

基于真-伪三连体神经网络的传感器 本体匹配方法

卢家伟, 刘昌锋

(兰州理工大学 机电工程学院, 兰州 730050)

摘要: 针对物联网系统中传感器本体交互所面临的异质性问题, 提出了一种基于真-伪三连体神经网络(GP-TNN)的传感器本体匹配方法。首先, 根据实体的出度入度提取核心实体, 进而提升训练样本的质量; 然后, 通过设计确定性的混合相似度度量方法, 用于生成高置信度的锚点匹配对, 以此避免训练过程中对参考对齐的依赖; 最后, 结合孪生网络和伪孪生网络对同类及异类标注属性的表征能力, 构建 GP-TNN 增强深层次语义特征挖掘能力。评估实验分别在本体对齐评测计划(OAEI)中 Benchmark 数据集以及 3 个真实的传感器本体上进行, 实验结果表明, 本文方法较对比方法能够更加有效地提高匹配结果的质量, 为物联网中传感器本体匹配任务提供了可靠的理论支持。

关键词: 计算机应用; 深度学习; 真-伪三连体网络; 传感器本体匹配; 本体对齐评测计划

中图分类号: TP391 **文献标志码:** A **文章编号:** 1671-5497(2026)09-2456-11

DOI: 10.13229/j.cnki.jdxbgxb.20250233

Sensor ontology matching method based on genuine-pseudo triplet neural network

LU Jia-wei, YAN Chang-feng

(School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: To address the heterogeneity problems in sensor ontology interaction within Internet of Things (IoT) systems, this paper proposes a sensor ontology matching method based on a Genuine-Pseudo Triplet Neural Network (GP-TNN). Firstly, core entities are extracted by analyzing the in-degree and out-degree of entities to enhance the quality of training samples. Secondly, a deterministic hybrid similarity measurement method is designed to generate high-confidence anchor matching pairs, thereby reducing the dependence on reference alignments during training. Finally, by leveraging the representation capabilities of Siamese and pseudo-Siamese networks for homogeneous and heterogeneous annotation attributes, the GP-TNN is constructed to enhance the extraction of deep semantic features. Evaluation experiments are conducted on the Benchmark dataset from the Ontology Alignment Evaluation Initiative (OAEI) and three

收稿日期: 2025-03-20.

基金项目: 国家自然科学基金项目(51165018); 甘肃省优秀研究生“创新之星”项目(2022CXZX-407).

作者简介: 卢家伟(1994-), 男, 博士研究生. 研究方向: 智能信息处理技术. E-mail: 18816200508@163.com

通信作者: 刘昌锋(1974-), 男, 研究员, 博士. 研究方向: 旋转机械智能运维技术, 人工智能.

E-mail: changf_yan@163.com

real-world sensor ontologies. Experimental results show that the proposed method can more effectively improve the quality of matching results than the comparative methods, and provides reliable theoretical support for the sensor ontology matching task in the Internet of Things.

Key words: computer application; deep learning; genuine-pseudo triplet neural network; sensor ontology matching; ontology alignment evaluation initiative

0 引言

伴随着物联网(Internet of Things, IoT)的快速发展,移动流量数据的急剧增长给传感器设备带来了诸多挑战,如高延迟和能源消耗增加^[1,2]。第六代通信技术(6G)和网络孪生等新兴网络技术正在成为促进复杂无线互联的关键支撑^[3],促使大量传感器设备在异构网络中进行数据交换。然而,由于不同网络环境下的数据标准、格式和语义定义存在差异,传感器数据往往缺乏统一的语义信息,导致数据的异质性问题。传感器本体通过机器可解释的方式组织传感器信息和知识,为解决数据异质性提供了一种有效手段。但由于本体构建过程中存在工程主观性,传感器本体自身也具有固有的异质性^[4],包括对相同概念使用不同的描述、对不同概念使用相同的描述,以及在特定领域构建不同粒度的层次结构等。这种异质性问题严重阻碍了传感器设备之间的交互与协作。

为解决上述问题,传感器本体匹配(Sensor ontology matching, SOM)成为一种关键技术,其能够有效识别待匹配实体之间的语义关系。近年来,深度学习在本体匹配中的应用已成为研究的热点^[5],研究者尝试利用深度神经网络自动提取概念的语义特征,以提高匹配的准确性和自动化程度。例如,Bento等^[6]将本体匹配任务定义为二分类问题,采用卷积神经网络(CNN)从字符嵌入中提取特征,并利用参考对齐数据训练匹配模型,从而减少对手工设定相似度阈值的依赖。Jiang等^[7]提出基于长短时记忆网络(LSTM)的匹配方法,结合概念的结构和语义信息进行对齐优化,有效缓解了梯度消失问题。Iyer等^[8]则采用监督学习的神经网络,引入双重注意力机制,以生成概念的上下文表示,并优化字符嵌入,使匹配结果更加精确。此外,Xue等^[9]采用孪生神经网络(Siamese Neural Network, SNN),同时对待匹配概念及其前两级超类进行字符嵌入,以学习概念的结构信息。

尽管这些方法在一定程度上提高了匹配质量,但仍然存在以下问题:①现有方法大多仅依赖单一的标注属性提取实体的语义特征,未能有效结合不同类型的标注属性信息,导致语义表征能力不足;②训练过程中高度依赖参考对齐数据,而在实际应用中,高质量的参考对齐往往难以获得,影响模型的适用性;③在处理高度复杂的异质本体时,匹配精度仍然有限。针对这些问题,本文提出了一种基于真-伪三连体神经网络(Genuine-pseudo triplet neural network, GP-TNN)的传感器本体匹配方法,结合核心实体选择策略、混合相似度度量方法以及GP-TNN的优势,综合优化匹配过程,以更有效地识别异质本体间的语义关系。

1 基本概念

1.1 传感器本体与传感器本体匹配

为便于本文的研究工作,传感器本体定义如下:

定义1 传感器本体可表示为一个三元组 $O = (C, P, I)^{[10,11]}$,其中: C 表示传感器本体中的类集合; P 表示传感器本体中的属性集合,其中包含数据属性和对象属性; I 表示传感器本体中的实例集合。

类、属性和实例统称为实体,它们各具有3种常见的标注属性,即标识符(ID)、标签(label)和注释(comment)。

为了解决传感器本体之间的异质性问题,需要进行传感器本体对齐,其定义如下:

定义2 传感器本体匹配过程可以表示为一个函数 $f(O_1, O_2, R, p, r)^{[12]}$,其中: O_1 和 O_2 分别表示源本体和目标本体; R 表示参考对齐集合,通常由领域专家提供; p 表示匹配过程中的参数集合; r 表示匹配过程中涉及的资源,通常指相关的背景知识和领域约束。

通过本体匹配过程,可生成对齐结果,其定义如下:

定义3 对齐结果可表示为一个四元组

$(e, e', n, L)^{[13]}$, 其中: e 和 e' 分别表示两个本体中的实体; n 表示相似度值, 其取值范围为 $[0, 1]$ 。当 $n=0$ 时, 表示两个实体完全不相似, 当 $n=1$ 时, 表示两个实体完全相似; L 表示语义关系, 本文中仅考虑等价关系。

1.2 相似度度量方法

相似度度量方法以两个本体 O_1 和 O_2 为输入, 生成一个大小为 $|O_1| \times |O_2|$ 的相似度矩阵 M , 矩阵中的每个元素表示本体 O_1 中第 i 个实体与本体 O_2 中第 j 个实体之间的相似度得分^[14]。

目前, 相似度度量方法可分为 3 类: 基于术语、语言学和结构的相似度度量方法, 然而, 基于结构的相似度度量方法计算复杂度较高, 导致计算开销增加^[9], 因此本文在构建训练数据的过程中不采用该方法。

1.2.1 基于术语的相似度度量方法

基于术语的相似度度量方法通过字符串相似性来计算两个实体的相似度。在本体匹配任务中, N -gram 相似度度量方法具有显著优势, 尤其当 $N=3$ 时效果较优^[15]。给定字符串 s_1 和 s_2 , N -gram 定义如下:

$$N\text{-gram}(s_1, s_2) = \frac{2 \times \text{comm}(s_1, s_2)}{N_{s_1} + N_{s_2}} \quad (1)$$

式中: $\text{comm}(s_1, s_2)$ 为 s_1 和 s_2 中共同的子串; N_{s_1} 和 N_{s_2} 分别为 s_1 和 s_2 中子串的数量。

另一个代表性的基于术语的相似度度量方法是 Cosine^[16], 它能够有效评估两个句子间的相似度。给定句子 D_1 和 D_2 , Cosine 相似度度量定义如下:

$$\text{Cos}(D_1, D_2) = \frac{V_1 \cdot V_2}{\|V_1\| \times \|V_2\|} \quad (2)$$

式中: V_1 和 V_2 分别为 D_1 和 D_2 的向量表示; $\|\cdot\|$ 为欧几里得距离。

1.2.2 基于语言学的相似度度量方法

Wu & Palmer 相似度度量方法^[17]作为典型的基于语言学的相似度度量方法, 其核心思想是利用待匹配实体在背景知识库中的公共父节点及其层次深度来计算相似度值。WordNet 是一个近义词词汇数据库, 其中的词汇按照清晰的层次结构进行组织, 常作为 Wu & Palmer 的背景知识库。给定两个实体 e_1 和 e_2 , Wu & Palmer 定义如下:

$$S_{\text{WP}} = \frac{2 \times \text{depth}(\text{LCA}(e_1, e_2))}{\text{depth}(e_1) + \text{depth}(e_2)} \quad (3)$$

式中: $\text{LCA}(e_1, e_2)$ 为 e_1 和 e_2 在 WordNet 层次结构中的最近公共祖先; $\text{depth}(e_1)$ 和 $\text{depth}(e_2)$ 分别为 e_1 和 e_2 在 WordNet 层次结构中的深度; $\text{depth}(\text{LCA}(e_1, e_2))$ 为最近公共祖先的深度。

2 基于 GP-TNN 的传感器本体匹配

基于 GP-TNN 的传感器本体匹配框架如图 1 所示, 蓝色和黑色连接线分别引导训练和测试过程(彩图见电子版, 以下同)。模型的训练过程包括以下步骤: 首先, 从待匹配的本体中提取核心实体; 其次, 通过混合相似度度量方法构建独立于参考对齐的训练数据集; 然后, 对训练数据集应用字符嵌入将字符转换为计算机可处理的向量表示; 最后, 将编码后的字符向量输入 GP-TNN 进行模型训练。

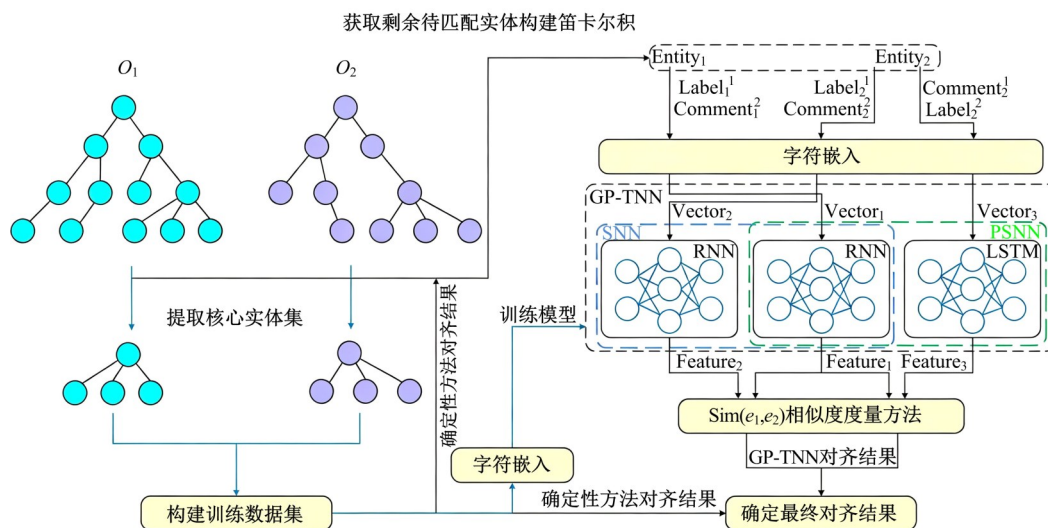


图 1 基于 GP-TNN 的传感器本体匹配框架

Fig. 1 Framework of sensor ontology matching based on GP-TNN

2.1 选取核心实体

度中心性 (Degree centrality, DC) 在网络中用于衡量节点的重要性,能够对节点重要性进行量化^[18]。即在传感器本体中,一实体的节点度越高,其DC越大,则该实体的重要性越高。为了获取具有代表性的训练数据,在目标本体中选择高DC的核心实体以构建训练集。具体而言,首先按照本体中实体节点的入度与出度之和进行降序排序,其次,选取排名前30%的实体作为核心

实体。

图2展示了经由概念模块划分的真实传感器本体SSN的核心类及其关系(绿色内容为SSN的核心本体SOSA的组成部分),其中圆角矩形表示类,实线连接子类,虚线描述类之间的关系,每个概念模块的名称位于左上角。图2中,Property类的DC为2,Condition类的DC为1,表明在选择核心实体时,Property将更可能被选为核心实体。

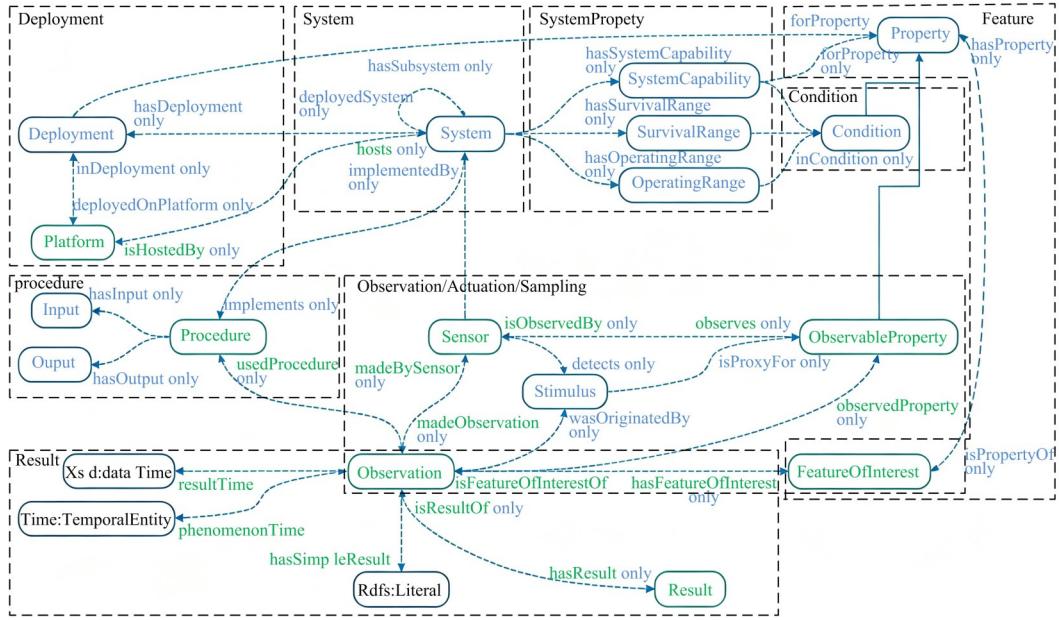


图2 SSN本体中核心类及其关系

Fig. 2 Core classes and their relationships in SSN ontology

2.2 构建训练数据集

为了提高模型训练数据的质量,本文在Lu等^[16]的研究基础上提出了一种确定性的混合相似度度量方法,该方法集成了多种相似度度量方法以识别潜在的对齐关系。与先前方法相比,本文简化了相似度度量的组合框架。具体而言,为了在超大规模本体中匹配核心实体时提升泛化能力并降低计算复杂度,此处采用顺序匹配策略。给定核心实体 e_1 和 e_2 ,以及它们各自的标注属性 $label_1$ 、 $label_2$ 、 $comment_1$ 和 $comment_2$,核心实体的相似度计算定义如下:

$$S(e_1, e_2) = \begin{cases} s_1 = N - \text{gram}(label_1, label_2), & \text{if } s_1 \geq T \\ \text{or } s_2 = \text{Sim}_{\text{WP}}(label_1, label_2), & \text{if } s_2 \geq T \\ \text{or } s_3 = \text{Cosine}(comment_1, comment_2), & \text{if } s_3 \geq T \end{cases} \quad (4)$$

式中: T 为给定阈值,为避免假阳性样本干扰模型训练,此处设为0.95。

根据式(4),本研究采用3种相似度度量方法按照自上而下的顺序策略计算两个实体之间的相似度值。当某一相似度度量方法确定等价关系后,后续的相似度计算将不再执行,从而有效避免高昂的计算开销。最终,将从核心实体中确定的对齐作为正样本。为了平衡数据集,防止训练过程中的偏差,正样本与负样本的数量需要保持一致。给定一个正样本 $A = (e, e', n, L)$,其中 e 来自 O_1 ,负样本构建通过将 e 替换为 O_1 中其他任意一个同类实体 e'' 来实现。

2.3 基于GP-TNN的传感器本体匹配过程

受孪生神经网络(Siamese neural network, SNN)^[19]的启发,GP-TNN通过内部子网络将输入映射到多维空间,并生成特征向量。在该空间

中,相似实体的特征向量将被训练得更加接近,实体之间的相似度由它们特征向量的相近程度进行衡量。GP-TNN由两种子网络结构组成:SNN和伪孪生神经网络(Pseudo-siamese neural network, PSNN)^[20]。GP-TNN的目标是同时通过待匹配实体的同类及异类标注属性识别它们间的对齐关系。

SNN由两个参数共享的子网络结构组成,通常用于识别同类输入之间的相似关系^[21,22]。为了在循环计算过程中有效捕捉实体间的关系及语义信息,本文在SNN的子网络中采用两个循环神经网络(Recurrent neural networks, RNNs)。鉴于本体中的label和comment字符串较短,RNN能够在学习较少参数的同时避免梯度消失,进而促进提升模型的学习能力。PSNN由两个参数不共享的子网络组成,通常用于评估不同类别输入之间的相似关系^[23]。在本研究中,PSNN采用RNN和长短时记忆网络(Long short-term memory, LSTM)作为其子网络。其中,LSTM通过门控机制调节记忆单元的状态,平衡短期信息更新与长期信息保持,从而能够在时间序列中有效聚焦关键语义信息,提升模型性能。如图1右侧GP-TNN模块所示,SNN和PSNN通过共享相同的RNN子网络,共同构成GP-TNN。

2.3.1 GP-TNN训练过程

为训练GP-TNN,过程包括以下步骤:首先,为了使模型能够学习来自同类和异类标注属性的语义信息,从每个正样本或负样本的实体中分别提取两组标注属性输入。第一组包括label₁¹、label₁¹和comment₁²,第二组包括comment₁²、comment₁²和label₁²,其中,上标表示组编号,下标则表示该标注属性属于源本体O₁或目标本体O₂。其次,为了使计算机能够识别输入实体,需要将其编码为数字向量。本文采用字符嵌入方法生成富含语义信息的数字向量(具体参见<https://github.com/minimaxir/char-embeddings>),该字符嵌入维度为300,且每个维度的取值范围为[0,1]。然后,将嵌入好的样本组输入至GP-TNN进行训练。最后,模型生成输出的每组语义特征向量,分别记为Feature₁、Feature₂和Feature₃,以前两个特征向量为例,它们间的距离定义如下:

$$\text{dis}(F_1, F_2) = \alpha \times \frac{\| \text{Feature}_1 - \text{Feature}_2 \|}{1 + \| \text{Feature}_1 - \text{Feature}_2 \|} \quad (5)$$

式中: $\|\cdot\|$ 为欧几里得范数; α 为膨胀系数,用于控制距离的敏感度范围。

有效的损失函数能够显著提升模型训练的效果。传统的对比损失函数^[24]主要针对双输入情况,而在本研究中,损失计算涉及3个输入,使得传统对比损失函数难以满足需求。因此,本文提出了一种新型损失函数,能够同时计算SNN和PSNN的损失,从而推进模型训练过程。

$$\text{Loss} = \sum_{i=0}^N \frac{y(d_i'^2 + d_i''^2) + (1-y)[(1-d_i')^2 + (1-d_i'')^2]}{2N} \quad (6)$$

式中: d_i' 和 d_i'' 分别为SNN和PSNN中待匹配实体间的距离; N 为样本组的数量; y 为样本的类型, $y=1$ 表示正样本, $y=0$ 表示负样本。

当输入为正样本时,损失函数中只剩下 $\sum_{i=0}^N \frac{y(d_i'^2 + d_i''^2)}{2N}$,如果语义特征空间中实际相似的实体之间的距离较大,则说明当前模型未能有效捕捉语义相似性,因此损失值增加,以引导模型进行优化调整。

类似地,当输入为负样本时,损失函数中只剩下 $\sum_{i=0}^N \frac{(1-y)[(1-d_i')^2 + (1-d_i'')^2]}{2N}$,若实际不

相似的实体在语义特征空间中的距离较小,则损失值随之增大,以促使模型进一步区分不相似实体。最终,在达到给定的训练轮次后,GP-TNN模型训练结束。

2.3.2 通过GP-TNN进行本体匹配

在获得训练完成的GP-TNN后,需要对待匹配本体中的余下实体进行匹配。如图1黑色连接线引导的过程所示,根据确定性方法的对齐结果,从O₁和O₂中分别取出两个同类尚未匹配的待匹配实体Entity₁和Entity₂,首先从该对实体中提取两组标注属性(即,{label₁¹,label₁¹,comment₁¹}和{comment₁²,comment₁²,label₁²});其次对这些标注属性执行字符嵌入,并将其输入至训练完成的GP-TNN中进行语义学习,以生成语义特征向量;然后,通过式(7)计算语义特征向量之间的相似度值,随即两组输入得到4个相似度值: $s(\text{label}_1^1, \text{label}_1^1)$, $s(\text{label}_1^1, \text{comment}_1^2)$, $s(\text{comment}_1^2, \text{comment}_1^2)$ 和 $s(\text{comment}_1^2, \text{label}_1^2)$;最终,给定阈值,基于上述4个相似度值评估等价关系,选取其中最大值,若该

值大于阈值时,即判定该对待匹配实体为等价关系。

基于距离的实体相似度度量公式定义为:

$$\text{sim}(e_1, e_2) = 1 - \text{dis}(F_1, F_2) \tag{7}$$

式中: e_1 和 e_2 为两个待匹配实体; F_1 和 F_2 分别为它们的语义特征向量,语义特征向量在空间上越接近,相似度越高,相似度值越大。

基于 GP-TNN 的传感器本体匹配方法的实现见算法 1,其中 i 和 j 分别表示本体 O_1 和 O_2 中待匹配实体的索引。

算法 1 基于 GP-TNN 的传感器本体匹配

Input: 待匹配实体的标注属性:label 和 comment

Output: 最终对齐结果

1. 移除停用词并进行大小写标准化;
2. 获取包含两类标注属性集合的集合 E, 包括
{label₂ⁱ, label₁ⁱ, comment₂ⁱ} 和
{comment₂^j, comment₁^j, label₂^j};
3. E 中的每个标注属性集合均通过字符嵌入转换为数字
向量表示;
4. for i=0; i<O₁.size(); i++ do
5. for j=0; j<O₂.size(); j++ do
6. 通过 GP-TNN 获取每个标注属性的特征向量;
7. 根据式(5)和(7), 计算 E 中每组标注属性的特征
向量间的相似度值;
8. end for
9. end for
10. 为每组实体的标注属性集合分别构建两个相似度矩
阵;
11. 对每个相似度矩阵进行筛选, 合并各矩阵中的对齐结
果, 生成最终的对齐结果。

3 实验结果与分析

实验中,首先采用 OAEI 提供的 Benchmark 测试集验证模型的有效性(具体参见 <https://oaei.ontologymatching.org>),该测试集涵盖多种本体异质类型,能够全面评估本文方法的匹配效果。Benchmark 数据集中的每个案例均包含一个目标本体及其参考对齐。表 1 将案例按照异质情况分成了 4 类,并对每类的实体数量和对齐数量进行了汇总。

为评估本文方法的匹配性能,选取传统本体匹配方法及 OAEI 中经典的本体匹配方法进行对比实验。传统本体匹配方法包括 Max 方法、Min 方法和 Average 方法。以 Max 方法为例,对于一对待匹配实体,Max 方法从式(4)中计算得到的 3

表 1 Benchmark 测试集中不同案例的简要描述
Table 1 Brief descriptions of different kinds of cases in Benchmark

案例	异质类型	实体数量	对齐数量
101	源本体	109	97
201—202	不同的词汇和语言特征	1 090	970
221—247	不同的结构特征	1 538	1 194
248—262	不同的词汇、语言和结构特征	5 610	4 325
总计		8 347	6 586

种相似度度量值中选择最大值作为该实体对的相似度值,如果该值大于设定阈值,则判定该实体对等价。OAEI 中经典的本体匹配方法包括 AML^[25]、LogMap 系列(LogMap、LogMapLt 和 LogMapBio)^[26]以及 XMap^[27]。各方法的简要介绍如下:

(1)AML:一种自动化本体匹配系统,主要基于词汇匹配技术,并强调利用外部资源作为背景知识。

(2)LogMap:一种可扩展的基于逻辑推理的本体匹配系统,支持实时用户交互以优化匹配过程。

(3)LogMapLt:LogMap 的轻量级变体,主要采用高效的字符串匹配技术。

(4)LogMapBio:LogMap 的另一变体,扩展了 BioPortal 作为动态中介本体,取代了对少量预选本体的依赖。

(5)XMap:通过 UMLS 和 WordNet 计算两个本体实体之间的相似程度,并结合词汇和上下文结构进行匹配。

此外,为了验证本文方法的适用性,使用 3 个真实的传感器本体进行测试:CSIRO、MMI 和 SSN (具体参见 <https://www.w3.org/2005/Incubator/ssn/wiki/images/4/42/SensorOntology20090320.owl.xml>(CSIRO);<https://mmisw.org/ont/mmi/device>(MMI);<https://www.w3.org/2005/Incubator/ssn/ssnx/ssn>(SSN))。其中,CSIRO 涉及环境传感器设备、监测过程与观测属性建模,强调设备配置与观测上下文之间的结构关系;MMI 描述海洋观测平台、传感器类型与测量参数,概念抽象程度较高,适用于遥感与多源观测数据的语义组织;SSN 涵盖与传感器相关的描述、概念及部分观测数据。相关匹配任务包括 CSIRO-MMI、CSIRO-SSN 和 MMI-SSN。

实验采用 1:1 对齐策略,即 O_1 中的每个实体只能与 O_2 中的一个实体匹配,反之亦然。各测试结果为 30 次独立测试的算数平均值。评价指标采用本体匹配领域常用的查准率(Precision)、查全率(Recall)和 F_1 值(F -measure),用于衡量方法在识别匹配实体对方面的准确性和完整性表现。

图 3 为本文方法在 Benchmark 中不同测试案例上的评价结果,由图 3 可知,在 202 和 248~262 号测试案例中,Recall 明显较低,其主要原因在于这些测试案例中的实体未包含 comment,导致模型无法从第二类标注属性(comment₂, comment₁, label₂)中提取语义信息,从而降低了 Recall。

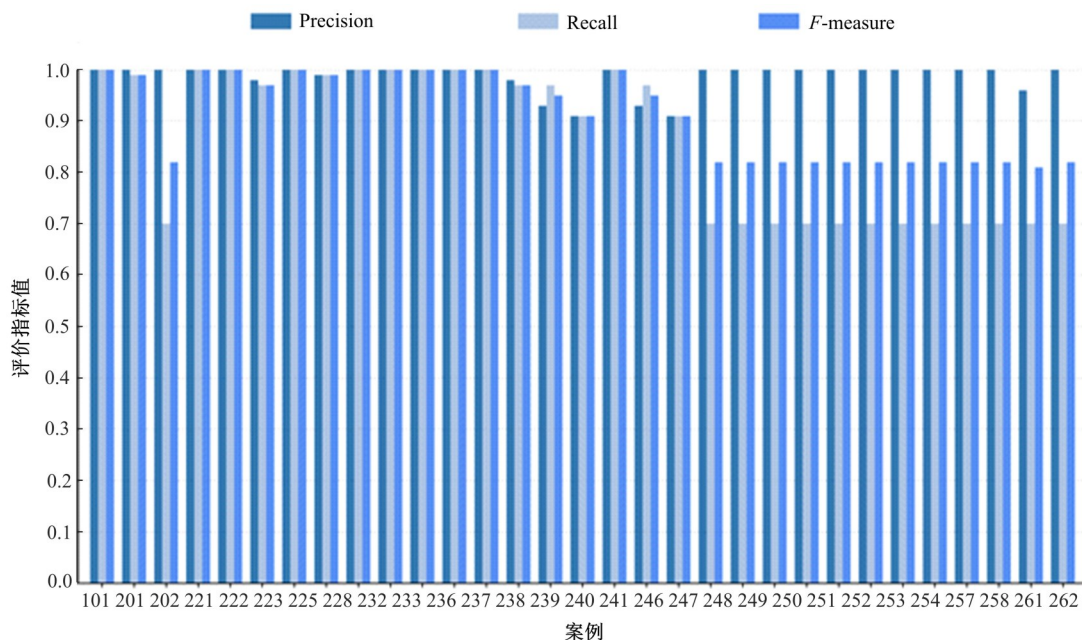


图 3 本文方法在 Benchmark 案例集上的评价结果

Fig. 3 Evaluation results of proposed method on benchmark

图 4 为传统方法与本文方法在 F -measure 上的对比。由图 4 可见,本文方法在 4 类案例上均优于传统的相似度度量方法。在 221~262 号测试案例中,Min 方法和 Average 方法均未能生成任何对齐结果,其原因在于不同的相似度度量方法在不同类型的异质问题上存在适应性差异,这两种方法分别忽略及削弱了由有效相似度度量方法计算得到的相似度值,导致最终的实体相似度值显著低于设定阈值,从而被过滤掉。虽然 Max 方法在 3 种传统方法中表现最佳,但其忽略了其他

相似度度量方法的参考价值,在复杂的匹配任务中仍然难以生成高质量的对齐结果。

表 2 为本文方法同 OAEI 中经典方法在 Benchmark 上关于 Precision 的对比结果。由表 2 可见,本文方法在 74.19% 异质案例中均获得了最高的值。尽管 AML 在平均 Precision 上比本文方法高 0.01,但其依赖于超过 10 种相似度度量方法作为约束,这些度量方法之间相互抑制,从而导致 Recall 下降,并会在调整相似度度量参数及调试匹配结果时增加计算复杂度。表 3 对比了各方法在 Recall 方面的表现,结果表明,本文方法在平均 Recall 上优于其他 4 种方法,仅略低于 LogMapLt。其主要原因在于本研究采用了较高的过滤阈值以确保对齐的 Precision,从而限制了 Recall 的进一步提升。

表 4 展示了各方法的综合对比结果。数据表明,本文方法在 F -measure 均值上达到了最高值,充分验证了其在处理各种异质案例方面的有效性。

图 5~图 7 分别展示了 4 种具有代表性的前

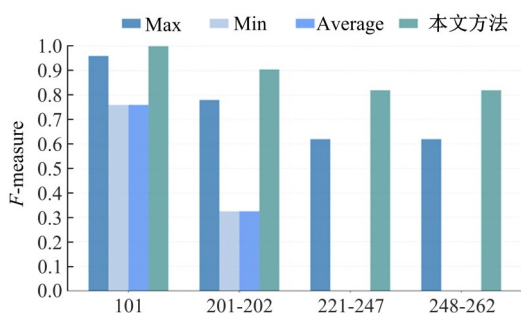


图 4 本文方法与传统匹配方法在 F -measure 上的比较

Fig. 4 Comparison of proposed method and traditional matching methods in terms of F -measure

表 2 本文方法与 OAEI 中经典方法在 Benchmark 数据集上的 Precision 比较

Table 2 Comparison of proposed method with the classical methods in OAEI on benchmark in terms of Precision

案例	AML	LogMap	LogMapLt	LogMapBio	XMap	本文方法	案例	AML	LogMap	LogMapLt	LogMapBio	XMap	本文方法
101	1.00	0.94	0.56	0.50	0.94	1.00	241	1.00	0.85	0.67	0.66	1.00	1.00
201	1.00	0.96	0.51	0.49	0.93	1.00	246	1.00	0.85	0.67	0.66	1.00	0.93
202	1.00	0.95	0.51	0.50	0.93	1.00	247	1.00	0.85	0.67	0.66	1.00	0.91
221	1.00	0.94	0.56	0.50	0.94	1.00	248	1.00	0.94	0.51	0.47	0.94	1.00
222	1.00	0.00	0.57	0.00	0.94	1.00	249	1.00	0.94	0.81	0.50	0.93	1.00
223	1.00	0.94	0.56	0.50	0.94	0.98	250	1.00	0.90	0.27	0.66	1.00	1.00
224	1.00	0.94	0.83	0.50	0.94	1.00	251	1.00	0.00	0.51	0.00	0.93	1.00
225	1.00	0.94	0.56	0.50	0.94	0.99	252	1.00	0.96	0.51	0.51	0.95	1.00
228	1.00	0.85	0.32	0.66	1.00	1.00	253	1.00	0.97	0.80	0.49	0.93	1.00
232	1.00	0.94	0.83	0.50	0.94	1.00	254	1.00	0.86	0.27	0.64	1.00	1.00
233	1.00	0.85	0.32	0.66	1.00	1.00	257	1.00	0.88	0.64	0.64	1.00	1.00
236	1.00	0.85	0.67	0.66	1.00	1.00	258	1.00	0.00	0.82	0.00	0.93	1.00
237	1.00	0.00	0.84	0.00	0.94	1.00	259	1.00	0.94	0.81	0.49	0.94	1.00
238	1.00	0.94	0.83	0.50	0.94	0.98	261	1.00	0.86	0.27	0.64	1.00	0.96
239	1.00	0.85	0.32	0.66	1.00	0.93	262	1.00	0.87	0.63	0.65	1.00	1.00
240	1.00	0.85	0.32	0.66	1.00	0.91	平均	1.00	0.79	0.58	0.50	0.96	0.99

表 3 本文方法与 OAEI 中经典方法在 Benchmark 数据集上的 Recall 比较

Table 3 Comparison of proposed method with classical methods in OAEI on benchmark in terms of Recall

案例	AML	LogMap	LogMapLt	LogMapBio	XMap	本文方法	案例	AML	LogMap	LogMapLt	LogMapBio	XMap	本文方法
101	0.00	0.96	0.99	0.56	1.00	1.00	241	1.00	1.00	1.00	1.00	1.00	1.00
201	0.28	0.76	0.78	0.44	0.64	0.99	246	1.00	1.00	1.00	1.00	1.00	0.97
202	0.28	0.76	0.78	0.44	0.65	0.70	247	1.00	1.00	1.00	1.00	1.00	0.91
221	0.34	0.96	0.99	0.56	1.00	1.00	248	0.29	0.74	0.78	0.42	0.68	0.70
222	0.33	0.00	0.99	0.00	0.67	1.00	249	0.27	0.75	0.79	0.44	0.72	0.70
223	0.34	0.95	0.99	0.56	1.00	0.97	250	0.83	0.79	0.79	0.79	0.79	0.70
224	0.34	0.95	0.99	0.56	1.00	1.00	251	0.27	0.00	0.78	0.00	0.45	0.70
225	0.34	0.96	0.99	0.56	1.00	0.99	252	0.28	0.76	0.79	0.45	0.65	0.70
228	1.00	1.00	1.00	1.00	1.00	1.00	253	0.27	0.76	0.78	0.44	0.62	0.70
232	0.34	0.95	0.99	0.56	1.00	1.00	254	0.83	0.79	0.79	0.79	0.79	0.70
233	1.00	1.00	1.00	1.00	1.00	1.00	257	0.82	0.79	0.79	0.79	0.79	0.70
236	1.00	1.00	1.00	1.00	1.00	1.00	258	0.28	0.00	0.79	0.00	0.46	0.70
237	0.33	0.00	0.99	0.00	0.67	1.00	259	0.28	0.75	0.78	0.44	0.71	0.70
238	0.34	0.96	0.99	0.56	1.00	0.97	261	0.83	0.79	0.79	0.79	0.79	0.70
239	1.00	1.00	1.00	1.00	1.00	0.97	262	0.83	0.79	0.79	0.79	0.79	0.70
240	1.00	1.00	1.00	1.00	1.00	0.91	平均	0.56	0.77	0.90	0.61	0.83	0.86

沿传感器本体匹配方法,即 ASMOV^[28]、CODI^[29]、SOBOM^[30]和 FuzzyAlign^[31],与本文方法在 Precision、Recall 和 *F*-measure 上的比较。从图 7 中可以明显观察到,本文方法是解决传感器数据异质性问题最具竞争力的方法,其在 3 个传感器匹配任务中的 *F*-measure 均优于其他 4 种传感器本体匹配方法。排名第 2 的是 FuzzyAlign,如图 5

所示,其在 Precision 方面较 ASMOV、CODI 和 SOBOM 具有显著优势,这主要得益于其在匹配过程中综合考虑了实体的语义特征、语言特性以及结构化信息等多个维度。然而,多种相似度度量方法的引入也产生彼此冲突的匹配结果,从而限制了其在 Recall 方面的提升。

本文方法在 Precision 均值上略逊于 Fuzzy-

表 4 本文方法与 OAEI 中经典方法在 Benchmark 数据集上的 F -measure 比较Table 4 Comparison of proposed method with classical methods in OAEI on benchmark in terms of F -measure

案例	AML	LogMap	LogMapLt	LogMapBio	XMap	本文方法	案例	AML	LogMap	LogMapLt	LogMapBio	XMap	本文方法
101	0.00	0.95	0.71	0.52	0.97	1.00	241	1.00	0.92	0.80	0.80	1.00	1.00
201	0.44	0.85	0.62	0.47	0.76	0.99	246	1.00	0.92	0.80	0.80	1.00	0.95
202	0.43	0.85	0.62	0.47	0.77	0.82	247	1.00	0.92	0.80	0.80	1.00	0.91
221	0.51	0.95	0.72	0.52	0.97	1.00	248	0.44	0.82	0.62	0.44	0.79	0.82
222	0.50	0.00	0.72	0.00	0.78	1.00	249	0.43	0.84	0.80	0.47	0.81	0.82
223	0.51	0.94	0.72	0.53	0.97	0.97	250	0.91	0.84	0.40	0.71	0.88	0.82
224	0.51	0.94	0.90	0.53	0.97	1.00	251	0.43	0.00	0.62	0.00	0.61	0.82
225	0.51	0.95	0.72	0.52	0.97	0.99	252	0.43	0.85	0.62	0.48	0.77	0.82
228	1.00	0.92	0.48	0.80	1.00	1.00	253	0.43	0.85	0.79	0.47	0.74	0.82
232	0.51	0.94	0.90	0.53	0.97	1.00	254	0.90	0.82	0.40	0.70	0.88	0.82
233	1.00	0.92	0.48	0.80	1.00	1.00	257	0.90	0.83	0.70	0.71	0.88	0.82
236	1.00	0.92	0.80	0.80	1.00	1.00	258	0.44	0.00	0.80	0.00	0.62	0.82
237	0.50	0.00	0.91	0.00	0.78	1.00	259	0.44	0.84	0.80	0.47	0.81	0.82
238	0.51	0.95	0.90	0.52	0.97	0.97	261	0.90	0.82	0.40	0.70	0.88	0.81
239	1.00	0.92	0.48	0.80	1.00	0.95	262	0.90	0.83	0.70	0.71	0.88	0.82
240	1.00	0.92	0.48	0.80	1.00	0.91	平均	0.66	0.77	0.68	0.54	0.88	0.91

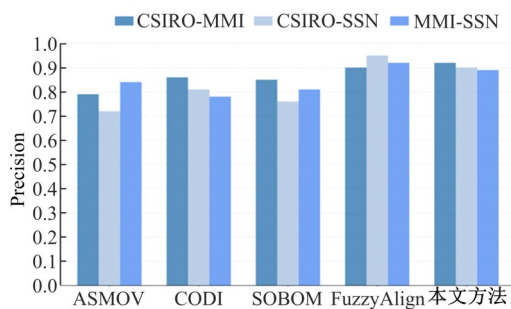


图 5 与前沿的传感器本体匹配方法在 Precision 上的比较

Fig. 5 Comparison with state-of-the-art sensor ontology matching methods in terms of Precision

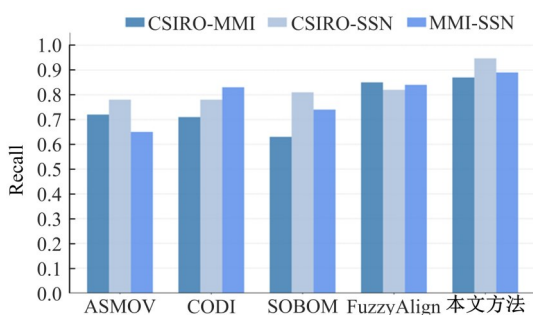
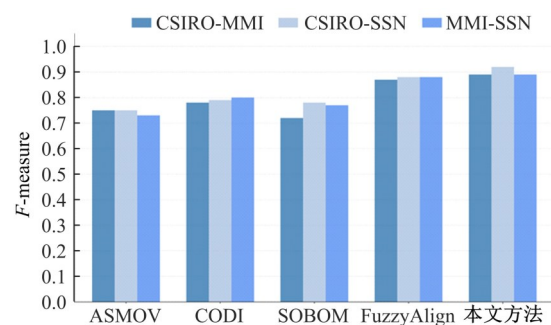


图 6 与前沿的传感器本体匹配方法在 Recall 上的比较

Fig. 6 Comparison with state-of-the-art sensor ontology matching methods in terms of Recall

Align, 这归因于核心实体集中存在假阳性匹配结果, 这部分错误的结果来源于 Wu&Palmer 方法在构造训练数据时引入的语义误差, 使得部分不应

匹配的实体对被错误识别为匹配对, 从而导致 Precision 下降。尽管如此, 本文通过确定性方法率先明确实体间的浅层局部信息和语义关联, 并引入 GP-TNN 进一步增强对两类标注属性关系的深层语义特征挖掘能力, 在保持 Recall 优势的同时, 有效缓解了训练数据中的语义偏差对 Precision 造成的影响, 从而实现更为均衡的匹配性能。

图 7 与前沿的传感器本体匹配方法在 F -measure 上的比较Fig. 7 Comparison with proposed and state-of-the-art sensor ontology matching methods in terms of F -measure

4 结束语

本文提出了一种基于 GP-TNN 的传感器本体匹配方法, 旨在解决 IoT 系统中的传感器本体异质问题, 从而提升系统的互操作性。该方法利

用确定性的混合相似度度量方法率先明确实体间的浅层局部信息和语义关联,初步覆盖高置信度的实体匹配对,在提供高质量训练数据的同时为提升 Recall 奠定基础;接着,引入 GP-TNN 对同类和异类标注属性进行深层次语义特征挖掘,进一步优化语义表示能力,从而实现更加全面且精准的本体匹配效果。为验证本文方法的有效性,在 Benchmark 数据集上将其与传统本体匹配方法及 OAEI 中的经典本体匹配方法进行对比,并在 3 组真实传感器本体匹配任务上进行适用性验证,结果表明,该方法能够有效完成各项本体匹配任务,这为解决 IoT 场景下的语义互操作难题提供了有效途径,也为异质语义信息融合提供了坚实的方法支撑。未来,为解决因标注属性缺失导致的语义信息不可获取问题,本文将进一步从语言学和结构层面深入挖掘实体的语义特征,研究基于语言学和结构的深度学习本体匹配技术。在此基础上,结合元匹配理论,对不同深度学习本体匹配技术的匹配结果进行加权融合,以提高最终匹配结果的质量。此外,本文将进一步提取实体的关键语义元素(涵盖术语、语言学和结构),并在同一实体特征向量中进行融合,以提升匹配效率,同时增强匹配结果的准确性和覆盖度,确保不同异质类型的传感器本体能够有效对齐。

参考文献:

- [1] Abdulazeez D H, Askar S K. A novel offloading mechanism leveraging fuzzy logic and deep reinforcement learning to improve IoT application performance in a three-layer architecture within the fog-cloud environment[J]. IEEE Access, 2024, 12: 39936-39952.
- [2] Chen X, Cao B, Ahmad I. Lightweight neural network-based viewport prediction for live VR streaming in wireless video sensor network[J]. Mobile Information Systems, 2021, 2021(1): 1-12.
- [3] Nivetha A, Preetha K S. A review on cyber-twin in sixth generation wireless networks: architecture, research challenges & issues[J]. Wireless Personal Communications, 2024, 138(3): 1815-1865.
- [4] Geng A, Lü Q. A multi-objective particle swarm optimization with density and distribution-based competitive mechanism for sensor ontology meta-matching[J]. Complex & Intelligent Systems, 2023, 9(1): 435-462.
- [5] Hamane Z, Samih A, Fennan A. Ontology matching using deep learning[J]. Journal of Theoretical and Applied Information Technology, 2023, 101(10): 4041-4058.
- [6] Bento A, Zouaq A, Gagnon M. Ontology matching using convolutional neural networks[C]//Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, 2020: 5648-5653.
- [7] Jiang C, Xue X. Matching biomedical ontologies with long short-term memory networks[C]//IEEE International Conference on Bioinformatics and Biomedicine, Seoul, Korea, 2020: 2484-2489.
- [8] Iyer V, Agarwal A, Kumar H. VeeAlign: a supervised deep learning approach to ontology alignment [C]//Proceedings of the 15th International Workshop on Ontology Matching, Athens, Greece, 2020: 216-224.
- [9] Xue X, Jiang C, Zhang J, et al. Matching sensor ontologies through Siamese neural networks without using reference alignment[J]. PeerJ Computer Science, 2021, 7: No. e602.
- [10] Wang P, Zou S, Liu J, et al. Matching biomedical ontologies with GCN-based feature propagation[J]. Mathematical Biosciences and Engineering, 2022, 19(8): 8479-8504.
- [11] 段阳, 侯力, 冷松. 金属切削加工知识图谱构建及应用[J]. 吉林大学学报: 工学版, 2021, 51(1): 122-133.
- [12] Duan Yang, Hou Li, Leng Song. Building and application of metal cutting knowledge graph[J]. Journal of Jilin University(Engineering and Technology Edition), 2021, 51(1): 122-133.
- [13] Xue X, Yang C, Jiang C, et al. Optimizing ontology alignment through linkage learning on entity correspondences[J]. Complexity, 2021, 2021(1): 1-12.
- [14] Xue X, Wang Y. Improving the efficiency of NSGA-II based ontology aligning technology[J], Data & Knowledge Engineering, 2017, 108: 1-14.
- [15] Chu S C, Xue X, Pan J S, et al. Optimizing ontology alignment in vector space[J]. Journal of Internet Technology, 2020, 21 (1): 15-22.
- [16] Mascardi V, Locoro A, Rosso P. Automatic ontology matching via upper ontologies: a systematic evaluation[J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 22(5): 609-623.
- [17] Lu J, Xue X, Lin G, et al. A new ontology meta-matching technique with a hybrid semantic similarity

- measure[C]//Advances in Intelligent Information Hiding and Multimedia Signal Processing, Changchun, China, 2020: 37-45.
- [17] Phan T T M, Duong A T, Tran P T M. WordNet-based computation of semantic similarity between two Vietnamese nouns[J]. *Technium Social Sciences Journal*, 2023, 43: 532-545.
- [18] Yu Y, Zhou B, Chen L, et al. Identifying important nodes in complex networks based on node propagation entropy[J]. *Entropy*, 2022, 24(2): 275-291.
- [19] Chicco D. Siamese neural networks: an overview[J]. *Artificial Neural Networks*, 2021, 2190: 73-94.
- [20] Ni W, Xu Q, Jiang Y, et al. PSNEA: Pseudo-siamese network for entity alignment between multi-modal knowledge graphs[C]//Proceedings of the 31st ACM International Conference on Multimedia, New York, USA, 2023: 3489-3497.
- [21] Li Y, Chen C L P, Zhang T. A survey on siamese network: methodologies, applications, and opportunities[J]. *IEEE Transactions on Artificial Intelligence*, 2022, 3(6): 994-1014.
- [22] 王神龙, 雍宇, 吴晨睿. 基于伪孪生神经网络的低纹理工业零件6D位姿估计[J]. *电子学报*, 2023, 51(1): 192-201.
Wang Shen-long, Yong Yu, Wu Chen-rui. 6D pose estimation of low texture industrial parts based on Pseudo-Siamese neural network[J]. *Acta Electronica Sinica*, 2023, 51(1): 192-201.
- [23] Gong Y, Yue Y, Ji W, et al. Cross-domain few-shot learning based on Pseudo-Siamese neural network[J]. *Scientific Reports*, 2023, 13(1): 1-14.
- [24] Xue X, Jiang C, Zhu H. Matching Ontologies Through Siamese Neural Network[C]//Mobile Multimedia Communications: 14th EAI International Conference, Online, 2021: 715-724.
- [25] Faria D, Pesquita C, Balasubramani B S, et al. OAEI 2016 results of AML[C]//Proceedings of the 11th International Workshop on Ontology Matching co-located with the 15th International Semantic Web Conference, Kobe, Japan, 2016: 138-145.
- [26] Jiménez-Ruiz E, Grau B C, Cross V. LogMap family participation in the OAEI 2017[C]//Proceedings of the 12th International Workshop on Ontology Matching Co-located with the 16th International Semantic Web Conference, Vienna, Austria, 2017: 153-157.
- [27] Djeddi W E, Khadir M T, Yahia S B. XMap results for OAEI 2016[C]//Proceedings of the 11th International Workshop on Ontology Matching co-located with the 15th International Semantic Web Conference, Kobe, Japan, 2016: 222-226.
- [28] Jean-Mary Y R, Shironoshita E P, Kabuka M R. Ontology matching with semantic verification[J]. *Journal of Web Semantics*, 2009, 7(3): 235-251.
- [29] Noessner J, Niepert M, Meilicke C, et al. Leveraging terminological structure for object reconciliation [C]//Extended Semantic Web Conference, Heraklion, Greece, 2010: 334-348.
- [30] Xu P, Wang Y, Cheng L, et al. Alignment results of SOBOM for OAEI 2010[C]//Proceedings of the 5th International Workshop on Ontology Matching, Shanghai, China, 2010: 203-211.
- [31] Fernandez S, Marsa-Maestre I, Velasco J R, et al. Ontology alignment architecture for semantic sensor web integration[J]. *Sensors*, 2013, 13(9): 12581-12604.