

物联网语义关联和决策方法的研究

冯建周^{1,2} 宋沙沙^{1,2} 孔令富^{1,2}

摘要 物联网系统包含大量的感知设备,产生大量孤立和异构的感知数据,形成数据孤岛.如何将不同感知设备产生的数据进行有效的语义关联、构建跨域的数据关联模型、屏蔽数据异构性、实现综合智能决策是物联网研究的关键问题.本文以物联网系统为研究对象,从语义入手,构建以环境-资源-行为为核心的物联网本体模型;基于关联数据方法构建以行为关联模型和资源关联模型为核心的物联网语义关联网络;提出一种基于事件的推理算法实现语义推理,从而更好地实现对物联网系统的智能决策;最后,通过两个智能家居系统的实例验证了本文方法的可行性,通过构建不同方法的查询实例验证了本文方法的优越性.

关键词 物联网,数据孤岛,领域本体,语义关联,数据推理

引用格式 冯建周,宋沙沙,孔令富.物联网语义关联和决策方法的研究.自动化学报,2016,42(11):1691-1701

DOI 10.16383/j.aas.2016.c150885

Research on Semantic Association and Decision Method of the Internet of Things

FENG Jian-Zhou^{1,2} SONG Sha-Sha^{1,2} KONG Ling-Fu^{1,2}

Abstract The Internet of Things (IoT) system contains a large number of sensing devices, as a result, massive isolated and heterogeneous data are generated and forming isolated islands of data. It is a key issue how to realize the effective semantic association of data produced by different sensing devices, and to construct a cross-domain data correlation model, so as to solve data heterogeneity and realize intelligent decision making. In this paper, with the IoT system as the research object, an IoT ontology model is built with environment-resource-behavior as the core. Based on linked data method, the IoT semantic linked network is constructed with the behavior linked model and the resource linked model as the core. And a reasoning algorithm based on event has been put forward to realize the semantic reasoning, so as to better realize the intelligent decision-making of the IoT system. At last, its feasibility is verified through two examples of smart home system, and its superiority is verified through building different query instances.

Key words The Internet of Things (IoT), data island, domain ontology, semantic association, data reasoning

Citation Feng Jian-Zhou, Song Sha-Sha, Kong Ling-Fu. Research on semantic association and decision method of the Internet of Things. *Acta Automatica Sinica*, 2016, 42(11): 1691-1701

随着物联网应用的不断发展,物联网系统的规模越来越庞大,不同系统间的关系也越来越复杂.越来越多的智能设备产生了海量、多源和异构的感知数据,给系统内和系统间的资源交互、数据关联和推理等带来了极大困难.以智能家居为例,室内的灯光、空调、暖气、电视、冰箱以及光强传感器、温度传感器、红外蜂鸣报警器、开关继电器等设备随时产生大量的异构多源数据,在表现形式、衡量单位、组织

结构等方面均不相同,彼此之间毫无关联.继而以智能家居为中心,还涉及到智能电力系统、智能供暖系统、智能娱乐系统、智能安防系统、智能医疗系统、智能交通系统等不同的物联网系统,不同系统的设备连接方式、通信协议、数据产生格式等方面均有较大差别.但是,以智能家居为纽带,各个系统间在实际应用过程中是互相联系和影响的.如何屏蔽感知数据的异构性和孤立性,实现数据互联和融合成为业内研究的热点问题.

针对以上问题,日益成熟的语义 Web 技术^[1-2]提供了新的解决思路.语义 Web 被看作是解决异构系统集成和协作问题的关键技术. Sheth 等^[3]提出语义传感器网络 (Semantic sensor web, SSW) 的概念,采用语义 Web 技术描述传感器网络,实现各种感知数据的有效集成. Coccoli 等^[4]提出了语义物联网 (Semantic web of things, SWOT) 的概念,试图以语义 Web 的方式描述物联网系统. De 等^[5]提出一种关于物联网服务的语义建模方法,采用 OWL-

收稿日期 2015-12-29 录用日期 2016-05-16
Manuscript received December 29, 2015; accepted May 16, 2016
河北省自然科学基金 (F2014203192), 燕山大学博士基金 (B845) 资助

Supported by Natural Science Foundation of Hebei Province (F2014203192) and Doctoral Fund of Yanshan University (B845)
本文责任编辑 夏元清

Recommended by Associate Editor XIA Yuan-Qing
1. 燕山大学信息科学与工程学院 秦皇岛 066004 2. 河北省虚拟技术与系统集成重点实验室 秦皇岛 066004

1. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004 2. Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao 066004

DL^[6] 建模物联网资源, 采用 OWL-S^[7] 建模物联网服务, 实现服务间的交互. W3C 语义传感器网络孵化群开发一个语义传感器网络 (Semantic sensor net, SSN) 本体^[8], 采用一个高层模型表示传感器设备、能力、观测和采集数据等内容, 但其对感知设备、观察和测量数据及平台方面的描述不能直接扩展到物联网领域的具体组件. 总之, 基于物联网系统的具体应用构建适合的领域本体模型, 从而屏蔽多源感知数据的语义异构性, 实现各种智能设备的互操作是当前研究物联网数据融合问题的一个重要的解决思路. 但是, 物联网系统远比无线传感器系统复杂: 1) 物联网系统往往跨多个应用领域, 拥有更多的智能设备, 除了大量的传感器, 还有 RFID、智能仪器、智能家电、智能机械等多种情况. 2) 物联网是一个高度动态的环境, 每时每刻都有新的节点加入或者原有节点离开. 随时会有新的不可预知的数据格式或者协议需要被系统支持和处理. 3) 物联网系统是一个高度受限的网络, 除了像传统的无线传感器网络一样受到时空等限制, 智能设备本身对系统的限制更加多样化、复杂化, 给物联网系统的语义建模带来了更多的挑战. 同时, 仅仅为物联网系统增加语义信息并不能完全解决各种智能物体间复杂的交互方式、行为预测、智能响应等问题. 为了能够实现智能交互和决策, 李戈等提出了一种基于环境的物联网服务建模方法, 充分考虑环境对物联网资源的影响, 将环境因素作为物联网服务的一部分进行建模^[9]. 徐杨等提出一个基于语义覆盖网的物联网信息资源描述、推理和应用模型, 以多智能体系统决策支持为应用基础, 研究新型物联网环境下的多智能体决策信息支持技术^[10]. 文献 [11] 为了实现人-机-物的动态协同, 进一步将面向服务的体系架构 SOA 和事件驱动架构 EDA 进行结合, 以服务的形式对物联网系统进行松散建模, 基于发布订阅机制实现感知信息的按需分发和事件驱动的服务动态协同. 以上方法虽然都考虑了物联网系统的环境影响等因素, 但是大多停留在时空环境等因素上, 对设备间的影响因素及对跨域物联网系统的数据关联等因素缺乏考虑.

为了实现异构物联网系统间的语义一致性和消除数据孤岛, 本文采用语义建模的方法构建了一种基于环境-资源-行为的领域本体模型, 实现了高度受限的行为和资源的语义描述. 采用 Linked data 思想实现物联网域内和域间的数据语义关联. 设计基于多种事件触发机制的物联网推理决策方法, 实现智能决策. 本文第 1 节阐述物联网领域本体模型的构建方式; 第 2 节讲述基于行为和资源的数据关联模型的构建方法; 第 3 节设计多种事件机制及其基于事件的推理过程; 最后构建两个智能家居实例

和多个查询实例, 测试本文算法的可行性和优越性.

1 物联网领域本体模型的构建

领域本体是对特定领域中概念和概念之间关系的精确描述, 通过将不同机构或组织对领域内概念及关系的异构描述映射到相同的领域本体, 提供互相理解的语义基础. 构建物联网系统的领域本体模型是实现物联网感知数据关联和语义融合的首要任务.

1.1 继承高层本体

为了提高领域本体模型的通用性和可识别度, 快速构建领域模型, 充分利用现有本体是值得推荐的方法. 尤其是业内得到公认的现有本体模型, 已经在很多项目中得到应用, 往往在关联数据云 (Linked data cloud) 中与其他本体实现了很好的语义关联, 用其作为新建领域本体的基础, 有助于新建本体模型与其他领域本体的关联, 快速构建自己的领域本体模型.

目前物联网 (或传感器网络) 本体较为权威的有 W3C 语义传感器网络孵化群构建的 SSN 本体、CSIRO 本体、MMI 本体、CESN 本体、A3ME 本体和 OntoSensor 本体等, 以粗粒度方式描述物联网领域的基本概念, 属于高层本体范畴, 不适合直接应用在特定的物联网系统, 但为具体领域的物联网本体建模提供很好的高层概念定义. 具体的领域本体完全可以继承这些本体中的概念, 从而实现更好的通用性. 例如在 SSN 本体中定义了 14 个模块 (Module), 包括 41 个类和 39 个对象属性, 并从 DUL 顶层本体直接继承了 11 个类和 14 个属性. SSN 中定义的类和属性具有很好的通用性, 例如 skeleton 模块中定义的 Sensor、FeatureOfInterest、Observation、Stimulus 等类和 hasProperty、observedBy 等属性; MeasuringCapability 模块中定义的 Accuracy、Frequency、Latency、Measurement-Range 等类和 hasMeasurementCapability 等属性. 本文的物联网领域本体构建在 DUL 顶层本体和 SSN 等行业高层本体的基础之上, 如图 1 所示.

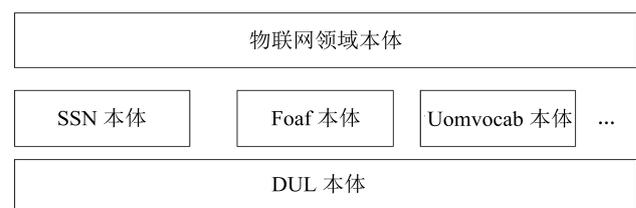


图 1 物联网领域本体的层次结构

Fig. 1 The hierarchical structure of IoT domain ontology

1.2 构建以环境—资源—行为为核心的物联网本体模型

物联网中的感知数据是在高度受限的环境中获取的, 本文从环境 (Environment)、资源 (Resource) 和行为 (Behavior) 三个方面定义语义本体模型, 简称 E-R-B 模型.

环境包括三种类型: 1) 时空环境, 包括感知信息所处的具体时间和位置, 例如采样时间、经纬度值等. 2) 自然环境, 即自然因素对感知信息的影响, 例如感知信息采集时的风速、温度、湿度、气压等自然因素. 3) 物理环境, 指感知信息所处的物理条件的影响, 例如设备之间的相互影响, 感知设备的性能因素等.

资源是物联网系统中所有具有感知能力或者提供各种智能服务的物理实体, 第一个特征是类型 (Type), 分为 Sensor 和 SmartDevice, 前者泛指各种感知设备, 可以有多个子类继承, 例如 TemperatureSensor、SpeedSensor 等. 后者泛指各种智能设备, 往往既具有感知能力又具有其他行为能力, 例如智能家居中的智能冰箱、智能打印机、智能空调等, 也对应多个子类; 第二个特征是功能, 泛指该资源具有的各种具体功能, 定义 Function 类及相关属性 hasFunction, 以动名词的形式定义具体功能, 例如一台智能洗衣机具有洗衣 (washClothes)、甩干 (dryClothes) 和定时 (fixTime) 等功能; 第三个特征是性能 (Performance), SSN 定义了 Measuring-Capability 模块, 包括精度 (Accuracy)、频率 (Frequency) 等类, 本文增加 accuracyValue、frequency-Value 等属性配合相关的类描述资源的性能.

行为 (Behavior) 是指感知设备每次获取信息的过程或者智能设备每次执行某个功能的过程, 例如温度传感器获取了一个温度值, 称为一个行为, 智能洗衣机甩干了衣服也是一次行为. 物联网中资源的行为是在高度受限的环境中产生的, 因此定义行为为一个五元组.

$$Behavior = \{Environment, Resource, Theme, Object, Result\} \tag{1}$$

式 (1) 中, Environment 是行为环境, 包括时空环境和其他环境影响因素. Resource 是行为主体, 即资源. 这里的资源不再是资源类型, 而是具体的资源实例. Theme 是行为主题, 与资源的功能值等同. Object 是行为对象, 即该行为的作用对象. Result 是行为结果, 对于传感器等资源的观测行为, 其结果是获得一个感知信息, 对于智能设备的行为结果则基于不同的功能有所不同. 环境、资源和行为构成物联网领域本体的核心, 其相互关系如图 2 所示.

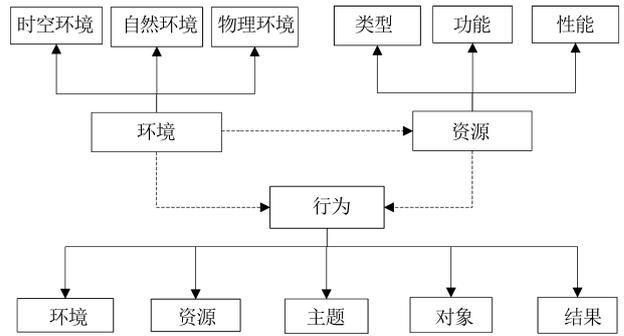


图 2 E-R-B 模型
Fig. 2 E-R-B model

2 感知数据的语义标注与关联

物联网感知设备的监测频率往往很高, 每隔几秒甚至更短的时间就会获取一次感知数据, 从而形成海量数据, 这些数据一般都是基于数据库存储. 如何为数据库存储的各种数据进行有效的语义标注以及如何构建各种孤立数据间的语义关联是实现数据融合的两个关键问题.

2.1 基于 D2R 技术的语义标注和关联

2.1.1 语义标注

当前的语义标注技术包括手工标注、半自动标注和自动标注^[12-13]等, D2R 作为一种从数据库到 RDF 三元组映射的半自动语义标注技术已经得到业内广泛关注. D2R 是一种将关系数据库数据发布为语义 Web 数据的工具, 主要包括 D2R Server、D2RQ Engine 以及 D2RQ Mapping 语言^[14]. D2R Server 是一个 HTTP Server, 主要功能是提供对 RDF 数据的查询访问接口. D2RQ Engine 的主要功能是使用一个可定制的 D2RQ Mapping 文件将关系型数据库中的数据转换成 RDF 格式. D2RQ Mapping 语言的主要功能是定义将关系型数据转换成 RDF 格式的 Mapping 规则^[15]. 为了能够将关系数据库中的各种数据进行语义标注, 本文采用自定义的 Mapping 文件方法, 将数据库的表名和字段名映射为自建领域本体中的类和属性, 实现物联网系统的语义标注, 为数据的语义关联和进一步的语义推理奠定基础.

2.1.2 语义关联

语义标注后的物联网系统并没有完全解决数据孤岛问题, 要想解决该问题, 需要基于语义标注实现相关数据的有效语义关联. 关联数据理论为不同系统间的数据语义互联提供了新的解决思路. 关联数据 (Linked data)^[16] 是一种利用 Web 在不同数据源之间创建语义关联的实践方法. 这里的不同数据源, 可以来自一个组织内部的不同系统, 也可以来自不

同组织的不同系统,它们的内容、存储地点以及存储方式都可以完全不同,但它们很可能存在着语义关联. Linked data 最大的特点是将不同的数据关联起来. Linked data 将任何事物都理解为资源,用 URIs 为资源命名,用户可以利用 HTTP URIs 查找资源,通过链接到其他的 URIs,可以发现更多的资源. 资源可以有多种描述,例如 HTML、XML、RDF 以及 JPEG. 其中 Web 数据主要通过 RDF 格式表示. RDF 将一个资源描述成一组三元组(主语、谓语、宾语).

通过将各种物联网设备感知的数据描述为 Linked data 形式,可以屏蔽不同设备、不同系统间的数据鸿沟,将各种数据孤岛有效关联,形成了统一的语义数据模型,为进一步的数据挖掘和推理奠定基础. D2R 提供了数据关联的实现方法,利用 `d2rq:refersToClassMap` 等元素实现不同数据集之间的关联关系.

2.2 基于行为和资源的数据关联模型

数据库中的感知数据经过 D2R 的 Mapping 映射后,会以 RDF 三元组的形式出现,三元组的主语是采用 URI 表示的资源¹,谓语是映射到本体的属性,宾语是具体值或者指向另外一个资源. 为了体现感知数据的高度受限性和系统的广泛互联性,以行为和资源(物联网资源)两个概念为核心构建行为关联模型和资源关联模型.

行为关联模型是以具体行为为主语构造的三元组集合,以式(1)为参照,从行为的环境、资源、对象、主题等多个角度构建语义关联. 除此之外,行为之间也可以构建语义关联关系,例如相同主题的行为、相同时间发生的行为、相同资源产生的行为等. 一个具体的行为模型如图 3 所示.

图 3 中,椭圆形表示关联数据中定义的资源,矩形表示文字或者数值信息,连线上的内容是 RDF 三元组的谓语,指领域本体中的属性. `behavior1` 是一个行为实例,以其为主语构建了 9 个 RDF 三元组,关联到其类型(`type`, 指向本体类 `Behavior`)、所处环境(`location`, 指向位置实例 `location1` 和 `measureTime`, 指向一个具体的测量时刻 `2015-11-05T17:00:06`)、行为对象(`featureOfInterest`, 指向一个房子的实例 `house3`)、主题(`theme`, 指向一个功能本体类 `MeasureTemperature`)、结果(`measureValue`, 指向一个具体数值 `12`) 和资源(`isMeasureBy`, 指向实施此次行为的资源主体,是一个传感器实例 `sensor4`). 同时,还基于不同的关系关联到其他行为,例如基于 `atSameTime` 关系关联到 `behavior9`, 基于 `bySameResource` 关系关联到 `behavior6`.

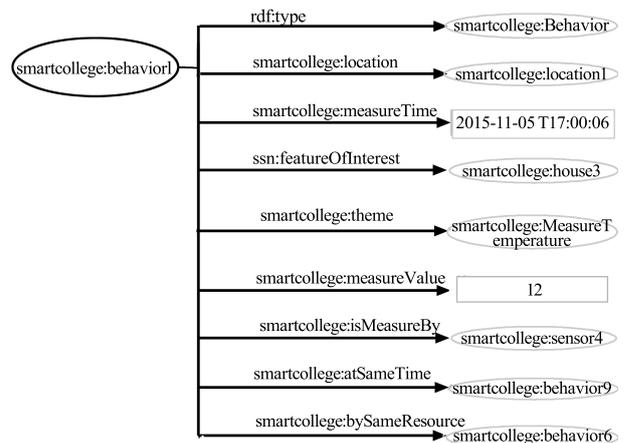


图 3 行为关联模型

Fig. 3 Behavior association model

资源关联模型是以具体物联网资源为主语构造的三元组集合. 资源关联到属性 `type`、`function` 和 `performance`, 同时,资源又是产生行为的主体,因此和行为具有关联关系,另外,资源之间由于功能相同,或者位置临近等关系也可以构造多种关联关系. 一个具体的资源关联模型如图 4 所示.

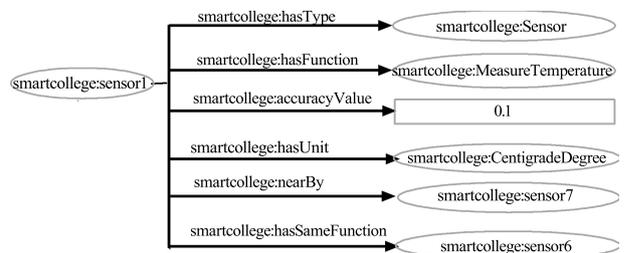


图 4 资源关联模型

Fig. 4 Resources association model

图 4 构造了 6 个 RDF 三元组,以资源实例 `sensor1` 为主语,关联到其类型(`hasType`, 指向一个本体类 `Sensor`)、功能(`hasFunction`, 指向一个功能本体类 `MeasureTemperature`)、测量精度(`accuracyValue`, 指向一个具体精度值 `0.1`)、数值单位(`hasUnit`, 指向一个本体类 `CentigradeDegree`). 此外,还与资源 `sensor7` 和 `sensor6` 基于关系 `nearby` 和 `hasSameFunction` 进行了关联.

通过构建行为关联模型,将感知数据放到一个动态的行为活动中,与行为的对象、主题、资源、环境关联到一起,大大丰富了数据的语义内涵,构建了数据的动态关联关系(例如数据在同一时刻产生等). 同时,将产生数据的资源通过各种属性和关联关系进行关联,进一步延伸了数据关联的背景,丰富了数据的语义内涵(例如通过关联资源属性,可知感知数

¹例如: <http://smartcollege.duapp.com/resource/city/Beijing>

据的监测精度等). 从而可以构造跨系统的语义关联模型, 屏蔽数据孤岛现象, 为推理决策奠定基础.

3 事件触发的推理机制

在将数据库中的数据构建了语义标注和数据关联之后, 已经形成了一个统一的语义数据模型, 如何充分利用该模型为用户决策服务是亟待解决的问题. 本文构建一种基于事件触发的推理决策模型. 事件是推理的前提, 也为推理提供信息资源, 一切推理皆围绕事件展开.

3.1 事件模型的定义

定义一个事件模型 Event 为一个 4 元组.

$$Event = \{Environment, Content, Result, Expect\} \quad (2)$$

其中, *Environment* 是事件所处的环境, 包括时间、地点和其他环境影响因素; *Content* 是事件的内容, 可以用三元组的形式表示为 $\{subject, predicate, object\}$, *subject* 是事件的主语, *predicate* 是事件的谓词, 表示行为动作, *object* 是事件的触发对象; *Result* 是事件带来的结果或影响; *Expect* 是事件的期望. 事件 (Event) 和行为 (Behavior) 有很多相似的特性, 但是事件不等同于行为, 事件的概念更加宽泛, 任何作用于物联网系统的内部或者外部影响都可以理解为事件, 行为是事件的子集.

物联网中的事件分为两种情况: 既定事件和非既定事件. 对于既定事件, 系统已经构建了相应的业务流程, 只要既定事件一发生, 就会触动预定义好的业务流程进行推理和决策. 对于非既定事件, 系统并未设计相应的业务流程, 只能以事件本身为事实因素查找可以调用的关联数据和推理规则. 非既定事件分为两种, 有期望的非既定事件和无期望的非既定事件. 对于有期望的非既定事件, 可以将期望内容作为查询内容, 进行反向推理, 查看是否能够推理出预期结果; 对于无期望的非既定事件, 只能进行前向推理, 希望挖掘出更多有意义的因素辅助决策.

3.2 基于事件的推理模型

物联网系统中的事件必然导致系统状态的变化, 例如一个关闭继电器开关的事件会导致灯光的熄灭, 但是这些都属于直观的状态变化, 由于物联网系统中各个不同组件的相互关联性, 一个事件的发生往往会导致一些隐式的状态变化, 产生蝴蝶效应, 但是这些隐式的状态变化需要构建推理机制才能发现. 当前的本体推理机有 racer、FaCT、pellet 和 jena 等, 本文采用 jena 作为推理引擎, jena 是美国 HP 实验室开发的一个开源 java 框架, 用来构建语

义 Web 和 Linked data 应用, 可以实现 RDF 数据的读取和写入, RDFS、OWL 等本体的操作, 基于 SPARQL 的数据查询和基于规则的推理等工作^[17]. 为了实现基于事件的推理, 本文从四种情况入手来设计推理算法.

3.2.1 既定事件推理

算法如下.

算法 1. 既定事件推理

步骤 1. 设定既定事件

$Event =$

$$\{Environment, Content, Result, Expect\}$$

步骤 2. 构建推理规则集

$$Rules = \{rule_1, rule_2, \dots, rule_n\}$$

步骤 3. 设定业务流程 BusinessProcess, 业务流程由一系列活动及其相互关系组成, 本文将业务流程描述为一个活动集合和一个关系集合的形式:

$$BusinessProcess = \{Activity, Relation\}$$

其中,

$$Activity = \{activity_1, activity_2, \dots, activity_m\}$$

$$Relation = \{relation_1, relation_2, \dots, relation_n\}$$

$activity_i$ ($1 \leq i \leq m$) 表示一个具体的活动, $relation_j$ ($1 \leq j \leq n$) 是建立在活动之间的关系, 包括顺序、并行、选择、随机等.

步骤 4. 当既定事件发生时, 整个事件以 RDF 三元组的形式进行表示, 并存入 jena 的模型 (Model) 中.

步骤 5. 触发业务流程, 将事件本身作为事实信息进行前向推理, 采用 RETE 算法^[18] 获取各种预定义的结果信息. 具体步骤可参考算法 3.

步骤 6. 如果既定事件中存在期望 Expect, 则从 Expect 出发进行反向链推理. 具体步骤可参考算法 2.

3.2.2 有期望的非既定事件推理

事件中的期望指事件希望得到的某种预期效果, 例如发生一个事件“小张约小王下午 2:00 去 A 电影院看电影”, 在这个事件中, “下午 2:00”和“A 电影院”是 *Environment*, “小张约小王看电影”是 *Content*, 其中“小张”是 *subject*, “小王”是 *object*, “约会看电影”是 *predicate*. 这个事件其实还可以添加一个期望 (Expect), 即“小王下午 2:00 空闲”. 对于事件期望信息, 采用 RDF 三元组形式表示, 并作为查询内容调用 jena 的反向推理机制, 推理算法如下.

算法 2. 有期望的非既定事件推理

步骤 1. 将事件期望 Expect 转化为 RDF 三元组表示, 并构建查询字符串 queryString;

步骤 2. 将本体文件存入 jena 模型 ontModel;

步骤 3. 将事实信息 RDF 文件存入 jena 模型 rdfModel;

步骤 4. 构建推理引擎 infModel, 装入推理规则集 rules、本体模型 ontModel 和事实信息集 rdf-Model;

步骤 5. 基于推理引擎和查询字符串实现 LP 算法反向链推理.

步骤 6. 得到验证结果, 推导出期望的真实性.

3.2.3 无期望的非既定事件推理

对于无期望的非既定事件, 只能调用前向推理规则进行推理, 推理算法如下.

算法 3. 无期望的非既定事件推理

步骤 1. 基于事件本身和相关事实构建 RDF 描述的事实库, 并存入 jena 模型;

步骤 2. 将领域本体存入 jena 本体模型;

步骤 3. 创建 jena 的推理引擎, 装入事实库、本体模型和规则集.

步骤 4. 采用 RETE 算法进行推理, 生成新的事实信息, 存入事实库, 重复执行该步骤直到不再产生新的推理结果, 推理结束;

3.2.4 多事件联合推理

除了上述由单独事件导致的推理过程, 还存在多个事件联合构造推理过程的情况. 例如上文的“小张约小王下午 2:00 去 A 电影院看电影”事件, 期望是“小王下午 2:00 空闲”, 但是在事实库中存在一条事实是“小王下午 2:00 有算法课”, 规则集中存在一条规则: $[rule: (?a \text{ smartcollege: hasNoLesson } ?b) \rightarrow (?a \text{ smartcollege: hasFreeTime } ?b)]$, a 指学生, b 是具体的时间, 即如果学生 a 在时间 b 没有课, 则 a 在时间 b 有空闲时间. 显然, 小王是学生, 基于反向链推理从 $(?a \text{ smartcollege: hasFreeTime } ?b)$ 出发, 会要求证实 $(?a \text{ smartcollege: hasNoCourse } ?b)$ 的存在, 而事实库中不存在该事实, 所以, 仅从该单一事件出发, 此期望不能实现. 但是再加入一个事件“心脏监测器发现宋老师心脏出现异常”, 该事件驱动前向推理, 得到一系列事实: “宋老师需要卧床休息”, “宋老师取消了下午的算法课”等. 因此事实库中更新了内容, 加入了“算法课取消”的事实, 经过再次推理, 可知之前的期望变成了现实. 由以上分析可见, 对于多事件联合推理情况, 可以综合运用之前的三种推理算法实现.

4 系统实现和评估

为了验证本文方法的可行性, 模拟构建一个多领域协作的智能家居系统, 包括智能供暖、智能交通、智能医疗等多个独立系统, 并以智能家居为纽带关联起来. 通过本文构建的领域本体、数据关联和推理理论, 屏蔽了数据孤岛, 实现了智能决策.

4.1 智能温控系统案例分析

本文首先构建了一个多领域协作的智能家居供暖温控系统. 该系统结合智能供暖和智能交通两大系统, 设定当主人在家时温度保持在 22°C , 当主人出门工作时供暖系统低温运行, 当主人下班返回家时, 温度要及时提升到人体适宜的 22°C . 在这个过程中, 室内的供暖系统配有水温传感器、水压传感器、室内温度传感器, 用于控制供暖流量的自动电磁阀等. 在智能交通系统中, 主人从公司到家的路上要经过 3 个测速传感器, 测速传感器可以获取实时的车流速度. 假设主人每天 17:00 准时下班, 从公司到家有 60 km 的车程, 希望综合各种因素实现智能温度调节, 确保主人回家前室温刚好提升到 22°C . 为了实现以上目标, 综合本文所提的方法开展如下工作.

1) 构建智能家居领域本体模型. 本文采用 protege 5.0 开发智能家居领域本体, 通过 Imported ontologies 模块输入高层本体 SSN² 和其他相关高层本体. 新建智能家居系统相关的多个资源本体类 (Class)、对象属性 (ObjectProperty) 和数据 (DataProperty) 属性. 该本体已经发布到 Webprotege 项目³, 相关的智能家居系统部分本体类和属性关系如图 5 所示.

2) 构建智能供暖系统感知数据的语义标注和数据关联. 智能供暖系统的数据库采用 Mysql 5.5 创建, 并在此基础上构建传感器信息表、水压数据表、水温数据表、室内温度数据表、室外温度数据表、测速信息表等多个数据表. 然后采用 D2R-Server-0.7 处理数据库数据. 构建 Mapping 文件实现语义标注和基于行为、资源的数据关联, 映射成虚拟的 RDF 三元组形式. 语义标注且关联后的模型如图 6 所示.

从图 6 可以看出, 原本毫无关联的室内温度、供暖水压、水温、车速等感知数据由于包含在行为模型中动态考虑, 出现了多条关联链条, 将其相互联系起来. 例如水温和水压由于来源于相同的测量对象而联系起来, 而三个测速感知数据由于位于同一条路线上而关联起来. 从而构成了一个关联数据网络, 并且将以 RDF 三元组事实库的形式出现在推理过程中.

²URI 为 <http://www.w3.org/2005/Incubator/ssn/ssnx/ssn>

³URI 为 <http://webprotege.stanford.edu/#Edit:projectId=6d7b0a34-a7c3-4fe8-95ed-2f464b3dad58>

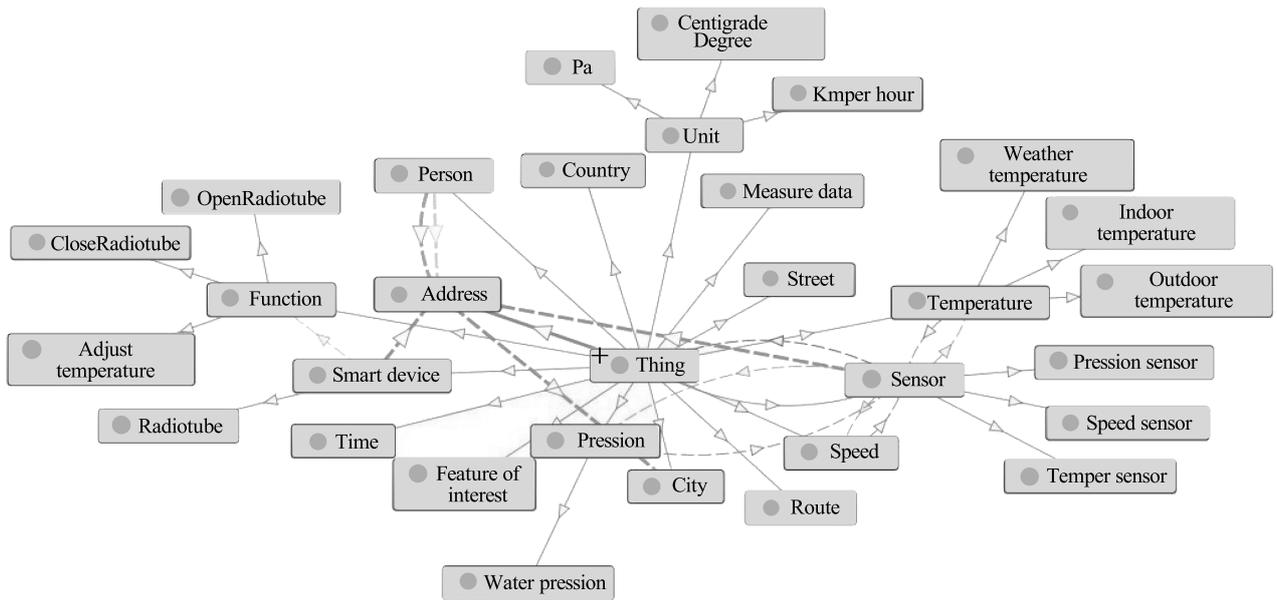


图 5 部分智能家居本体图
 Fig. 5 Part of the smart home ontology figure

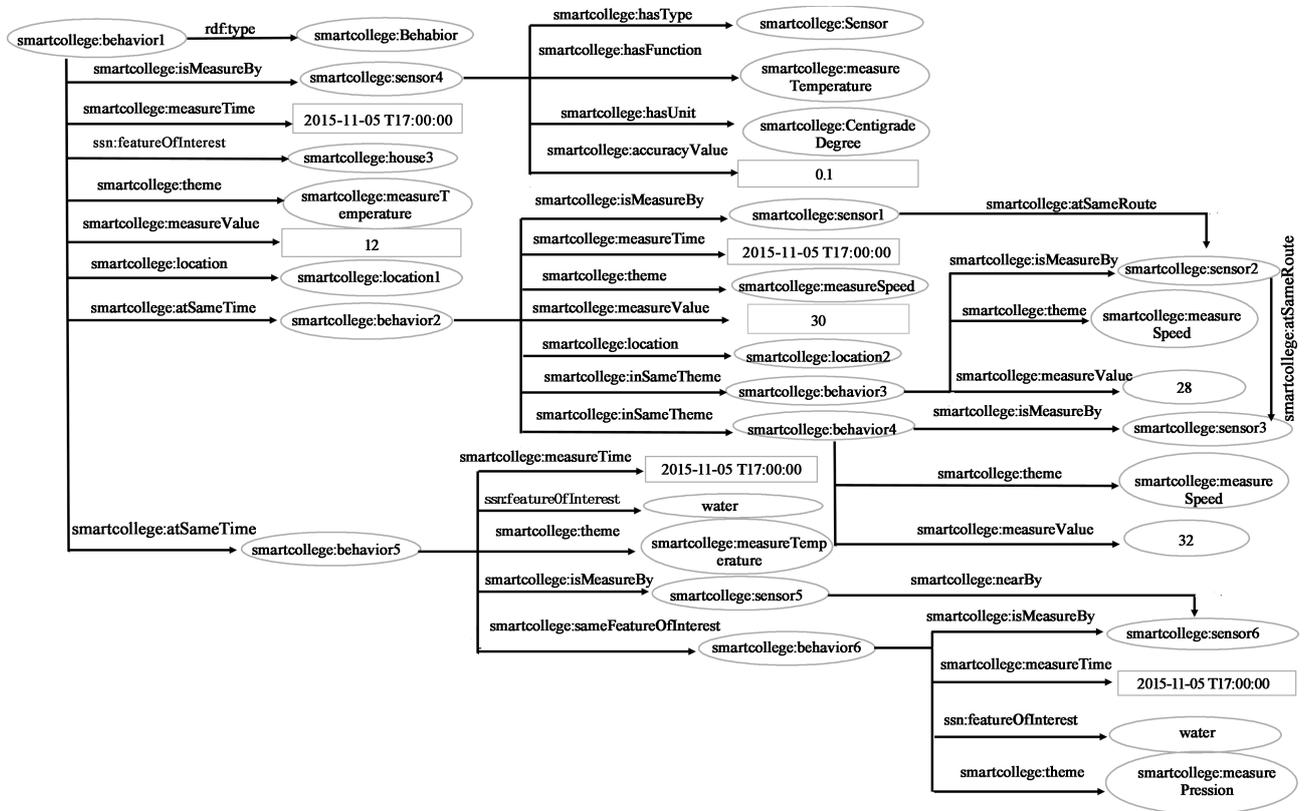


图 6 智能家居数据关联模型
 Fig. 6 Data association model of smart home

3) 设计既定事件的业务流程. 案例涉及的事件是既定事件“时间到达 17:00”, 即每天系统时间一到 17:00 就触发该事件推理. 为了实现以上目标, 设计顺序的业务流程如下:

步骤 1. 预测主人到家时间 *arriveTime*.

$$\begin{aligned} arriveTime &= 17:00 + \frac{Distance}{aveSpeed}, \\ aveSpeed &= \frac{\sum_{i=1}^n Speed_i}{n} \end{aligned} \quad (3)$$

其中, *Distance* 是从公司到家的路程距离, *aveSpeed* 是平均车速, 为多个测速传感器所测速度 *Speed_i* 平均值.

步骤 2. 模拟计算当前环境下的升温速度 *incrRate*.

$$incrRate = R \times \frac{wTemp \times wPression}{inTemp - outTemp} \quad (4)$$

其中, *wTemp* 是水温, *wPression* 是水压, *inTemp* 是室内温度, *outTemp* 是室外温度, *R* 是指定系数.

步骤 3. 计算开始升温时间 *startTime*.

$$startTime = arriveTime - \frac{22 - inTemp}{incrRate} \quad (5)$$

4) 智能供暖系统的推理过程. 本系统安装 jena 2.13 版本, 结合 Eclipse 开发环境开发智能供暖推理系统. 具体流程如下:

步骤 1. 创建推理模型.

1) 创建 RDF 模型, 将生成的 Mapping 文件 *smartcollege.n3* 装入模型.

```
ModelID2RQ m = new ModelID2RQ ("file:/D:/smartcollege.n3");
```

2) 构建本体模型, 将创建的领域本体 *smartcollege-ontology.owl* 装入模型.

```
OntModel o = ModelFactory.createOntologyModel();
```

```
File myFile = new File ("data/smartcollege-ontology.owl");
```

```
o.read(new FileInputStream (myFile), "");
```

3) 创建 OWL 推理模型, 绑定 RDF 模型 *m* 和本体模型 *o*.

```
Reasoner reasoner = ReasonerRegistry.getOWLReasoner();
```

```
Reasoner screasoner = reasoner.bindSchema(o);
```

```
InfModel infmodel = ModelFactory.createInfModel(screasoner,m);
```

步骤 2. 建立推理规则集. 为了能够顺利推理出

供暖系统的升温时间, 调用推理规则集中的相关规则, 举例如下.

规则 1: 如果位于地址 *y* 的速度传感器 *s* 测得该时刻的速度值为 *m*, *p* 在该时刻经过 *y*, 则推理出 *p* 的速度为 *m*.

```
[r1: (?x1 rdf: type smartcollege: Behavior),
(?x1 smartcollege: location ?y), (?x1 smartcollege: isMeasureBy ?s), (?x1 smartcollege: theme smartcollege: MeasureSpeed), (?x1 smartcollege: featureOfInterest smartcollege: Car), (?x1 smartcollege: measureValue ?m), (?x2 rdf: type smartcollege: Event), (?x2 smartcollege: subject ?p), (?p rdf: type smartcollege: Car), (?p smartcollege: pass ?y), (?x1 smartcollege: atSameTime ?x2) → (?p smartcollege: speedValue ?m)]
```

规则 2: 如果地址 *d* 位于城市 *c* 且城市 *c* 在某时刻天气预报温度为 *z*, 则推理出 *d* 的室外温度为 *z*.

```
[r2: (?d rdf: type smartcollege: Address),
(?d smartcollege: city ?c), (?x rdf: type smartcollege: Behavior), (?x smartcollege: location ?c), (?x smartcollege: measureTime ?t), (?x smartcollege: theme smartcollege: ForecastWeather), (?x smartcollege: temperatureValue ?z) → (?d smartcollege: outdoorTemperatureValue ?z)]
```

步骤 3. 基于推理规则构建 *GenericRuleReasoner* 推理机.

```
Reasoner reasoner = new GenericRuleReasoner(Rule.parseRules(rule));
```

```
InfModel infmodel = ModelFactory.createInfModel(reasoner, m);
```

步骤 4. 执行推理. 获取主人的平均车速、供暖系统的加温速度和加温时刻等, 推理结果如图 7 所示.



```

<terminated> smartdorm [Java Application] D:\MyEclipse\Common\binary\com.sun
aveSpeed 30
incrRate 1.2000000000000002
arriveTime 19:0:48
startTime 18:37
System start countdown 1:37:00

```

图 7 Jena 推理结果

Fig. 7 Jena reasoning result

图 7 中, 计算出平均车速 *aveSpeed* 为 30 km/h. 由于路程全长 60 km, 计算出主人到家时间 *arriveTime* 为 19:00, 基于室内外温差和供暖管道水温水压计算得到的升温速率 *incrRate* 是 1.2°C/minute, 推理出升温时间 *startTime* 为 18:37.

4.2 多事件联合推理案例分析

上面的案例建立在单事件触发的数据关联模型

上,为了证明本文方法的完备性,以上节提到的“约会看电影”案例为例,构造一个多事件联合推理模型.在案例中,“小张约小王下午 2:00 去 A 电影院看电影”是一个非既定事件,该事件存在一个期望“小王下午 2:00 空闲”.另外,“心率传感器于当天 12:06 监测到宋老师的心率为 110”触发一个事件,“血压传感器于当天 12:06 监测到宋老师血压为 160/100”又是一个事件,这两个事件为非既定事件,且没有期望信息.

1) 基于事件期望的反向推理.对于第一个事件,由于存在期望“小王下午 2:00 空闲”,将其设置为查询语句($?a$ smartcollege: hasFreeTime $?b$) 在事实库中查询,由于事实库中无对应的事实,在规则库中发现一条规则如下:

规则 3: 学生 b 在某一时间 x 没有课,则推理出 b 在时间 x 空闲.

[$r3: (?b$ rdf: type smartcollege: Student), ($?x$ rdf: type smartcollege: Time), ($?b$ smartcollege: hasNoCourse $?x$) \rightarrow ($?b$ smartcollege: hasFreeTime $?x$)]

因此触发反向链推理,查询事实库中是否存在“小王在下午 2:00 没有课”这个事实.结果该事实也不存在,事实库中只有“小王下午 2:00 有算法课”的事实.因此可以推理出期望无法实现.

2) 基于事件的正向推理.由于另外存在心率和血压监测事件的发生,触发了云端的医学诊断系统,得到诊断结果是“宋老师生病了,健康状态为 3 级,建议卧床休息”(假设健康状态分为 5 级,其中第 5 级为正常).从而生成事实:(song smartcollege: healthStatus 3) 和 (song smartcollege: medicalAdvice smartcollege: BedRest).继而触发了一条推理规则如下:

规则 4: 如果课程 y 由教师 a 讲授并且 a 的当前医嘱为卧床休息,则推理出 a 将取消课程 y .

[$r4: (?a$ rdf: type smartcollege: Teacher), ($?y$ rdf: type smartcollege: Course), ($?b$ rdf: type smartcollege: Student), ($?x1$ rdf: type smartcollege: Event), ($?x1$ smartcollege: subject $?b$), ($?b$ smartcollege: hasCourse $?y$), ($?y$ smartcollege: isToughtBy $?a$), ($?x2$ rdf: type smartcollege: Event), ($?x2$ smartcollege: subject $?a$), ($?a$ smartcollege: medicalAdvice smartcollege: BedRest), ($?x1$ smartcollege: hasSameDay $?x2$) \rightarrow ($?a$ smartcollege: cancelCourse $?y$)]

由于事实库中出现了新的事实“宋老师取消了算法课”,触动了另一条推理规则:

规则 5: 学生 b 有一门课程 y ,课程 y 的上课时间是 t ,教师 a 取消了课程 y ,则推理出学生 b 在时

间 t 没有课.

[$r5: (?b$ rdf: type smartcollege: Student), ($?a$ rdf: type smartcollege: Teacher), ($?y$ rdf: type smartcollege: Course), ($?b$ smartcollege: hasCourse $?y$), ($?y$ smartcollege: CourseTime $?t$), ($?a$ smartcollege: cancelCourse $?y$) \rightarrow ($?b$ smartcollege: hasNoCourse $?t$)]

该规则生成了新的事实“小王下午 2:00 没课”.从而触发推理规则 3,推出小王下午 2:00 空闲.事件的期望实现了.

以上案例涉及的领域本体同样存在于之前构建的智能家居本体中,涉及的数据关联模型与案例 1 类似.通过以上多事件联合推理过程,验证了本文方法的完整性和通用性.

4.3 实验对比分析

为了测试本文方法的优越性,在以上实验的基础上,以智能家居为实验背景,引入智能供水、智能电网、智能医疗等多个实例,测试数据基于行业特征随机产生,存储在多张数据表中,存储数据量达 50 万条,然后基于语义标注实现数据关联,设计了 100 条推理规则,用来实现基于事件的推理决策.由于推理问题最终可以以查询的形式表达,例如上面的推理升温时刻的例子可以转化为查询语句:“select $?arriveTime$, $?incrRate$, $heatingTime$, $?startTime$ where $?startTime + ?heatingTime = ?arriveTime$ ” ($heatingTime$ 是加热所需时间).本文设计了 50 个查询实例,分别采用本文方法与缺乏数据关联或事件推理的方法进行查询推理,并设计了查得率和正确率两个性能指标,查得率是能够得出查询结果的实例数量和总查询量的比率,正确率是查询得到正确结果的实例数量和总查询量的比率.表 1 从查得率和正确率两个方面对不同方法的性能进行了比较.

表 1 不同方法的查得率和正确率比较表
Table 1 The comparison table of detection rate and accuracy

方法编号	语义标注和数据关联	事件推理	查得率 (%)	正确率 (%)
1	无	无	36	24
2	有	无	72	60
3	有	有	98	98

由表 1 可知,对于没有设置数据关联和事件推理的查询方法,只能基于数据表和表间的外键关联等少量关联信息进行查询,50 个实例有 18 个得到查询结果(查得率 36%),其中 12 个实例的查询结

果正确 (正确率 24%)。对于增加了语义标注和数据关联的查询方法, 查得率和正确率得到了较大提高。本文所提的带有数据关联和事件推理方法的查得率达到 98% (只有一个查询实例没有查到结果), 正确率也达到 98%。可见本文方法对于屏蔽数据孤岛、提高推理决策效果的优越性。

5 结论

本文旨在解决物联网系统异构及其产生的大量感知数据造成的数据孤岛问题, 试图屏蔽数据孤岛, 实现跨系统的数据关联, 实现智能决策。本文从语义 Web 入手, 通过构建领域本体模型实现物联网感知数据的语义标注, 采用 Linked data 思想实现数据间的语义关联, 采用基于事件的推理机制实现语义推理和决策, 最后通过智能家居领域的两个案例, 实现了语义标注, 数据关联和基于 jena 的推理决策, 验证了本文方法的可行性; 通过设计多个测试实例, 从查得率和正确率两个方面验证了本文方法的优越性。

致谢

感谢韩露和陈依冰在论文实验中作出的贡献。

References

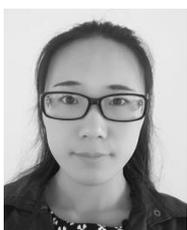
- Berners-Lee T, Hendler J, Lassila O. The semantic web. *Scientific American*, 2001, **284**(5): 34–43
- Nebot V, Berlanga R. Finding association rules in semantic web data. *Knowledge-Based Systems*, 2012, **25**(1): 51–62
- Sheth A, Henson C, Sahoo S S. Semantic sensor web. *IEEE Internet Computing*, 2008, **12**(4): 78–83
- Coccoli M, Torre I. Interacting with annotated objects in a Semantic Web of Things application. *Journal of Visual Languages and Computing*, 2014, **25**(6): 1012–1020
- De S, Barnaghi P, Bauer M, Meissner S. Service modelling for the Internet of Things. In: Proceedings of the 2011 Federated Conference on Computer Science and Information Systems (FedCSIS). Szczecin, Poland: IEEE, 2011. 949–955
- Grau B C, Horrocks I, Motik B, Parsia B, Patel-Schneider P, Sattler U. OWL 2: the next step for OWL. *Web Semantics: Science, Services, and Agents on the World Wide Web*, 2008, **6**(4): 309–322
- OWL-S: semantic markup for web services [Online], available: <http://www.w3.org/Submission/OWL-S>, November 22, 2004
- Compton M, Barnaghi P, Bermudez L, García-Castro R, Corcho O, Cox S, Graybeal J, Hauswirth M, Henson C, Herzog A, Huang V, Janowicz K, Kelsey W D, Le Phuoc D, Lefort L, Leggieri M, Neuhaus H, Nikolov A, Page K, Passant A, Sheth A, Taylor K. The SSN ontology of the W3C semantic sensor network incubator group. *Web Semantics: Science, Services, and Agents on the World Wide Web*, 2012, **17**: 25–32
- Li Ge, Wei Qiang, Li Li-Xing, Jin Zhi, Xu Yan, Zheng Li-Wei. Environment based modeling approach for services in the Internet of Things. *Science China: Information Science*, 2013, **43**(10): 1198–1218
(李戈, 魏强, 李力行, 金芝, 许焱, 郑丽伟. 物联网服务建模: 一种基于环境建模的方法. 中国科学: 信息科学, 2013, **43**(10): 1198–1218)
- Xu Yang, Wang Xiao-Feng, He Qing-Yi. Internet of Things based information support system for multi-agent decision. *Journal of Software*, 2014, **25**(10): 2325–2345
(徐杨, 王晓峰, 何清漪. 物联网环境下多智能体决策信息支持技术. 软件学报, 2014, **25**(10): 2325–2345)
- Qiao Xiu-Quan, Zhang Yang, Wu Bu-Dan, Cheng Bo, Zhao Shuai, Ma Hua-Dong, Chen Jun-Liang. An EDSOA-based services provisioning approach for Internet of Things. *Science China: Information Science*, 2013, **43**(10): 1219–1243
(乔秀全, 章洋, 吴步丹, 程渤, 赵帅, 马华东, 陈俊亮. 事件驱动、面向服务的物联网服务提供方法. 中国科学: 信息科学, 2013, **43**(10): 1219–1243)
- Cuzzola J, Jovanović J, Bagheria E, Gašević D. Evolutionary fine-tuning of automated semantic annotation systems. *Expert Systems with Applications*, 2015, **42**(20): 6864–6877
- Tosi D, Morasca S. Supporting the semi-automatic semantic annotation of web services: a systematic literature review. *Information and Software Technology*, 2015, **61**: 16–32
- The D2R Project [Online], available: <http://d2rq.org/>, March 12, 2012
- Zhang Jing, Ma Chun-E. How to use D2R to publish linked data [Online], available: http://www.ibm.com/developerworks/cn/web/1003_zhangjing_d2r/index.html, April 25, 2010
(张静, 马春娥. 如何利用 D2R 发布 Linked data [Online], available: http://www.ibm.com/developerworks/cn/web/1003_zhangjing_d2r/index.html, 2010 年 4 月 25 日)
- The Linked Data Project [Online], available: <http://linkeddata.org/>, April 16, 2015
- The Jena Project [Online], available: <http://jena.apache.org/>, March 23, 2015
- Gu Xiao-Dong, Gao Yang. Rete algorithm: current issues and future challenge. *Computer Science*, 2012, **39**(11): 8–12, 33
(顾小东, 高阳. Rete 算法: 研究现状与挑战. 计算机科学, 2012, **39**(11): 8–12, 33)



冯建周 燕山大学信息科学与工程学院副教授. 2013 年获燕山大学计算机应用技术专业博士学位. 主要研究方向为语义 Web, 服务计算. 本文通信作者.

E-mail: fjzwxh@ysu.edu.cn

(FENG Jian-Zhou Associate professor at the School of Information Science and Engineering, Yanshan University. He received his Ph.D. degree from computer application technology at Yanshan University in 2013. His research interest covers semantic web and service computing. Corresponding author of this paper.)



宋沙沙 燕山大学信息科学与工程学院硕士研究生. 2015 年获得沈阳化工大学计算机科学与技术学院学士学位. 主要研究方向为语义 Web.

E-mail: song@stumail.ysu.edu.cn

(SONG Sha-Sha Master student at

the School of Information Science and Engineering, Yanshan University. She received her bachelor degree from the School of Computer Science and Technology, Shenyang University of Chemical Technology in 2015. Her main research interest is semantic web.)



孔令富 燕山大学信息科学与工程学院教授. 1995 年获得哈尔滨工业大学机电电子工程专业博士学位. 主要研究方向为智能信息处理, 物联网工程.

E-mail: lfkong@ysu.edu.cn

(KONG Ling-Fu Professor at the School of Information Science and Engineering, Yanshan University. He received his Ph.D. degree in machinery and electronics engineering at Harbin Institute of Technology in 1995. His research interest covers intelligent information processing and the Internet of Things engineering.)