

知识图谱精化研究综述*

谢刚^{1,2}

(1.中国科学院自动化所,北京 100080;2.贵州师范大学 大数据与计算机科学学院,贵州 贵阳 550001)

摘要: 知识图谱是一种在移动互联网大时代下产生的新型知识表示方法,而精化是知识图谱应用研究的主要内容之一,其主要任务是知识图谱补全和错误检测等,在信息检索、机器人、智能问答等领域有着重要的应用前景。因此,对知识图谱精化进行研究具有十分重要的意义。对当前知识图谱精化方法进行了较为全面、深入的总结,并对知识图谱未来的主要研究方向进行了展望。

关键词: 知识图谱;精化;补全;错误检测

中图分类号: TP391

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.180696

中文引用格式: 谢刚. 知识图谱精化研究综述[J]. 电子技术应用, 2018, 44(9): 29-33, 38.

英文引用格式: Xie Gang. Review of knowledge graph refinement[J]. Application of Electronic Technique, 2018, 44(9): 29-33, 38.

Review of knowledge graph refinement

Xie Gang^{1,2}

(1. Institute of Automation, Chinese Academy of Science, Beijing 100080, China;

2. School of Big Data and Computer Science, Guizhou Normal University, Guiyang 550001, China)

Abstract: Knowledge graph is a new knowledge representation method in the mobile Internet era. Knowledge graph refinement is one of the main contents in application of knowledge graph and its main goal are completion and error detection. At the same time, knowledge graph refinement is usually applied in areas of the information retrieval, robot, intelligent answer and so on. At result, studies on knowledge graph refinement had great significance. This paper addresses the methods of knowledge graph refinement and finally presents challenges of know-ledge graph.

Key words: knowledge graph; refinement; completion; error detection

0 引言

随着链接开放数据源(如 DBpedia)的出现以及谷歌在 2012 年提出知识图谱的概念,全球掀起了研究知识图谱的热潮,涌现出了大量的知识图谱构建技术^[1-5],并构建了各种知识图谱,这些知识图谱要么是开放的,要么是公司私有的,如 Freebase^[2]、维基数据(Wikidata)^[3]、DBpedia^[4]、YAGO^[5]等,但无论采用哪种技术,构造出来的知识图谱都不完美^[6]。随着研究的深入,越来越多的研究者开始关注知识图谱的覆盖率和正确率。而提高知识图谱的覆盖率和正确率是知识图谱精化的主要目的,对知识图谱进行精化具有十分重要的意义。

近年来,该领域的研究进展非常迅速,涌现出了一大批研究成果,已经研发出了多种知识图谱精化方法,这些方法主要集中在讨论知识图谱补全^[7-28]和知识图谱错误探测^[29-34]两个方面,这也是本文从这两个方面进行综述的原因。

本文的贡献是:(1)讨论各种知识图谱精化方法;(2)

为国内同行介绍本领域的最新研究成果,了解该领域的研究进展,从而推动我国在该领域的发展。

1 知识图谱精化相关概念

1.1 知识图谱的概念

“知识图谱”是一种描述真实世界客观存在的实体、概念及它们之间关联关系的语义网络。可以利用知识图谱开发语义检索和自动问答等应用^[1]。知识图谱的结构如图 1 所示。可见,知识图谱是一个有向图,由模式(schema)图和数据图构成。其中,模式图描述类之间的关系;数据图描述实体之间的关系。图 1 描述的知识(事实)如下:

- (1)李四是一个教师
- (2)北京是一个城市
- (3)中国是一个国家
- (4)李四的出生地为北京
- (5)北京位于中国
- (6)李四的国籍是中国

1.2 知识图谱构建与知识图谱精化

知识图谱构建是使用各种技术从无到有构造知识

* 基金项目:国家自然科学基金(61309006,61563011)

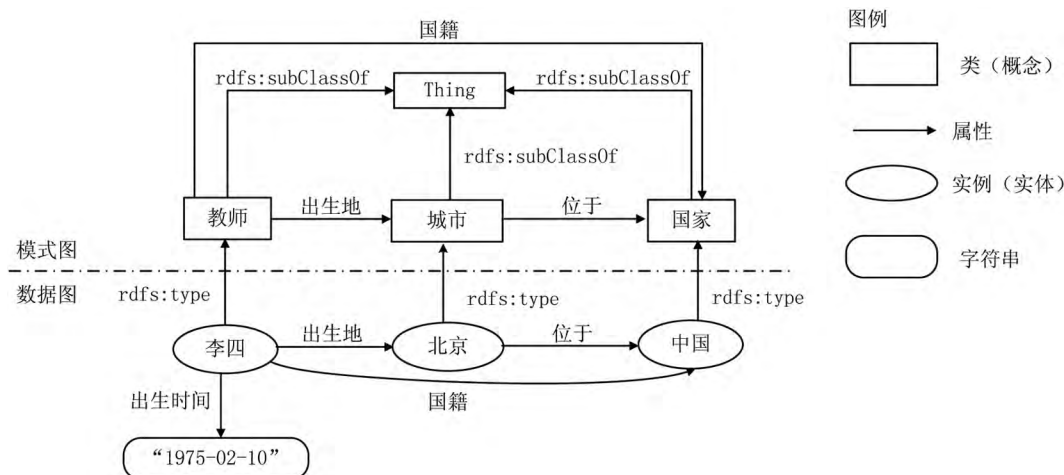


图1 知识图谱示意图

图谱,而知识图谱精化是使用各种技术对知识图谱进行完善。可见,要构建一个完美的知识图谱,需要经过多个精化步骤。因此,知识图谱构建和知识图谱精化是相辅相成、不可分割的。另外,本文将关系、文字和类型称为精化目标。

2 常用的知识图谱补全方法

知识图谱补全的目的是利用已有信息,预测丢失的实体、类型和实体间的关系,从而提高知识图谱的覆盖率。它是知识图谱精化的主要任务之一,其对应的精化目标包括实体、类型和实体间的关系。但根据已有文献,发现目前该方面的研究主要集中在对类型和实体间的关系进行精化。

本节根据知识图谱补全使用的数据源,将知识图谱补全方法分为知识图谱内部补全和知识图谱外部补全两大类。其中,知识图谱内部补全方法是指仅使用知识图谱本身预测丢失信息的方法总称,知识图谱外部补全方法是指除使用知识图谱本身以外,还使用其他数据源(如文本语料)来预测丢失信息的方法总称。下面将从这两个方面对知识图谱错误探测进行综述。

2.1 知识图谱内部补全方法

为了揭示内部补全方法因精化目标的不同而不同,本小节将根据精化目标的不同,把内部补全方法分成实体类型内部补全和关系内部预测两类进行综述。

2.1.1 实体类型内部补全

实体类型内部补全就是利用知识图谱本身已有的实体、实体类型和实体关系预测丢失的实体类型。

在机器学习领域,常用多分类方法对实体类型进行补全。其中,PAULHEIM H等人^[7-8]提出了一种基于条件概率的补全算法 SDType,这种算法的思想是通过实体所具有的关系预测实体类型。SDType 算法的评价矩阵是正确率(precision)、召回率和新增类型数目。但这种算法的缺点是假设关系之间是相互独立的,而现实世界中这种假设在很多情况下是不成立的,并且该算法没有用类

型的层次结构。利用 SDType 算法,已经为知识图谱 DBpedia 新增了 3.4 亿条类型语句。KROMPAB D 等人^[9]利用张量分解预测实体类型,这种方法的思想是把知识图谱表示成一个实体-实体-关系的三维张量,然后通过张量分解的方法实现类型补全。该方法的评价矩阵是正确率、召回率和正确率-召回率曲线。张香玲等人^[10]提出了一种由谓词和谓词及谓词和类型的相互作用补全实体类型的模型,在该模型中,为了解决类型语义漂移,使用 PMI 技术设计一个有效的谓词-类型推理图及基于图上的随机游走算法。该模型的评价矩阵是正确率和召回率。SLEEMAN J 等人^[11]将主题模型用在关系预测中,这种方法的思想是首先将实体表示成文档,应用 LDA 抽取文档的主题,然后通过分析主题和实体类型的共现关系,根据分析结果,将实体类型指派给主题对应的实体。该方法的评价矩阵是正确率和召回率。

在数据挖掘领域,利用关联规则预测知识图谱丢失的信息。PAULHEIM H 等人^[12]基于数据冗余信息使用关联规则来预测 DBpedia 中丢失的类型。这种方法的评价矩阵为正确率和增加的类型数。

2.1.2 关系内部预测

按照相同的思路,在机器学习领域,也把预测关系的存在与否看成是一个二分类问题。其中,SOCHER R 等人^[13]提出一种通过训练张量神经网络预测新关系的方法。例如:如果一个人出生在德国,那么该方法就能根据这个关系预测他的国籍是德国。这种方法的评价矩阵是精确率(accuracy),已被用于 Freebase 和 WordNet 中。BAIER S 等人^[14]也提出了类似的方法,但他们在预测过程增加了模式知识,以提高关系预测的性能。不同的是该方法的评价矩阵是正确率-召回率曲线面积和 ROC 曲线面积。类似地,ZHAO Y 等人^[15]通过将关系嵌入到一个低维空间中预测 Freebase 中关系的存在,这种方法的评价矩阵是正确率。

同样地,在数据挖掘领域,将关联规则挖掘也用于预

测关系。其中, KIM J 等人^[16]提出了一种利用关联规则预测 DBpedia 中实体关系的方法。这种方法只能预测来自于维基百科分类中的实体关系, 其评价矩阵是正确率和增加的关系数目。KOLTHOFF C 等人^[17]利用关联规则挖掘思想查找意义丰富的关系链来预测关系, 该方法的评价矩阵是正确率和召回率。

2.2 知识图谱外部补全方法

与知识图谱外部补全方法类似, 为了揭示外部补全方法因精化目标的不同而不同, 本小节将根据精化目标的不同, 把外部补全方法分成实体类型外部补全和关系外部预测两类进行综述。

2.2.1 实体类型外部补全

实体类型外部补全就是利用知识图谱本身和外部数据来预测丢失的实体类型。根据已有文献分析, 实体类型外部补全方法的研究主要集中在机器学习和自然语言处理领域。

在机器学习领域, 主要将外部数据表示成实体特征进行分类。因为维基百科页之间的链接没有约束, 所以维基百科网页之间的链接比知识图谱中相应实体的链接要多。因此, NUZZOLESE A G 等人^[18]利用维基百科链接图和 KNN 分类算法来预测知识图谱中的实体类型。如果一个知识图谱包含到维基百科的链接, 那么就以相关页的分类为基础, 将维基百科网页之间的链接表示成特征向量, 这种方法的评价矩阵是正确率和召回率。APRIOSIO A P 等人^[19]将 DBpedia 各种语言版本中的实体类型作为特征来预测丢失的类型, 该方法使用不同距离公式的 K-NN 分类器, 综合应用这些不同的距离公式, 得到了最好的结果。这种方法的评价矩阵是正确率和召回率。SLEEMAN J 等人^[20]将支持向量机用于 DBpedia 和 Freebase 中的实体类型预测。为了提高覆盖率和正确率, 作者利用知识图谱间的内部链接和其他知识图谱的属性对知识图谱实例进行分类, 这种方法的评价矩阵为正确率和召回率。

在自然语言处理领域, KLIEGR T^[21]等人使用了不同语言的摘要来进行实体类型预测, 从而大大提高知识图谱的覆盖率和正确率, 这种方法的评价矩阵是正确率和召回率。

2.2.2 关系外部预测

关系外部预测就是利用知识图谱本身和外部数据来预测丢失的实体关系。

一部分研究者利用远程监督法和自然语言处理方法对大规模文本语料库进行处理以预测实体关系, 其思路为: 首先, 通过命名实体识别将知识图谱中的实体链接到语料库(如维基百科)中; 然后, 以知识图谱已有的关系为基础, 找到与关系对应的文本模式, 例如, “author”关系对应的文本模式为“Y’s book X”; 最后, 利用已找到的文本模式去发现语料库中的新关系。其中, APRIOSIO A P 等人^[22]将远程监督法用于预测 DBpedia 中的关系, 该方

法将维基百科作为语料库, 并且将正确率和召回率作为评价矩阵。GERBER D 等人^[23]也提出了类似的方法, 并开发了一个 RdfLiveNews 原型。在该原型中, 利用新闻的 RSS 来解决 DBpedia 的时效性, 即判断预测到的新关系在 DBpedia 中属于过时的关系还是丢失的关系。这种方法使用的评价矩阵是正确率、召回率和精确率。

一部分研究者利用 Web 搜索引擎填充知识图谱^[24]。和上述研究类似, 这种方法首先找到关系对应的词汇, 然后使用这些词汇形成搜索语句以填充丢失的关系值。显然, 该方法使用整个网络作为语料库, 并使用信息提取和抽取技术进行知识图谱的补全。这种方法使用的评价矩阵是正确率、召回率和排名。

一部分研究者直接从网站的表格中抽取关系^[25-26]。其中, HOGAN A 等人^[25]提出一种从维基百科表格中抽取关系的方法。他们认为维基百科表格中共存的两个实体共享知识图谱中的一条边, 为了补全这些边, 首先使用已有关系从表格中抽取候选实体集, 然后对候选实体子集进行标注, 最后基于已标注的候选实体子集, 使用分类算法来识别知识图谱中真正成立的关系, 这种方法使用的评价矩阵是正确率和召回率。RITZE D 等人^[26]将上述方法扩展到任意的 HTML 表格中, 该方法的不足是不仅要求表的列必须与 DBpedia 本体中的属性匹配, 而且要求行也要与 DBpedia 中的实体匹配。这种方法使用的评价矩阵是正确率和召回率。

一些研究者认为许多自动构建的知识图谱包含很多到其他知识图谱的链接, 可以利用这些链接对知识图谱进行融合。其中, DUTTA A 等人^[27]提出一种在知识图谱之间建立概率映射的方法。这种方法首先以类型和属性的分布概率为基础, 创建知识图谱之间的映射, 然后利用该映射得到知识图谱中丢失的事实, 最后, 在两个知识图谱使用的类型系统之间建立映射。这样就可以用一个知识图谱的类型去预测另一个知识图谱的类型。该方法利用黄金标准进行评估, 其评价矩阵是正确率和召回率。

另外, WANG Q 等人^[28]利用耦合的路径排序算法补全知识图谱。这种方法首先设计了一个聚类算法自动发现彼此高度相关的关系, 然后采用多任务学习策略对这些关系的预测进行耦合, 这样是为了能够利用关系之间的联系和共享隐式数据。该方法使用的评价矩阵是平均正确率和平均倒数排名(Mean Reciprocal Rank)。

3 常用的知识图谱错误探测方法

与知识图谱补全方法不同, 知识图谱错误探测的目的是利用已有信息, 识别图中的错误信息, 同样, 本节也将错误探测分成内部和外部两类。

3.1 知识图谱错误内部探测方法

目前错误内部探测方法主要集中在文字值错误和链接错误上, 因此本部分只对这两类方法进行综述。

3.1.1 文字值错误内部检测

异常检测(Outlier detection)的目的是识别一个数据

集中与大多数数据偏离的实例,即特征显著的数据。由于异常检测在许多情况下仅处理数值型数据,因此数值型文字自然成为这些方法处理的对象。其中, WIENAND D 等人^[29]将不同的单变量异常值检测方法(如四分位范围或核密度估计)用于 DBpedia 中,该方法使用正确率和新增文字数作为评价矩阵。

为了降低自然异常的影响, FLEISCHHACKER D 等人^[30]对文献[29]的方法进行了扩展,将实例集分成更小的子集,从而提高识别的正确率。这种方法还能使用其他知识图谱预测交叉检测异常,是内部检测和外部检测方法的混合。

3.1.2 知识图谱链接错误内部检测

PAULHEIM H^[31]指出异常检测不仅可用于数值型数据,还可用于知识图谱的内部链接。他首先将链接表示成多维特征向量,然后利用标准的异常检测技术(如局部异常因素检测、基于簇的异常检测)指派异常分数,基于这些异常分数和所有链接的整体分布情况,能够识别出不合理的链接。LI H 等人^[32]使用概率模型学习属性之间的数学关系(如小于、大于),例如,一个人的出生日期必须在死亡日期之前。如果知识图谱中有关系与这些关系不符,那么就说明该关系是错误的。

3.2 知识图谱错误外部探测

知识图谱错误外部探测就是除了利用知识图谱本身外,还利用外部的资源来检测错误。外部探测方法主要集中在错误关系探测和错误文字值探测两方面。所以,本小节将从这两个方面进行综述。

3.2.1 错误关系外部检测

错误关系外部检测就是除了利用知识图谱本身外,还利用外部的资源来检测错误的实体间关系。其中, PAULHEIM H 等人认为在知识图谱构造过程中大量的错误都是由一个共同的原因(如错误的映射或程序错误)造成的,因此,只需检测少量的样本,就会发现大量错误的语句。于是他们提出了一种识别不一致性的自动化聚类方法^[33],该方法只需要给人提供代表性的样本即可,从而解决了上述的规模问题。

3.2.2 错误文字值外部检测

文献[34]提出了一种使用知识图谱链接探测错误数字值的自动方法,作者利用相同资源的链接和单个资源中属性之间的不同匹配函数来识别错误。他们认为如果多个外部资源与知识图谱中的一个事实发生冲突,那么就认为该事实是错的。

4 讨论

通过文献发现,将知识图谱精化方法分成知识图谱补全和知识图谱错误探测两大类是严谨的。因为目前基本不存在一个方法同时解决知识图谱补全和知识图谱错误探测。唯一的例外是文献[8],该文献既能进行知识图谱补全又能进行知识图谱错误探测。但它实际上是两个方法,分别是 SDType 和 SDValidate,因为这两个方法

不是一个整体,而是独立存在的。其中 SDType 负责进行补全, SDValidate 负责进行错误探测。在知识图谱精化方面,为什么大量的研究成果都只用在一方面,这个原因还不太明确。但在客观世界中,知识图谱补全和知识图谱错误探测这两个过程是相辅相成的。除了将补全和错误检测严格区别以外,还发现多数方法只能处理一种精化目标,同时处理多种精化目标的方法相当少。因此,将每类精化任务按照精化目标进行分类这也是严谨的。

在知识图谱补全方面,本文所介绍的方法都是对已有实体的类型或关系进行补全。经文献分析,目前没有方法能够增加新的实体,这种实体集扩展方法属于 NLP 领域,但这种方法对于进一步提高知识图谱覆盖率非常有用,尤其可以减少长尾实体。可见,研究增加新实体的方法也将是知识图谱精化的一个新方向。

在知识图谱错误探测方面,所有方法都输出一个潜在错误的语句列表。但据笔者所知,只有文献[33]能从错误列表中发现知识图谱模式的错误。因为模式是知识图谱的一个基础构建,模式的错误就会造成实体的关系错误。可见,探测模式错误也将是知识图谱精化的一个新方向。

在评价矩阵方面,发现大量的方法将正确率和召回率作为主要的评价矩阵,偶尔也有方法使用 ROC 曲线、精确率或均方根误差;在评估方法方面,发现有一半以上的评估方法只使用 DBpedia 这样一种知识图谱,这样的评估结果的作用非常有限。因为大多数的研究只对特定的知识图谱有用,但知识图谱根据特征的不同而不同。因此,对于只用一种知识图谱评估的方法来说,有以下问题值得研究:(1)能否在不用特征的知识图谱上有同样的性能;(2)在精化过程中是否用了知识图谱本身的特征,如是否隐含地使用 DBpedia 实体和对应的维基百科页之间的链接;(3)是否过度拟合图谱的特定特征。另外,还发现只有少数评价方法对计算性能进行评估。但在大规模知识图谱阶段,计算性能这个指标是一个不可忽视的维度。为了将来有一个可比较的知识图谱评价方法,需要选一个既在数量上可比较、也在计算性能上可比较的基准(benchmark)。目前这样的研究工作在语义网络的其他领域(如模式和实例匹配、推理和问答系统)已经开展。可见,知识图谱精化的通用评价方法将是知识图谱精化的另一个方向。

5 结论

多年来,许多研究者提出了各种知识图谱精化方法,取得了丰硕的研究成果。由此可以预见,知识图谱精化研究将是一个有着非常广阔研究前景的领域。

本文对知识图谱精化方法进行了综述。综述结果表明,该分类标准是严谨的。知识图谱精化涉及机器学习、统计学和 NLP 相关知识和技术,是一个综合的研究方向。几乎没有一个精化方法能同时提高知识图谱的完备性和正确率,也没有方法对多个精化目标进行精化,即

还没有一个改善知识图谱质量的整体解决方案。在评价方面,多数评价方法通常都是在一个特定的知识图谱上进行评价,这使得难以对它们的性能进行比较。

综上所述,虽然知识图谱精化已经取得了丰硕的研究成果,并且已成功应用于许多领域,但仍然还不成熟,依然有很大的挑战。将来可从以下几个方面对知识图谱精化进行深入的研究:(1)改善知识图谱质量的整体解决方案;(2)知识图谱扩展性的研究;(3)知识图谱通用的评价;(4)未知领域知识图谱的构建。随着大规模网络知识图谱的出现,知识图谱的扩展和自动化的知识图谱精化将是该领域未来发展的趋势。

参考文献

- [1] SINGHAL A. Introducing the knowledge graph: things, not strings[EB/OL]. (2012-05-xx)[2018-02-12]. <https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>.
- [2] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]. New York: ACM, 2008: 1247-1250.
- [3] VRANDECIC D, KROTZSCH M. Wikidata: a free collaborative knowledge base[J]. Communications of the ACM, 2014, 57(10): 78-85.
- [4] LEHMANN J, ISELE R, JAKOB M, et al. DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia[J]. Semantic Web, 2015, 6(2): 167-195.
- [5] REBELE T, SUCHANEK F, HOFFART J, et al. YAGO: a multilingual knowledge base from Wikipedia, Wordnet, and Geonames[M]. Berlin Heidelberg: Springer International Publishing, 2016.
- [6] GABRILOVICH E, USUNIER N. Constructing and mining web-scale knowledge graphs[C]. New York: ACM, 2016: 1195-1197.
- [7] PAULHEIM H, BIZER C. Type inference on noisy RDF data[M]. West Berlin: Springer, 2013.
- [8] PAULHEIM H, BIZER C. Improving the quality of linked data using statistical distributions[J]. International Journal on Semantic Web and Information Systems(IJSWIS), 2014, 10(2): 63-86.
- [9] KROMPAß D, NICKEL M, TRESP V. Large-scale factorization of type-constrained multi-relational data[C]. New York: IEEE, 2015: 18-24.
- [10] 张香玲, 陈跃国, 毛文祥, 等. 基于随机游走的实体类型补全方法[J]. 计算机学报, 2017(10): 2352-2366.
- [11] SLEEMAN J, FININ T, JOSHI A. Topic modeling for RDF graphs[EB/OL]. (2015-01-01) [2018-02-12]. http://ceur-ws.org/Vol-1467/LD4IE2015_Sleeman.pdf.
- [12] PAULHEIM H. Browsing linked open data with autocomplete[EB/OL]. (2012-05-03)[2018-02-12]. https://km.aifb.kit.edu/sites/swc/2012/submissions/swc2012_submission_15.pdf.
- [13] SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]. Newry: Curran Associates, Inc., 2013: 926-934.
- [14] BAIER S, TRESP V. Type-constrained representation learning in knowledge graphs[C]. New York: Springer-Verlag, Inc., 2015: 640-655.
- [15] ZHAO Y, GAO S, GALLINARI P, et al. Knowledge base completion by learning pairwise-interaction differentiated embeddings[J]. Data Mining & Knowledge Discovery, 2015, 29(5): 1486-1504.
- [16] KIM J, KIM E K, WON Y, et al. The association rule mining system for acquiring knowledge of DBpedia from Wikipedia categories[EB/OL]. [2018-02-12]. http://semanticweb.kaist.ac.kr/home/images/a/a8/The_Association_Rule_Mining_System_for_Acquiring_Knowledge_of_DBpedia_from_Wikipedia_Categories.pdf.
- [17] KOLTHOFF C, DUTTA A. Semantic relation composition in large scale knowledge bases[EB/OL]. (2015-08-01) [2018-02-12]. https://ub-madoc.bib.uni-mannheim.de/39756/1/LD4IE2015_Kolthoff.pdf.
- [18] NUZZOLESE A G, PRESUTTI A G V, CIANCARINI P. Type inference through the analysis of Wikipedia links[EB/OL]. (2012-10-03)[2018-02-12]. <http://www.ceur-ws.org/Vol-937/LDOW2012-paper-13.pdf>.
- [19] APROSIO A P, GIULIANO C, LAVELLI A. Automatic expansion of dbpedia exploiting wikipedia cross-language information[M]. Berlin Heidelberg: Springer, 2013.
- [20] SLEEMAN J, FININ T. Type prediction for efficient coreference resolution in heterogeneous semantic graphs[C]. Piscataway, NJ: IEEE Computer Society, 2013: 78-85.
- [21] KLIEGR T. Linked hypernyms: enriching DBpedia with targeted hypernym discovery[J]. Web Semantics Science Services & Agents on the World Wide Web, 2015, 31: 59-69.
- [22] APROSIO A P, GIULIANO C, LAVELLI A. Extending the coverage of DBpedia properties using distant supervision over Wikipedia[EB/OL]. [2018-02-12]. http://www.ceur-ws.org/Vol-1064/Aprosio_Extending.pdf.
- [23] GERBER D, HELLMANN S, SORU T, et al. Real-time RDF extraction from unstructured data streams[C]. New York: Springer-Verlag, Inc., 2013: 135-150.
- [24] WEST R, GABRILOVICH E, MURPHY K, et al. Knowledge base completion via search-based question answering[C]. New York: ACM, 2014: 515-526.
- [25] HOGAN A, MILEO A. Triplifying Wikipedia's tables[EB/OL]. (2013-10-04)[2018-02-12]. http://www.ceur-ws.org/Vol-1057/MunozEtAl_LD4IE2013.pdf.
- [26] RITZE D, LEHMBERG O, BIZER C. Matching HTML

(下转第 38 页)

序的攻击。双核安全架构可以实现安全环境与正常环境的隔离,一般物理攻击无法突破隔离技术的防御能力。如果设备密钥受到物理攻击的危害,可以通过执行密钥更新协议来部署新设备密钥。

4 结束语

本文利用 PUF 能够将密钥隐含地嵌入设备而不需要物理存储的特性,构建了双核隔离 SoC 架构的可信根 PRoT,并以 PRoT 为信任锚点设计了信任链和密钥更新协议。实验结果表明,PRoT 能够以较小 TCB 运行于 Zynq-7000 AP SoC 中,通过对 PRoT 中部分 TPM 命令和 Seal/Unseal 服务的性能评估,以及对初始种子和唯一设备密钥的安全性分析,证明了 PUF 能够为嵌入式系统提供安全灵活的可信根。

参考文献

- [1] 孙春子.基于双核的系统安全框架的研究与实现[D].南京:南京理工大学,2008.
- [2] 利文浩.面向多域计算环境安全的系统软件研究[D].上海:上海交通大学,2015.
- [3] ALVES T.TrustZone:integrated hardware and software security[J].White Paper,2004.
- [4] ARM.Securing the system with TrustZone ready program [EB/OL].[2013-12-19].<http://www.arm.com/products/security-on-arm/trustzone-ready/index.php>.
- [5] ARENO M,PLUSQUELLIC J.Securing trusted execution environments with PUF generated secret keys[C].IEEE International Conference on Trust,Security and Privacy in Computing and Communications.IEEE Computer Society,2012:1188-1193.
- [6] FENG W,FENG D,WEI G,et al.TEEM:a user-oriented

trusted mobile device for multi-platform security applications[M].Trust and Trustworthy Computing.Springer Berlin Heidelberg,2013:133-141.

- [7] KANG H,HORI Y,KATASHITA T,et al.Cryptographic key generation from PUF data using efficient fuzzy extractors[C].International Conference on Advanced Communication Technology.IEEE,2014:23-26.
- [8] 谭良,徐志伟.基于可信计算平台的信任链传递研究进展[J].计算机科学,2008,35(10):15-18.
- [9] KUMAR A,MISHRA R S,KASHWAN K R.Challenge-response generation using RO-PUF with reduced hardware[C].International Conference on Advances in Computing, Communications and Informatics.IEEE,2016.
- [10] MORELOS Z R.Encoder/decoder for binary BCH codes in C (Version 3.1)[EB/OL].[2017-12-19].http://www.rajivchakravorty.com/source-code/uncertainty/multimedia-sim/html/bch_8c-source.html.
- [11] RSA Data Security Inc.RSAREF[EB/OL].[2017-12-19].<http://www.homeport.org/adam/crypto/rsaref.phtml>.
- [12] TCG.Trusted platform module library part 1: architecture, Family 2.0, Level 00 Revision 01.07[Z].2014.
- [13] 吴少刚,许华.可信嵌入式芯片启动加载程序 iPMON 的设计[J].计算机工程与设计,2008,29(1):5-8.

(收稿日期:2017-12-19)

作者简介:

吴缙(1994-),通信作者,男,硕士研究生,主要研究方向:专用集成电路设计技术,E-mail:woshi57890@163.com。

徐金甫(1965-),男,博士,教授,主要研究方向:专用集成电路设计技术。

(上接第 33 页)

- tables to DBpedia[C].New York:ACM,2015:1-6.
- [27] DUTTA A,MEILICKE C,PONZETTO S P.A probabilistic approach for Integrating heterogeneous knowledge sources[M].Switzerland:Springer,2014.
- [28] WANG Q,LIU J,LUO Y,et al.Knowledge base completion via coupled path ranking[C].Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics,2016:1308-1318.
- [29] WIENAND D,PAULHEIM H.Detecting incorrect numerical data in DBpedia[M].Berlin:Springer International Publishing,2014.
- [30] FLEISCHHACKER D,PAULHEIM H,BRYL V,et al.Detecting errors in numerical linked data using cross-checked outlier detection[M].Berlin:Springer International Publishing,2014.

- [31] PAULHEIM H.Identifying wrong links between datasets by multi-dimensional outlier detection[EB/OL].(2014-04-05)[2018-02-12].http://www.ida.liu.se/~patla00/conferences/WoDOOM14/papers/wodoom14_5.pdf.
- [32] LI H,LI Y,XU F,et al.Probabilistic error detecting in numerical linked data[M].Cham:Springer,2015.
- [33] PAULHEIM H,GANGEMI A.Serving DBpedia with DOLCE-More than just adding a cherry on top[M].Berlin Heidelberg:Springer,2015:180-196.
- [34] LIU S,AQUIN M D,MOTTA E.Towards linked data fact validation through measuring consensus[EB/OL].(2015-08-01)[2018-02-12].http://videolectu.res.net/site/normal_dl/tag=961758/eswc2015_liu_linked_data_01.pdf.

(收稿日期:2018-03-13)

作者简介:

谢刚(1980-),男,博士,副教授,主要研究方向:知识表示、人工智能。