# 基于Python课程知识图谱的构建与应用

知识图谱是一个结构化的语义知识库，用于以符号形式描述物理世界中的概念及其相互关系，其基本组成单位是“实体—关系—实体”和“实体—属性—值”两种三元组[1]。知识图谱已经广泛应用于信息检索、文本分类、问答系统和机器翻译,以及网络学习等领域[4-6].python课程知识图谱旨在构建面向python类课程的知识图谱,提供一种形式化的、显式的python课程知识表示和知识组织模型。

对于知识图谱,相关文献[7-10]论述了其内涵、发展历史、构建方法、应用和挑战等.具体地,Li等[７]介绍了知识图谱的发展历程和发展方向,阐述了知识图谱的数据来源和构建 方法,以及基于知识图谱的知识推理和用户搜索意图挖掘. Zhu等[8]论述了文本知识图谱、视觉知识图谱和多模态知识 图谱的构建方法和应用领域.Liu等[9]分析了知识图谱构建 的研究现状、技术框架、关键问题,以及在信息检索和问答系 统中的应用等.另外,Guan等[10]介绍了基于知识图谱的知识推理的研究进展,以及在垂直信息检索和问答系统中的应用等。

对于课程知识图谱,相关文献[11-14]分析了课程知识图谱 构建方法,以及基于课程知识图谱的教学改革等.Zhang [11] 分析了课程知识本体和课程知识图谱构建方法.课程知识本体中的概念关系包括概括关系、特化关系、兄弟关系、等价关 系、成分关系和实例关系.Liu等[12]和 Xie等[13]介绍了基于 知识图谱的信息论课程教学改革方案和数字媒体课程教学改 革方法. 对于课程本体,相关文献[15-20]介绍了课程知识本体的构 建方法,以及课程知识本体在课程自主学习系统和课程资源 查询系统的应用等.具体地,Zhong [１５]论述了课程内容结构 本体模型和课程知识点本体模型构建方法.课程知识点本体 中课程知识点的关系包括继承关系、整体部分关系、实例关 系、依赖关系和平行关系等.Zeng等[１６]介绍了面向概念的数 学知识获取方法,构建了数学概念知识体系.数学知识包括 数学概念知识、数学断言知识和数学实例知识.数学领域上 层本体包括概念定义、断言和实例.He等[１７]将概念图引入 离散数学课程自主学习系统中.概念图由概念及其关系构 成,概念关系包括分类和特例等.Li等[１８]以计算机网络课程 为例,阐述 了 E-Learning课 程 体 系 结 构 和 课 程 本 体 构 建 方 法.概念关系包括依赖关系、泛化关系、整体部分关系、约束 关系和功能关联关系.Jiang [１９]论述了基于 WordNet和维基 百科的中文领域本体即初等几何本体的构建方法.另外,Lv 等[２０]介绍了课程知识本体建模和本体推理方法,设计了课程 资源查询系统.该系统包括本体库、常用查询知识库和查询 处理器等。

通过综合分析以上研究成果可以看出，国内外学者对本体构建进行了较深入 的研究、考虑的内容比较全面，但也存在一些不足之处：还没有形成一个泛化能力较强的本体模型，无法实现自动化。因此，业内公认的本体构建方法都需要在领域专家的指导和参与下完成。

[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,(3):582-600.

[４] WU X．From BigDatatoBigKnowledge:HACE＋BigKE[J]．JournalofComputerScience,２０１６,４３(７):３Ｇ６．

[５] XIONG F,GAOJ．EntityAlignmentforCrossＧLingualKnowＧ ledgeGraphwithGraphConvolutionalNetworks[C]∥The２８th InternationalJointConferenceon ArtificialIntelligence．２０１９: ６４８０Ｇ６４８１．

[６] WANG X,WANG D,XU C,etal．ExplainableReasoningover KnowledgeGraphsforRecommendation[C]∥TheAAAIConＧ ferenceonArtificialIntelligence．２０１９:５３２９Ｇ５３３６．

[７] LIT,WANGC,LIH．DevelopmentandConstructionofKnowＧ ledgeGraph[J]．JournalofNanjing UniversityofScienceand Technology,２０１７,４１(１):２２Ｇ３４．

[８] ZHU M,PAOB,XUC．ResearchProgressonDevelopmentand ConstructionofKnowledgeGraph[J]．JournalofNanjingUniＧ versityofInformationScienceandTechnology(NaturalScience Edition),２０１７,９(６):５７５Ｇ５８２．

[９] LIU Q,LIY,DUAN H,etal．KnowledgeGraphConstruction Techniques [J]．JournalofComputerResearchand DevelopＧ ment,２０１６,５３(３):５８２Ｇ６００．

[１０]GUAN S,JIN X,JIA Y,et al．Knowledge Reasoning Over KnowledgeGraph:A Survey[J]．JournalofSoftware,２０１８, ２９(１０):２９６６Ｇ２９９４．

[１１]ZHANG M．Research on Construction ofCourse Knowledge GraphandSearchTechnology[D]．Wuhan:WuhanUniversity, ２０１６．

[１２]LIUZ,LIZ．ResearchonInformationTheoryTeachingReform Basedon KnowledgeGraph Theory[J]．ComputerKnowledge andTechnology,２０１８,１４(１２):１２５Ｇ１２７．

[１３]XIEZ,LIU Y．ResearchonTeachingReformofDigitalMedia Knowledgeby MeansofKnowledgeGraph Modeling[J]．SoftＧ wareGuild,２０１７,１６(１１):２３０Ｇ２３２．

[１４]WANGL．ReconstructionofMOOCCoursesbasedonMultimoＧ dalKnowledgeMapfromthePerspectiveofDeepLearning[J]． ModernEducationTechnology,２０１８,２８(１０):１００Ｇ１０６．

[１５]ZHONG Y．OntologyＧbasedCurriculum KnowledgePoint MoＧ delingof MajorofInformation ManagementandInformation System [J]．InformationResearch,２０１３(８):９４Ｇ９８

[１６]ZENG Q,CAO C,SUI Y,etal．Research on OntologyＧbased MathematicalKnowledgeAcquisitionandKnowledgeHeritance Mechanism [J]．Microelectronics& Computer,２００３,２０(９):１９Ｇ ２７．

[１７]HEZ,ZHUANG Y．DiscreteMathematicsCourseAutonomous LearningSystemBasedonConceptMap[J]．HigherEducation ofSciences,２０１８(１):９０Ｇ９５．

[１８]LIH,YANG G．CourseDevelopmentofeＧLearningbasedon Ontology[J]．ComputerEngineeringandDesign,２０１０,３１(４): ８８１Ｇ８８４．

[１９]JIANG Y．Construction and Application of OntologyＧbased Mathematics Knowledge Base [D]．Chengdu:University of ElectronicScienceandTechnologyofChina,２０１１．

[２０]LVJ,YU X．Ontology ModelingandReasoningforCurriculum Knowledge[J]．ComputerEngineering,２０１１,３７(４):６１Ｇ６３．

知识图谱的构建方式可以分为自顶向下和自底向上两种。自底向上是指从开放的网络中抽取信息并向上归为一类，这种方式比较适合构建通用的知识图谱。而自顶向下的构建方式主要应用于构建领域知识图谱。领域知识图谱涉及到的知识相对固定且变化较少，并且知识的精度和深度要求更高，因此，采用自顶向下的方式，可以确保知识图谱的整体质量[23]。

自顶向下方式一般要求首先完成领域本体构建[24]。1993 年, Gruber[6]将本体定义为“一种概念化的精确的规格说明”。1998 年,Studer[7]进一步扩充了本体的概念,将其定义为“共享概念模型的明确形式化规范说明”。本体主要是用来描述某个领域内的概念和概 念之间的关系,使得它们在共享的范围内具有大家共同认可的、明确的、唯一的定义。

本体构建的过程相当繁琐,而且构建过程往往因各自领域和具体工程的不同而有所不同[8]。针对课程领域的本体构建，杨东明等[5]在研究面向初等数学的知识点关系提取时，以领域概念作为学科知识本体的基本单位，并定义了概念之间的依赖、被依赖、属于、包含、拥有、被拥有、同义、反义关系。潘颖等[6]认为课程知识的基本单位是知识点，因此在构建课程知识本体时将知识点定义为类，并通过父类和子类机制来描述知识之间的层次关系；为了反映课程的教学要求，他们还定义了“课程教学大纲”类及其“实验”子类。姜强等[8]认为应从章、节、知识点和学习对象四部分来设计知识模型。黄焕[7]将课程的知识表示为章、节和知识点三种级别的知识单元，并定义了包含、顺序和相关关系。

通过综合分析以上研究成果可以看出，国内外学者对本体构建进行了较深入 的研究、考虑的内容比较全面，但也存在一些不足之处：还没有形成一个泛化能力较强的本体模型，无法实现自动化。因此，业内公认的本体构建方法都需要在领域专家的指导和参与下完成。

在通过本体构建完成了课程知识图谱的顶层设计后，需要利用信息抽取技术，从相关语料中抽取信息，以填充知识图谱。抽取任务包括实体抽取、关系抽取和属性抽取。

实体抽取也称命名实体识别,是从文本中自动识别出命名实体.与之相似的任务是实体集扩充,该任务指的 是根据种子实体集,从文本中抽取出相同类别的新实体.对于实体集扩充,Bootstrapping 方法是最直接的想法[40], 此方法根据种子实体从文本中抽取出特征模板,然后利用这些模板从文本中抽取出新的实体,再根据新实体从 文本中抽取新的特征模板.反复迭代此过程,便可以抽取出目标概念下大量的新实体.这种方法最大的问题是语义漂移[41],即:随着迭代次数的增加,扩充的新实体会逐渐偏移原来的类别.

关系抽取指的是从文本中抽取出实体和实体之间的关系,这样才能将零散的实体联系起来.关系抽取的算 法可以分为基于规则的方法和基于机器学习的方法.基于规则的方法需要人工制定较多规则且难以全面;基于机器学习的方法又可以分为有监督、半监督和无监督这 3 类:有监督的方法需要大量质量较好的标注数据,半 监督的方法需要少量标注数据,无监督的方法不需要标注数据.使用有监督的方法进行关系抽取,可以看作是多 分类问题,预先将每个关系定义为一个类别,然后将句子中实体之间的关系划分到预先定义的类别中;较多采用的半监督算法有 Bootstrapping 算法、协同训练算法和标注传播算法[42],其中,Bootstrapping 方法中的远程监督 方法目前最受学界关注,该方法首先将每种关系的少量三元组作为种子集,然后回标出同时包含种子集中三元 组的两个实体的句子作为表征这一关系的训练数据,再从文本中找出符合这一关系的其他句子,这些句子中的 实体和实体之间则很有可能也满足这一关系.该方法最大的问题是训练数据噪声,即:包含两个实体的句子的语 义关系可能会有很多种,但是回标的时候所有句子被当成了一种语义关系.本文对有监督、半监督和无监督的 方法都进行了相关实验.

属性抽取则是从文本中抽取出实体的属性信息。由于可以将实体的属性视为实体和属性间的一种名词性关系,因此也可以将属性抽取问题视为关系抽取问题[43]

[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,(3):582-600.

[４] WU X．From BigDatatoBigKnowledge:HACE＋BigKE[J]．JournalofComputerScience,２０１６,４３(７):３Ｇ６．

[５] XIONG F,GAOJ．EntityAlignmentforCrossＧLingualKnowＧ ledgeGraphwithGraphConvolutionalNetworks[C]∥The２８th InternationalJointConferenceon ArtificialIntelligence．２０１９: ６４８０Ｇ６４８１．

[６] WANG X,WANG D,XU C,etal．ExplainableReasoningover KnowledgeGraphsforRecommendation[C]∥TheAAAIConＧ ferenceonArtificialIntelligence．２０１９:５３２９Ｇ５３３６．