

引文格式:

林映利, 李红辉, 张春, 等. 基于知识图谱的协议转换方法研究 [J]. 集成技术, 2023, 12(1): 26-41.

Lin YL, Li HH, Zhang C, et al. Research of protocol conversion based on knowledge graph [J]. Journal of Integration Technology, 2023, 12(1): 26-41.

基于知识图谱的协议转换方法研究

林映利¹ 李红辉^{1,2*} 张春^{1,2} 闫佳和¹ 王梓博¹

¹(北京交通大学计算机与信息技术学院 北京 100044)

²(高速铁路网络管理教育部工程研究中心 北京 100044)

摘要 互联网+政务大数据具有跨领域、多协议、难融合的特点。在大数据采集汇聚过程中, 存在多种协议转换的需求, 要求网关能够实现统一协议适配转换, 为多源异构数据汇聚和数据融合提供数据支撑。传统的协议转换方法通常针对特定的协议转换需求设计的, 可扩展性较差, 不适用于多种协议转换。该文拟通过研究分析协议报文结构及协议转换特点, 提出一种协议转换知识图谱的构建方法。通过构建图谱的模式层以及数据层, 建立含有协议报文结构和报文字段映射关系的协议转换知识图谱。并在此基础上提出一种基于知识图谱的协议转换方法, 以实现不同协议之间的报文转换。同时, 通过协议转换应用实例以及与已有协议转换方法的对比实验, 验证了该方法的有效性。

关键词 互联网+政务大数据; 协议转换; 知识图谱; 协议报文

中图分类号 TN 915.04; TP 182 **文献标志码** A **doi**: 10.12146/j.issn.2095-3135.20220730001

Research of Protocol Conversion Based on Knowledge Graph

LIN Yingli¹ LI Honghui^{1,2*} ZHANG Chun^{1,2} YAN Jiahe¹ WANG Zibo¹

¹(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

²(Research Center for High-speed Railway Network Management, Ministry of Education, Beijing 100044, China)

*Corresponding Author: hhli@bjtu.edu.cn

Abstract Internet+government affairs big data has the characteristics of cross-domain, multi-protocol and difficult convergence. In the process of big data collection and aggregation, there is a need for a variety of protocol conversions, which require a gateway to realize a unified protocol adaptation conversion and provide the data support for multi-source heterogeneous data aggregation and data fusion. Traditional protocol conversion methods are usually designed for specific protocol conversion requirements, and their scalability is poor, so they are not suitable for the situation where there are multiple protocol conversion requirements.

收稿日期: 2022-07-30 修回日期: 2022-10-20

基金项目: 国家重点研发计划项目(2019YFB2102500)

作者简介: 林映利, 硕士研究生, 研究方向为知识图谱、协议转换; 李红辉(通讯作者), 教授, 研究方向为大数据技术与工程、人工智能、轨道交通信息技术等, E-mail: hhli@bjtu.edu.cn; 张春, 研究员, 研究方向为大数据智能分析、人工智能核心技术的研发与应用等; 闫佳和, 博士研究生, 研究方向为数据融合; 王梓博, 硕士研究生, 研究方向为知识图谱。

By studying and analyzing the structure of protocol messages and the characteristics of protocol conversion, this paper proposes a method to construct the knowledge graph of the protocol conversion. By constructing the schema layer and instance layer of the graph, a knowledge graph of protocol conversion containing the structure of protocol messages and the mapping relationship between message fields is established. On this basis, a protocol conversion method based on the knowledge graph is proposed to realize the packet conversion between different protocols. The effectiveness of the proposed method is verified by an application example of protocol conversion and a comparison experiment with the existing protocol conversion methods.

Keywords Internet+government affairs big data; protocol conversion; knowledge graph; protocol message

Funding This work is supported by National Key Research and Development Program of China (2019YFB2102500)

1 引言

互联网+政务大数据来源于多种异构信息。一种政务应用场景往往涉及多种不同来源的信息,如在城市外围或地质灾害多发地区布置地质灾害监测传感器,采集监测数据并传输到应急指挥中心,从而实现地质灾害实时动态监测和预警。在信息采集与传输的过程中,涉及物联网多传感器多协议间的数据传递。

信息来源不同,信息传输协议也不同,互联网+政务大数据呈现跨领域、多协议、难融合的特点。在大数据采集汇聚过程中,存在多种协议转换的需求。因此,要求网关能够实现统一协议适配转换,为异构数据关联和数据融合提供数据支持。

在传统协议转换研究方面,宗娜^[1]和 Zhou 等^[2]研究了实现 HTTP 协议与 CoAP 协议转换的理论与方法;赵小虎等^[3]设计了面向煤矿井下的多协议融合网关,实现了 CAN、ZigBee 等协议到以太网协议的转换。在通用协议转换研究方面,Saito 等^[4]提出一种物联网的应用协议转换方法,并制定了转换规则的语句和语法。

传统的协议转换方法通常针对特定的协议转换需求设计的,对于新的协议转换需求,需要重

新设计新的协议转换方法。在不同的政务应用场景中,传统的协议转换方法并不能灵活满足多种协议转换的需求。

知识图谱是人工智能领域的重要分支技术,最早由谷歌公司于 2012 年提出。知识图谱按应用场景可分为通用知识图谱和领域知识图谱两大类^[5]。其中,领域知识图谱以领域专业知识为主,不同领域知识图谱的构建与应用已有较多研究成果。在医学领域,张德政等^[6]将知识图谱应用于中医知识领域,构建了中医核心知识图谱;杨帅等^[7]针对 COVID-19 构建了科学文献、中西药治疗等知识图谱,实现了知识图谱的融合。在网络安全领域,廉龙颖^[8]和 Wang 等^[9]分别构建了网络空间安全知识图谱和网络安全社会工程领域知识图谱,为网络安全分析与信息检索提供了依据。而在协议知识领域,协议知识图谱的相关研究较少。徐健^[10]基于知识图谱构建了 5G 协议知识库,以解决 5G 协议知识面广、不易高效查询等问题。以字段为实体、字段映射关系为实体间关系构建知识图谱,可以体现协议转换过程中报文字段间的映射关系。但在协议转换领域,暂无协议转换相关知识图谱的研究。

本文拟通过研究协议报文结构及协议转换特点,提出一种协议转换知识图谱的构建方法。

通过构建图谱的模式层及数据层，建立含有协议报文结构和报文字段映射关系的协议转换知识图谱。在此基础上，提出一种基于知识图谱的协议转换方法，以实现不同协议之间的报文转换，并通过协议转换应用实例、与已有协议转换方法的对比实验，验证本文方法的有效性。

2 相关研究现状

2.1 知识图谱相关研究

知识图谱由知识及知识之间的关系组成^[11]，其基本组成单位是“实体-关系-实体”三元组及“实体-属性-属性值”实体与属性值对。知识图谱能够将不同种类的信息连接，形成关系网络，提供从“关系”的角度分析问题的条件^[12]。

知识图谱按应用场景可分为通用知识图谱(Universal Knowledge Graph, UKG)和领域知识图谱(Domain Knowledge Graph, DKG)。UKG以百科知识图谱、常识知识图谱为主。常见的百科知识图谱有DBpedia、YAGO、Freebase、Wikidata等，常识知识图谱有Cyc、ConceptNet、NELL等^[13]。DKG以各领域的专业知识为主，典型的应用场景有国防、金融、公安、工业、医疗等。如在医疗领域，CPubMed-KG^[14]是目前国内最大的全开放中文医学知识图谱，以疾病为中心的三元组规模达到了390万；在网络安全分析领域，CyberTwitter^[15]是用于发现和分析Twitter的网络安全情报，其作为一个开放情报来源，以tweet的形式分析实时信息并进行更新，提取关于各种可能威胁的情报；但在通信协议领域，协议相关知识图谱研究和应用较少。

知识图谱的构建技术包括知识抽取、知识融合、知识加工和知识更新等。其中，知识抽取是构建知识图谱的关键技术之一，包括实体识别、关系抽取和属性抽取。目前，在实体识别方面，主流方法是基于神经网络的命名实体识别。

Li等^[16]提出了基于BERT-BiLSTM-ATT-CRF模型的中文命名实体识别方法，且性能较好；谢腾等^[17]提出了一种基于BERT-BiLSTM-CRF模型的中文实体识别方法。目前，命名实体识别(Named Entity Recognition, NER)常用的模型有BERT-CRF、BiLSTM-CRF等。

2.2 协议转换相关研究

当使用不同协议的设备进行通信和数据交换时，需要进行协议转换。现有的协议转换方法可分为两种：具体的协议转换方法和通用的协议转换方法。因协议转换的多样性与复杂性，协议转换研究领域针对具体的协议转换方法的研究较多。例如，针对以太网HTTP/HTTP2协议与物联网CoAP协议间的转换，宗娜^[1]与Zhou等^[2]分别进行了相应的转换代理和互通特性的研究；为克服物联网多协议领域的差异问题，Zainudin等^[18]研究了一种多协议网关，以实现物联网中CoAP、MQTT和Websocket的多协议域之间的通信。但上述协议转换方法都是针对具体协议具体转换设计的解决方案，适用场景单一，所支持的协议转换类型和转换方式无法更改^[19]，程序不可扩展、泛化性差，无法满足政务大数据采集汇聚过程中的灵活需求。

目前，通用的协议转换方法的研究较少。Saito等^[4]提出一种物联网的应用协议转换方法，根据转换规则，将协议报文转换为Middle format，从而实现CoAP、MQTT、XMPP和SMTP等物联网应用层协议间的互相转换。Saito等提出的基于规则的协议转换，将各协议与Middle format的映射关系体现在规则中，随着协议数量的增加，不同协议转换为Middle format的规则也相应增加。

苗力心^[19]结合软件定义思想，将协议转换研究重心下移至物理链路层，提出了基于通用描述符的协议转换方法。通过构建与所有异构协议存在映射关系的协议映射元作为协议转换的中间

级, 并存储协议关键字段与映射元的映射关系。该协议映射元方法需为每个字段构建一个映射关系表, 当系统内有 n 种协议时, 需构建 $k \times n$ 个映射表, 其中, k 为协议中关键字段的数量。

上述通用协议转换方法将协议信息转换为中间格式, 用规则文件或映射关系表存储不同协议字段与中间格式的映射关系, 其规则文件或映射关系表会随着新协议的加入而增加。而本文提出的基于知识图谱的协议转换方法, 将字段作为实体, 通过实体间的关系可直接体现不同协议字段间的映射关系, 无须将协议转换为中间格式。协议间的映射关系通过知识图谱进行存储, 可在原知识图谱的基础上, 增加新的协议转换映射关系, 从而对协议转换知识图谱进行扩展与更新。

3 协议转换知识图谱构建方法

3.1 构建目的与方法

本文从 HTTP 与 CoAP 的转换着手, 针对应用层协议的转换进行研究。应用层协议报文一般由头部字段和可选的数据负载组成。协议转换是在保证报文功能不变的情况下, 进行两种不同协议报文之间字段信息的等价转换。因此, 协议转

换知识图谱应该以字段为主要实体, 以字段间的关系体现协议间的转换映射。

知识图谱的逻辑结构由模式层和数据层组成。模式层是知识图谱的核心层次, 其位于数据层之上, 作为构建知识图谱的实体及关系的知识框架。数据层存储具体知识, 以“实体-关系-实体”三元组及“实体-属性-属性值”实体与属性值对作为知识的基本表达方式。

协议转换所需知识大多来源于协议相关文档, 对相关专业的专业性和准确性要求较高。本文采用自顶向下的方式构建协议知识图谱。首先, 确定协议本体的设计范围, 根据协议转换的知识和经验设计协议本体, 构建知识图谱模式层, 所构建的模式层为知识图谱提供了数据框架。然后, 从数据源中提取实体, 根据数据框架确定实体间关系, 构建得到知识图谱数据层。所构建的数据层为知识图谱的主要组成部分。协议转换知识图谱的构建方法如图 1 所示。

3.2 协议本体设计

本体侧重于概念及其相互关系之间的形式化表达, 不涉及具体的实体与属性。协议本体架构作为协议知识图谱的知识框架, 即协议知识图谱的模式层, 规范了实体对象间的关系。为构建知

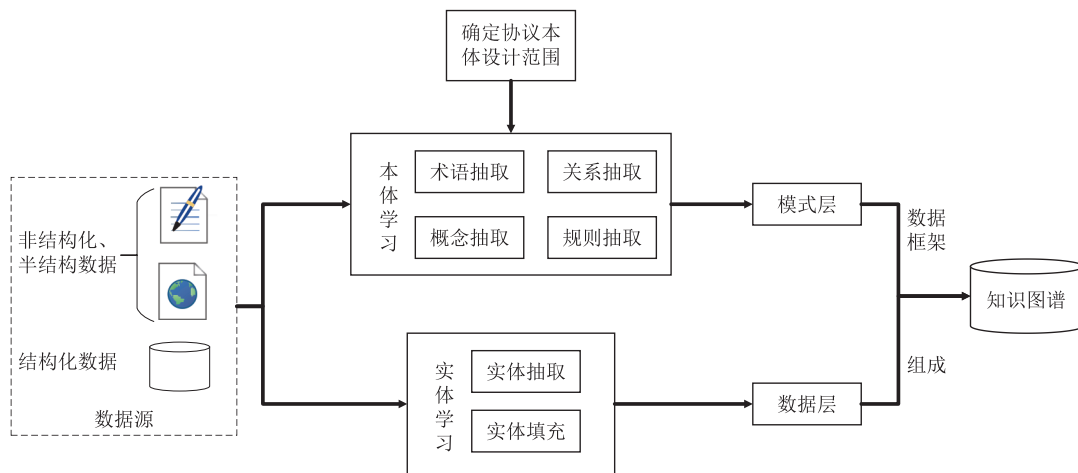


图 1 自顶向下的协议知识图谱构建方法

Fig. 1 Up-to-down construction method of knowledge graph of protocol

识图谱的模式层，需确定协议本体的设计范围。协议转换涉及两个不同协议间报文信息的交换，本体类别层次结构应考虑协议、报文、字段以及负载数据等概念。

3.2.1 协议本体类设计

基于协议核心概念，设计出 5 个核心类：Protocol, Message, Fields, Value, Payload Data。核心类还可进一步细化出更多的子类。本体中类的部分层次结构如图 2 所示。

Protocol 类用于描述协议转换知识图谱当前可支持的协议；Message 类用于描述协议拥有的报文类型，其中报文类型根据报文结构进行划分；Fields 类用于描述报文中包含的字段，Fields 类实例之间的映射关系是协议转换的核心之一；Value 类用于描述字段取值，某些字段可能拥有多个固定取值。在协议转换的过程中，Value 类的实例之间可能存在映射关系；Payload Data 类用于描述报文中的负载数据信息。

3.2.2 协议本体属性设计

本体类的属性可分为对象属性与数据属性。对象属性用于描述类与类之间的关系，本文设计了 8 种协议核心概念之间的关系，8 种协议本体的对象属性设计如表 1 所示。由表 1 可知，协议本体的对象属性需体现两个方面：

(1) 核心概念间的包含关系。如 Protocol 类与 Message 类之间可建立“has_message”的对象属性，表示协议拥有哪些报文类型；Message

类与 Fields 类之间可建立“has_field”的对象属性，表示该报文与字段的包含关系。

(2) 不同协议间转换的映射关系。如 Message 类的内部可建立“message_mapping”（报文映射）的对象属性，表示不同协议间报文的映射关系；Fields 类的内部可建立“direct_mapping”（直接映射）和“indirect_mapping”（间接映射），表示协议转换时不同协议字段间的直接映射、间接映射关系。

表 1 定义类的对象属性

Table 1 Defines the object properties of the class

类 (起始域)	关系 (→)	类 (对象范围)
Protocol	has_message	Message
Protocol	has_payload	Payload Data
Message	has_field	Fields
Message	message_mapping	Message
Fields	has_value	Value
Fields	direct_mapping	Fields
Fields	indirect_mapping	Fields
Value	value_mapping	Value

数据属性用于描述类的内部属性，针对 5 个核心类，本实验分别设计了相应的数据属性如表 2 所示。

3.2.3 协议本体架构

通过定义协议本体类、类的对象属性与数据属性，可最终得到协议本体。协议本体类与类的对象属性组成了协议本体架构，如图 3 所示。将协议本体架构作为协议知识图谱的模

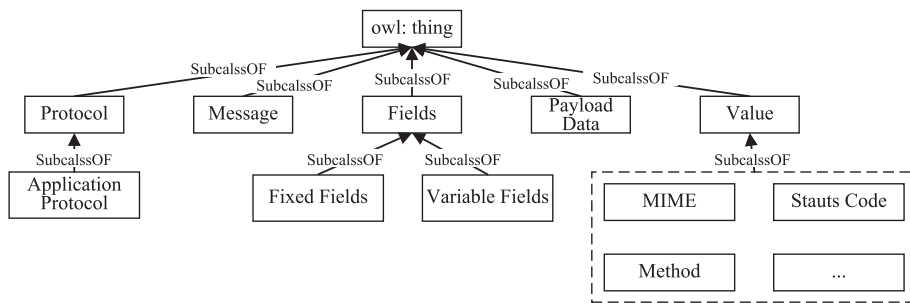


图 2 本体类别层次

Fig. 2 Hierarchy of classes in ontology

表 2 定义类的数据属性

Table 2 Defines the data properties of the class

核心类	数据属性	解释
Protocol	ProtocolName	协议名称, 如 HTTP、CoAP、MQTT、XMPP 等
	ProtocolTransportType	协议传输类型, 表示应用层协议是基于何种协议进行传输, 如 TCP 或 UDP
	ProtocolVersion	协议版本
Message	MessageName	报文名称
	MessageDataFormat	报文数据格式, 表示协议报文所采用的数据格式, 如文本、二进制
	MessageHasPayload	报文是否携带负载
Fields	FieldName	字段名称
	FieldType	字段类型, 表示该字段为固定字段或可选字段
	FieldDefaultValue	字段缺省值, 表示固定字段默认的取值
Value	Value	字段取值
	ValueType	值数据类型, 表示字段取值的数据类型
Payload Data	PayloadName	负载数据名称
	PayloadMediaType	负载数据媒体类型, 表示负载数据允许的媒体类型, 如二进制、文本、XML 等
	PayloadMAXLength	负载数据最大值, 表示协议允许的负载数据的最大值
	PayloadIdentifier	负载标识符, 表示负载数据起始的标志

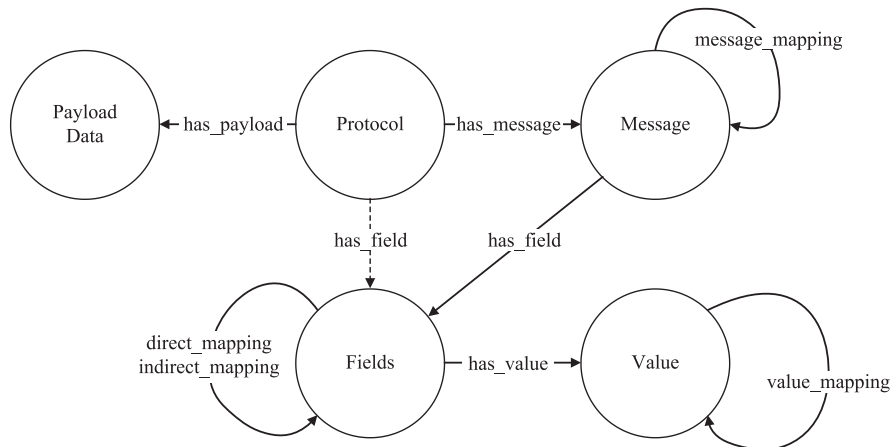


图 3 协议本体架构

Fig. 3 Protocol ontology architecture

式层, 规范了不同类的实体对象间的关系, 如 Protocol 类的实例与 Message 类的实例间可建立“has_message”关系; Message 类的实例之间可建立“message_mapping”关系; Fields 类的实例之间可建立“direct_mapping”或“indirect_mapping”关系。

3.3 协议知识抽取

协议转换相关知识主要来源于协议标准文

档、协议转换标准文档及相关网页搜索结果等。本文构建协议知识图谱采用的数据属于半结构化数据和非结构化数据, 部分数据来源如表 3 所示。

对于半结构化数据, 采取 Request+Xpath 的方式, 从 web 页面中提取; 对于非结构化数据, 利用 label studio 对数据进行实体标注, 选择 BiLSTM+CRF 模型进行实体抽取。以 HTTP

表 3 数据来源

Table 3 Data sources

类型	来源
协议标准文档	RFC 9110 ^[20] 、RFC 9112 ^[21] 、RFC 7252 ^[22] 等
协议转换标准文档	RFC 8075 ^[23] 等
相关网页搜索	百度百科、MDN Web Docs 等

和 CoAP 协议为例，实体抽取的数据集如表 4 所示。本文对相关数据进行实体抽取，最终得到相关实体 1 079 个，将提取到的实体匹配填充到协议本体中，作为本体实例。

表 4 协议实体数据集

Table 4 The dataset of the protocol entity

数据	来源	行数	词数
半结构化数据	相关网页		1 041
非结构化数据	标准文档	10 986	67 776
	相关网页	2 290	11 349

3.4 知识图谱构建实现

由于 Protégé 用于数据分析和可视化的功能有限，而 Neo4j 可便捷地检索、遍历、导航连接数据，且 Neo4j CQL 易于阅读和学习，因此，本文采用基于图模型的 Neo4j 数据库作为数据存储层。将提取的实体及实体间的关系等数据信息整理成结构化数据，以三元组的形式存储在 CSV 文件^[24]中，并调用 py2neo 库将 CSV 文件数据存入 Neo4j 数据库中。

以 HTTP 与 CoAP 在协议转换知识图谱中的部分映射关系为例，图谱可视化如图 4 所示，节点颜色的图例在图中左上角。图 4 展示了 HTTP、CoAP 及其部分字段、报文、值等实体和实体间的关系，如 HTTP 中的“status-code”字段实体与 CoAP 中的“code”字段实体存在“indirect_mapping”关系，“status-code”字段

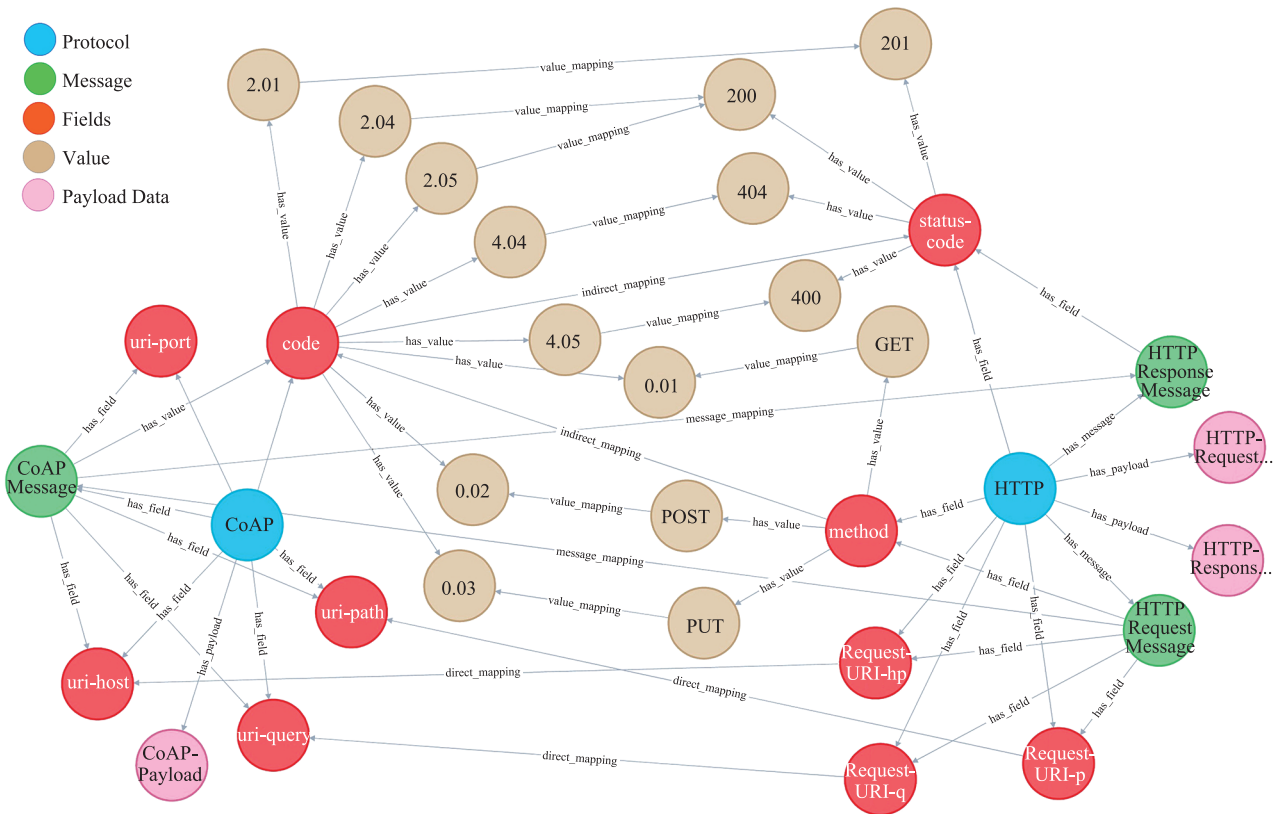


图 4 协议转换知识图谱(示例)

Fig. 4 Knowledge graph of protocol conversion (example)

的值“201”与“code”字段的值“2.01”存在“value_mapping”关系。

4 基于知识图谱的协议转换方法

4.1 基于知识图谱的通用协议转换框架

如图 5 所示, 基于知识图谱的通用协议转换框架由知识层和转换层组成。其中, 转换层负责将接收到的协议报文转换成另一种协议的报文, 包括报文解析、协议转换及报文生成等步骤, 该协议转换过程需要知识层提供相关知识, 以辅助实现协议转换; 知识层由协议转换知识图谱组成, 协议转换知识图谱以协议标准文档等作为数据源, 存储各协议报文结构、字段信息和字段间映射规则等知识, 该层负责为转换层提供知识、数据支撑。

4.2 基于知识图谱的协议转换方法

本节将介绍一种基于知识图谱的协议转换方法。为方便描述, 本文将被转换的报文称为原报

文, 原报文对应的协议称为原协议, 转换后得到的报文称为目标报文, 目标报文对应的协议称为目标协议。

4.2.1 转换流程设计

协议转换流程如图 6 所示, 接收到原报文后, 首先, 通过解析得到含原报文信息的 JSON 数据; 然后, 对 JSON 数据中的字段进行映射, 得到对应的目标报文字段以及字段值, 对目标报文剩下的固定字段进行缺省值补全, 得到含有目标报文信息的 JSON 数据; 最后, 根据含有目标报文信息的 JSON 数据生成目标报文并发送。

4.2.2 数据格式设计

在网关上进行协议转换, 网关接收原报文后, 解析得到原报文字段值和负载数据(若报文存在负载数据)等原报文信息, 将原报文信息以 JSON 格式进行保存。其中, 储存协议报文信息的 JSON 数据格式由协议名称键值对、字段与值的键值对及可选的切分标记键值对、负载数据键

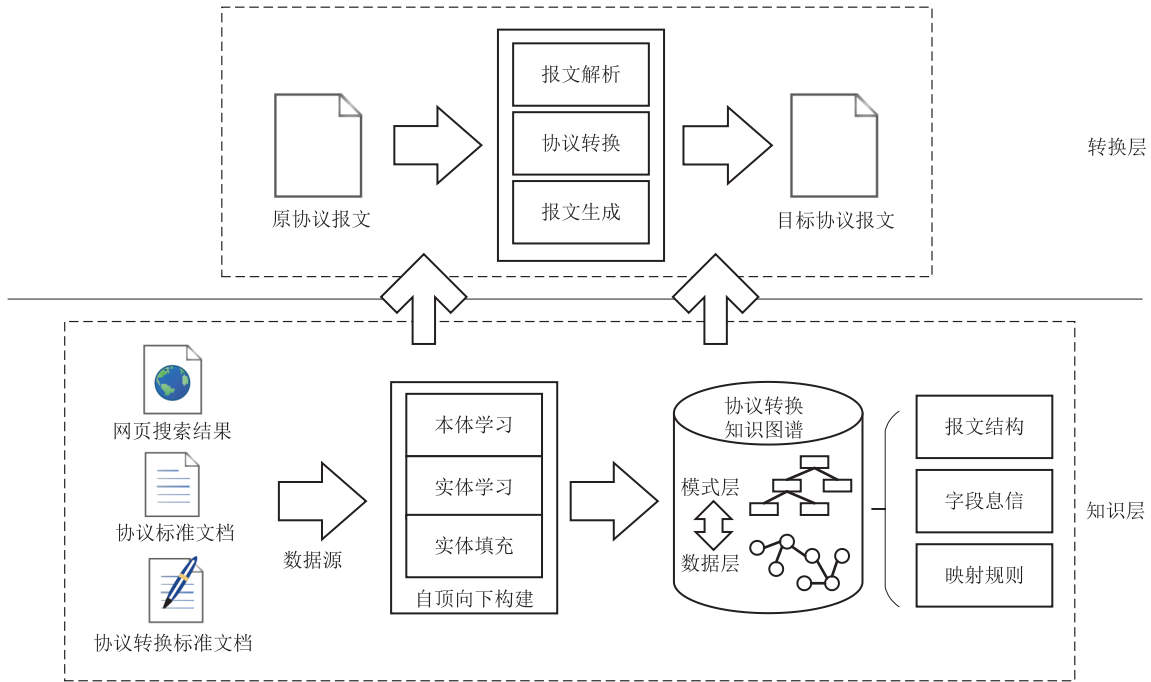


图 5 基于知识图谱的应用层协议转换框架

Fig. 5 Application layer protocol conversion model based on knowledge graph

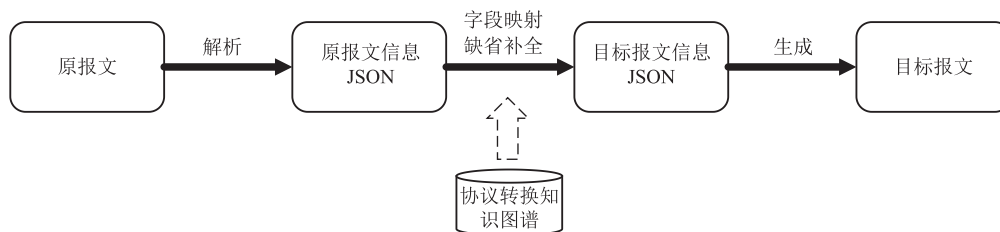


图6 协议转换流程

Fig. 6 Process of protocol conversion

值对组成，具体内容如下：

```

    {
      "protocol":<protocolName>
      "<fieldName>": <fieldValue>
      "<fieldName>":["<fieldValue>,<fieldValue>,..."]
      ...
      ("splitID":<ID>)
      ("<PayloadData>":<data>)
    }
  
```

4.2.3 协议转换方法设计

协议转换实现方法的具体步骤如下：

(1) 原报文解析

网关接收原报文并对其进行解析，得到原报文字段值等原报文信息后，按照规定的JSON格式进行保存。

(2) 原协议与目标协议的映射

在协议转换知识图谱中，若原报文的类型与目标报文的类型存在“message_mapping”关系，则原报文与目标报文之间的转换成立，原协议字段与目标协议字段间可能存在两种映射：“direct_mapping”与“indirect_mapping”。

“direct_mapping”与“indirect_mapping”。

对于原协议字段与目标协议字段，若存在直接映射关系，则其原协议字段值即目标协议字段值，如图7所示；若存在间接映射关系，则其字段值之间存在“value_mapping”（值映射）关系，由原协议字段值可以得到对应的目标协议字段值，如图8所示。

由协议转换知识图谱可知目标协议的最大负载信息。若原报文存在负载，且实际负载小于目标协议最大负载，则原报文负载即目标协议报文

负载。同理，若实际负载大于目标协议最大负载，则按照目标协议最大负载的大小，将原报文负载切割成多段，分别作为目标报文负载。

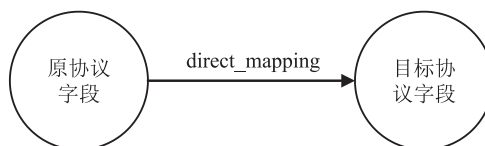


图7 字段间直接映射关系

Fig. 7 Direct mapping between fields

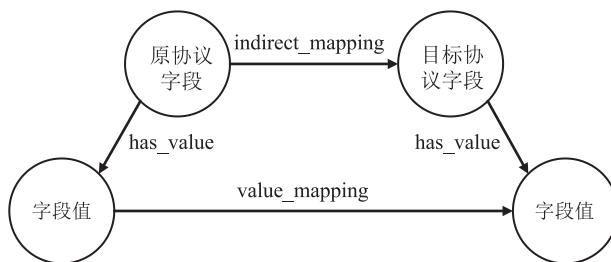


图8 字段间接映射关系

Fig. 8 Indirect mapping between fields

(3) 目标报文生成

根据协议转换知识图谱中字段间的映射关系，可以得到目标报文中与原协议有映射关系的字段与字段值。但一般情况下，原报文字段与目标报文字段并非都存在映射关系，仅通过字段映射得到的目标协议报文信息并不完整。因此，需在协议转换知识图谱中，查询该目标协议报文中所有字段属性为固定字段的字段，检查目标协议报文信息中是否包含所有固定字段，若缺少某固定字段，则将其缺省值作为字段值，补充相应键值对。由此可以得到含有目标报文信息的JSON

数据, 网关根据其携带的目标协议报文信息生成目标报文并发送。

5 方法应用: HTTP 协议与 CoAP 协议转换

5.1 HTTP 客户端与 CoAP 服务器的通信过程

实现协议转换的网关等同于原协议服务器, 负责接受来自原协议客户端的请求, 并作为目标协议的客户端, 将请求发送至目标协议服务器; 或等同于原协议客户端, 接收原协议服务器的响应, 并作为目标协议的服务器, 将响应发送至目标协议客户端。

以 HTTP 协议与 CoAP 协议的转换为例。首先, HTTP 客户端向网关发送 POST 请求, 网关将 HTTP 请求报文转换为 CoAP 协议报文, 并发送给 CoAP 服务器; 然后, CoAP 服务器向网关发送 CoAP 响应, 网关将 CoAP 协议报文转换为 HTTP 响应报文, 并发送给 HTTP 客户端。在上述过程中, CoAP 端均采用 CON 模式进行传输。HTTP 客户端通过网关与 CoAP 服务器通信的过程如图 9 所示。

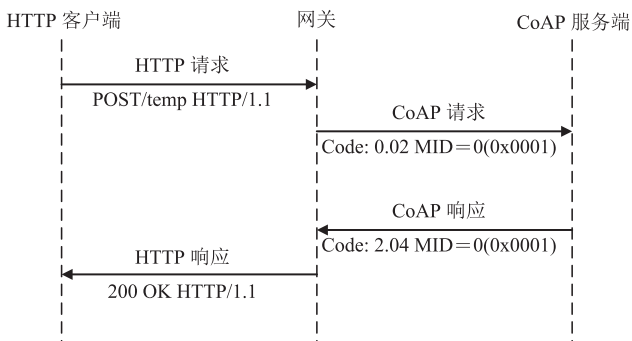


图 9 HTTP 客户端与 CoAP 服务器通信过程

Fig. 9 Communication between HTTP client and CoAP sever

5.2 协议转换过程

基于知识图谱的协议转换算法流程图如图 10 所示。

将 HTTP 请求报文转换为 CoAP 报文的过程如下:

(1) HTTP 请求报文解析

从 HTTP 客户端接收报文后, 解析得到由协议类型的键值对、字段名与字段值的键值对和负载键值对组成的 HTTP 请求报文信息的 src.json。HTTP 请求报文示例与解析结果如图 11 所示。

(2) HTTP 到 CoAP 的映射

分别查询该 HTTP 请求报文中的字段与目标协议 CoAP 的字段, 是否存在直接映射或者间接映射的关系。例如, Content-Type 字段与 Content-Format 选项存在间接映射关系, Content-Type 字段的值为 “text/plain”, 值映射后得到 Content-Format 选项的值为 “text/plain;charset=utf-8”。由协议转换知识图谱可知 CoAP 的最大负载信息, 若 HTTP 请求报文负载小于 CoAP 的最大负载, 则 HTTP 请求报文负载即为 CoAP 报文负载; 若 HTTP 请求报文负载大于 CoAP 的最大负载, 则将原负载内容按 CoAP 最大负载的大小切割成多段, 分别作为 CoAP 报文负载。HTTP 到 CoAP 的映射结果如图 12 所示。

(3) CoAP 报文生成

查询该 CoAP 报文中所有字段属性为固定字段的字段。若缺少某固定字段, 则将其缺省值作为字段值, 补充相应键值对, 得到含有 CoAP 协议报文信息的 dst.json。如转换后的中间结果缺少 CoAP 的 Ver 字段, 其缺省值为 1, 补全后得到协议转换结果如图 13 所示。根据解析 dst.json 得到的 CoAP 报文信息, 生成 CoAP 报文并发送给 CoAP 服务器。

5.3 性能评估

5.3.1 实验环境

通过分析 HTTP 与 CoAP 协议转换执行时间与内存占用, 对本文方法进行性能评估, 实验环境如表 5 所示。

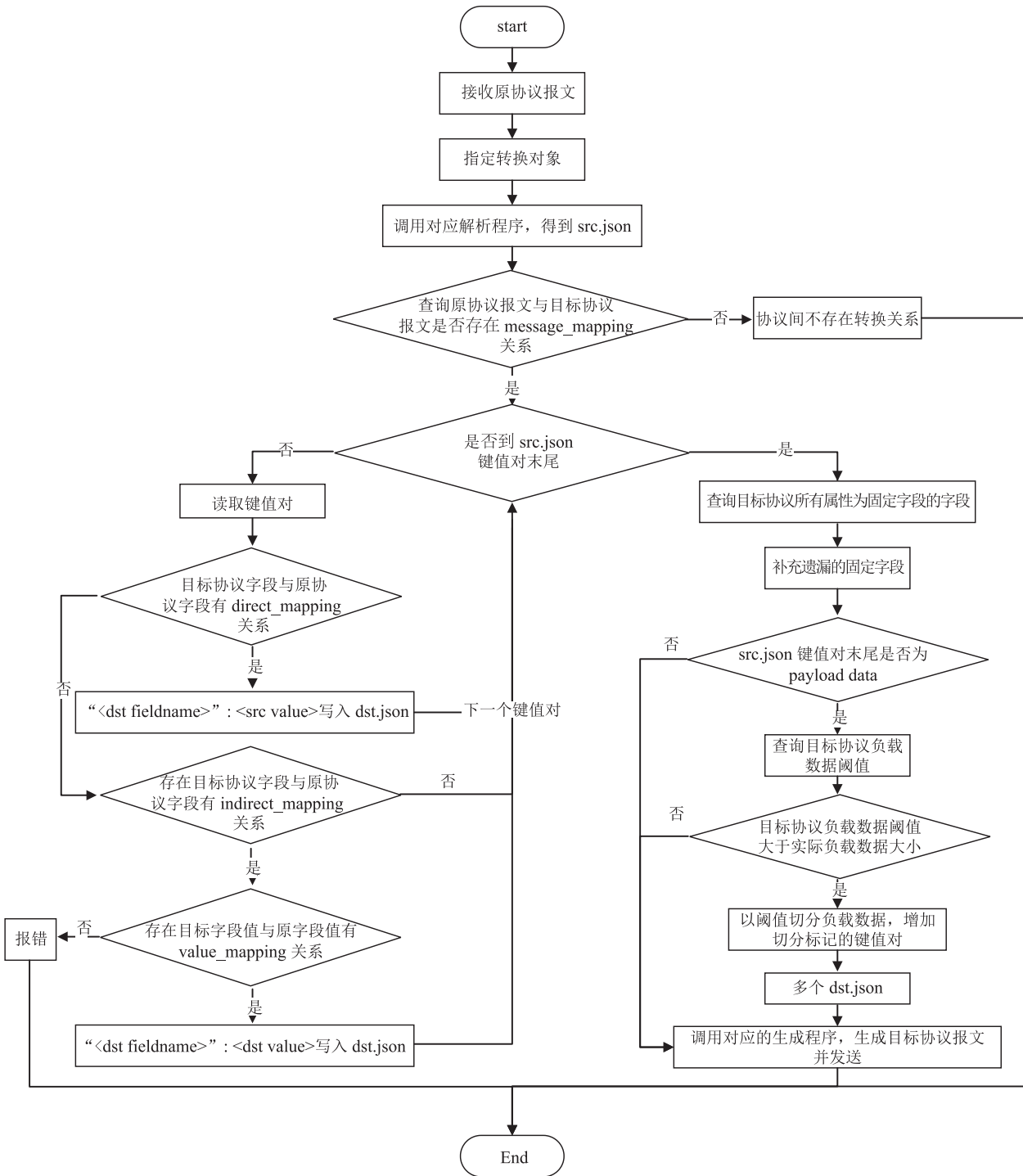


图 10 基于知识图谱的协议转换算法流程图

Fig. 10 The flow chart of protocol conversion algorithm based on knowledge graph

5.3.2 实验过程设计

从接收 HTTP 客户端请求到返回请求响应为一个协议转换周期。HTTP 客户端分别执行 100

次 PUT、GET、POST 请求，统计实验过程中周期平均执行时间、最大执行时间、最小执行时间，并统计每个周期中 HTTP 请求转换为 CoAP

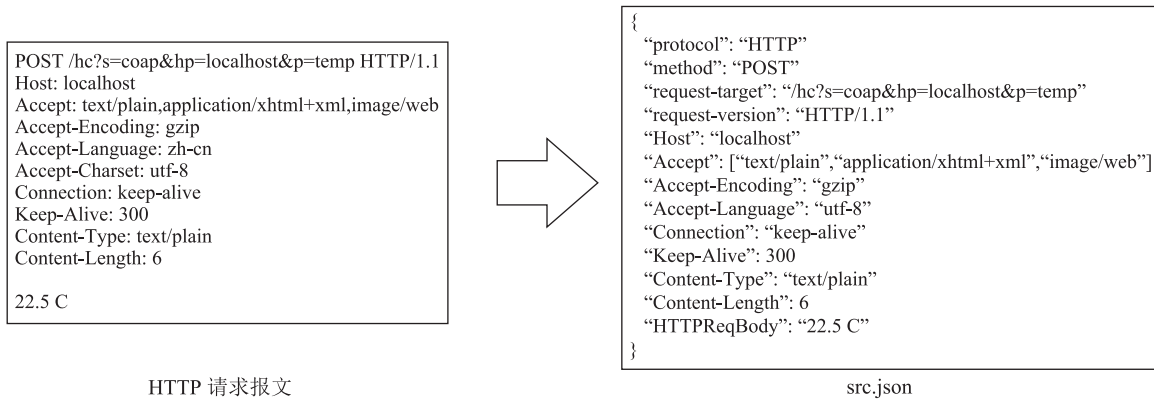


图 11 HTTP 请求报文与解析结果

Fig. 11 HTTP request message and parsing result

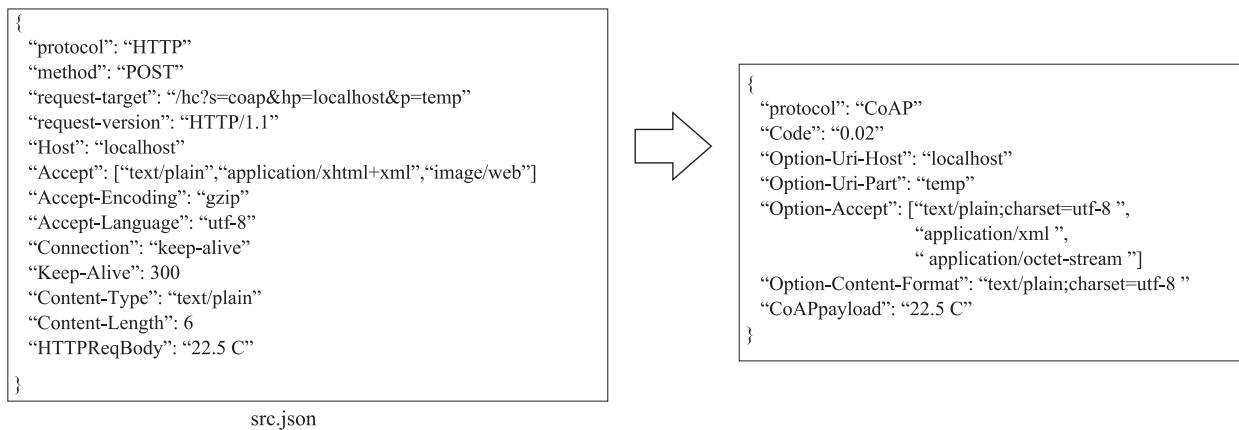


图 12 HTTP 到 CoAP 的映射结果

Fig. 12 Mapping result of HTTP to CoAP

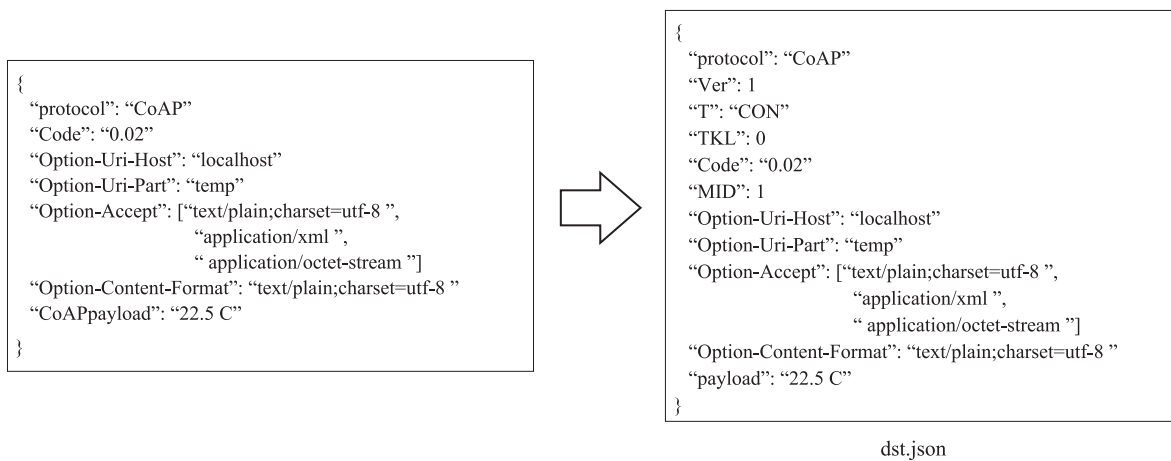


图 13 协议转换结果

Fig. 13 Protocol conversion result

表5 实验环境

Table 5 Experimental environment

实验环境	版本信息
PC	Windows 11, 内存 16 GB, 处理器 i5-12500H
Python	3.10.8
PyCharm	2022.2.1
Neo4j Desktop	1.5.0
Neo4j	4.4.11
Wireshark	3.6.8

请求和 CoAP 响应转换为 HTTP 响应的平均执行时间、最大执行时间、最小执行时间。执行时间通过 Wireshark 抓取到的每个周期 HTTP、CoAP 协议的请求报文和相应响应报文时间之间的差值进行计算。

通过使用 python 第三方库 memory_profiler, 监控 HTTP 客户端连续发送 PUT 请求时转换程序的内存占用情况。由 memory_profiler 生成内存占用随时间变化的曲线图, 可直观地观测到协议转换程序的内存占用情况。

5.3.3 实验结果及分析

(1) 执行时间

协议转换执行时间的统计结果如表 6 所示, 协议转换周期的平均执行时间约 32 ms, 最小执行

表6 协议转换平均、最大、最小执行时间

Table 6 Average, maximum, minimum execution time of protocol conversion

请求方法	协议转换过程	t_{trans} (ms)		
		平均执行时间	最小执行时间	最大执行时间
PUT	Total	32.911	17.185	47.508
	HTTP→CoAP	5.601	4.406	19.481
	CoAP→HTTP	27.310	12.779	29.574
GET	Total	31.512	17.001	34.074
	HTTP→CoAP	4.765	4.072	6.271
	CoAP→HTTP	26.747	12.473	29.309
POST	Total	31.898	17.238	37.754
	HTTP→CoAP	4.916	4.246	6.846
	CoAP→HTTP	26.982	12.722	32.857

时间约 17 ms, 最大执行时间 < 48 ms; HTTP 请求转换为 CoAP 请求的平均执行时间 < 6 ms, 最小执行时间约 4 ms, 最大执行时间 < 20 ms; CoAP 响应转换 HTTP 响应的平均执行时间约 27 ms, 最小执行时间约 12 ms, 最大执行时间 < 33 ms。

宗娜^[1]给出了 HTTP 与 CoAP 转换代理实现 100 次 GET、POST 请求的平均执行时间, 其与本方法实现 HTTP 与 CoAP 协议转换周期的平均执行时间的对比如表 7 所示。由表 7 可知, 本方法所实现的 HTTP 与 CoAP 的协议转换用时更少, 转换效率更高。

表7 平均执行时间对比

Table 7 Comparison of average execution time

相关文献	GET (ms)	POST (ms)
文献 ^[1]	56.3	69.22
本文	31.512	31.898

PUT、GET、POST 请求的协议转换周期中, 请求与响应平均执行时间占比如图 14 所示。执行 100 次 PUT、GET、POST 请求过程中, CoAP 响应转换 HTTP 响应的平均执行时间分别占总转换执行时间的 82.98%、84.88%、84.59%, HTTP 请求转换为 CoAP 请求的平均执行时间占总转换执行时间的 17.02%、15.12%、15.41%。CoAP 响应转换为 HTTP 响应的过程中, 调用 HTTP 相关外部库进行报文的发送, 占

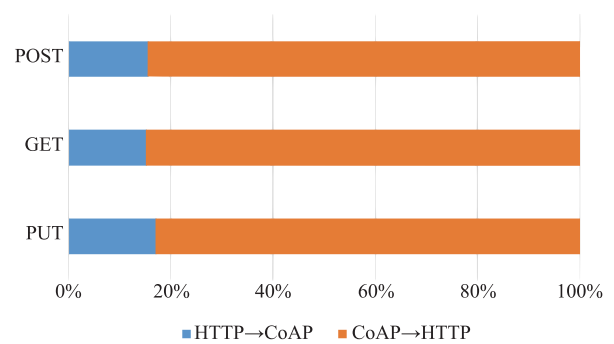


图14 请求与响应的平均执行时间占比

Fig. 14 Proportion of average execution time of requests and responses

用了大量时间, 导致 CoAP 响应转换 HTTP 响应的平均执行时间占比较高。

(2) 内存占用情况

协议转换程序执行过程中内存占用变化情况如图 15 所示。由图 15 可知, 程序开始执行时, 内存占用首先上升至 80 MiB 左右, 随着请求次数的增加, 内存占用最终保持在 90 MiB 左右, 占系统总运行内存的 0.55%。由此可知, 运行协议转换程序对于系统内存的占用不高, 系统性能良好。

6 讨论与分析

互联网+政务大数据来源于多种异构信息, 具有跨领域、多协议、难融合的特点。在大数据采集汇聚的过程中, 存在多种协议转换需求, 需要网关实现统一的协议适配转换。本文提出了一种协议转换知识图谱的构建方法, 建立含有协议报文结构和报文字段映射关系的协议转换知识图谱。并在此基础上, 提出了一种基于知识图谱的协议转换方法, 以实现不同协议之间的报文转换。

在协议转换研究领域, 基于知识图谱的协议

转换是一个较新的研究方向。将知识图谱、协议转换等作为关键词, 在 CNKI、Google Scholar、IEEE 等数据库中进行搜索, 发现国内外暂无类似研究成果和参考文献。

本文对协议的结构层级进行研究与分析, 依据含义将其划分为 5 类核心概念: Protocol、Message、Fields、Value 和 Payload Data, 并依此设计了协议本体。本文提出的基于知识图谱的协议转换方法, 可在已有的协议转换知识图谱的基础上, 增加新协议的报文字段信息与新的协议转换映射关系, 对协议转换知识图谱进行扩展与更新, 具有良好的泛用性和可扩展性, 能够灵活适应多种应用层协议转换需求。在协议转换执行时间与内存占用方面, 将本文提出的方法与现有案例方法进行性能测试与评估分析, 验证了本文方法的有效性。但本文仍存在以下不足: (1) 解析与生成报文受到具体程序实现的限制; (2) 缺少对应的成熟数据集, 图谱构建较为困难。

本文创新性地提出了一种协议转换知识图谱的构建方法和一种基于知识图谱的协议转换方法。这些方法基于 HTTP、CoAP 协议转换, 可能存在一定的局限性, 未来将研究更多的应用层协议的报文结构与映射关系, 以完善协议转换知

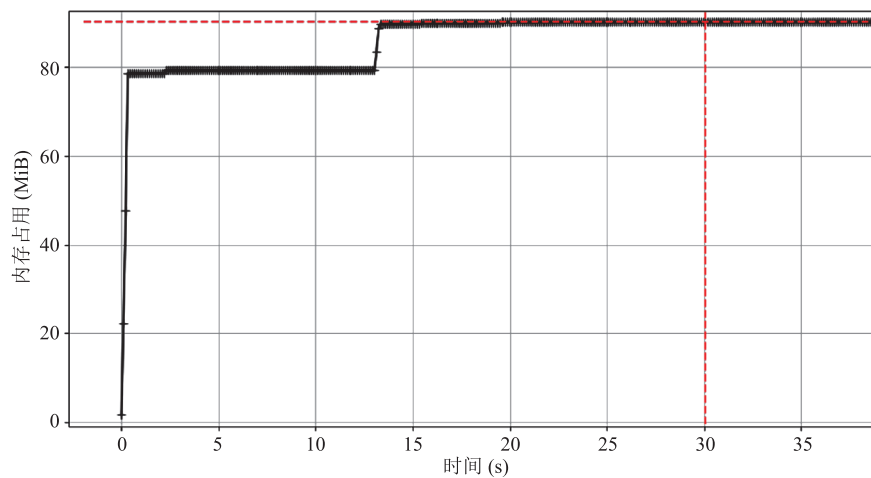


图 15 协议转换程序内存占用情况

Fig. 15 Memory usage of protocol conversion

识图谱, 并对利用协议转换知识图谱解析和生成应用层协议报文开展相关研究。

参 考 文 献

- [1] 宗娜. M2M 通信中 HTTP 与 CoAP 的协议转换代理的研究与实现 [D]. 北京: 北京邮电大学, 2016.
Zong N. Research and implementation of protocol conversion proxy between HTTP and CoAP in M2M communication [D]. Beijing: Beijing University of Posts and Telecommunications, 2016.
- [2] Zhou JP, Wei GY. Two patterns in conversion between HTTP2 and CoAP: request-reponse and publish-subscribe [C] // Proceedings of the 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, 2019: 1178-1184.
- [3] 赵小虎, 王宽, 沈雪茹, 等. 面向煤矿井下的多协议融合网关设计 [J]. 工矿自动化, 2019, 45(1): 6-12.
Zhao XH, Wang K, Shen XR, et al. Design of multi protocol fusion gateway for coal mine [J]. Industrial and Mine Automation, 2019, 45(1): 6-12.
- [4] Saito K, Nishi H. Application protocol conversion corresponding to various IoT protocols [C] // Proceedings of the 46th Annual Conference of the IEEE Industrial Electronics Society, 2020: 5219-5225.
- [5] 顾丹阳, 李明倩, 权冀川, 等. 基于本体的主战武器装备知识图谱构建 [J]. 指挥控制与仿真, 2021, 43(6): 14-20.
Gu DY, Li MQ, Quan JC, et al. Ontology based knowledge graph construction of main battle weapon equipment [J] Command Control & Simulation, 2021, 43(6): 14-20.
- [6] 张德政, 谢永红, 李曼, 等. 基于本体的中医知识图谱构建 [J]. 情报工程, 2017, 3(1): 35-42.
Zhang DZ, Xie YH, Li M, et al. Construction of knowledge graph of traditional Chinese medicine based on the ontology [J]. Information Engineering, 2017, 3(1): 35-42.
- [7] 杨帅, 王小红, 赵志刚, 等. COVID-19 知识图谱构建与应用研究 [J]. 青岛大学学报(工程技术版), 2021, 36(4): 22-29.
Yang S, Wang XH, Zhao ZG, et al. COVID-19 knowledge graph construction and application research [J]. Journal of Qingdao University (Engineering & Technology Edition), 2021, 36(4): 22-29.
- [8] 廉龙颖. 基于本体的网络空间安全知识图谱的构建方法 [J]. 黑龙江科技大学学报, 2021, 31(2): 254-258.
Lian LY. Ontology-based construction method for cyberspace security knowledge graph [J]. Journal of Heilongjiang University of Science & Technology, 2021, 31(2): 254-258.
- [9] Wang ZG, Zhu HS, Liu PP, et al. Social engineering in cybersecurity: a domain ontology and knowledge graph application examples [J]. Cybersecurity, 2021, 4(1): 31.
- [10] 徐健. 基于知识图谱构建 5G 协议知识库 [J]. 移动通信, 2020, 44(8): 73-79.
Xu J. A knowledge graph-based 5G protocol knowledge base construction [J]. Mobile Communication, 2020, 44(8): 73-79.
- [11] 李涛, 王次臣, 李华康. 知识图谱的发展与构建 [J]. 南京理工大学学报, 2017, 41(1): 22-34.
Li T, Wang CC, Li HK. Development and construction of knowledge graph [J]. Journal of Nanjing University of Science and Technology, 2017, 41(1): 22-34.
- [12] 凡天娣, 景旭文, 肖志建, 等. 基于本体的船舶焊接工艺知识图谱构建 [J]. 电焊机, 2019, 49(12): 8-13.
Fan TD, Jing XW, Xiao ZJ, et al. Ontology-based construction of ship welding procedure knowledge graph [J]. Electric Welding Machine, 2019, 49(12): 8-13.
- [13] 董兴芝. 面向智能高铁安全保障的知识图谱构建及应用关键技术研究 [D]. 北京: 中国铁道科学研究院, 2022.
Dong XZ. Research on key technology of knowledge graph construction and its application for intelligent high-speed railway safety assurance [D]. Beijing: China Academy of Railway Sciences,

- 2022.
- [14] 陈清财. 大规模中文开放医学知识图谱CPubMed-KG [EB/OL]. (2022-03-25)[2022-10-08]. <http://cs.hitsz.edu.cn/info/1044/5586.htm>.
Chen QC. Large-scale Chinese open medical knowledge graph CPubMed-KG [EB/OL]. (2022-03-25)[2022-10-08]. <http://cs.hitsz.edu.cn/info/1044/5586.htm>.
- [15] Mittal S, Das PK, Mulwad V, et al. CyberTwitter: using Twitter to generate alerts for cybersecurity threats and vulnerabilities [C] // Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2016: 860-867.
- [16] Li DY, Tu YF, Zhou XS, et al. End-to-end Chinese entity recognition based on BERT-BiLSTM-ATT-CRF [J]. ZTE Communications, 2022. DOI: 10.12142/ZTECOM.2022S1005.
- [17] 谢腾, 杨俊安, 刘辉. 基于 BERT-BiLSTM-CRF 模型的中文实体识别 [J]. 计算机系统应用, 2020, 29(7): 48-55.
Xie T, Yang JA, Liu H. Chinese entity recognition based on BERT-BiLSTM-CRF model [J]. Computer Systems applications, 2020, 29(7): 48-55.
- [18] Zainudin A, Syaifudin MF, Syahroni N. Design and implementation of node gateway with MQTT and CoAP protocol for IoT applications [C] // Proceedings of the 4th International Conference on Information Technology, Information Systems and Electrical Engineering, 2019: 155-159.
- [19] 苗力心. 软件定义协议转换关键技术研究 [D]. 郑州: 战略支援部队信息工程大学, 2020.
Miao LX. Research on key technologies of software-defined protocol conversion [D]. Zhengzhou: Information Engineering University, 2020.
- [20] RFC 9110. HTTP Semantics [S].
- [21] RFC 9112. HTTP/1.1 [S].
- [22] RFC 7252. The Constrained Application Protocol (CoAP) [S].
- [23] RFC 8075. Guidelines for Mapping Implementations: HTTP to the Constrained Application Protocol (CoAP) [S].
- [24] 张润梅, 杨超, 尹蕾, 等. 徽派建筑知识图谱的半自动化构建 [J]. 安徽建筑大学学报, 2021, 29(5): 13-19.
Zhang RM, Yang C, Yin L, et al. Semi-automated build of Huizhou architectural knowledge graph [J]. Journal of Anhui Jianzhu University, 2021, 29(5): 13-19.