并保持元数据中字段命名和知识体系中属性名称一致。由于底层的基础库既要不断从下方各业务系统中获取数据，也要被多个上层数据库或数据仓库访问，还经常承担着下级不同业务系统间共享数据的职责，面临高速的更新和插入操作，必须保持其可扩展性和并发性。因此，在根据知识体系来设计基础库时，应遵从以下原则：实体的属性放在一个基础数据库中，实体间关系应放在另外的基础数据库中单独存储，不可简单的把实体间关系的属性简单粗暴的作为字段插入到某个实体的记录中。这跟主题库根据知识体系中的实体间关系把尽量多维度的数据关联起来不同。

指导数据接入的标准化：数据治理将从不同来源接入的数据分别存入不同的归集库，以便后续对数据质量溯源。基于以上设计好的元数据，在接入过程中需要指定归集库中每个表格的每个字段对应于哪个实体或实体关系的属性。否则不同来源的数据中包含的字段五花八门，后续的基础库、主题库、专题库建设无法有效识别关联数据，更加不可能自动化的构建新的指标和智能模型。

指导数据的聚合：根据城市知识体系中每类实体包含的属性，可以将对应于同一实体属性的字段从不同的归集库聚合到同一个基础库。此外，利用城市知识体系中实体间的关系，可以将不同基础库的内容进一步融合到主题库和专题库中。尤其在基于时空范围的聚合方式中，由于不存在统一的实体标识，更是需要依靠城市知识体系中定义的实体间关系来聚合数据。

如图 7 所示，从不同来源接入私家车、司机、违章处罚和驾照等多种数据，建立了 n 个归集库分别存储。根据交通领域的城市知识体系，这些归集库中的表将分别被打上交通领域物实体“私家车”类、人实体“司机”类、事件实体“违章处罚”类和物实体“驾照”类的标签，并根据城市知识体系的规范命名相关表格为“私家车基础信息表、司机基础信息表、违章处罚记录表和驾照信息表”。按照城市知识体系，跟司机和私家车这两个实体相关的很多专业属性（比如司机的违章扣分、驾照类别）并不包含在相关表里，因此，前两个表被命名为基础信息表。

在接入数据后，将每个表的每个字段映射到知识体系的某个属性上。例如，归集库 1 中违章处罚记录表中的车牌号字段对应于私家车类实体的专业属性“车牌号属性”。在归集库 2 中的驾照信息表中，指定身份证号字段对应于司机类实体的基础属性中的“身份证属性”。

随后开始构建基础库。注意，司机基础库只存储司机这类实体的属性，司机买卖车辆的记录属于人-物的购买和出让关系，应该单独建立人-物关系数据库，并在其中建立司机与车辆的买卖数据表，每一条记录表示一个人物实体间购买或者出让的关系，包含司机的 ID、车的 ID 以及价格、成交时间、过户时间和税款等购买或出让关系的属性。因为一个司机可以买卖多辆车，一辆车也会先后被不同的人买入和卖出，这些属性不能简单的作为字段放在司机或者车的基础信息库里。同样道理，车停靠在停车场属于地物的承载关系，相关记录也应该存储在地物关系信息库中，并在该库中建立车辆停放表。该表中每一条记录表示一次车辆和停车场的承载关系，包括车的 ID、停车场 ID、时间、价格等关系属性。

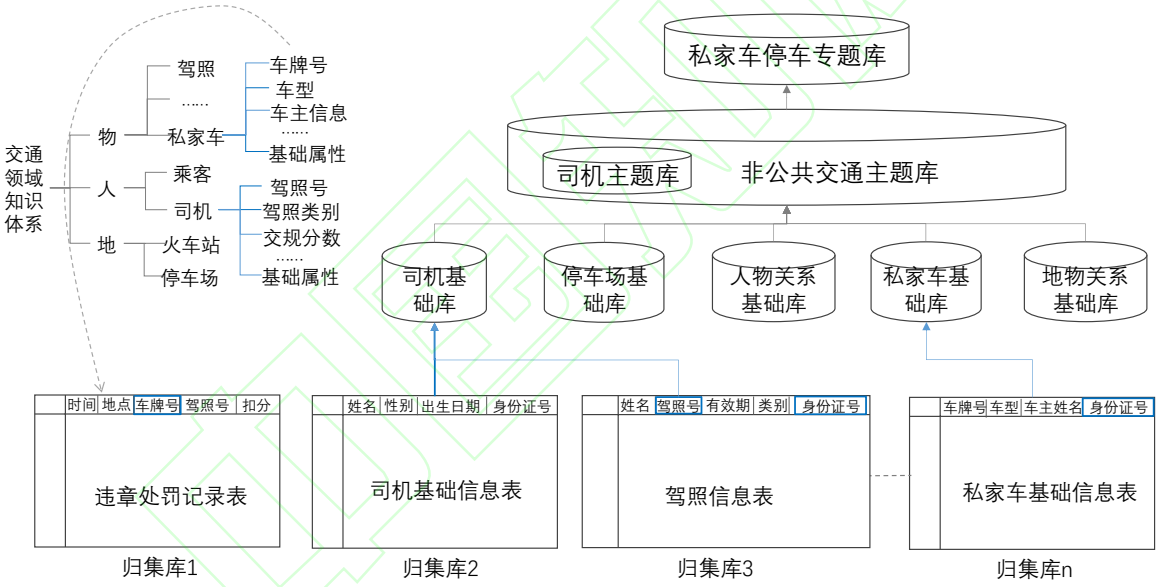

图 7 基于城市知识体系的数据治理过程

Fig. 7 Data Governance Process Based on Urban Knowledge System

完成以上操作后，根据交通领域城市知识体系中司机包含的属性，系统可以自动推荐归集库中被映射到司机属性的数据字段，通过人工确认后便可快速建立司机的基础库。如图 7 所示，司机基础信息表的数据可以全量导入到司机基础库中；通过司机的身份证号，在归集库 3 中的驾照信息也会作为司机实体的属性被关联到司机基础库。同理，根据城市知识体系中私家车实体的属性和数据接入过程中对数据表及字段打好的标签，可以快速构建私家车基础库、停车场基础库等。以后涉及到司机或者车辆的基础信息，可以直接从两个对应的基础库中获取，无需再次访问下面的归集库和业务系统。在归集来自不同业务系统的数据后，这些基础库有时也承担着向下为不同业务系统共享数据的职责。例如，一个司机的电话号码通过一个业务系统修改后，各个业务系统都可以从基础库得到最新的司机电话。

进一步，根据交通领域知识体系中非公共交通包含的实体和实体间关系可以构建非公共交通主题库。例如，基于司机跟车辆之间存在“驾驶关系”和“拥有关系”，利用司机的身份证号可

以关联到其拥有和驾驶过的车辆信息，以及在驾驶这些车辆过程中的违章处罚信息，这些信息也可以一并导入到以司机为中心的主题库中。以后，只要涉及到跟司机相关的信息，均可从司机主题库中调取，无需多次访问不同的基础库。

6.2 指导指标和特征的生成

作为各应用的直接展示点以及模型的重要输入项，指标体系一直是各业务和系统的重点关注事项，该体系的质量直接决定了整个智能城市应用的价值。要了解指标体系为什么需要城市知识体系的指导，以及城市知识体系如何指导指标体系的构建，首先要知道指标体系的设计原则。

1）构建指标体系的原则。设计重要、且可行的指标体系作为关注点需要综合考虑以下三点原则：

专业性：需要深厚的行业知识，知道某个行业需要重点关注哪些指标，每个指标揭示的意义，以及指标变化带来的启示和可能的行动点。例如，前面提到的公交车空载率指标是反应一个城市中公交系统资源利用率的一个维度。空载率高则意味着公交车几乎没人坐，存在大量的公交车辆资源闲置。应在满足市民出行需求的前提下，优化公交线路和班次，减少车辆和人员投入。反之，则证明公交车非常拥挤，人满为患、资源紧缺，应该考虑加大公交系统的资源投入和线路扩容。

针对性：指标体系的设计要结合本地的实际情况和需求。以交通出行领域为例，很多城市并没有地铁，关于地铁的一系列指标都不适用。很多小型城市火车站的人流量也并不大，因此，在很多大城市中高度关注的交通枢纽人流量等指标在小型城市并不需要。但这些小城市有可能是旅游城市，对于绿色交通领域，如共享单车、新能源充电桩、旅游观光车等方面，却非常关注，需要设计一套针对本地的交通出行指标体系，而不能照搬北京、上海等大城市的相关指标体系。

可行性：设计好的指标是否能够及时产生，还需要评估相关数据的可获得性、质量和性价比。有时当地并没有提前记录计算指标需要的数据，使得指标无法产生；或者数据的质量太差，更新频度不及时，导致计算的指标失去应有的价值；或者为了获得所需数据需要安装昂贵的设备并增加繁重的运营成本，虽然方案可行，但产生该指标的性价比非常低，也不可行。

2）指标的构建为什么需要城市知识体系的指导。基于以上指标设计原则，考虑到实际存在地域差异性、业务发展性和空间复杂性，不存在一套固定、通用的指标体系能满足现实的需求，人们必须具备为不同城市、基于当下需求、即时设计合理指标体系的能力。

地域差异性：由于不同城市存在一定的差异，很难有一套全国通用的指标体系能满足不同城市的针对性需求。例如，有的城市没有地铁，有的城市关注高铁站，有的城市火车站并不是热点，反而关注长途客运或者水路客运。

业务发展性：随着智能城市业务的发展，工作重心和应用需求会发生变化，指标体系也会不断演变。很多新生事物，如无人配送、互联网医院、共享汽车等，之前并未出现过，更不会有现成的指标可用，需要及时的设计新指标来反应业务的本质。

空间复杂性：指标体系的空间理论上是无穷大的，各种数据的连接会产生组合爆炸，无法提前穷尽储备起来。因此，我们必须要寻找那些变幻莫测现象背后不变的本质。

长期以来，指标体系的构建缺乏规范和方法论，主要依靠少数人的经验，导致各智能城市项目上水平参差不齐。因此，在智能城市应用中，指标体系的构建急需理论的支撑和规范的指导。城市知识体系就是描述城市中不断演变现象背后的本质规律，是业务知识的高度凝炼和泛化表达，它用有限、稳固的空间来衍生无穷、变化的可能性。

实体类别有限：由于城市知识体系中设计实体的类别具有高度的概括性和通用性，大类上只有“人、地、事、物、组织”五类，在每个领域里还有一些细分类别，实体类别是有限空间。由于同类实体有共性，其实体属性的数量也是有限空间。很多新生事物只是某类实体的某个属性多了一个取值可能，并没有超过原来的知识体系范畴。例如，共享单车其实还是自行车这个实体类，只是这个自行车这个实体的功能属性，多了一个共享的取值可能。其它属性，如型号、重量、尺寸等仍可沿用自行车这个实体的属性。因此，属性数量有限，但其具体的取值可以不断添加。

实体间的关系有限：在城市知识体系设计的时候充分考虑了普适性和概括性，抽象出来的实

体间的关系具有很强的通用性和泛化能力。例如，无论是加油站和私家车，还是充电站和新能源汽车，它们之间都是承载关系。无论是绿色公交和专用车道，还是普通车辆和机动车道，它们之间都是通行关系。

3）城市知识体系帮助指标体系构建。城市知识体系在增强指标体系的专业性、提升指标体系的针对性、落实指标体系的可行性 3 个方面发挥重要作用，帮助快速构建各城市、各领域所需的指标体系。

增强专业性：城市知识体系帮助我们快速掌握专业知识，增强指标构建的专业性。城市知识体系告诉人们城市各领域分别有哪些人、地、事、物、组织实体，这些实体有哪些重要的属性以及实体间有哪些重要关系。这些经过提炼的实体和实体间关系揭示了多变业务背后的本质，不会随着时间轻易改变，因此，这些实体属性、实体关系的属性就是指标的直接来源和指标构思的起点。

以交通领域为例，该领域有“道路”这类地实体，道路有等级、长度、限速、当前时速、车道功能等重要属性。因此，我们可以围绕这些属性非常轻松的设计出一个城市中跟道路相关的指标，如道路总条数、总长度、各等级道路的条数、各等级道路的累计长度，并计算道路在不同等级上的数量分布和长度分布（一般 0 级路表示城际间高速、1 级路为城内快速道路、2 级路为主干道、3 级以下为细支道路）。这些指标从一个角度揭示了城市交通基础设施的资源分布情况。同理，通过城市知识体系，我们知道交通领域的物有“私家车”这类实体，因此可以计算私家车的总数量作为指标。

又因为城市知识体系中“车”跟“道路”有“通行”关系，我们可以把私家车的总数跟各等级的道路累计里程数相结合(尤其是 0 级高速和 1 级快速路的长度)，用道路长度除以车辆总数，得到单位车辆可使用的各级道路长度，便可从一个角度反应出一个城市里拥有的道路是否能够满足相应的车流。当然还需要其它指标的组合才能得出更加完备的结论。

提升针对性：由于各个城市的特点和需求不一样，我们需要为一些城市设计一些新的指标。此时，城市知识体系的泛化能力将发挥巨大的作用。

例如，前面提到的小型旅游城市在交通领域关注绿色出行主题，共享自行车是重点关注的物实体。如图 8 所示，由于自行车和私家车都属于交通领域物这类实体，且在该领域物-地存在“通行关系”，因此，私家车和自行车跟道路都有“通行关系”。按照上面计算单位车辆拥有的快速路长度指标，我们也可以类比泛化，计算单位自行车可以使用的自行车道里程数，并以此作为衡量绿色交通的基础设施资源是否够用的指标之一。同理，基于城市知识体系，我们还可以泛化出单位公交车可用公交车道里程数等等。这样即便我们去了一个关注绿色交通的旅游城市，也能很快基于城市知识体系设计出适合它的指标体系。

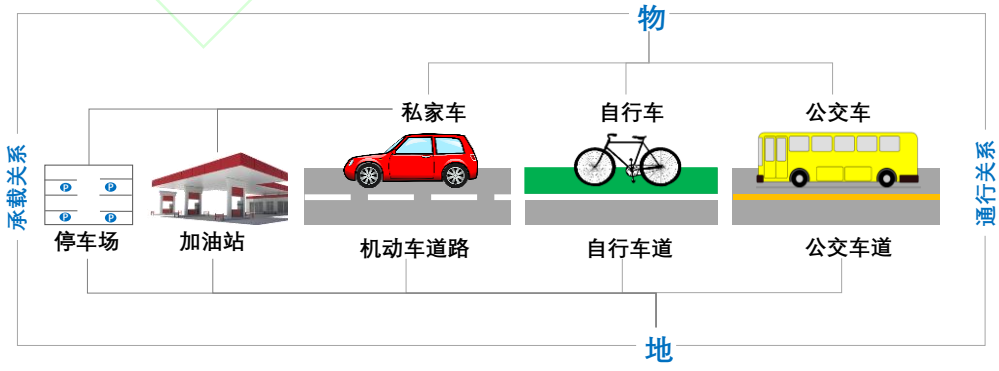

图 8 基于城市知识体系中物-地关系的指标泛化示例

Fig.8 Generalization of Object-Land Relationship Based on Urban Knowledge System

进一步，通过城市知识体系得知地-物之间还有“承载关系”，如私家车跟停车场、加油站的关系。因此可以用以上同样的方法计算出单位车辆可用的停车位数量以及单个加油站供给的私家

车数量。更进一步我们还可以推演出类似的指标，如单个充电桩供给的新能源汽车数量、单个加气站供给的天然气公交车数量等等。即便随着城市的不断发展，业务的不断演进，有新的“地”和“物”实体不断涌现，但他们之间的关系并未超出城市知识体系中定义的内容，可以持续沿用，因此，只要掌握了城市知识体系，就可以很快设计出合理的指标。

落实可行性：在落实指标是否可行时需要作两方面的判断，一是该指标在本项目上是否已经被其他人设计并计算过，前人已经计算过的指标应可直接判定可行。另一方面，如果之前没有计算过，该指标需要的数据在该项目上是否存在，如果有，数据质量如何。没有城市知识体系这两个问题都无法判断。

关于第一方面，由于不同人的变量命名习惯不同，使得即便已经有人计算了同样的指标，后面的人也不得而知，失去了最好的可行性验证的机会。参看本文第 3 小节的第 2）点中的详细介绍和示例。

关于第二方面，由于不同数据库对于表和字段的命名规则各不相同，在某个项目上计算指标所需的元素跟数据表具体字段的对应关系并不能直接传递至其它项目，用于判断新项目上该指标需要的数据是否存在且质量可行。只有把数据的字段跟城市知识体系的属性关联起来，我们才能判断计算指标需要的数据是否存在且可行。以上面的“单位自行车可以使用的自行车道里程数”指标为例，在计算自行车道总里程时，它需要道路这个实体的功能属性（为“自行车道”）和长度属性。在数据治理时，必须对接入的数据按照城市知识体系打好标签，哪个字段是道路的功能属性，哪个是道路的长度属性，否则后续的工作都无法开展。在构建基础库和主题库时，字段的命名规则更加应该按照城市知识体系的内容来命名，这样后续的工作都可以自动化开展起来。

综合以上描述，从城市知识体系中产生指标体系的思路如下：根据城市知识体系产生关键指标：按照单个实体属性、两个实体间关系的属性再到 3 个及以上的实体间关系的属性的顺序逐步生成重要指标。大部分属性都可根据三者及以下实体的关系来完成。在看单个实体属性时，可以通过总和以及在相应取值空间的分布来产生有意义的指标。在利用两者以上实体关系产生指标时，可以利用边的属性直接产生指标，也可以根据几个实体的关系，对各自属性进行简单的加减乘除运算得到有价值的指标。

根据知识体系和已有指标泛化新指标：通过将实体归属于相应类别，利用抽象的实体间关系来泛化衍生新实体间的关系。如，根据私家车-停车场的承载关系，泛化出新能源汽车和充电站之间的关系，由此可以类比“单位车辆拥有的停车位”产生“单位新能源车拥有的充电桩”的指标。

根据城市知识体系查找已有指标，判断新指标所需数据的可行性。

6.3 指导智能模型的构建

智能模型用于分析计算某类实体的某种属性、某类实体间关系的属性或者定义的某个指标。无论针对的目标是以上 3 种情况中的哪一种，城市知识体系都可指导智能模型结构的设计、特征的选择和已有模型的利用。

1）指导模型结构设计和特征选择。在需要对某类实体的某种属性进行分析计算时，可以利用该实体的其他属性作为辅助信息，也可以利用实体间的关系来借用与之关联的其他实体的属性，以及实体间关系上的属性。同理，当需要对某类实体间关系的属性预测时，可以利用该类关系的其它属性、关系关联的两类实体的属性作为输入。实体间关系可以作为模型结构设计的参考，如地-地之间的空间关系、连接关系和人-人之间的协作、血缘关系等同类关系可以构造图模型，人-地的访问关系等相互关系都可以构成二分图模型。

举一个简单的例子，如希望预测某条道路上的通行速度，其本质是对物-地实体间的“通行关系”的速度属性预测，此时，智能模型可以考虑“通行关系”的其它属性，比如时间点属性作为输入。通过该时间点属性还可以把前几个时间段的速度以及过去几天同时间段的速度作为模型的输入。此外，模型还需要利用该关系连接的地类实体（道路）的属性和物类实体（车辆）的属性作为输入。例如，道路的车道数、限速、曲直比等属性以及即将到达该道路的车辆数量。如果

采用简单的模型，可把前面提到属性作为特征输入，把通行速度作为输出，构建一个回归模型。如果想做更加精准的预测，需要把区域内的道路放到一起考虑，因为道路之间有地-地的连接关系，会相互影响。道路和道路之间的连接关系可以作为模型结构的重要参考，比如基于路网来设计一个图预测模型，把道路看成边，道路两端的端点看成节点，或者把道路看作节点，一条道路通往另一条道路的流量看成边。

再举一个更为复杂的例子，如图 9 所示，如果要预测未来到达红色区域的人流量，其本质是预测人-地两类实体之间的访问关系的数量属性。对于步行前来的人流量，根据城市知识体系的指引，地-地之间存在相邻的“空间关系”，要考虑该红色区域周边地区过来的人流。由于并非所有人都通过步行进入该区域，还涉及到乘坐公交车、地铁，或者私家车前往该区域的人流量。这些流量本质上是通过人-物实体间在交通领域中的“承载关系”（如人乘坐公交车）或“驾驶”关系（如人驾驶私家车），加上地-物之间的“承载关系”（如停车场承载私家车，地铁站承载地铁车辆等），以及地-地之间的从属关系（如某地铁站从属于该红色区域）来实现的人对该区域的“访问关系”。

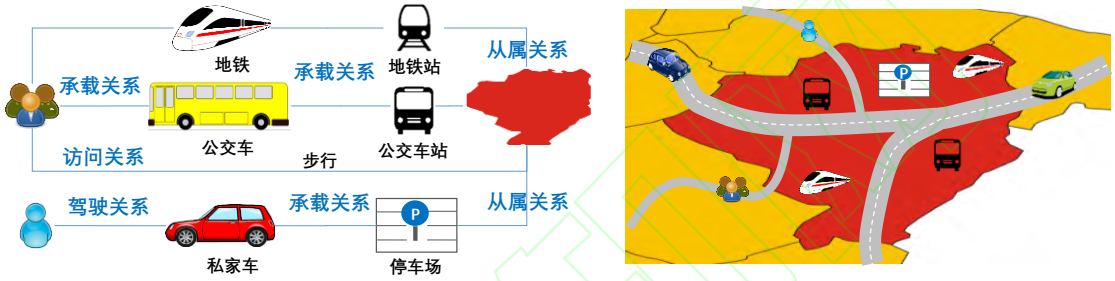

图 9 基于城市知识体系的模型设计和特征选择示例

Fig.9 Model Design and Feature Selection Based on Urban Knowledge System

有了以上基于知识体系的问题解析后，人们自然会想到找该区域的地铁站、公交车站和停车场，分析通过地铁、公交和自驾车前来该区域的人流量，最后把这些流量跟步行人员的流量相加即得到到达红色区域的总流量。如果想进入更深度的思考，根据城市知识体系，地铁站被地铁线路通过地-地之间的“连接关系”构成了一张网，因此可以引入某些图模型来预测整个地铁网络各个站点的人流量，把地铁站看作图中的节点，连接站点的地铁线路看成边，并把图 9 左半部分中的承载关系、乘坐关系的属性拿来作为模型的输入。

2）充分利用已有模型

在定义模型时，指定每个输入和输出跟实体属性或实体间关系属性的对应关系。当一个模型设定好之后，当有用户需要设计新的模型时，一旦选定了模型种类并设定了输入和输出，可以很快查询到之前是否有类似或者完全相同的模型已经被定义好了。相似的模型可以为新模型提供参考，拓宽设计思路；一样的模型可以直接复用，无需再次设计。

7 结 语

城市知识体系为基于知识的智能城市应用提供规范和理论支撑，指导数据治理过程，帮助从业者快速掌握行业知识，有效组织和表达知识，设计重要指标、智能模型和特征，挖掘、发挥知识的价值，并让不同应用产生的知识不断沉淀、复用和融合，让智能城市可以更快、更好的发展。

参考文献

[1] Zheng Yu. Introduction to Urban Computing[J]. Geomatics and Information Science of Wuhan University, 2015, 40(1): 1-13(郑宇. 城市计算概述[J]. 武汉大学学报·信息科学版, 2015, 40(1): 1-13)

[2] Zheng Yu. Urban Computing: Driving Smart Cities with Big Data and AI[J]. Communications of China Computer Federation, 2018, 14(1): 8-17(郑宇. 城市计算: 用大数据和AI驱动智能城市[J].中国计算

机学会通讯, 2018, 14(1): 8-17)
[3] Yu Zheng. Urban Computing[M]. Cambridge, UK: MIT Press, 2019
[4] Zheng Yu. Intelligent City Operating System[J]. Communications of China Computer Federation, 2020, 16(12): 39-44(郑宇. 智能城市操作系统[J]. 中国计算机学会通讯, 2020, 16(12): 39-44)
[5] Zheng Yu. Big Data + Artificial Intelligence Promote the Reform of Municipal Governance Model [J]. Faren Magazine, 2021(3): 60-63(郑宇.大数据+人工智能推动市域治理模式变革[J].法人,2021(3): 60-63)
[6] Zheng Yu. Unified Urban Governance Models[J]. Geomatics and Information Science of Wuhan University, 2022, 47(1): 19-25(郑宇. 城市治理一网统管[J]. 武汉大学学报·信息科学版, 2022, 47(1): 19-25)

Author:ZHENG Yu, PhD, he is the Vice President of JD.COM and the president of JD Intelligent Cities Research. Before Joining JD.COM, he was a senior research manager at Microsoft Research. He was the Editor-in-Chief of ACM Transactions on Intelligent Systems and Technology from 2015 to 2021, and has served as the program co-chair of Industrial Track at ICDE 2014, CIKM 2017 and IJCAI 2019. He was also a keynote speaker of AAAI 2019, KDD 2019 Plenary Keynote Panel, IJCAI 2019 Industrial Days and MDM 2021. His monograph, entitled Urban Computing, has been used as the first text book in this field. In 2013, he was named one of the Top Innovators under 35 by MIT Technology Review (TR35) and featured by Time for his research on urban computing. In 2016, he was named an ACM Distinguished Scientist and elevated to an IEEE Fellow in 2020 for his contributions to spatiotemporal data mining and urban computing. He is also a chair professor at Shanghai Jiaotong University and an adjunct professor at Nanjing University.
E-mail: msyuzheng@outlook.com

Foundation support: The National Key Research and Development Program of China（2019YFB2101805）; the National Natural Science Foundation of China（62076191）.