

本体摘要技术综述

丁悦航, 于洪涛, 黄瑞阳, 李英乐

(国家数字交换系统工程技术研究中心, 河南 郑州 450002)

摘要: 本体是知识共享的重要工具, 也是知识图谱的上层结构。随着数据规模和本体复杂性的增长, 本体理解与应用日益困难。本体摘要作为缩小本体规模的一项技术, 为加速本体理解与应用提供了技术支持。给出了本体摘要的不同定义; 对本体摘要方法进行比较分析; 介绍了本体摘要评估的指标体系; 最后提出本体摘要的进一步研究方向。

关键词: 知识图谱; 本体; 本体摘要; 本体评估

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2018081

Ontology summarization technology survey

DING Yuehang, YU Hongtao, HUANG Ruiyang, LI Yingle

National Digital Switching System Engineering & Technological R & D Center, Zhengzhou 450002, China

Abstract: Ontology is an important tool for knowledge sharing, while it is also the upper structure of knowledge graph. With the explosion of data and ontology complexity, ontology understanding and application are becoming more and more difficult. As a technique to shrink ontology scale, ontology summarization accelerated ontology understanding and application technologically. Different definitions of ontology summarization were shown, ontology summarization methods were compared and analyzed, ontology summarization evaluation parameter system was introduced, and at last, possible future research area was given.

Key words: knowledge graph, ontology, ontology summarization, ontology evaluation

1 引言

近年来, 知识图谱作为可推理的结构化数据集, 已经越来越多地应用于智能语义搜索、移动个人助理和深度问答系统^[1]。本体作为知识图谱的上层结构, 在知识推理和数据层的构建中发挥着重要作用。随着数据规模和本体复杂性的飞速增长, 本体的理解 and 应用难度日益增大。本体摘要作为生成原本体缩略版本的一种方式, 能够有效降低本体理解的难度, 加快本体应用的速度,

从而有效缓解上述问题。

本体摘要的思想源于文本摘要。文本摘要即对文本中的句子, 制定重要性评估方法, 从而选出最重要的若干个句子作为摘要结果。本体摘要由 Zhang 等^[2]于 2007 年提出, 其通过对 RDF (resource defined framework) 句子进行重要性评估, 选出若干重要的 RDF 句子作为摘要结果。随后, 本体摘要根据目的衍生出用户导向和任务导向的本体摘要。用户导向的本体摘要是让人更方便地理解本体的内涵, 从而便于进一步重用本体,

收稿日期: 2018-08-27; 修回日期: 2018-09-29

通信作者: 丁悦航, 739815262@qq.com

项目基金: 国家自然科学基金资助项目 (No.61521003)

Foundation Item: The National Natural Science Foundation of China (No.61521003)

因此，基于此目的的方法大多抽取本体中的重要概念作为摘要结果。任务导向的本体摘要是利用摘要本体完成任务，加快任务的完成速度，因此，基于此目的的方法大多在不影响应用的前提下删除原本体中冗余的信息。

在本体领域中，与本体摘要功能相似的 2 个技术是本体分割和本体模块化。本体分割^[3]将一个大本体分割成若干个子本体，每个子本体包含原始本体的一个子主题^[4]。本体模块化^[5]用于简化本体重用：将原始本体中需要重用的部分模块化，从而重用本体的一部分而非整个本体^[6]。这 2 种技术都能缩小本体规模，促进本体理解，但二者均基于划分技术，没有针对性地保留重要信息。本体摘要与本体领域其他概念的关系如图 1 所示。

2 本体摘要的定义

本体摘要可以在保留本体语义的前提下有效解决本体体积过大导致的查询效率低、内涵难以理解等问题。近年来，国内外已有多位学者对本体摘要进行了研究，然而不同学者对本体摘要的理解不尽相同，给出的定义也存在差异。

Zhang 等^[2]参照文本自动摘要的定义^[7]，首先提出本体摘要的概念：本体摘要是针对某个特殊用户或任务，从本体中提取知识，构建原本体缩略版本的过程。Zhang 是最早提出本体摘要这一概念的学者，指出本体摘要分为用户驱动和任务驱动 2 种形式。本体摘要概念的提出对其后各领域学者对该问题的研究提供了参考借鉴。

Li 等^[8]认为，相较于本体的其他抽取技术，本体摘要应该保留整个本体中的重要信息，并且是自动生成的。他们参照文本摘要的定义^[9]给出了本体摘要的定义：本体摘要是自动生成给定本体缩略版本的过程，用于为用户提供重要信息。可见，Li 等的定义更偏重于用户导向的本体摘要。

Deng 等^[10]认为本体摘要是将原始本体映射到规模更小本体的过程。他们将本体摘要分为无损摘要和有损摘要 2 种。摘要本体必须满足以下 2 个条件：1) 摘要本体的所有信息都由原本体映射而来；2) 摘要本体的规模必须小于原本体。设原本体为 O ，若通过映射 f 生成的摘要本体 O^c 满足 $\forall \alpha(O \models \alpha \Rightarrow O^c \models f(\alpha))$ ，则 O^c 为无损摘要。生成无损摘要的目的是推理，因此要在尽可能不损失信息的前提下降低推理复杂度。

3 本体摘要方法

本体摘要方法按照目的分为用户驱动和任务驱动 2 类。用户驱动的本体摘要致力于满足用户快速理解大规模本体空间的需要，因此倾向于提取本体中的重要概念和关系。任务驱动的本体摘要致力于在尽可能不丢失信息的前提下把本体缩小到满足问答系统任务的必要大小，因此倾向于删减本体中的不必要信息。

图 2 展示的是一个示例本体。其中，上半部分用实线绘制，表示模式层，即概念及其之间的关系，对应描述逻辑中的 TBox；下半部分用虚线绘制，表示数据层，即个体及其之间的关系，对应描述逻辑中的 ABox。TBox 即关于描述领域的概念术语的断言，用于定义概念、概念间的关系、

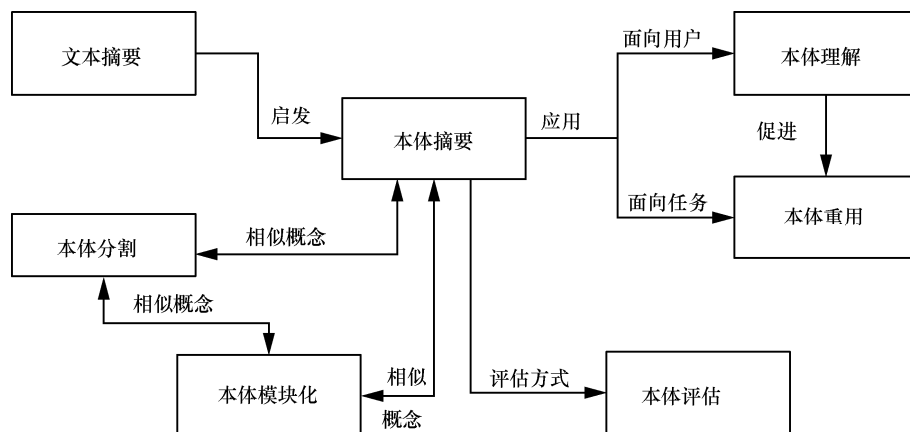


图 1 本体摘要与各个概念的关系

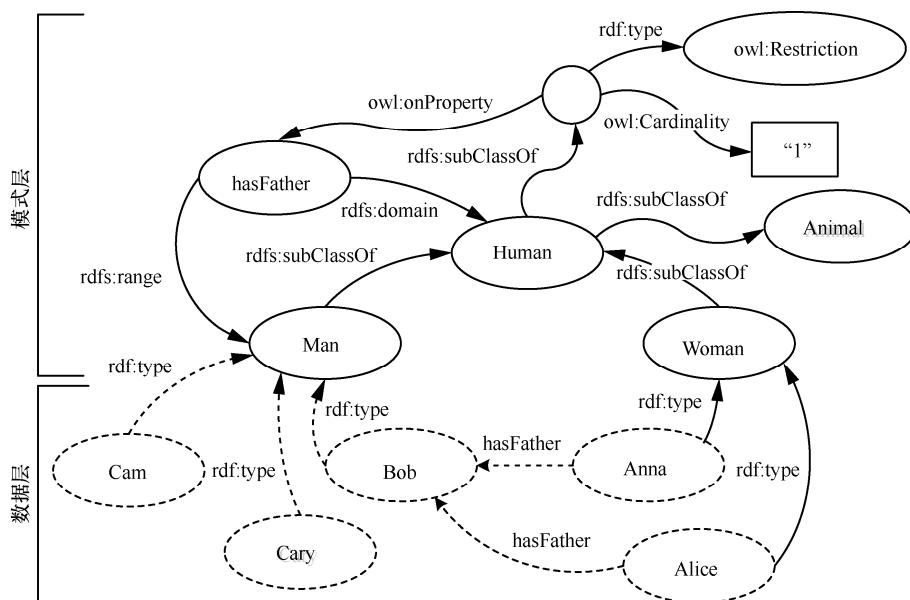


图 2 示例本体

角色间的关系；ABox 即包含应用领域外延知识的知识库，包括实体的概念断言和关系断言^[11]。接下来，以示例本体为例，说明不同本体摘要方法的具体过程。

3.1 面向任务的摘要方法

Alani 等^[12]通过删除问答系统中没有用到的本体信息来加快问答速度。具体地，只保留具有对应实例的类和属性，问答对中涉及的类和属性，以及已有属性的定义域和值域所在的类。通过将摘要后的本体用于问答系统验证了摘要本体的有效性。然而，这种本体摘要方式仅针对特定的应用，无法保证摘要本体的语义完整性和一致性。

图 3 是示例本体经过 Alani 等算法处理后的摘

要，算法针对模式层进行摘要。数据层中的 Allen 和 Bob 属于模式层中的概念 Man，因此保留模式层中的 Man。同理，数据层中的 Anna 和 Bella 属于模式层中的概念 Woman，因此也保留概念 Woman。此外，数据层中属性 hasFather 的定义域和值域分别为 Human 和 Man，因此保留 hasFather 属性和 Human 概念。此外，图中匿名节点的作用在于，限制每个人有且只有一个父亲，因此也要保留这种限制。Animal 这一概念没有对应实体，因此去除 Animal 这一概念。

Achille 等^[13]通过缩小 ABox 来提升本体推理任务的速度。推理可以转化为一种一致性判别，因此，摘要后的 ABox 应具有快速判别一致性的

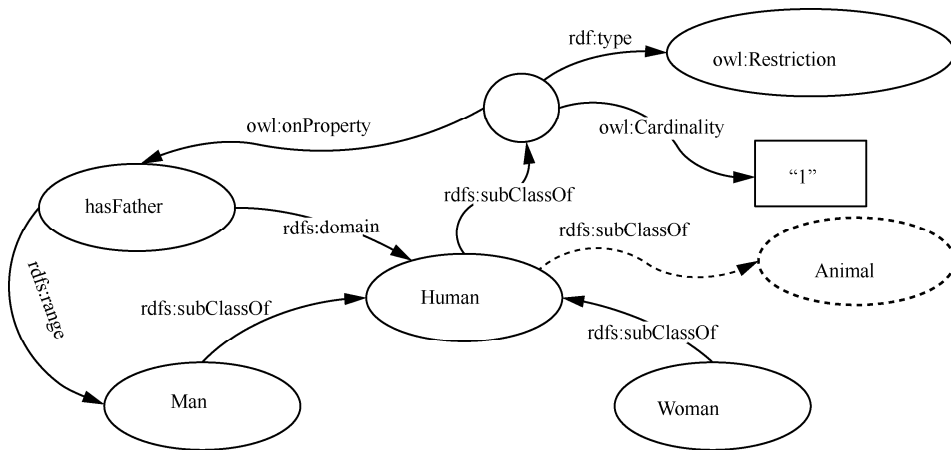


图 3 Alani 等算法摘要后的示例本体模式层

性质。具体地，针对基于描述逻辑的本体，通过映射函数 f 将 ABox 中的若干相似个体映射为摘要本体中的同一个个体，以此简化 ABox 的规模，输出其摘要。对 ABox 的摘要进行分割后分别判别一致性，对于不符合一致性的部分，回溯到原本体进行一致性判别，从而在一定程度上解决了对基于描述逻辑的本体推理困难问题。此方法的一大优势是，当原本体发生变化时，摘要本体也随之变化，无须重复计算。然而，这种摘要方法无法确切地给出一致的 ABox 摘要，只能用于特定的推理任务。

图 4 摘要部分展示的是示例本体经过 Achille 等算法处理后的摘要本体，此算法仅针对数据层，此摘要保证映射前后元素间的关系不变。设映射函数为 f ，有 $f(\text{Bob}) = B, f(\text{Anna}) = f(\text{Alice}) = A, f(\text{Cam}) = f(\text{Cary}) = C$ 。从图中可以看出，映射前有 hasFather 关系的 2 个实体，映射后仍然有这种关系。图中灰色椭圆覆盖的部分为摘要本体分割后的一个子本体，每个子本体是摘要本体中的一个闭连通子集，对应原本体中的若干个闭连通子集。

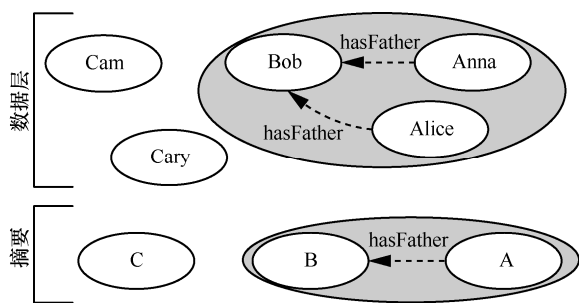


图 4 Achille 等算法摘要前后的数据层

Deng 等^[10]给出了本体无损摘要的理论框架。通过合并概念断言和关系断言减小 ABox 中断言的个数，从而达到缩小 ABox 的目的。首先获取所有实例集合并查找其中的等价关系，然后构建并存储等价实例集合，将相互等价的实例用一个新的实例代替。相对于其他任务导向的方法，这种摘要方法在没有减少原本体中任何信息的前提下实现了本体摘要，适用于本体推理、问答系统等多种任务。然而，此方法存在一个问题：当原本体发生变化时，原本相互等价的概念或实例可能不再等价。因此，原本体一旦发生变化，就需要重新计算其摘要本体，增加了开销。

图 5 摘要部分是经过 Jun Fang 等算法处理后的本体摘要，该算法仅针对数据层，需合并等价概念断言和等价关系断言。若两概念断言中的实体交换位置不改变本体语义，那么认为两实体等价。图 5 所示的本体中有 $\text{Man}(\text{Cam})$ 和 $\text{Man}(\text{Cary})$ ，Cam 和 Cary 这 2 个实体的交换不改变本体语义，因此两实体等价。同理，若 2 个关系断言中的实体交换位置不改变本体语义，那么两实体等价。在图 5 所示的本体中，Anna 和 Alice 等价。将相互等价的实体合并为一个新的实体，并记录新实体代表的等价元素，就完成了摘要。

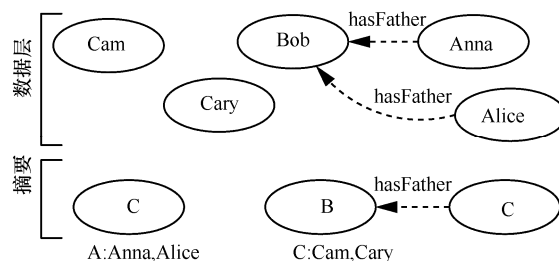


图 5 Jun Fang 等算法摘要前后的数据层

3.2 面向用户的摘要方法

Zhang 等^[2]以 RDF 语句为本体摘要的基本单元。该方法将 RDF 句子看作节点，RDF 句子之间的关系看作边，将本体转化为一个有向图。节点 v_1 、 v_2 之间形成连边的条件是：若 v_1 代表的 RDF 句子的客体也是 v_2 代表的 RDF 句子的主体，那么存在一条从 v_1 到 v_2 的连边。若 v_1 代表的 RDF 句子的主体也是 v_2 代表的 RDF 句子的主体，那么分别存在一条从 v_1 到 v_2 和从 v_2 到 v_1 的连边。对于转化后的有向图，计算每个节点在图中的“中心性”以决定节点的相对重要性。选取中心性最高的若干个 RDF 句子作为摘要结果。节点中心性通过若干参数的线性加权得到。用到的参数有：1) 度中心性^[14]，其中，入度中心性衡量了一个节点的普及程度，出度中心性衡量了一个节点的权威性；2) 介数中心性^[15]，这一参数即图中所有节点对之间的最短路径中经过当前节点的数量，衡量了节点的桥接性；3) 特征向量中心性^[16]，这一参数是度中心性的扩展，某个节点的特征向量中心性取决于节点的度与其邻接点的重要性。

Zhang 等将摘要本体中的连接分为对等连接 (coordinate link) 和序列连接 (sequential link)，在图 6 中分别由 C 和 S 标识，算法仅考虑模式层。

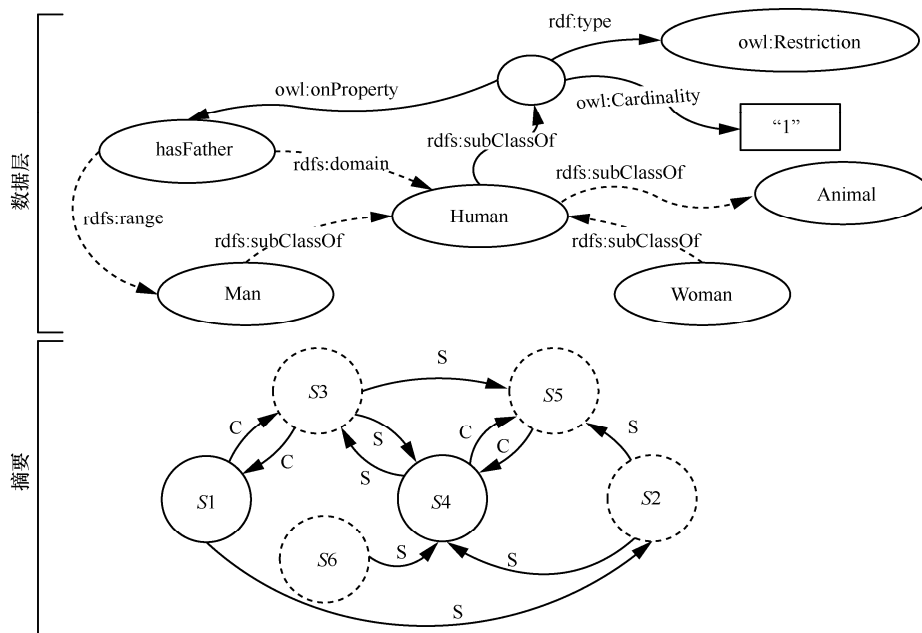


图 6 基于 RDF 句子的算法摘要的图结构

图 6 摘要部分, 每个节点代表一个 RDF 句子转化而来的 SPO 三元组 (Subject, Predicate, Object), 边代表三元组之间的联系。若某个三元组 A 的尾实体是另一个三元组 B 的头实体, 那么图结构中存在一条从 A 到 B 的序列连接。若某个三元组 A 和另一三元组 B 的头实体相同, 那么摘要本体中, A 和 B 之间存在 2 条对等连接。图 6 数据层中, 不同线型的关系对应的三元组对应摘要中不同线型的节点。转化后, 利用复杂网络中的各项参数进行节点重要性评估, 从而选出最重要的若干个节点, 即 RDF 句子。

然而, 单纯地将本体转化为图, 用结构参数生成本体摘要可能会忽略概念层面的语义信息^[17]。

Peroni 等^[18]抽取核心概念作为本体摘要结果。

该方法将概念看作节点, 概念间的类属关系 (subClassOf) 看作连边, 将本体转化为一个有向图。针对每个节点定义若干表示其不同特性的参数, 线性加权后作为最终的重要性评价指标, 选取重要性最高的若干个概念作为摘要结果。方法用到的参数有: 1) 命名简洁性^[19], 参数偏好标签简洁的概念, 为核心概念的衡量增添了语义信息; 2) 基础等级。这一参数评估概念在本体分类学中的中心性; 3) 密度, 这一参数偏好具有较多属性和分类关系的概念; 4) 覆盖, 这一参数的目的在于不忽视本体的任何一个重要部分; 5) 热度, 这一参数强调了最为常用的若干概念。

如图 7 所示, 算法基于粗体部分所示的本体网络进行重要概念的选取。仅保留本体中的概念和类属关系构成网络, 并根据概念标签和复杂网络理论

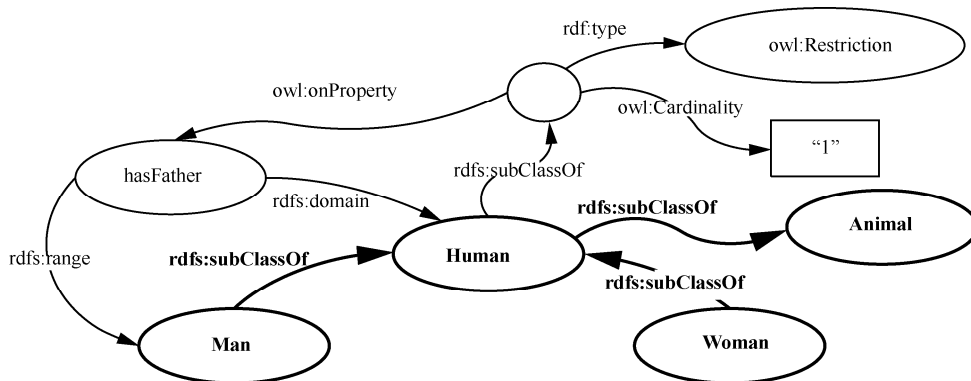


图 7 Peroni 等算法摘要后的模式层

中的各项指标完成重要概念的选取。

Zhang 等^[20]在之前工作^[2]的基础上添加了术语的语义信息。他们仍然以 RDF 句子为单位进行摘要。与之前不同的是，在评估句子重要性的过程中，他们将 RDF 句子转化为 SPO (Subject, Predicate, Object) 三元组，构造句子-术语二分图，并在二分图上计算节点的结构重要性和语义重要性。其中，结构重要性是全球语义网中与当前句子中 SPO 三元组相关的实体数量。语义重要性是当前句子对应的 SPO 三元组在全球语义网中出现次数的加权平均。将 2 种重要性线性加权后进行重排序，得到 RDF 句子的重要性排序，选取若干个重要性最高的 RDF 句子作为本体摘要结果。

图 8 是算法用于计算节点重要性的二分图。左侧的 S_i 表示本体模式层的 RDF 句子，与图 6 中的 S_i 有相同含义；右侧节点表示概念。若某个 RDF 句子 S_i 中包含某个概念 C_j ，那么二分图中存在一条从 S_i 到 C_j 的边。

上述方法利用不同的转化方式将本体转化为图，设定若干参数选取重要程度最高的节点。然而，从本体摘要的定义来看，其目的是产生相较之前本体规模更小的子本体。而上述方法输出的要么是孤立的概念，要么是彼此间可能联系并不密切的 RDF 语句，没有输出可用的子本体。

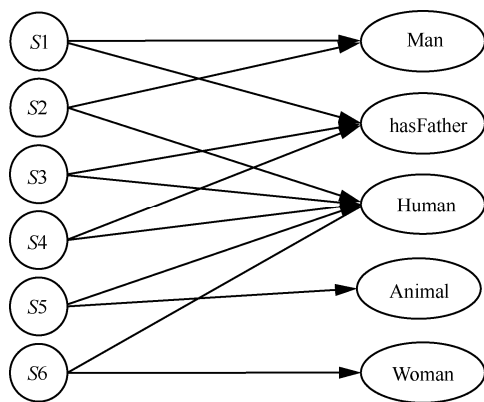


图 8 基于二分图的算法摘要的图结构

Pires 等^[21]构造子本体作为摘要结果。将本体中的概念看作节点，关系看作有向边，首先通过度中心性和频繁性的线性加权选取重要概念节点，然后对概念节点进行分组，最后通过添加组间节点连接各个组，形成原本体对应的子本体。

Queiroz-sousa 等^[22]在 Pires 等工作的基础上构

造用户个性化子本体作为摘要结果。将本体中的概念看作节点，关系看作有向边，首先通过度中心性和接近中心性的线性加权选取重要概念节点，然后通过拓宽路径算法选取连接重要节点的必要节点，进而生成若干条经过重要概念节点的重要路径，从而生成原本体对应的子本体。

Troullinou 等^[23]在之前工作^[24-25]的基础上，结合数据层和模式层设计了摘要方法。具体地，首先利用数据层的数据计算本体中边的相对基数；然后根据模式层中节点所连边的类型和相对基数计算节点的重要程度，从而选取重要概念节点；最后设计算法选取尽可能不相交的、连接重要概念节点的路径，形成原本体对应的子本体。

上述 3 种方法的思想都是首先将本体中的概念看作节点，关系看作边，将本体转化为图。然后通过多种参数的加权和计算节点重要性，从而筛选出重要节点。最后通过算法选取路径连接重要节点，形成摘要图。3 种方法形成的摘要图都与图 7 形式相同。上述方法的区别在于，Pires 等通过分组和连接两步选取路径，Queiroz-sousa 等通过用户给定的参数生成个性化摘要，Troullinou 等在选取重要节点时同时利用了数据层和模式层的信息。

4 评价准则

Li 等^[26]参照本体评估的方式，给出了本体摘要评估的 3 种方式：1) 基于应用的评估，摘要的质量与利用该摘要应用的表现成正比；2) 基于黄金标准的评估，通过对比人工构建的标准摘要和机器生成摘要的相似性进行评估，相似性越高，说明机器的摘要越准确；3) 基于语料覆盖度的评估，根据摘要对本体的语义覆盖度进行评估。其中，面向用户的摘要大多通过基于黄金准则的方法进行评估，具体地，比较用户和算法对概念的排序相似度，或对比用户构建本体和算法生成本体的相似度。面向任务的摘要大多通过基于应用的方法进行评估，具体地，基于摘要本体进行推理、指导问答系统等，通过运行结果评价摘要本体质量。由于基于语料覆盖度的本体摘要评估仅适用于评价手工构建的本体^[27]，因此目前尚无基于语料覆盖度的本体摘要评估。

接下来,详细介绍基于应用的评估和基于黄金准则的评估。

4.1 基于应用的评估

Alani 等^[12]将摘要前后的本体分别应用于问答系统,并通过对比二者返回答案的一致性判断本体摘要水平。基于摘要本体的问答系统返回的答案与原问答系统返回的答案越相似,说明摘要结果越好。

曲琛等^[28]对摘要前后的本体进行推理一致性检测,若对摘要前后的本体进行推理的结果完全相同,说明二者在语义上是等价的。

事实上,问答系统返回答案的过程也可转化为推理过程。具体地,将待判断断言的否定放入问答系统进行推理,若得出不一致的结论,说明断言是正确的。因此,可通过推理机^[29]完成基于应用的评估。推理机是对知识进行解释的程序,根据知识的语义,按一定策略找到的知识进行解释执行^[30]。目前常用的本体推理机及其推理类型如表1所示^[31-36]。

表1 典型推理机及其推理类型

推理机	推理方法
Jess ^[31]	基于规则
Jena ^[32]	基于规则
Pellet ^[33]	基于描述逻辑
Racer ^[34]	基于描述逻辑
FaCT ⁺⁺ ^[35]	基于描述逻辑
FuzzyDL ^[36]	基于描述逻辑

4.2 基于黄金准则的评价

1) 摘要结果为 RDF 句子

Zhang 等^[2]通过对比专家选出的 RDF 句子和算法选出的 RDF 句子来评估摘要。具体地,算法摘要与专家评估结果的相似度通过式(1)计算。

$$\text{sim}(M, S, \{J_1, \dots, J_n\}) = \frac{M(S, J_1) + \dots + M(S, J_n)}{n} \quad (1)$$

其中, M 是词汇重叠度量, S 是算法生成的摘要, $J_1, \dots, J_n$ 是专家构建的摘要, 即选出的具有代表性的 RDF 句子集合, n 是专家个数。

词汇重叠度量的计算公式如下。

$$M(S, J) = \frac{|T_{\text{complete}}| + 0.5|T_{\text{partial}}| + |T_{\text{sibling}}|}{|T|} \quad (2)$$

其中, T 是 J 中包含的术语集合, $T_{\text{complete}} \subseteq T$ 是 J 与 S 中相同术语的集合, $T_{\text{partial}} \subseteq T - T_{\text{complete}}$ 是 J 中上层或下层术语属于 S 的术语的集合, $T_{\text{sibling}} \subseteq T - T_{\text{complete}} - T_{\text{partial}}$ 是 J 中同层术语属于 S 的术语的集合。

2) 摘要结果为重要概念

Peroni 等^[18]根据八位专家选出的本体中的 20 个重要概念, 评估算法的质量。首先, 对比八位专家给出的本体中的重要概念集合 $C_i, i=1, 2, \dots, 8$, 并将超过半数专家认同的概念作为标准概念, 记标准概念集合为 C 。随后, 将算法摘要出的重要概念 A 与标准概念进行比较, 通过式(3)算出其与标准概念的匹配度 m 。

$$m = \frac{|A \cap C|}{|C|} \quad (3)$$

Li 等^[26]从 Zhang 等^[2]算法生成的 RDF 句子中抽取重要概念, 与 Peroni 等算法的结果进行对比。结果表明, 虽然 2 种算法摘要的方法和对象不同 (Zhang 等以 RDF 句子为摘要单元, 而 Peroni 等以概念为摘要单元), 但返回的概念结果相似。然而, Peroni 等的算法包含更多名称简洁的概念和更普及的概念。这是因为, Zhang 等的算法仅考虑了本体的结构信息, 而 Peroni 等还考虑了命名简洁性和词汇普及性。

Queiroz-sousa 等^[22]利用同样的方法评估了自己的算法, 并与 Li 等的评估结果进行了对比, 对比结果如表2^[22]所示。原文中对 Zhang 等算法在 financial 本体下的匹配度计算有误, 表2中已进行更正。

表2 3种本体摘要算法对比

本体	与标准概念的匹配度 ^[18]	与标准概念的匹配度 ^[2]	与标准概念的匹配度 ^[22]
biosphere	80.0%	60.0%	90.0%
financial	66.7%	55.6%	66.7%
aktors portal	75.0%	$n/a\%$	62.5%

3) 摘要结果为子本体

Troullinou 等^[24]参照 Zhang 等^[2]提出的词汇重叠度量计算公式, 提出了本体相似度计算公式, 来衡量算法生成的摘要本体 S 与专家构建的黄金

本体 A 的相似度。相似度计算公式如下。

$$Sim(S, A) = \frac{|K_{match}| + 0.6 \sum_1^{K_{sub}} \frac{1}{depth} + 0.3 \sum_1^{K_{super}} \frac{1}{depth}}{|K|} \quad (4)$$

其中, K 是 A 中包含的术语集合, $K_{match} \subseteq K$ 是 A 与 S 中相同术语的集合, $K_{sub} \subseteq K - \frac{K_{match}}{K_{super}} \subseteq$

$K - K_{match}$ 分别是 A 中上层或下层术语属于 S 的术语的集合, $depth$ 是理想类与生成摘要标识类之间的距离。

Troullinou 等通过本体相似度对比了 Peroni 等的算法、Queiroz-sousa 等的算法, 以及他们之前提出的 RDF Digest 算法^[24], 对比结果如表 3 所示。

4.3 其他评价方法

Zhang 等^[20]通过定性评估的方式将摘要结果与其之前的方法^[2]进行对比。因为算法用到了全球语义网中的信息, 对专家而言, 记住语义网中

的所有信息是不现实的。因此, 他们针对本体信息绘制相应的摘要图进行评估。

Deng 等^[10]从空间复杂度和时间复杂度两方面评估无损摘要的质量。空间复杂度即给定一个本体, 摘要算法在保留其语义的前提下能将其缩减到的最小规模。用压缩本体中的公理数量与原本本体中的公理数量作比, 来衡量其空间复杂度。具体计算公式如式(5)所示。

$$\begin{aligned} & compressibility \\ &= \frac{number_of_axioms_inO^{ec}}{number_of_axioms_inO} \end{aligned} \quad (5)$$

时间复杂度即寻找本体中等价关系花费的时间和压缩本体花费的时间。

表 4 给出了本文所述算法的各项性能对比。

5 结束语

目前, 大多数本体摘要方法为用户导向的摘要^[38], 对任务导向摘要方法的研究仍处于初步阶

表 3

4 种本体摘要算法对比

本体	与标准概念的匹配度			
	文献[18]	文献[22]	文献[24]	文献[23]
biosphere	64.9%	64.5%	63.1%	64.7%
financial	54.0%	61.9%	65.0%	64.8%
aktors portal	42.6%	49.4%	62.2%	55.0%

表 4

本体摘要算法及评估方式对比

方法	目标	输出结果	利用语义信息	摘要对象	评估方式	备注
文献[2]	用户导向	RDF 句子	否	模式层	专家评测后, 用 vocabulary overlap 评价机器给出的结果	借鉴文本摘要方法
文献[20]	用户导向	RDF 句子	是	模式层	定性评估, 根据 2 个例子进行分析	构造句子-术语二分图, 并利用全球语义网中的信息进行重要性评估
文献[18]	用户导向	核心概念	是	模式层	专家评测与机器结果对比	考虑了概念的普及程度
文献[37]	用户导向	核心概念	否	模式层	n/a	选取网络中的重要节点
文献[21-22]	用户导向	子本体	否	模式层	专家评测与机器结果对比	先选取核心概念, 再通过路径连接
文献[23-25]	用户导向	子本体	否	数据层/模式层	比较参考本体和生成本体的相似度	利用了数据层和模式层的信息进行摘要
文献[10,28]	任务导向	子本体	否	数据层	评估空间复杂性和时间复杂性, 即摘要前后本体大小变化和一致性检测需要的时间	通过合并等价关系缩小本体, 摘要前后本体蕴含的信息量不变
文献[12]	任务导向	子本体	否	模式层	将摘要前后的本体应用于问答系统, 根据答案质量判断本体摘要水平	仅针对特定应用, 无法保证摘要本体的语义完整性和一致性
文献[13]	任务导向	子本体	否	数据层	对摘要前后的本体进行推理, 若结果一致, 说明未破坏本体信息	原本体发生变化时, 摘要本体也随之变化

段。用户导向的摘要旨在输出更为简洁的、便于用户理解的摘要信息，其目前有3种输出形式：RDF句子、重要概念集合，以及图形化的子本体。任务导向的摘要旨在提供高效的摘要应用服务，其目前输出形式为子本体。本文对本体摘要的研究现状进行了梳理，总结了用户导向和任务导向的摘要方法，分类介绍了摘要质量的评估方法。摘要的未来研究趋势主要集中在以下几个方面。

1) 目前没有统一的摘要的评价指标，如何根据本体的内在特性，提出一种统一的摘要评价标准是该领域发展的关键点。

2) 现有摘要方法都是利用某种规则，从现有本体中抽取关键信息作为摘要结果。如何根据现有的摘要信息，抽象出本体中没有显式出现的摘要词语，是一个新的研究方向。

3) 现有摘要方法没有利用机器学习算法，可以考虑将机器学习和深度学习算法与摘要结合，提出新的摘要方法。

参考文献:

- [1] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582-600.
LIU Q, LI Y, YIN H, et al. Knowledge graph construction techniques[J]. Journal of Computer Research and Development, 2016, 53(3): 582-600.
- [2] ZHANG X, CHEN G, QU Y Z. Ontology summarization based on RDF sentence graph[C]. International Conference on World Wide Web. ACM, 2007: 707-716.
- [3] WANG R, YU X, LI Y, et al. Ontology partition method for semantic query routing[C]. International Conference on Intelligent Pervasive Computing. IEEE, 2008: 86-89.
- [4] STUCKENSCHMIDT H, KLEIN M. Structure-based partitioning of large concept hierarchies[C]. International Semantic Web Conference. 2004: 289-303.
- [5] D'AQUIN M, SCHLICHT A, STUCKENSCHMIDT H, et al. Ontology modularization for knowledge selection: experiments and evaluations[C]. International Conference Database and Expert Systems Applications. 2007: 874-883.
- [6] D'AQUIN M, SCHLICHT A, STUCKENSCHMIDT H, et al. Modular ontologies: concepts, theories and techniques for knowledge modularization[J]. Lecture Notes in Computer Science, 2009, 5445(4): 67-89.
- [7] MANI I. Automatic Summarization[M]. John Benjamins Publishing Company, 2001.
- [8] LI N, MOTTA E, D'AQUIN M. Ontology summarization: an analysis and an evaluation[J]. International Workshop on Evaluation of Semantic Technologies, 2010.
- [9] ERKAN, RADEV, DRAGOMIR R. LexRank: graph-based lexical centrality as salience in text summarization[J]. Journal of Qiqihar Junior Teachers College, 2011, 22:2004.
- [10] 邓玲玉. 基于重要度计算的语义物联网本体摘要方法研究[D]. 大连: 大连海事大学, 2017.
DENG L. Research on method of importance computing based semantic web of things ontology summarization[D]. Dalian: Dalian Maritime University, 2017.
- [11] 张灵峰, 夏战锋, 彭志平. 基于Tbox和Abox的描述逻辑推理研究[J]. 计算机技术与发展, 2010, 20(11): 122-125.
ZHANG L, XIA Z, PENG Z. Research on description logic's reasoning based on Tbox and Abox[J]. Computer Technology and Development, 2010, 20(11): 122-125.
- [12] ALANI H, HARRIS S, O'NEIL B. Winnowing ontologies based on application use[C]. European Conference on the Semantic Web: Research and Applications. 2006: 185-199.
- [13] FOKOUE A, KERSHENBAUM A, MA L, et al. The summary ABox: cutting ontologies down to size[C]. The Semantic Web - ISWC 2006. 2006: 343-356.
- [14] MIKA P. Social Networks and the Semantic Web[C]. IEEE/WIC/ACM International Conference on Web Intelligence. 2004: 285-291.
- [15] FREEMAN L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1997: 35-41.
- [16] ZAFARANI R, ABBASI M A, LIU H. Social media mining: an introduction[M]. Cambridge University Press, 2014.
- [17] Ontology selection: ontology evaluation on the real semantic Web.
- [18] PERONI S, MOTTA E, D'AQUIN M. identifying key concepts in an ontology, through the integration of cognitive principles with statistical and topological measures[C]. Asian Semantic Web Conference on the Semantic Web. 2008.
- [19] ROSCH E. Principles of categorization[J]. Readings in Cognitive Science, 1988:312-322.
- [20] ZHANG X, CHENG G, GE W Y, et al. Summarizing vocabularies in the global semantic Web[J]. Journal of Computer Science and Technology, 2009, 24(1):165-174.
- [21] PIRES C E, SOUSA P, KEDAD Z, et al. Summarizing ontology-based schemas in PDMS[J]. Icdew, 2010:239-244.
- [22] QUEIROZ-SOUSA P O, SALGADO A C, PIRES C E. A method for building personalized ontology summaries[J]. Journal of Information and Data Management, 2013, 4(3): 236.
- [23] TROULLINO G, KONDLAKIS H, DASKALAKI E, et al. Ontology understanding without tears: the summarization approach[J]. Semantic Web, 2016, 8(6).
- [24] TROULLINO G, KONDLAKIS H, DASKALAKI E, et al. RDF digest: efficient summarization of rdf/s kbs[C]. European Semantic Web Conference. 2015: 119-134.
- [25] TROULLINO G, KONDLAKIS H, DASKALAKI E, et al. RDF digest: ontology exploration using summaries[C]. The ISWC 2015 Posters & Demonstrations Track Co-located with the 14th International Semantic Web Conference (ISWC-2015). 2015.
- [26] LI N, MOTTA E. Evaluations of user-driven ontology summarization[J]. Lecture Notes in Computer Science, 2010, 6317(1): 544-553.
- [27] 宋丹辉. 本体评价研究综述[J]. 情报理论与实践, 2011,

- 34(9):118-122.
- SONG D. Ontology evaluation and research survey[J]// Information Studies:Theory & Application, 2011, 34(9):118-122.
- [28] 曲琛. 语义物联网本体的摘要方法研究[D]. 大连海事大学, 2014.
- QU C. Research on summarization method of ontology in semantic web of things[D]// Dalian Maritime University. 2014.
- [29] SINGH S, KARWAYUN R. A Comparative Study of Inference Engines[C]// International Conference on Information Technology: New Generations. IEEE, 2010:53-57.
- [30] 潘超, 古辉. 本体推理机及应用[J]. 计算机系统应用, 2010, 19(9):163-167.
- PAN C. GU H. Ontology reasoned and its application[J]//Computer System&Applications, 2010, 19(9):163-167.
- [31] FRIEDMAN-HILL E J. JESS, the rule engine for the Java platform[J]. Sandia National Laboratories, 2006, 48(2): 340-354.
- [32] CARROLL J J, DICKINSON I, DOLLIN C, et al. Jena: implementing the semantic Web recommendations[C]//International World Wide Web Conference on Alternate Track Papers & Posters. ACM, 2004: 74-83.
- [33] SIRIN E, PARSIA B, GRAU B C, et al. Pellet: A practical OWL-DL reasoner[J]. Web Semantics Science Services & Agents on the World Wide Web, 2007, 5(2):51-53.
- [34] HAARSLEV V, MÖLLER R. Racer: a core inference engine for the semantic Web[J]. EON, 2003(October):27-36.
- [35] TSARKOV D, HORROCKS I. FaCT++, description logic reasoner: system description[C]//International Joint Conference on Automated Reasoning. 2006:292-297.
- [36] BOBILLO F, STRACCIA U. FuzzyDL: An expressive fuzzy description logic reasoner[C]//IEEE International Conference on Fuzzy Systems. 2008:923-930.
- [37] GRAVES A, ADALI S, HENDLER J. A method to rank nodes in an RDF Graph[C]// Poster and Demonstration Session at the, International Semantic Web Conference(DBLP). 2008:309-320.
- [38] POURIYEH S, ALLAHYARI M, LIU Q, et al. Graph-based Ontology Summarization: a survey[J]. arXiv:1805.06051, 2018.

[作者简介]



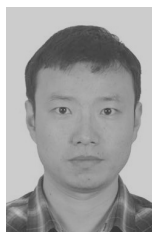
丁悦航（1995-），女，山东菏泽人，国家数字交换系统工程技术研究中心硕士生，主要研究方向为复杂网络、知识图谱。



于洪涛（1970-），男，辽宁丹东人，博士，国家数字交换系统工程技术研究中心研究员，主要研究方向为网络大数据分析 & 处理。



黄瑞阳（1986-），男，福建漳州人，博士，国家数字交换系统工程技术研究中心助理研究员，主要研究方向为文本挖掘、图挖掘。



李英乐（1985-），男，河北衡水人，硕士，国家数字交换系统工程技术研究中心助理研究员，主要研究方向为网络大数据分析 & 处理。