

## 基于知识图谱的家政服务课程推荐融合模型

邹蕊玲, 朱郑州

引用本文

邹蕊玲, 朱郑州. [基于知识图谱的家政服务课程推荐融合模型](#)[J]. 计算机科学, 2024, 51(2): 47-54.

ZOU Chunling, ZHU Zhengzhou. [Fusion Model of Housekeeping Service Course Recommendation Based on Knowledge Graph](#) [J]. Computer Science, 2024, 51(2): 47-54.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

### [基于知识图谱与用户兴趣的推荐算法](#)

Knowledge Graph and User Interest Based Recommendation Algorithm

计算机科学, 2024, 51(2): 55-62. <https://doi.org/10.11896/jsjcx.221200169>

### [基于生成式对抗网络和正类无标签学习的知识图谱补全算法](#)

Knowledge Graph Completion Algorithm Based on Generative Adversarial Network and Positive and Unlabeled Learning

计算机科学, 2024, 51(1): 310-315. <https://doi.org/10.11896/jsjcx.230300006>

### [基于知识图谱的兴趣捕捉推荐算法](#)

Interest Capturing Recommendation Based on Knowledge Graph

计算机科学, 2024, 51(1): 133-142. <https://doi.org/10.11896/jsjcx.230500133>

### [DL<sup>+</sup>:一种增强型双层知识图谱推理框架](#)

DL<sup>+</sup>:An Enhanced Double-layer Framework for Knowledge Graph Reasoning

计算机科学, 2023, 50(12): 302-313. <https://doi.org/10.11896/jsjcx.221000170>

### [基于知识图残差注意力网络的推荐方法](#)

Recommendation Method Based on Knowledge Graph Residual Attention Networks

计算机科学, 2023, 50(11A): 220900180-7. <https://doi.org/10.11896/jsjcx.220900180>

# 基于知识图谱的家政服务课程推荐融合模型

邹莼玲 朱郑州

北京大学软件与微电子学院 北京 102600

(zouchunling@pku.edu.cn)

**摘要** 针对家政服务人员家政课程在线学习需求的增加,而现有的家政服务课程在线学习网站存在资源较少、课程不够系统化和不具有课程推荐功能等状况,使得家政服务相关从业人员的在线学习门槛变高。通过分析现有的家政服务课程在线学习网站,提出构建家政服务课程知识图谱,并将家政服务课程知识图谱与推荐算法进行融合,设计了一种融合深度学习技术的规则与水波偏好传播相结合的 R-RippleNet 家政服务课程推荐模型。R-RippleNet 模型的使用对象包括老学员和新学员,老学员部分是基于水波偏好传播模型进行课程推荐,新学员部分则基于规则模型进行课程推荐。实验结果表明,老学员使用 R-RippleNet 模型的 AUC 值为 95%,ACC 值为 89%,F1 值为 89%,新学员使用 R-RippleNet 模型的总体精确率均值为 77%,NDCG 均值为 93%。

**关键词:** 融合模型;知识图谱;家政服务;课程推荐;图数据库

**中图分类号** TP391

## Fusion Model of Housekeeping Service Course Recommendation Based on Knowledge Graph

ZOU Chunling and ZHU Zhengzhou

School of Software and Microelectronics, Peking University, Beijing 102600, China

**Abstract** Housekeeping service practitioners' demand for online learning of housekeeping service courses has increased. However, the existing online learning websites of housekeeping service courses have few resources, insufficient systematic courses and no course recommendation function, which makes the threshold of online learning for housekeeping service practitioners become higher. Based on the analysis of the existing online learning websites of housekeeping service courses, this paper proposes to construct the knowledge graph of housekeeping service courses, and integrates the knowledge graph of housekeeping service courses with the recommendation algorithm, and designs an R-RippleNet recommendation model for housekeeping service courses that combines the rules of deep learning technology and the water-wave preference propagation. The objects used by R-RippleNet model include old students and new students. The old students make course recommendation based on the water wave preference propagation model, while the new students make course recommendation based on the rule model. Experimental results show that the AUC value of old trainees using R-RippleNet model is 95%, ACC value is 89%, F1 value is 89%, the mean of the overall accuracy rate of new trainees using R-RippleNet model is 77%, the mean of NDCG is 93%.

**Keywords** Fusion model, Knowledge graph, Housekeeping service, Course recommendation, Graph database

## 1 引言

目前国内的家政服务培训形式分为线上和线下。进行线下培训的家政公司都有其特定的培训体系,把培训内容从简单到复杂进行讲解,并带有实际操作,但不足之处在于大部分费用较高,少部分获得国家补贴的免费培训也需要学员拿出集中的时间,到指定地点培训,存在耗时耗力的缺点;线上的家政服务培训访问方式主要以 Web 网页为主,具有随时随地都可以学习的优势,满足家政服务行业从业人员弹性学习的需求,不足之处是不能进行现场详细的实际操作,但培训师可以通过线上视频教学演示,来弥补不能线下实际操作的不足。

目前互联网技术日新月异,加快了线上形式家政服务培训的发展。调查 10 家家政服务在线学习网站发现,大部分课程比较系统全面,部分基本能满足从业人员对职业素养和专业技能的家政服务培训需求,只是少部分没有检索功能以及资源较少。由于家政服务是传统行业,目前家政服务在线学习网站还没有课程推荐模块。现有家政服务在线学习网站的相关现状如表 1 所列。

知识图谱构建过程<sup>[1-8]</sup>包括知识来源、抽取、表示、融合、推理以及存储。职业教育知识图谱<sup>[9-14]</sup>以“职业相关知识学习”为核心,整理职业、课程层级结构和关系,构建职业相关的全部知识点之间的关联,再链接相关教学资源,最终通过可视化的图谱形象地展示完整的职业知识架构。

表1 家政服务在线学习网站的相关现状

Table 1 Related status of housekeeping service online learning website

来源	优势	不足	适应程度
中国家政服务培训网	课程比较系统全面	最近更新是2014年,没有系统推荐功能	低
58职教	课程比较系统全面	没有详细检索功能,没有系统推荐功能	中
腾讯课堂	来源较多,各个家政培训公司都可以上传资源	没有推荐功能,资源较少	低
维佳课堂	课程比较系统全面	没有系统推荐功能	中
中青家政在线学习	课程比较系统全面	没有系统推荐功能	中
风声云职教	师资阵容强大,课程比较系统全面	没有系统推荐功能	中
壹管家培训学院	课程比较系统全面	没有系统推荐功能	中
家政培训考务平台	课程比较系统全面	部分课程仅适用于黑龙江地区,没有系统推荐功能	中
河北省家政服务培训网	课程比较系统全面	没有系统推荐功能	中
中华网校	课程比较系统全面,有在线测试功能	没有系统推荐功能	中

知识图谱特征学习<sup>[15-17]</sup>与推荐系统有依次学习、交替学习和联合学习3种结合模式。其中依次学习中的特征学习与推荐系统两个模块是独立的,训练开销较大,但推荐模块更新较快,比较经典的算法有DKN<sup>[18]</sup>;交替学习可以使知识图谱与推荐系统的信息互通,有助于提升系统泛化能力,比较经典的算法有MKR<sup>[19]</sup>;联合学习能够将推荐系统模块的监督信息反馈到知识图谱特征学习中,能进行端到端的学习,但需要较大的开销才能确定目标函数的结合方式和权重,比较经典的算法有CKE<sup>[20]</sup>和RippleNet<sup>[21]</sup>。

综上,基于知识图谱的家政服务课程推荐融合模型首先对现有的家政服务课程在线学习网站进行分析,其次构建家政服务课程知识图谱,再次将家政服务课程知识图谱与推荐算法进行联合学习,最后设计出一种融合深度学习技术的规则与水波偏好传播相结合的R-RippleNet家政服务课程推荐模型。

## 2 家政服务课程知识图谱构建

### 2.1 数据来源及整理

数据来源是5本《家政服务员》系列培训教材、培训过程中录制的4840个家政服务课程视频和110498条课程评分数据。由于最新的培训教材是2020年出版的,目前还没有电子版,先将5本培训教材目录进行扫描,形成PDF文档并依次进行命名;再使用OCR在线文字识别转换工具,将图片中的文字转换成文档;最后,人工将文档按照书名整理成csv文件并校验后形成教材目录数据源。

### 2.2 实体抽取

#### 2.2.1 实体类别描述

《家政服务员》的教材目录分为篇、职业模块、培训课程和学习单元,根据教材目录,人工将实体分为职业类别、职业技能、课程和学习资源;其中篇与职业类别对应,形式化为PRC;职业模块与职业技能(PRS)对应,培训课程与课程(COU)对应,学习单元与学习资源(RES)对应,还有培训师(PTT)和等级(RAT)两大类实体;综上,职业类别、职业技能、课程、学习资源、培训师以及等级,一共六大类实体。

#### 2.2.2 基于关键词的教材目录实体识别

通过规则的方法对教材目录中涉及的实体按照实体类别进行识别。首先定义实体分类函数,目的是将教材目录内容与关键词匹配,得到相应的实体类别;其次输入已经整理好的教材目录csv文件;再次运行实体分类函数并进行循环;

最后输出教材目录实体分类结果。

#### 2.2.3 基于非完全匹配的文件名称实体识别

培训过程中录制的家政服务课程和学习资源文件名称都是培训师手动输入的,与教材目录会存在一定的误差,通过FuzzyWuzzy非完全匹配函数fuzz.partial\_ratio对文件名称与教材目录中家政服务课程和家政服务学习资源的实体进行计算。

两个实体分别用 $e_i$ 和 $e_j$ 表示,它们的长度用 $l_i$ 和 $l_j$ 表示,用 $l_n$ 表示 $e_i$ 和 $e_j$ 中相同部分长度,用 $g(e_i, e_j)$ 表示两个实体中较短实体的差异字数,用 $l(e_i, e_j)$ 表示两个实体中的较短实体长度,用fuzz.partial\_ratio( $e_i, e_j$ )表示 $e_i$ 和 $e_j$ 两个实体的相似度。两个实体中较短实体的差异字数 $g(e_i, e_j)$ 与 $l_i, l_j$ 和 $l_n$ 之间的关系如式(1)所示,两个实体中的较短实体长度 $l(e_i, e_j)$ 与 $l_i$ 和 $l_j$ 之间的关系如式(2)所示,相似度fuzz.partial\_ratio( $e_i, e_j$ )与 $g(e_i, e_j)$ 和 $l(e_i, e_j)$ 之间的关系的计算式如式(3)所示。

$$g(e_i, e_j) = \begin{cases} l_i - l_n, & \text{if } l_i < l_j \\ l_j - l_n, & \text{if } l_i \geq l_j \end{cases} \quad (1)$$

$$l(e_i, e_j) = \begin{cases} l_i, & \text{if } l_i < l_j \\ l_j, & \text{if } l_i \geq l_j \end{cases} \quad (2)$$

$$\text{fuzz.partial\_ratio}(e_i, e_j) = \frac{(1 - g(e_i, e_j)) * 100}{l(e_i, e_j)} \quad (3)$$

#### 2.2.4 基于集合性质的培训师实体识别

文件名称数据中有对应的培训师名称,培训师在进行注册时会进行名称查重,故培训师名称不会出现重复。而集合(set)是一个无序不重复的元素序列。直接读取文件名称数据,对数据中的培训师一列进行集合操作,将重复出现的同一个培训师保留一个在集合中。

#### 2.2.5 实体抽取结果

从教材目录中抽取367个概念实体,从文件名称中抽取6365个物理实体,从培训师中抽取80个物理实体,人工校验后得到6812个实体。

## 2.3 关系补全

### 2.3.1 关系类别描述

关系指实体与实体之间的关系,其中职业技能属于职业类别,学习资源属于课程。职业类别本身拥有自己的职业等级,职业技能本身也拥有自己的技能等级。课程本身有自己的课程等级,有其授课教师(培训师),还有与其相似的课程。学习资源本身有自己的培训师和资源等级,可能会有前驱

学习资源,也有与其相似的学习资源。培训师有自己培训的职业类别,也有自己培训的职业技能,还有自己教授的培训课程。不同的职业类别有自己需要掌握的职业技能,对应就有自己需要学习的课程。综上,家政服务课程知识图谱有 15 类关系。

### 2.3.2 实体关系补充

由于没有完整的三元组数据,首先根据关系类别描述的内容,手工给教材目录中实体与实体之间添加相应的关系;其次基于已有的教材目录中的完整的  $(h, r, t)$  三元组,通过基于 TransH 模型链接预测(Link Prediction)的方法补充文件名称中的关系。TransH 模型的评分函数如式(4)所示,损失函数如式(5)所示。

$$f_r(h, t) = \| (h - w_r^T h w_r) + d_r - (t - w_r^T t w_r) \|_2^2 \quad (4)$$

$$\mathcal{L} = \sum_{(h, r, t) \in \Delta} \sum_{(h', r', t') \in \Delta' \setminus (h, r, t)} [f_r(h, t) + \gamma - f_r(h', t')]_+ \quad (5)$$

### 2.3.3 关系补充结果

通过人工方式给教材目录添加 214 个关系,通过链接预测给文件名称中补充了 142200 个关系,通过人工方式给教材目录中部分实体与培训师之间添加 2200 个关系,一共补充了 144614 个关系。

## 2.4 知识融合

### 2.4.1 实体消歧

文件名称中不同培训师的相同名称课程会有名称重复,不同等级之间的课程也会有名称重复,对课程名称进行重命名来解决课程名称重复的问题。将文件名称  $F_i$  的标准课程名称  $A_i$  与等级  $B_i$ 、培训师名称  $C_i$  进行组合,生成新的课程名称  $N_i$ ;如果文件名称不是标准课程名称,需要先通过非完全匹配的相似度计算,转换成标准课程名称;等级在括号中

表示,培训师在破折号后面表示,形成新的较为规范的 1520 个课程名称,课程重命名的计算式如式(6)所示:

$$N_i = \begin{cases} A_i(B_i) - C_i, & \text{if } F_i \neq A_i \\ F_i(B_i) - C_i, & \text{if } F_i = A_i \end{cases} \quad (6)$$

学习资源名称与课程名称一样,也存在歧义的问题,故也将学习资源与等级、培训师名称进行组合,并去掉其他冗余信息,其中等级在括号中表示,培训师在破折号后面表示,形成新的较为规范的 4840 个学习资源并进行人工校验。

### 2.4.2 实体对齐

实体对齐(Entity Resolution)的目的是验证两个实体是否为同一个事物或对象,在实体对齐的过程中,可能删除冗余实体或者找到相同实体。具体为,已知在知识图谱中存在 EqualTo 关系,如果需要确定  $e_1, e_2$  两个实体为同一个对象的可能性,在这个过程中需要使用 KGE 的得分函数对三元组  $(e_1, EqualTo, e_2)$  进行打分。

教材目录中的等级是概念实体,分为预学、初级、中级、高级和特级;文件名称中的等级是物理实体,也分为预学、初级、中级、高级和特级;将教材目录中涉及的等级与文件名称中涉及的等级进行实体对齐,并进行人工校验。

教材目录中的家政服务课程与职业类别构成需要学习课程的关系,与培训师形成课程的老师关系,不能与文件名称中的课程进行实体对齐。教材目录中的概念实体学习资源与其他实体不存在相关关系,可以将其去掉,并进行人工校验。

## 2.5 知识存储

先给 6566 个实体和 15 种关系添加 ID,然后将实体和三元组导入 Neo4j 图数据库,得到家政服务课程知识图谱示例,如图 1 所示。

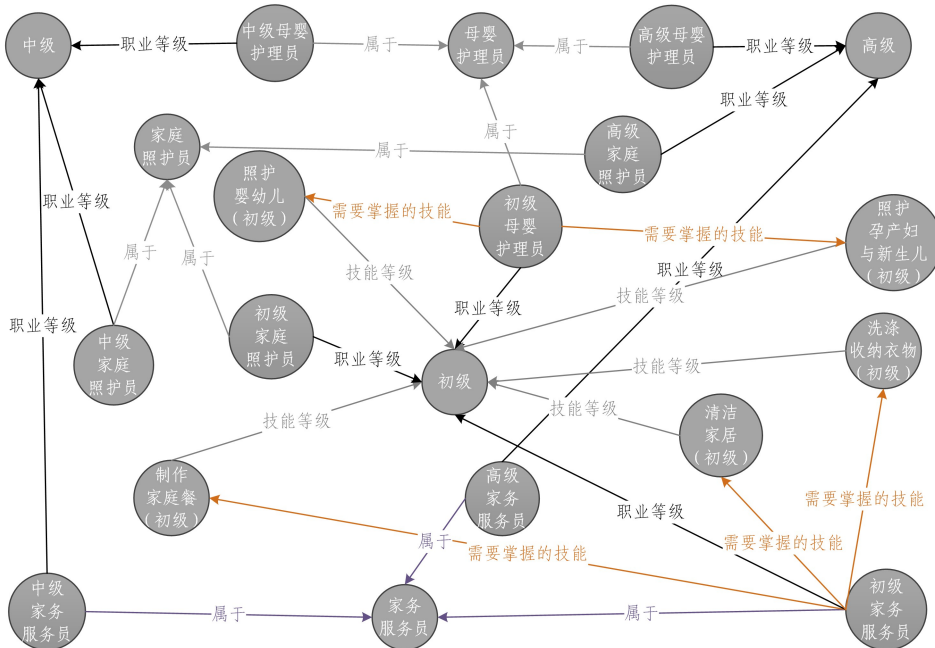


图 1 家政服务课程知识图谱示例

Fig.1 Example of KG of housekeeping service course

## 3 R-RippleNet 课程推荐模型设计

课程推荐模型是基于家政服务课程知识图谱进行设计

的,同时融合了学员历史学习时对课程的评分记录,通过学员的种子课程在家政服务课程知识图谱中的偏好传播情况,不断发现学员潜在想要学习的家政服务员相关课程,进而对

学员进行课程推荐。这样做不但可以有效解决传统推荐模型冷启动及数据稀疏问题,同时还避免了依次学习推荐模型无法解决的端到端的问题。

虽然偏好传播模型可以解决数据稀疏问题,但对于没有任何学习记录的新学员并不适用。对于新学员采用规则方法进行课程推荐,在学员注册时,让其选择自己想要从事的职业类别,不同等级职业类别对应不同的职业名称,根据职业名称结合家政服务课程知识图谱中的相关课程进行相应的规则推荐。因此,家政服务课程推荐采用规则与水波相结合的 R-RippleNet 模型。

R-RippleNet 模型包括两部分,分别是供老学员使用的基于水波的课程推荐模型和供新学员使用的基于规则的课程推荐模型,其中基于规则的课程推荐模型需要使用家政服务课程知识图谱数据、所有学员历史课程评分数据和学员注册时填写的职业名称数据,基于水波的课程推荐模型需要使用家政服务课程知识图谱数据、学员  $u$  的历史课程评分数据。新学员由于没有自己的历史课程评分数据,因此不适用基于水波的课程推荐模型,而对于老学员,基于水波的课程推荐模型的课程推荐效果比基于规则的课程推荐模型更好。R-RippleNet 模型两部分使用数据的情况如表 2 所列。

表 2 R-RippleNet 模型两部分使用数据情况的对比

Table 2 Comparison of two parts of R-RippleNet model using data

模型名称	数据			
	家政服务课程知识图谱	所有学员历史课程评分数据	学员 $u$ 的历史课程评分数据	注册时填写的职业名称
基于规则的课程推荐模型	是	是	否	是
基于水波的课程推荐模型	是	否	是	否

学员使用 R-RippleNet 模型学习时,家政服务课程推荐模型的流程如下。

Step1 通过获取学员学习情况记录表来判断学员的状态,没有学习情况记录就是新学员,有学习情况记录就是老学员。

Step2 如果是新学员,获取学员注册时填写的职业名称并与家政服务课程知识图谱联合,基于规则进行课程推荐,并生成课程推荐列表。

Step3 如果是老学员,获取学员学习时产生的历史家政服务课程评分数据与家政服务课程知识图谱基于水波偏好传播进行课程推荐,并生成课程推荐列表。

Step4 将生成课程推荐列表展示给学员。家政服务课程推荐模型的流程图如图 2 所示。

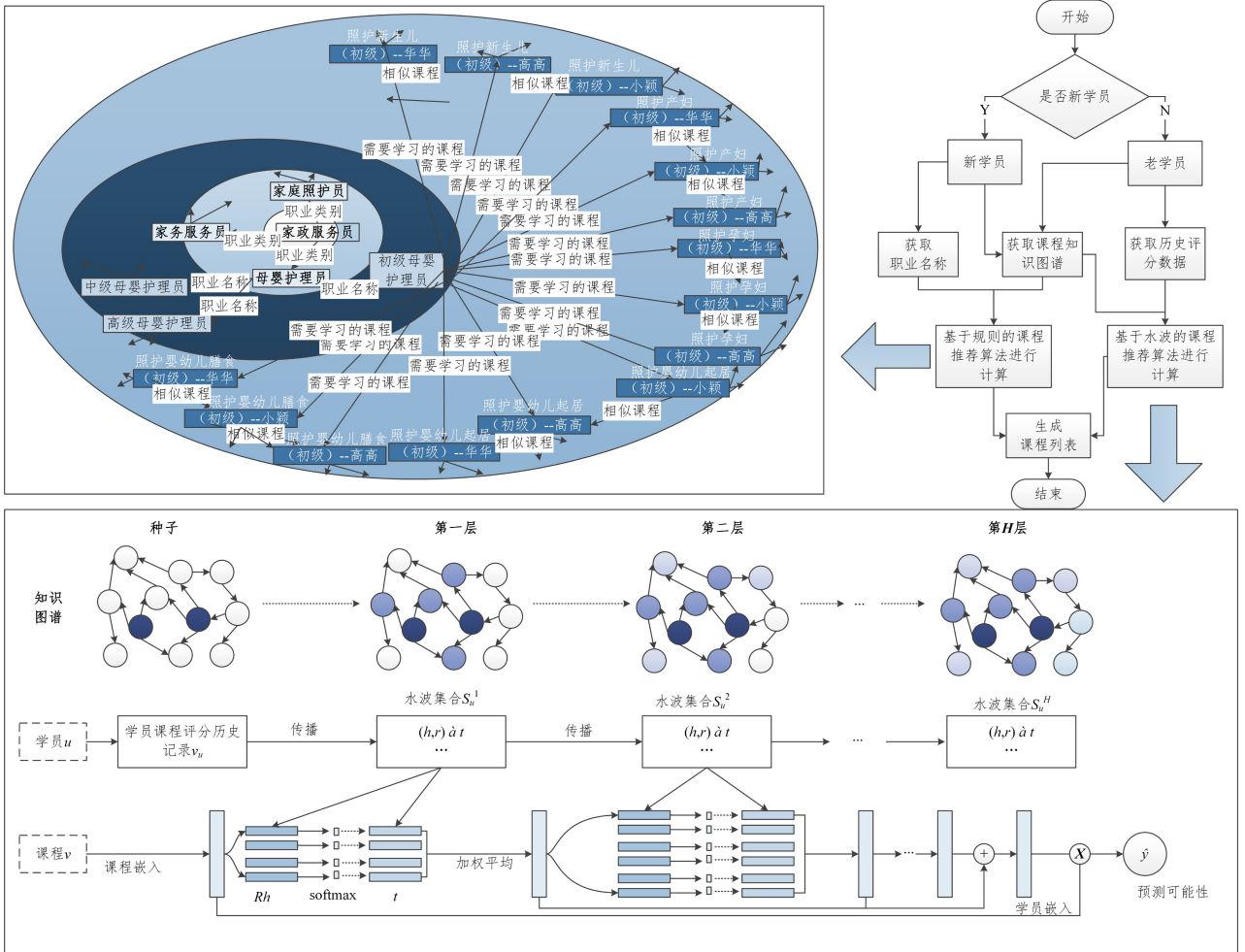


图 2 R-RippleNet 课程推荐流程图

Fig. 2 Flowchart of R-RippleNet course recommendation

### 3.1 基于水波的课程推荐模型

R-RippleNet 课程推荐模型供老学员使用的部分是基于水波偏好传播的课程推荐模型,偏好传播原理是借鉴了水波传播过程,以学员评分的家政服务课程为种子,在家政服务课程知识图谱上,用种子课程向外一圈一圈地扩散到其他家政

服务课程上,这个过程被称为家政服务课程偏好传播。偏好传播中外层的家政服务课程同样属于学员潜在的偏好课程,因此在刻画学员时,同样需要考虑到外层的家政服务课程,并将其作为学员的偏好。基于偏好传播的家政服务课程知识图谱推荐过程示例如图3所示。

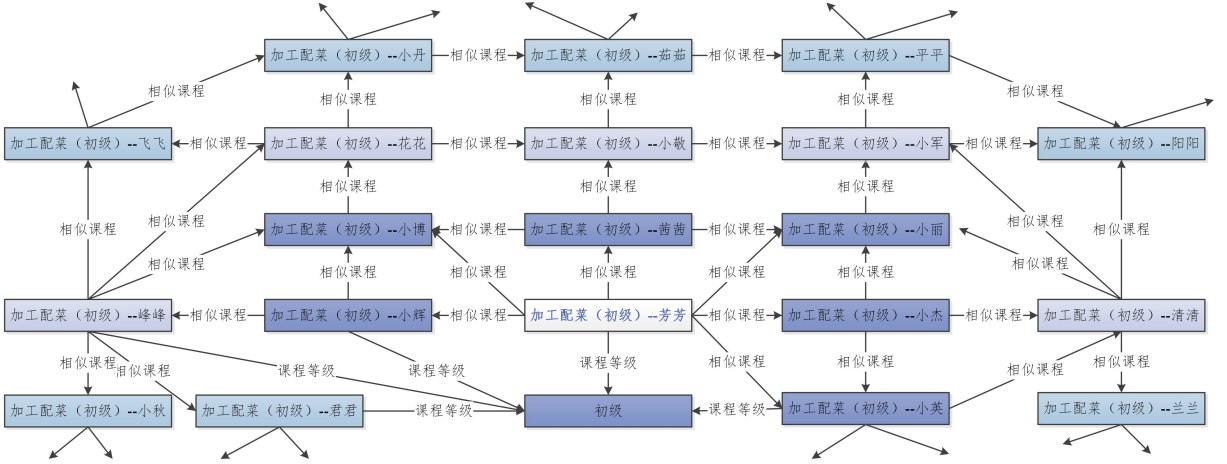


图3 基于水波偏好传播的家政服务课程知识图谱推荐示例

Fig. 3 Recommended example of KG for housekeeping service courses based on water wave preference propagation

课程偏好传播在家政服务课程知识图谱推荐示例中的具体表现如下:白色矩形部分为种子课程,深紫色文字部分为种子课程扩散的第一圈,浅紫色文字部分为种子课程扩散的第二圈,依次往外均为种子课程扩散的家政服务课程。基于水波偏好传播的课程推荐模型的推荐过程如下。

Step1 先获取学员  $u$  的历史评分的家政服务课程作为种子集合  $S$ (用这些评分课程作为学员的已有的偏好信息,不考虑评分的高低)。

Step2 判断学员  $u$  的种子集合与输入家政服务课程  $v$  的关系。家政服务课程  $v$  如果是学员种子集合中的一员,则作为训练数据的正例,即点击家政服务课程  $v$  的概率为 1,如果不是种子集合中的一员,即负例,即点击该家政服务课程  $v$  的概率为 0,随机选择除评分课程外的家政服务课程作为负例。

Step3 创建学员  $u$  相关的偏好传播的水波集合,从初始种子向外扩展一次,用  $(h, r, t)$  表示。为了避免水波集合过大,本文将最大扩展跳数设置为 2,即种子课程只进行两次扩散;在图3所示的家政服务课程知识图谱推荐示例中,从白色矩形部分 1 跳先扩散到深紫色矩形部分,2 跳再扩散到浅紫色矩形部分;另一方面,构建的知识图谱都是有向图,这里只考虑课程节点的出度,防止双向扩散带来更大的系统开销。

Step4 根据嵌入向量内积,计算课程  $v$  与第一层水波集合上的  $(h, r)$  的归一化相似度,假设图谱  $hr=t$ ,根据归一化后的相似度  $p_i$ ,对第一层水波集合中的  $(t)$  进行加权求和,得到的结果作为这一层的输出;在上述过程中,对于水波集合中的每一个  $(h, r, t)$ ,用家政服务课程知识图谱中的  $(h * r)$  与课程  $v$  相乘得到课程与步骤 Step4 中每个家政服务课程知识图谱中  $(h_i, r_i)$  的相关性得分,再通过 softmax 计算相关性得分的归一化相似度  $p_i$ ,计算式如式(7)所示:

$$p_i = \text{softmax}(v^T R_i h_i) = \frac{\exp(v^T R_i h_i)}{\sum_{(h, r, t) \in S_u^1} \exp(v^T R h)} \quad (7)$$

通过对所有  $t$  进行加权求和,得到第一层输出结果  $o_u^1$ ,计算式如式(8)所示:

$$o_u^1 = \sum_{(h, r, t) \in S_u^1} p_i t_i \quad (8)$$

Step5 重复 Step3 和 Step4,将第一层的水波集合的尾实体  $t$  作为第二层的头实体  $h$ ,取出第二层的水波集合,然后用第二层的水波集合与前一次输出的家政服务课程相似度及加权表示作为输出。假设一共重复  $H$  次,将多次的扩散结果相加,得到学员  $u$  的最终表示  $o$ ,计算式如式(9)所示:

$$o = o_u^1 + o_u^2 + \dots + o_u^H \quad (9)$$

Step6 最终的预测是通过家政服务课程  $v$  的嵌入向量和学员  $u$  的表示进行内积得到,家政服务课程嵌入向量可以通过嵌入层查找得到。上文将多次水波集合得到的输出作为学员表示,再把两次的水波集合输出  $o$  累加作为最终的学员  $u$  表示。最终的预测值是通过家政服务课程  $v$  与学员  $u$  内积后,再通过一个激活函数 sigmoid 得到概率,计算式如式(10)所示:

$$\hat{y}_{uv} = \sigma(u^T v) = \frac{1}{1 + e^{-u^T v}} \quad (10)$$

### 3.2 基于规则的课程推荐模型

R-RippleNet 课程推荐模型供新学员使用的部分是基于规则的课程推荐模型,会根据学员注册时填写的职业名称推荐相应的起始课程。图2中,基于规则的课程推荐算法进行计算时,设学员选择的职业类别为母婴护理员,将会给出初级母婴护理员、中级母婴护理员和高级母婴护理员 3 个相关的职业名称选项列表。假设学员从选项列表中选择初级母婴护理员,将会给出职业名称为初级母婴护理员在课程知识图谱上进行 1 跳的所有课程,并对所有课程按照课程评分排名

前 5 的课程生成系统课程推荐列表,学员可以从相关课程列表中选择一种进行学习。推荐过程如下。

Step1 先获取课程的历史评分,再计算所有课程的平均评分并按照降序存入集合  $R$  中,以便根据评分结果的高低得出相关课程推荐结果列表。

Step2 获取学员  $u$  在注册时填写的职业名称信息。

Step3 结合家政服务课程知识图谱,构建与学员  $u$  填写职业名称相关的第一次偏好传播的水波集合,即以职业名称信息向外扩展,用  $(h, r, t)$  表示。这里只进行一次扩散,获取职业名称相关的所有需要学习的课程并存入集合  $C$  中。

Step4 将课程评分数据集合  $R$  中的课程  $v_i$  与集合  $C$  中的所有课程进行余弦相似度计算,如果相似度计算结果为 1,则将添加课程  $v_i$  到课程推荐列表  $L$  中;如果相似度计算结果小于 1,则不添加到课程推荐列表中。上述过程中,课程评分数据集合  $R$  中的课程名称  $v_i$  与职业名称相关,需要将学习课程集合  $C$  中的课程名称  $v_j$  进行余弦相似度计算,如式(11)所示:

$$\text{similarity}(v_i, v_j) = \frac{(\sum_{i=1}^n (v_i \times v_j))}{\sqrt{(v_i)^2} \times \sqrt{(v_j)^2}} \quad (11)$$

Step5 查看课程推荐列表  $L$  中的课程数量,如果课程数量小于 5,重复 Step4;如果课程数量等于 5,结束计算,并将课程推荐列表中的 5 种课程作为新用户推荐结果展示给学员,供学员选择其中一种进行学习。

## 4 实验分析及结果

由于模型使用对象有老学员和新学员,其中老学员可以通过已有的课程评分数据进行实验,并将 R-RippleNet 模型分别与依次学习的经典模型 DKN、交替学习的经典模型 MKR 以及联合学习的经典模型 RippleNet 进行对比实验。而新学员没有学习情况记录,需要其在完成注册以后,根据其填写的职业名称通过 R-RippleNet 模型产生课程推荐列表,并将课程推荐列表发送给新学员进行测试。

### 4.1 实验数据集及环境

实验包含 110498 条老学员家政服务课程评分数据,6566 个实体(节点),15 种关系,144 614 个家政服务课程知识图谱关系数据。新学员根据实验评价指标进行测试并收集数据,测试分 7 天进行,计划每天 30 人,共计 210 人。

老学员课程推荐实验在个人电脑上进行,Windows10 系统,Intel CORE i5 8th Gen 处理器,使用 JetBrains PyCharm Community Edition 2019. 2. 3 进行测试。新学员课程推荐实验,通过个人电脑将课程推荐列表链接发送给新学员。

### 4.2 实验数据集及环境

老学员课程推荐实验选取 AUC, ACC 和 F1 这 3 个指标评价家政服务课程推荐结果。AUC 指 ROC 曲线下方的面积,可以通过 AUC 来判断课程推荐模型的优劣, AUC 的值为 1,说明模型堪称完美, AUC 的值为 0.85~0.95,说明模型效果很好, AUC 的值为 0.75~0.85,说明模型效果一般,准确率 ACC 如式(12)所示, F1 值如式(13)所示:

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$F1 = \frac{2TP^2}{2TP + FP + FN} \quad (13)$$

其中,  $TP$  代表系统推荐了该家政服务课程,且学员真实学习了该家政服务课程;  $FP$  代表系统推荐了该家政服务课程,但学员没有学习该家政服务课程;  $FN$  代表系统没有推荐该家政服务课程,但学员真实学习了该家政服务课程;  $TN$  代表系统没有推荐该家政服务课程,且学员没有学习该家政服务课程。

新学员推荐评价指标包括精确率和归一化折损累计增益。其中,通过精确率(Precision)来评价模型首次推荐的成功率,精确率表示推荐成功的家政服务课程占全部推荐家政服务的课程比率,用  $u$  表示学员,用  $U$  表示所有参与测试的学员,  $u \in U$ ,用  $P(u)$  表示所有推荐家政服务课程,用  $T(u)$  表示学员查看的家政服务课程,精确率如式(14)所示:

$$\text{Precision} = \frac{\sum_{u \in U} P(u) \cap T(u)}{\sum_{u \in U} P(u)} \quad (14)$$

通过归一化折损累计增益(Normalized Discounted Cumulative Gain, NDCG)评价学员对家政服务课程推荐列表的满意程度, NDCG 的计算过程如下:

Step1 需要给家政服务课程推荐列表中的每一种家政服务课程  $i$  赋予一个相关性分数  $G$ ,用  $r(i)$  表示  $i$  的相关性得分,如式(15)所示:

$$G = r(i) \quad (15)$$

Step2 对家政服务课程推荐列表中的  $k$  个家政服务课程  $i$  的相关性分数  $r(i)$  进行累加,得到 CG,如式(16)所示:

$$\text{CG@K} = \sum_i^k r(i) \quad (16)$$

Step3 根据家政服务课程排名前后对每一个推荐家政服务课程  $i$  的相关性得分  $r(i)$  比上  $\log_2(i+1)$  进行增益增高或者折损得到 DCG,如式(17)所示:

$$\text{DCG@K} = \sum_i^k \frac{r(i)}{\log_2(i+1)} \quad (17)$$

Step4 计算理想情况下推荐家政服务课程的增益得到 IDCG。本文中每个家政服务课程推荐列表均有 5 种课程,根据 DCG,排名越靠前增益越高,那么理想情况下,假设排在第一位到第五位的相关性分数依次为 5, 4, 3, 2, 1, 通过计算,得到 IDCG 的值为 10.27。IDCG 详细的计算过程如表 3 所列。

表 3 IDCG 的计算过程

$i$	$G$	$\text{CG@K}$	$\text{DCG@K}$
1	5	1.00	5.00
2	4	1.58	2.52
3	3	2.00	1.50
4	2	2.32	0.86
5	1	2.58	0.39
合计	15	9.49	10.27

Step5 通过计算 DCG 与 IDCG 的比值得到 NDCG,如式(18)所示:

$$NDCG@K = \frac{DCG@K}{IDCG@K} \quad (18)$$

### 4.3 实验方案

老学员对 R-RippleNet 模型 AUC, ACC 和 F1 的实验方案如下:将 110498 条学员家政服务课程评分数据按照 6:2:2 的比例划分成实验集、验证集和测试集;将 6566 个实体(节点),15 种关系,144614 个家政服务课程知识图谱关系数据加入测试数据。

新学员对 R-RippleNet 模型的精确率测试方案如下:要求参与测试的学员首先完成注册,然后学习 5 种家政服务课程,每个家政服务课程都会相应有一个家政服务课程推荐列表,如果学员查看列表中的某个家政服务课程,则记为 1,否则记为 0。在理想情况下,学员在学习 5 种家政服务课程过程中,均查看了家政服务课程推荐列表中的家政服务课程,通过式(14)计算得到精确率的值为 1。

新学员对 R-RippleNet 模型推荐满意度的测试方案如下:要求报名参加测试的学员需要将家政服务课程推荐列表中 5 种家政服务课程依次进行查看,根据学员查看家政服务课程的顺序,对家政服务课程进行相关性分数标记,理想情况下在推荐列表中排第一位的家政服务课程相关性分数为 5,但学员在最后才查看该家政服务课程,则该课程的相关性分数为 1。

### 4.4 实验参数设置

实验中所有模型均采用 Google 开源的 TensorFlow 进行实验环境搭建,实体和关系嵌入的维度为 16, L2 正则项的权值为  $10 \times 10^{-7}$ ,学习率为 0.02, batch size 大小设置为 1024,迭代 10 次,其中 R-RippleNet 模型最大跳数为 2, KGE 项的权重为 0.01,每一跳的波纹大小为 32。

### 4.5 实验结果分析

将本文提出的 R-RippleNet 模型与 DKN, MKR 和 RippleNet 这 3 个单一模型进行对比实验,实验结果如表 4 所列。

表 4 课程推荐模型对比实验结果

Table 4 Comparative experimental results of course recommendation models (%)

指标	算法				
	R-RippleNet	RippleNet	MKR	DKN	
老学员	AUC	95	95	93	75
	ACC	89	89	87	54
	F1	89	89	87	70
新学员	Precision	77	不支持	不支持	不支持
	NDCG	92	不支持	不支持	不支持

由表 4 可知,在已有的课程评分数据集上 R-RippleNet 模型与 RippleNet 模型的实验结果数值一致,其中 AUC 值达到了 95%,分别比 MKR 和 DKN 提高了 2%和 20%;AAC 值达到了 89%,分别比 MKR 和 DKN 提高了 2%和 35%;F1 值达到了 89%,分别比 MKR 和 DKN 提高了 2%和 19%;DKN, MKR 和 RippleNet 这 3 个单一模型中, RippleNet 模型的效果最好,因此选用 RippleNet 模型作为 R-RippleNet 模型的基模型。

测试分 7 天进行,计划每天 30 人,实际前 6 天每天都是

30 人,第 7 天是 29 人,共计 209 人。由于参与测试的学员都是新学员,因此 R-RippleNet 课程推荐模型的总体精确率只有 77%,7 天精确率为 [74%, 81%], NDCG 值总体为 92%,7 天 NDCG 值为 [91%, 93%]。

**结束语** 通过构建家政服务课程知识图谱,并将其用于家政服务课程推荐模型中,具体为将家政服务课程知识图谱中的实体向量、关系向量与学员向量、课程向量进行联合学习,得到 R-RippleNet 课程推荐模型,并对模型进行测试,且测试结果正常,符合预期。对于新学员使用课程推荐模型的精确率只有 0.77,有待于进一步提高。下一步将继续在知识图谱的构建和推荐模型等领域进行研究。

### 参考文献

- [1] QIAN L F, CUI X L. Reserch on Construction Method of Domain Knowledge Graph Based on Transfer Learning [J]. Journal of Modern Information, 2022, 42(3): 31-39.
- [2] LIANG J R, E H H, SONG M N. Method of Domain Knowledge Graph Construction Based on Property Graph Model [J]. Computer Science, 2022, 49(2): 174-181.
- [3] YUE L X, LIU Z Q, XU H Y. Domain Knowledge Mapping Construction Based on Interactive Visualization [J]. Information Science, 2020, 38(6): 145-150.
- [4] LIU Y C, LI H Y. Survey on Domain Knowledge Graph Research [J]. Computer Systems & Applications, 2020, 29(6): 1-12.
- [5] HANG T T, FENG J, LU J M. Knowledge Graph Construction Techniques: Taxonomy, Survey and Future Directions [J]. Computer Science, 2021, 48(2): 175-189.
- [6] JUAN S, ORA L. Designing and Building Enterprise Knowledge Graphs [M]. Morgan & Claypool Publishers, 2021: 19-96.
- [7] LI Z Y. Research on Few-shot Knowledge Graph Model Based on Adaptive Attention [D]. Dalian: Dalian University of Technology, 2022.
- [8] YU H, ZHANG J, WU M H, et al. A framework for rapid construction and application of domain knowledge graphs [J]. CAAI Transactions on Intelligent Systems, 2021, 16(5): 871-884.
- [9] DONG Y B, HOU X. A Knowledge Graph for Curriculum System [C] // 2018 International Conference on Education Reform and Management Science (ERMS2018). Atlantis Press, 2018: 448-452.
- [10] LI Z, ZHOU D D, WANG Y. Research of Educational Knowledge Graph from the Perspective of "Artificial Intelligence+": Connotation, Technical Framework and Application [J]. Journal of Distance Education, 2019, 37(4): 42-53.
- [11] LI Z, ZHOU D D. Research on Conceptual Model and Construction Method of Educational Knowledge Graph [J]. e-Education Research, 2019, 40(8): 78-86, 113.
- [12] CHEN K, TAN Y L. Knowledge Map of Ideological and Political Courses in China Based on CiteSpace [J]. Heilongjiang Researches on Higher Education, 2020, 38(2): 128-132.
- [13] XIE R, ZHU W P. Domain Knowledge Graph of Artificial Intelligence Course and Its Innovative Teaching Method [J]. Soft-

ware Guide, 2021, 20(12):179-186.

[14] ZHONG Z, TANG Y W, ZHONG S C, et al. Research on Constructing Model of Educational Knowledge Map Supported by Artificial Intelligence[J]. e-Education Research, 2020, 41(4): 62-70.

[15] SUN Z, WANG H L. Overview on the Advance of the Research on Named Entity Recognition[J]. Data Analysis and Knowledge Discovery, 2010(6):42-47.

[16] LI D M, ZHANG Y, LI D Y, et al. Review of Entity Relation Extraction Methods [J]. Journal of Computer Research and Development, 2020, 57(7):1424-1448.

[17] ANTOINE B, NICOLAS U, ALBERTO G D, et al. Translating Embeddings for Modeling Multi-relational Data[C]// Advances in Neural Information Processing Systems 26, vol. 4; 27th Annual Conference on Neural Information Processing Systems 2013. Lake Tahoe, Nevada, USA; Neural Information Processing Systems, 2013; 2799-2807.

[18] WANG H W, ZHANG F Z, XIE X, et al. DKN: Deep Knowledge-Aware Network for News Recommendation [EB/OL]. <https://arxiv.org/abs/1801.08284>; 2018.

[19] WANG H W, ZHANG F Z, ZHAO M, et al. Multi-Task Feature Learning for Knowledge Graph Enhanced Recommendation [EB/OL]. <https://arxiv.org/abs/1901.08907>; 2019.

[20] ZHANG F Z, NICHOLAS J Y, LIAN D F, et al. Collaborative Knowledge Base Embedding for Recommender Systems [EB/OL]. <http://www.kdd.org/kdd2016/subtopic/view/collaborative-knowledge-base-embedding-for-recommender-systems>.

[21] WANG H W, ZHANG F Z, ZHAO M, et al. RippleNet: Propagating User Preferences on the Knowledge Graph for Recommender Systems[EB/OL]. <https://arxiv.org/pdf/1803.03467.pdf>.



**ZOU Chunling**, born in 1991, master. Her main research interests include construction of domain knowledge map and recommendation model.



**ZHU Zhengzhou**, born in 1979, Ph.D., associate professor. His main research interests include personalized recommendation and educational big data in big data environment.

(责任编辑:喻黎)

## 2023 年“CCF 会员发展优秀奖”揭晓

作为一个以个人会员为主的会员制学术社团, CCF 的很多新会员是通过导师、同学、同事、朋友或亲属等 CCF 老会员的推荐了解、认同并加入 CCF 的, CCF 热忱欢迎新会员的加入, 同时对老会员在会员发展中做出的努力和贡献表示衷心的感谢! CCF 将永久记录所有推荐者的推荐贡献, 并将在 CCF 颁奖典礼上授予年度推荐会员数 Top10 的会员“2023 年度会员发展优秀奖”。

2023 年有 2795 个会员参与了推荐, 共推荐新加入的专业会员 3326 人, 学生会会员 7129 人。

2023 年 CCF 会员推荐会员 Top10 名单

序号	姓名	单位名称	推荐会员指数(含学生会会员)
1	武春岭	重庆电子工程职业学院	135.5
2	齐竹云	鹏城实验室	126.25
3	吴帆	上海交通大学	118.75
4	李建林	南京信息职业技术学院	87.25
5	鲍淑娣	宁波工程学院	61.25
6	秦瑶	河南工业大学	58
7	於志文	哈尔滨工程大学	56.75
8	白硕	恒生电子股份有限公司	55.25
9	卢明波	呼伦贝尔学院	54.75
10	崔立真	山东大学	53.5
	邱兆文	黑龙江拓盟科技有限公司	53.5

1)数据来源:CCF 会员推荐会员系统

2)统计时间:2023 年 1 月 1 日—2023 年 12 月 31 日

3)按推荐会员总数排序(4 个学生会会员折算 1 个专业会员)

欢迎大家积极推荐身边的专业人士加入 CCF, 让越来越多的专业人士在 CCF 平台受益, 共同发展壮大我们自己的学术共同体。