

# 基于知识图谱的 个性化习题推荐系统设计研究<sup>\*</sup>

王冬青,殷红岩

(华南师范大学 教育信息技术学院,广东 广州 510631)

**摘 要:**针对学习者的认知差异构建个性化推荐系统是“互联网+教育”环境下的重要研究内容之一。现有的习题推荐大多采用协同过滤或基于项目反应理论的自适应推荐算法,基于题目的知识点和难度特征提供个性化推荐服务,而忽略语义信息丰富的非结构化学科知识。本文引入知识图谱从语义层次描述与呈现学科知识,结合习题的知识点及难度,基于学习者的历史作答数据判断学生的知识点掌握状态,并设计了基于知识图谱的个性化习题推荐系统,初步验证表明,推荐结果具有较好的可解释性与用户满意度,能够为学习者提供更加精准、高效的习题推荐服务。

**关键词:**知识图谱;习题;个性化推荐

**中图分类号:**G40-057

**文献标志码:**A

**文章编号:**1673-8454(2019)17-0081-06

## 一、引言

《国家中长期教育改革和发展规划纲要(2010-2020年)》提出减轻中小学习者的课业负担,并要求设计适当的课程难度。由于升学等因素对成绩的需求,现在的学习者采用“题海战术”依然常见。国际 K12 在线学习协会认为个性化学习能够满足学习者自身的需求和学习兴趣,<sup>[1]</sup>《地平线报告:2015 高等教育版》将个性化学习定位为“艰难的挑战”,并提出个性化学习是针对个体学习者特定的学习需求、兴趣、意愿或文化背景而推出的一系列教育项目、学习经验、教学方法和学术支持策略。<sup>[2]</sup>《教育信息化十年发展规划(2011-2020 年)》提出:“要为每一名学习者提供个性化学习的信息化环境和服务。”<sup>[3]</sup>随着信息技术的发展、网络资源数量的增加,学习者的个性化学习需求也更加强烈。传统的教师统一练习、布置作业难以满足学习者的个性化学习需求,为了实现学习者的个性化练习,本文针对中小学个性化学习的特点,设计并实现了基于知识图谱的习题个性化推荐系统,采集学习者的测评数据,结合学科知识丰富的语义信息,考虑习题的知识点、难度、知识结构等因素,为学习者推荐符合其学习状态的习题,使学习者的练习更有针对性,提高学习者的学习效率。

## 二、相关研究现状

### 1. 知识图谱研究现状

2012 年,Google 公司首次提出知识图谱的概念,是指其用于提升搜索引擎性能的知识库。知识图谱是一种以实体、属性、关系来描述的知识体系,以可视化手段展示知识。各大互联网公司和高校纷纷建立起自己的知识图谱,如百度的“知心”、搜狗的“知立方”、华东理工大学构建的“知识魔方”、复旦大学 GDM 实验室开发的“知识工场”和清华大学构建的基础教育知识图谱等。知识图谱在教育领域中主要有以下研究:①学科知识图谱的构建方法研究。杨玉基、许斌等提出了一种准确高效的知识图谱构建方法——“四步法”,并用这种方法在较短时间内高效地构建了中国基础教育九门学科的知识图谱。<sup>[4]</sup>钟亮通过网络爬虫对百度百科中与“化学”主题相关的词条信息进行抓取,利用分词、实体识别、实体关系识别构建化学学科知识图谱,并对结果进行了评价。<sup>[5]</sup>②利用知识图谱为学习者提供资源导航、个性化推荐等功能。例如可汗学院建立的数学学科知识图谱,形成知识网络,学习者可以直观地看到自己的学习进度与掌握状态,并实现学习资源导航的功能。吴强强、陈昊鹏等以软件工程专业为例,基于 MOOC 平台设计了简单而高效的专业课程知识图谱,并基于学习者的历史课程学习数

<sup>\*</sup> 基金项目:本文系国家自然科学基金青年科学基金项目“基于情境感知的智慧教学动态生成性数据采集模型与交互可视分析机制研究”(课题编号:71701071)。

据,进行学习质量评估和课程推荐。<sup>[6]</sup>Haiping Zhu、Feng Tian 等验证了学习者在四种学习情境下的八种学习路径偏好,提出了一种基于知识图谱的多约束学习路径推荐模型并验证有效性,使学习者能够在有限的时间内平衡学习目标和学习时间。<sup>[7]</sup>

## 2. 个性化习题推荐方法研究现状

目前大多数个性化习题推荐系统都是采用协同过滤算法与基于项目反应理论的自适应算法,主要根据用户相似度、知识点与习题难度进行推荐。孙权、贺细平将协同过滤用于在线测评系统,比较了基于用户与基于项目的协同过滤算法的有效性;<sup>[8]</sup>姜盼对协同过滤算法进行优化,依据用户相似度提供习题个性化推荐服务;<sup>[9]</sup>单瑞婷、朱天宇等改进了协同过滤方法,将认知诊断结果与用户相似度结合,基于习题的知识点、难度等为学习者提供个性化推荐;<sup>[10][11]</sup>柏茂林基于项目反应理论改进相似用户模型,将能力水平相近的用户作为相似用户,采用协同过滤算法为用户提供个性化习题推荐。<sup>[12]</sup>这些个性化习题推荐方法虽然关注了学习者之间学习经历的相似度或知识点的掌握情况,但并未考虑学科知识中丰富的语义关系。除了习题的所属知识点、难度等属性外,知识点间还具有丰富的语义关系,基于这些语义关系能够更加系统地描述与呈现学科知识,更加准确地分析学习者的知识掌握情况,从而更加精准地为学习者推荐个性化习题。

因此本文基于学习者的习题作答情况判断其知识点掌握状态,构建面向学科知识的知识图谱并将其进行交互可视化,提出了一种基于知识图谱的个性化习题推荐方法,设计了个性化习题推荐系统原型,初步案例分析表明本文提出的个性化习题推荐方法具有很好的可解释性与用户满意度。

## 三、基于知识图谱的个性化习题推荐系统设计

本文所构建的基于知识图谱的个性化习题推荐系统,主要包含测评数据采集、知识图谱构建、知识图谱可视化、个性化推荐四个模块,如图 1 所示。基于知识图谱的个性化习题推荐系统的工作流程是从测评数据采集模块中获取学生的习题作答详情,据此判断学习者的知识点掌握状态,然后结合知识图谱寻找学习者的薄弱知识点,最后利用个性化推荐模块找到符合学习者知识点掌握状态的习题推荐给学习者。

### 1. 学习者知识点掌握状态的判定

为学习者推荐个性化习题,首先需要判断学习者当前的知识点掌握状态,测评数据是判断学习者知识点掌握状态的依据。当前学校教育中学习的测评侧重于学习

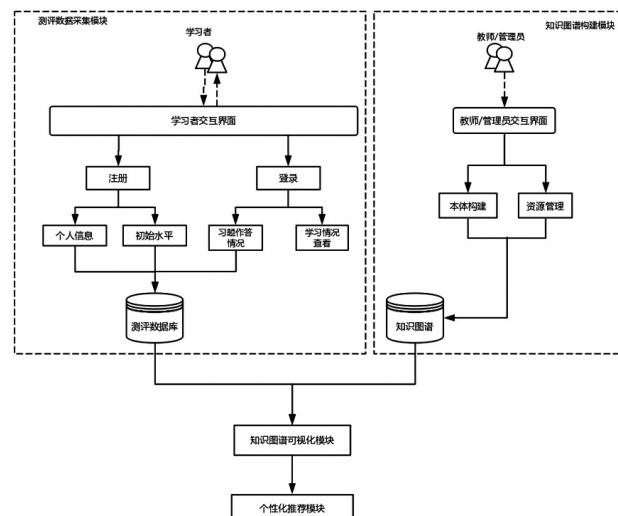


图 1 习题个性化推荐系统总体架构

者的结果表现,如测评得分、习题正确率等,较少关注学习者的学习内容掌握情况。牟智佳、俞显发现现有的智能测试平台仍存在测评目的异化的问题,即测评目的在于选拔和甄别,而忽视测评的诊断、反馈和促进功能。<sup>[13]</sup>

学习者的知识点掌握状态是习题推荐的基础,为了更加精确地诊断学习者的学习内容掌握情况,为其提供更加精准的推荐服务,本文基于云测评环境采集了学习者的基本信息、习题作答等数据,以判断学习者的初始学习水平及知识点掌握状态。系统在学习者首次登录时要求学习者填写自己的姓名、性别、年级、学校、班级、家庭背景等信息,并将这些信息进行存储与分析,以便了解学习者的个人学习背景,准确分析学习者的情况。除此之外,还采集了学习者的习题作答数据,包含了学习者每次进行习题作答时产生的习题、是否正确等数据。

学习者首次登录系统时没有做题记录,无法获知学习者的初始学习水平,因此无法判断学习者的知识点掌握状态,不能进行个性化推荐。因此,根据学习者的学习背景,如根据学校、年级、班级等为学习者生成测试习题,根据学习者的作答结果分析学习者的初始学习水平。

学习者的知识点掌握状态是习题推荐中最重要的数据,能够发现学习者的薄弱知识点。首先要采集学习者的习题作答数据,然后通过以下步骤计算学习者的知识点掌握状态,如图 2 所示。

①判断该学习者对于高难度习题的正答率是否大于或等于 0.85,若是,则判定该学习者在该知识点上处于较高水平,认为学习者已经掌握了该知识点,可以进行下一知识点的学习;若否,则进行步骤②。

②判断该学习者对于该知识点下中等难度习题的正答率是否大于等于 0.85,若是,则判定学习者已经掌

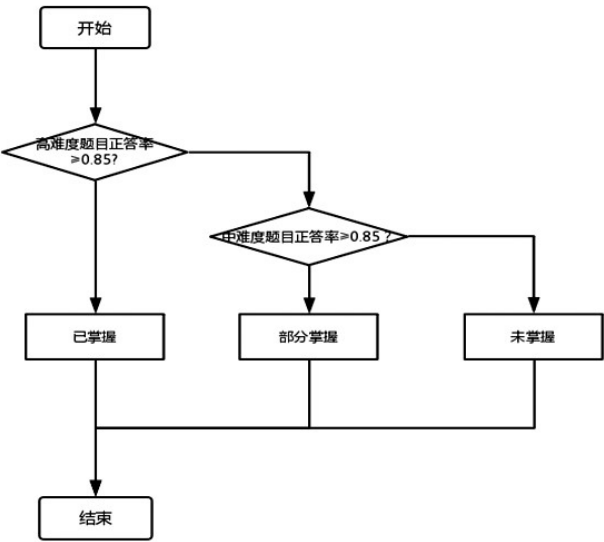


图2 知识点掌握状态判断流程

握该知识点的基础内容,还需进一步学习直至掌握该知识点;若否则认为该学习者尚未掌握该知识点。

2.知识图谱构建与可视化

(1)知识图谱构建

知识图谱是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。<sup>[14]</sup>采用知识图谱呈现学科知识,能够对知识点及知识点间的关系进行结构化描述,建立起语义化的学科知识结构,能够更好地帮助学习者了解学科知识体系。知识图谱的逻辑结构包含数据层和模式层。数据层是以事实为基本单位存储知识,如果以“实体-关系-实体”或“实体-属性-值”的三元组作为事实的表达方式,则存储在数据库中的数据形成庞大的实体关系网络。逻辑层是知识图谱的核心,通常以本体库管理数据层的事实知识,以本体规范实体、关系、属性及其表达。<sup>[15]</sup>利用本体表示学科知识,能够实现知识的重用与共享,能让学习者更加全面地了解学科知识体系。

本体构建是知识图谱构建过程中的重要步骤。本体是领域知识的结构化规范,它描述了概念及概念间的关系。本文采用领域知识本体常用的构建方法“七步法”,利用美国斯坦福大学开发的本体编辑器 protégé 构建了学科知识本体。

①本体层次结构模型构建

根据中小学学科知识结构特点,本文构建学科本体层次模型如图3所示。共分4个层级——学科、课程、章节知识、知识点。其中学科是第一个层级,包含若干个课程,课程中又包含若干个章节知识,知识点是教学活动中的基本单元,不能再进一步细分。

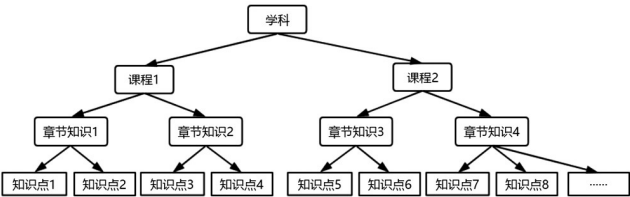


图3 学科本体层次

②数据属性

仅用知识点名称无法全面描述知识的特征,数据属性用来描述知识点的自身属性,除了知识点的标识符、名称、所属年级外,学习目标定义了该知识点学习者应该达到的掌握水平;每个知识点下包含若干资源,如微课、习题等,同时提供资源的链接地址,将学习资源映射至知识本体中。数据属性及具体内容如表1所示。

表1 数据属性及具体内容

数据属性	英文名称	具体内容
标识符	Identifier	该知识点的唯一标识符
名称	Title	该知识点的名称
所属年级	Grade	该知识点所属年级
学习目标	Mastery	该知识点学习者应达到的学习目标:记忆、理解、应用、分析、评价、创造
资源类型	Resource Type	该知识点下包含的学习资源类型,如微课、习题等
资源地址	Url	该知识点下的学习资源链接地址

③关系属性

除了知识点的自身属性外,知识点间的关系属性描述知识点间的逻辑关系。除属于、包含、被包含等关系外,前驱、后继是知识点之间的重要关系,它定义了知识的学习顺序。如知识点A与知识点B是前驱关系,则必须要学习完知识点A后才能够学习知识点B;知识点B与知识点A是后继关系,则学习完知识点A后可以学习知识点B。

(2)知识图谱可视化

在教育教学中应用知识图谱能够帮助学习者系统地查看学科知识,明确自己在知识系统中的定位、处于知识系统的哪个层级;利用已经构建的学科知识图谱,结合学习者的测评数据,分析学习者的知识点掌握状态,生成学习者个人知识图谱,为学习者提供学科知识体系与知识点掌握状态可视化显示。根据联通主义的理念,采用树状结构、层层递进的方式将学科知识图谱可视化呈现给学习者,使学习者能够更加全面直观地了解所学内容的知识体系。并用不同颜色表示学习者对不同



知识点的掌握程度,如绿色表示学习者已经掌握该知识点、蓝色表示学习者已经部分掌握、红色表示学习者尚未掌握,如图4所示,帮助学习者检测自己的学习状态和知识点掌握情况,从而提高学习效率。点击相应的知识点能够显示学习者对该知识点习题的详细作答情况,统计学习者每种难度习题的作答数量与正确率,使学习者更加详细地了解自己在每个知识点的历史做题情况,从而更加明确地判断自己在该知识点的掌握情况。知识图谱能够帮助学习者自我检测知识与技能掌握情况,同时也为习题的个性化推荐奠定基础。

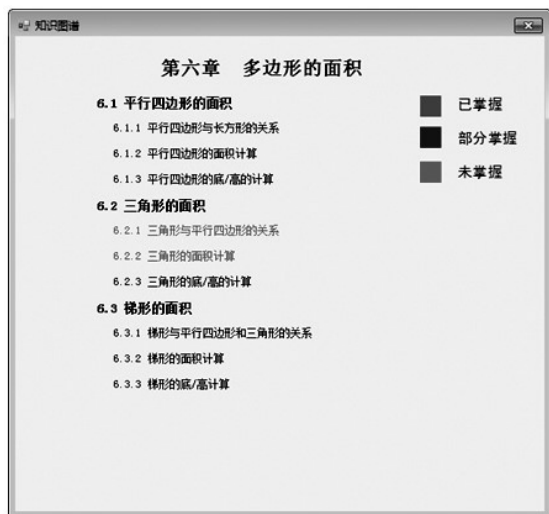


图4 知识图谱可视化

### 3.基于知识图谱的习题个性化推荐方法

维果斯基提出“最近发展区理论”,提出学习者的发展有两个水平:一是学习者现有的发展水平,指学习者在独立活动时能够达到的水平;二是学习者可能的发展水平,指学习者在指导下能够达到的水平。这两者之间的差距就是最近发展区。教学应该在学习者的最近发展区内进行,这样才能调动学习者的积极性,发挥学习者的学习潜能。同理,习题的推荐也应在学习者的最近发展区内进行,使学习者能够快速且有效地掌握。针对中小学习者的个性化练习,主要是针对学习者的知识点掌握情况,发现学习者的薄弱知识点,分析知识点掌握问题,为学习者推荐其最近发展区内的习题,帮助学习者更高效地掌握知识点,提高学习者的知识技能。个性化推荐引擎是推荐系统的核心部分,本文基于学习者的题目作答情况分析,设计并提出了基于知识图谱的个性化习题推荐方法。

本文基于知识图谱中学习者的知识点掌握情况及语义化的学科知识图谱实现习题的个性化推荐。首先依据知识图谱及学习者的题目作答详情,判断学习者的知

识点掌握状态,确定薄弱知识点,为学习者推荐适合其知识点掌握水平的习题。基于知识图谱的个性化习题推荐流程及伪代码如图5、图6所示,具体推荐过程描述如下。

算法:基于知识图谱的个性化习题推荐方法  
输入: skm(学生作答详情); knowledgeMap  
输出: 个性化习题推荐列表

```

if (skm.highRate < 0.85) //高难度习题正确率是否大于0.85
{
    List<Question> Q1 = QuestionsDao.getQuestionFromKeyword(knowledge.keyword);
    if (skm.middleRate < 0.85) //中难度习题正确率是否大于0.85
    {
        if (skm.lowRate < 0.85) //低难度习题正确率是否大于0.85
        {
            Knowledge preKnowledge = KnowledgeMap.getPreLevelKnowledge(knowledge.preLevel);
            result = getRecommendQueueList(preKnowledge);
            //在知识图谱中查找前驱知识点,前驱知识点已掌握,推荐本知识点下的低难度习题
            if (result.Count == 0) {
                foreach (Question q in Q1)
                {
                    if (q.difficulty < 0.4)
                        result.Add(q);
                }
            }
            else return 前驱知识点习题; //尚未掌握前驱知识点,推荐前驱知识点习题
        }
        else //已掌握本知识点下的低难度习题,推荐本知识点的中难度习题
        {
            foreach (Question q in Q1)
            {
                if (q.difficulty > 0.4 && q.difficulty < 0.7)
                    result.Add(q);
            }
        }
    }
    else //尚未掌握该知识点,推荐本知识点下高难度习题
    {
        foreach (Question q in Q1)
        {
            if (q.difficulty > 0.7)
                result.Add(q);
        }
    }
}
else return 后继知识点习题列表; //已掌握该知识点,在知识图谱中查找后继知识点,推荐后继知识点习题

```

图5 基于知识图谱的个性化推荐方法伪代码

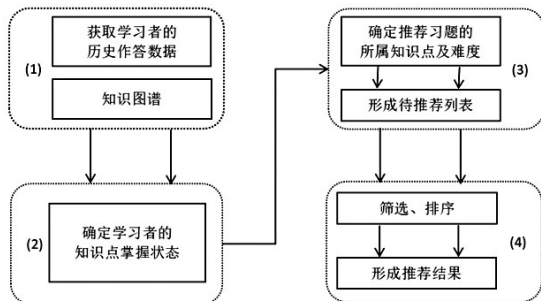


图6 基于知识图谱的个性化习题推荐流程图

#### (1)获取学习者的历史习题作答详情与知识图谱

①对于首次登录的学习者,获取学习者的初始水平测试的题目作答情况。

②对于非首次登录的学习者,可以从学习者的测评数据库中获取学习者的知识点答题情况。

#### (2)判断学习者的知识点掌握状态

结合知识图谱对学习者的答题情况进行分析,得到学习者的知识点掌握状态,确定学习者当前的薄弱知识点。

#### (3)学习者特征与习题的匹配

将学习者的掌握状态与语义化的知识图谱结合,分析学习者未掌握该知识点是否因为前驱知识点未掌握或是由于该知识点下的基础未掌握,从而确定目标习题所属的知识点与难度。

①若该学习者已经掌握该知识点,表明学习者已具备学习下一知识点的能力,则向学习者推荐后继知识点的习题。

②若学习者对该知识点的掌握程度为部分掌握,则

认为该学习者还需要进一步学习该知识点,为学习者推荐该知识点下的高难度习题,帮助学习者进一步提升对该知识点的掌握程度。

③若学习者尚未掌握该知识点,则判断学习者在该知识点下的低难度习题正答率是否大于 0.85,若是,则认为该知识点下习题处于初级水平,还需要进一步学习,因此向其推荐本知识点下的中难度习题;若否,则需要结合知识图谱判断学习者前驱知识点的掌握状态,若前驱知识点未达到已掌握状态,则认为该学习者前驱知识点仍需进一步学习,以达到学习本知识点的能力,因此向其推荐前驱知识点的习题;若前驱知识点已掌握,则认为该学习者已达到学习本知识点的能力,因此为其推荐本知识点下的低难度习题。

根据已确定的习题知识点及难度从题库中抽取习题,形成待推荐列表。

(4)习题推荐

从待推荐列表中筛选出该学习者从未作答过的习题,并根据习题难度进行排序,将排序后的习题推荐给学习者。

学习者登录 Edu WPS 学生端,选择目标科目进行习题推荐,在习题作答前,学习者可以依据自己的学习情况选择是否学习推荐习题的相关微课,最终形成含交互式控件的个性化习题推荐文档,学习者直接在文档中作答,将答案填至交互式控件中并提交结果,学习者提交答案后能够查看正确答案,了解自己的作答情况,并有针对性地选择学习内容。且学习者提交习题的答案后学习者的作答情况也会即时更新,为下一次的推荐做准备。

四、案例分析

本节选取两个案例分析展示基于知识图谱的个性化习题推荐系统对两个学习者的习题推荐情况。学习者 A 与学习者 B 对于《空间与图形》这一章的知识点掌握程度如图 7、图 8 所示。从图中可以看出学习者 A 已经掌握本章的前 3 个知识点,尚未学习《三角形的面积》一章;学习者 B 已经掌握本章的前 4 个知识点,对知识点“三角形的面积计算”处于未掌握状态。

在实际推荐中,对于学习者 A,由于其已经掌握了“平行四边形的底/高的计算”这一知识点,认为其已经具备了学习下一知识点的能力,因此为其推荐下一知识点下的低难度习题,推荐结果如图 9 所示;对于学习者 B,由于其尚未掌握“三角形的面积计算”知识点,且已经掌握了上一知识点,但对本知识点需要进一步学习,因此为其推荐知识点“三角形的面积计算”下的中难度习题,推荐结果如图 10 所示。



图 7 学习者 A 知识图谱

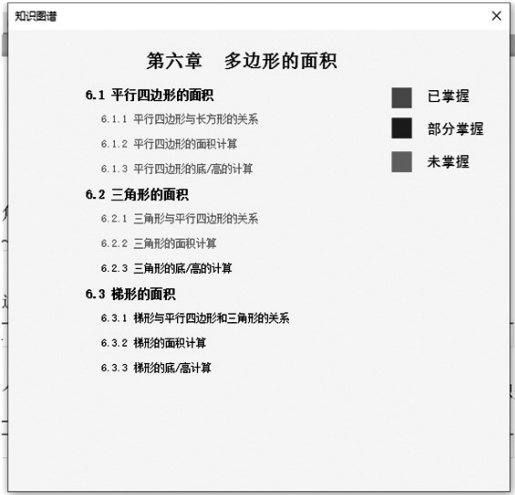


图 8 学习者 B 知识图谱

智能推荐试题汇总

- 1、三角形的面积等于平行四边形面积的一半。 ( )  
C T C F
- 2、两个花园的周长相等,它们的面积也一定相等。 ( )  
C T C F
- 3、一个三角形和一个平行四边形等底等高,如果平行四边形的面积是 86 平方分米,那么这个三角形边形的面积是\_\_\_\_\_平方米。
- 4、一个平行四边形的面积是 120 平方分米,与它等底等高的三角形的面积是\_\_\_\_\_平方米。
- 5、两个\_\_\_\_\_的三角形能拼成一个平行四边形。

图 9 学习者 A 推荐结果

## 智能推荐试题汇总

- 1、把三角形的底和高都扩大到原来的3倍,三角形的面积( )。  
A. 扩大到原来的3倍 B. 扩大到原来的6倍 C. 扩大到原来的9倍 D. 不变  
C A C B C C C D
- 2、一个三角形的底是4分米,高是2分米,它的面积是( )。  
A. 2平方分米 B. 4平方分米 C. 1平方分米 D. 4分米  
C A C B C C C D
- 3、一个直角三角形的三条边的长度是3cm、4cm和5cm,这个三角形的面积是( ) cm<sup>2</sup>。  
A. 12 B. 6 C. 20 D. 10  
C A C B C C C D
- 4、一个等腰直角三角形,腰长16厘米,面积是( )平方厘米。  
A.256 B.64 C.128 D.160  
C A C B C C C D
- 5、一个等腰直角三角形的直角边长2cm,这个三角形的面积是\_\_\_\_\_。

图 10 学习者 B 推荐结果

另外,我们调研了学习者 A、B 对系统推荐结果的满意度。学习者 A、B 认为知识图谱可视化呈现能够帮助他们更加明确知识结构,了解自己的知识点掌握情况,从而有针对性地进行学习;且系统推荐结果符合其知识点掌握情况,能够提升练习效率。并且学生 A 认为在习题作答前先学习推荐的微课资源能够提升练习的准确性。学习者 A、B 均对知识图谱可视化呈现效果及习题推荐结果有较高的满意度。

从该案例分析中可以看出,依照本文提出的基于知识图谱的习题个性化推荐方法,能够帮助学习者明晰自己的知识点掌握情况,针对学习者的知识点掌握状态为其推荐相应的习题,推荐结果具有较好的可解释性与用户满意度。

## 五、结束语

随着信息技术的发展,传统练习方式难以满足学习者的个性化学习的需求,为了使学习者的练习能够更加针对薄弱知识点,更加个性化,本文基于云协作测评环境对学习者的测评数据进行采集与分析,构建知识图谱并可视化显示,然后以此为基础为学习者推荐个性化习题。知识图谱的构建能够帮助学习者直观了解自己的学习进度,并为学习者提供更加符合其学习状态的习题。随着“人工智能+教育”的发展,知识图谱在教育教学中的应用将会更加广泛。

## 参考文献:

- [1]Edsurge.Five Key Elements of Personalized Learning[EB/OL].<https://www.edsurge.com/n/2015-04-01-five-key-elements-of-personalized-learning>,2016-01-23.
- [2]L·约翰逊,S·亚当斯贝克尔等.新媒体联盟地平线报告(2015 高等教育版)[J].北京广播电视大学学报,2015(s1):1-18.
- [3]孔晶,郭玉翠,郭光武.技术支持的个性化学习:促进学生发展的新趋势[J].中国电化教育,2016(4):88-94.
- [4]杨玉基,许斌,胡家威,全美涵,张鹏,郑莉.一种准确而高效的领域知识图谱构建方法[J].软件学报,2018(10).
- [5]钟亮.面向百度百科的化学知识图谱构建方法研究[J].软件导刊,2017(8):168-170.
- [6]吴强强,陈昊鹏,赵子豪,蒋建伟.基于 MOOC 平台数据和知识图谱的学习路径推荐——以软件工程专业为例[J].工业和信息化教育,2017(11):33-38.
- [7]Haiping Zhu,Feng Tian,Ke Wu,et al.A Multi-constraint Learning Path Recommendation Algorithm Based on Knowledge Map[J].Knowledge-Based Systems,2018(143):102-114.
- [8]孙权,贺细平.协同过滤算法在 ACM 在线评测推荐系统中的应用研究[J].电脑与信息技术,2015(6):11-14.
- [9]姜盼.基于协同过滤的习题推荐系统设计与实现[D].长江大学,2017.
- [10]单瑞婷,罗益承,孙翼.基于认知诊断的协同过滤试题推荐[J].计算机系统应用,2018(3):136-142.
- [11]朱天宇,黄振亚,陈恩红,刘淇,吴润泽,吴乐,苏喻,陈志刚,胡国平.基于认知诊断的个性化试题推荐方法[J].计算机学报,2017(1):176-191.
- [12]柏茂林.基于协同过滤的数学习题个性化推荐系统的设计与实现[D].渤海大学,2018.
- [13]牟智佳,俞显.教育大数据背景下智能测评研究的现实审视与发展趋向[J].中国远程教育,2018(5):55-62.
- [14]刘峤,李杨,段宏,刘瑶,秦志光.知识图谱构建技术综述[J].计算机研究与发展,2016(3):582-600.
- [15]徐增林,盛泳潘,贺丽荣,王雅芳.知识图谱技术综述[J].电子科技大学学报,2016(4):589-606.

(编辑:王天鹏)