doi: 10.7690/bgzdh.2024.11.012

知识图谱构建下基于近邻的军事职业教育课程推荐模式

周立军[1]，吕海燕[1]，刘　凯[2]
（1. 海军航空大学航空基础学院，山东 烟台 264001；2. 海军航空大学航空作战勤务学院，山东 烟台 264001）

摘要：针对当前军事职业教育中官兵选课存在过程缺乏引导、决策不科学等问题，基于知识图谱与神经网络的在线课程推荐模型，提出一种分析知识图谱技术和基于协同过滤的推荐系统工作原理，运用神经网络和大数据技术进行设计，并进行实验设计和结果分析。结果表明：该模型能兼顾学习者的兴趣及其自身相关的因素，为解决上述问题提供了借鉴。

关键词：知识图谱；军事职业教育；推荐系统；神经网络
中图分类号：TJ0　　**文献标志码：**A

Curriculum Recommendation Mode of Military Vocational Education Based on Nearest Neighbor Under Knowledge Mapping

Zhou Lijun[1], LYU Haiyan[1], Liu Kai[2]
(1. *College of Aviation Foundation*, *Naval Aviation University*, *Yantai* 264001, *China*;
2. *College of Air Combat Service*, *Naval Aviation University*, *Yantai* 264001, *China*)

Abstract: In view of the lack of guidance and unscientific decision-making in the course selection process of officers and soldiers in the current military vocational education, based on the online course recommendation model of knowledge mapping and neural network, this paper proposes an analysis of the working principle of knowledge mapping technology and recommendation system based on collaborative filtering, uses neural network and big data technology to design, and carries out experimental design and result analysis. The results show that the model can take into account the learners' interests and the factors related to themselves, which provides a reference for solving the above problems.

Keywords: knowledge map; military vocational education; recommendation system; neural network

0　引言

随着新时代军事教育方针的贯彻，着眼加快推进军队院校教育、部队训练实践、军事职业教育三位一体新型人才培养体系建设，已成为我军军事人才培养的主要途径[1]。军事职业教育课程以军职在线课程为主要形式，具有规模大、模式开放、时空自由等特点，已经成为官兵获取知识技能的主要渠道。由于我军军事职业教育起步较晚，现有的军职在线平台还无法为官兵提供精准高效的课程推荐服务；因此，如何减少官兵选课成本，为官兵提供个性化、多样化的课程服务，以提高网络课程资源的使用效率，是目前军队高职院校亟待解决的问题。

知识图谱是科学知识图谱的别称，在一些情况下，把它叫作“知识域映射地图”[2]。它综合了多个学科的知识，利用各种技术，以图解的形式呈现出知识的内在结构和规律，在实际应用和科研中起着举足轻重的作用。在大数据时代，知识图谱是一种高效、准确的关系表达手段，可以把各种实体的信息联系起来，从而构成一个庞大的知识关系网络。基于课程的海量信息，建立了网上课程的知识图谱，并将其应用于课程推荐领域中的语义网络。成为解决上述问题的有效途径。

1　知识图谱的构建技术

1.1　知识图谱概念

知识图谱是一种用来表示实体世界中的概念和它们之间联系的结构语义知识库[3]，主要通过 RDF 模型描述。其基本组成单位是“主体-属性-客体”三元组[4]，表示为 G=(subject，property，object)，其中，主/客体表示数据中的节点，属性表示数据中的关系。如图 1 所示，一个 RDF 三元组可以用 2 个点和 1 条带标签的有向边图表示。

图 1　RDF 示例

收稿日期：2024-06-17；**修回日期：**2024-07-20
基金项目：海军航空大学教学改革成果培育项目(HKDXJG2020022)
第一作者：周立军(1982—)，男，湖南人，硕士。

在互联网中，可以通过URI来标识资源和属性；因此，可以通过链接的前后衔接关系来唯一标识网页中的一个 RDF 三元组。如：<http://jwgl.hjhy.mtn/Student/Inclass><http://course.educa.com#select><http://jwgl.hjhy.mtn/Course/curricula>。

在人工智能和大数据技术的推动下，知识图谱成为描述网络环境中实体之间关系的最切实际的表达形式；同时，知识图谱可以实现多源异构数据的整合与利用[5-6]。在知识图谱的结合下，推荐系统可精准计算学生和课程、学生与学生、课程与课程之间的关联性，便于向用户提供精准智能的课程推荐服务。

1.2 知识图谱构建过程

知识图谱的整体技术流程如图 2 所示。

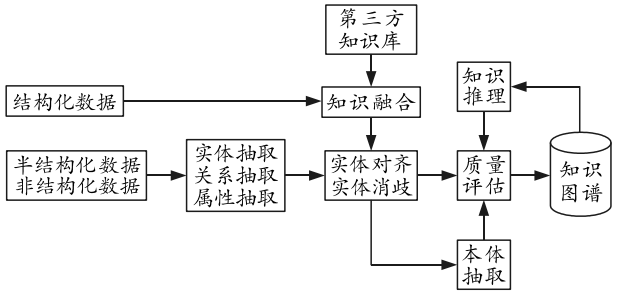

图 2 知识图谱构建技术流程

知识图谱的构建及迭代过程用部分代表，主要由 3 个步骤组成：

1) 知识提取。

从各种资料来源中获取实体、实体之间的关系和实体的属性，即 RDF 资料的获取。知识抽取中最重要的任务是关系抽取，关系抽取是在实体抽取基础上，通过语义分析判断实体间的关系(父子、依赖、包含等)，可以利用机器学习算法实现关系抽取。当前，深度学习网络已经成为关系抽取领域的一个热门话题[7]，基于深度学习的序列标注方法，进行自然语言处理中的分词、词性标注、命名实体、抽取关键词等；在此基础上，对知识图谱进行了关联提取，并将所标记的谓语用作实体关系。根据上述，将数据信息进行训练，避免了由独立任务引起的错误扩散；但它的弊端在于，采用了顺序标注来提取实体和关联，结果 3 个元素之间存在着重叠，从而增加了知识融合的难度。

2) 知识融合。

在获取新的知识后，要把已有的知识和新的知识相结合，使各种表现形式的实体相互结合。该过程涉及“实体对齐”“实体消歧”等问题。实体对齐又叫实体匹配，就像在多词同义法中，要判断 2 个或更多的实体在不同的数据来源中是不是指同一个物体，举例来说，“大基”和“大学计算机基础”这 2 个指代都指向同一个实体。实体消歧则指同一个称呼同时代表多个实体，需要靠其他属性消除歧义，例如，“计算机程序设计”在不同教学层次的数据源中可能代表 2 门不同的课。

3) 知识加工。

新知识要经过一系列加工过程，获得结构化、网络化的知识体系才能进入知识库。新的融合知识可以从逻辑推理中获得。知识处理包括本体构建、知识推理和知识评价 3 大部分。

2 基于协同过滤的推荐系统

2.1 推荐系统

在大数据背景下，服务器处理的数据量十分庞杂，数据冗余高，价值密度低，需要通过数据挖掘技术将有价值的数据挖掘出来。数据挖掘的核心就是机器学习，依靠机器构建有效模型，通过该模型从大量数据中识别出有价值的数据进行推荐。

与传统的搜索引擎相比，推荐系统并不需要用户特别的要求，而是其需要根据用户的历史信息，选择合适的算法进行模型化，并将其推荐给用户。如此，就可以积极地将自己感兴趣的内容介绍给用户；因此，推荐系统与搜索引擎可以功能互补[8]。

基于近邻的推荐算法，是解决有参照系背景下用户/物品推荐问题的经典方法，其典型代表就是协同过滤算法，在学术界和工业界应用十分广泛[9]。该算法主要通过用户的历史行为数据发现用户喜欢的物品，并对这些偏好进行度量和打分；然后，通过对用户对同一商品的评价和喜好来有效测验每个用户之间是否存在相同性，而针对有共同喜好的用户来说，可以主动给其推荐好的物品，如图 3 所示。

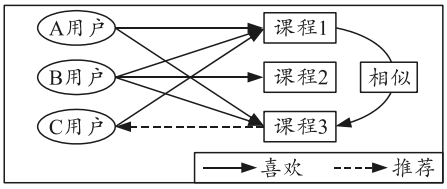

图 3 基于物品的协同过滤推荐

从课程推荐系统来看，利用协同过滤算法，计算课程之间的相似度，然后向用户推荐他之前喜欢课程的相似课程，从而达到推荐的目的。协同过滤算法利用数学统计的方法搜索出目标课程中的几个最类似的课程，并对其进行预测，最后给出对应的

推荐课程。

2.2　协同过滤算法流程

1) 计算课程之间的相似度。

包含 n 个学员的学员集可表示为 $U=\{u_1, u_2,\cdots, u_n\}$，包含 m 门课程的课程集表示为 $I=\{i_1, i_2,\cdots, i_m\}$。可以根据每名学员对课程的评分情况，构建每名学员的评分矩阵 $\boldsymbol{C}_{m\times1}$。然后，可以通过获取用户选课情况表来建立 n 名用户和 m 门课程的倒排表，形成 $n\times m$ 矩阵。然后，根据用户选课的倒查表，可以构建 $m\times m$ 同现矩阵(即同时喜欢 2 门课程的用户数)。

假设 $N(i)$表示喜欢课程 i 的用户数，$N(i)\cap N(j)$ 表示同时喜欢课程 i 和课程 j 的用户数，$\mathrm{sim}(i, j)$表示课程之间的相似度，其计算公式为：

$$\mathrm{sim}(i, j)=N(i)\bigcap N(j)/N(i)。\tag{1}$$

由此，可以构建课程之间的相似度矩阵 $\boldsymbol{W}_{m\times m}$，有 $w_{ij}=\mathrm{sim}(i, j)$。

2) 计算推荐结果。

得到课程的相似度矩阵之后，通过下式计算用户 u 对课程 i 的兴趣：

$$P_{(u,i)}=\sum_{j\in S(i,K)\cap N(u)} w_{ij}\cdot r_{uj}。\tag{2}$$

式中：$N(u)$为用户喜欢的课程集合；$S(i,k)$为课程 i 最相似的 k 门课程的集合；w_{ij} 为课程 i 和 j 的相似程度；r_{uj} 为用户 u 对课程 j 的兴趣(对于隐反馈数据集，如果用户 u 对课程 j 有过行为，即可令 $r_{uj}=1$)

在此基础上，对某一用户的课程推荐结果 $T_{m\times1}$ 即可用相似度矩阵 $\boldsymbol{W}_{m\times m}$ 和评分矩阵 $\boldsymbol{C}_{m\times1}$ 的乘积表示。

$$\boldsymbol{T}_{m\times1}=\boldsymbol{W}_{m\times m}\bullet\boldsymbol{C}_{m\times1}。\tag{3}$$

上式的含义是，与用户历史上感兴趣的课程越相似，越有可能在用户的推荐列表中获得较高的排名。基于每个用户的推荐结果列向量 $\boldsymbol{T}_{m\times1}$ 可组成 $m\times n$ 矩阵 $\boldsymbol{T}$，其元素用 t_{tc} 表示，即用户 t 选择课程 c 的推荐概率。

3) 惩罚热门课程。

上述计算课程相似度矩阵 $\boldsymbol{W}_{m\times m}$ 中存在一个缺陷，如果课程 j 过于热门，选择它的用户很多并进行了评分，则所计算出的 $\mathrm{sim}(i, j)$就会很大，从而造成任何一门课程都和热门课程有很大的形似度。因此，可以采用下面修正后的公式计算相似度，降低热门课程的影响。

$$\mathrm{sim}_{cf}(i, j)=N(i)\bigcap N(j)\Big/\sqrt{N(i)\bullet N(j)}。\tag{4}$$

2.3　推荐系统的评价方法

衡量推荐项目准确度常用的判据包括精确率(Precision)和召回率(Recall)[10]。精确率是指预测结果中符合实际值的比例。召回率是指正确分类的数量与所有“应该”被正确分类的数量的比例。计算公式为：

$$\mathrm{Precision}=TP/(TP+FP)；\ \mathrm{Recall}=TP/(TP+FN)。\tag{5}$$

式中：TP 表示实际为正样本，预测结果也为正样本；FP 表示实际为负样本，预测结果为正样本；FN 表示实际为正样本，预测结果为负样本。

实际应用中，协同过滤算法还出现了各种各样的问题，比如有冷启动问题、系统扩展性问题等缺陷；因此，需要针对不同的真实应用环境，采用基于内容推荐算法、神经网络推荐算法等其他一系列算法。

3　基于知识图谱与神经网络的推荐模型

3.1　神经网络

神经网络是模拟生物神经网络的一种计算模式，它是由许多神经元的结点互相连接形成的，层间的节点彼此相连，同一层次的节点间没有任何联系。可以使用神经网络来拟合任意函数，可将其当作一个黑盒，只需要提供充分的训练数据 X，就能对对应的功能进行拟合 f，对任意数据 x，$f(x)\approx$期望的 y。

在神经网络模型中，最常用的模型之一就是单隐层前馈神经网络(single-hidden layer feedforward neural networks，SLFNs)，从图 4 可以看出，它的典型例子有多层感知器(multilayer perceptron，MLP)、BP 神经网络等[11]。SLFNs 中的输入层结点数目与样本特征空间维数基本一致；因此，在网络学习之前，往往会先对样本进行预处理，而常规的方法是对样本进行标准化与降维。标准化可以使得采样平均值接近于 0，并且使得样本的方差值基本一致，从而在统计学上可以加快学生的学习速度。当特征的维度较高时，数据可能包含冗余信息及噪声信息，模型训练将会耗费大量时间；并且，在实际应用中，会给模型辨识带来错误，使模型的正确率下降，而采用特征降维法可以减小由于冗余信息引起的错误，进而提高了模型的正确率。

实际应用证明，神经网络的分类、近似等性能都与神经网络的构造息息相关；因此，网络的信息处理能力与网络的连接强度、网络的拓扑关系密切。在神经网络中，输入、隐、输出 3 个节点的选

取与优化是一个非常关键的问题。如果隐层节点数量过少，网络的学习能力就会下降，网络的性能也会下降，从而无法达到预期的学习结果；如果隐层的节点数量过多，则会产生过度拟合，从而影响到网络对未知样本的预测和分类，也就是网络的泛化性能下降。

在知识图谱背景下，需要将图像、文本等语义信息表示为低维稠密的实体向量，即知识表示学习(Embedding)，然后再结合神经网络进行训练、计算模型。

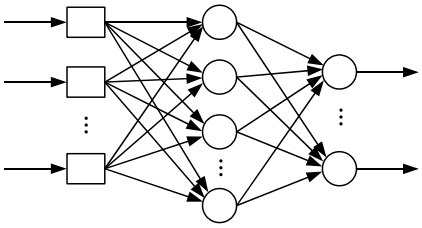

图 4 单隐层前馈神经网络模型

3.2 基于知识图谱与神经网络的推荐模型

笔者提出一种基于知识图谱和神经网络的推荐模型，该模型共包含 4 部分：

1) 向量的输入。把学生和课程中的稀疏优先向量作为输入，并将其转换成低维密度矢量的表达形式。

2) 深层特性的学习。以学员向量和课程向量 Embedding 为输入，利用 MLP 挖掘学员和课程的特征，得到学员和课程之间的非线性特征表示。MLP 模块使用 ReLu 作为激活函数，即 $f(\boldsymbol{x})=\max(0, \boldsymbol{x})$。从前一层神经网络传入神经元的输入向量 $\boldsymbol{x}$，使用 ReLu 激活功能的神经元将输出 $\max(0, \boldsymbol{W}^{\mathrm{T}}x+b)$，到下一层次的神经元，或者输出到整个神经元。上式中，$\boldsymbol{W}$ 为权值矩阵，其元素 w_{ij} 代表隐藏层第 i 个神经元和输入层第 j 个神经元间的权值，b 为偏置项。MLP 最终的输出记为 Φ_M。

3) 知识图谱特性的学习。在此基础上，提出了一种基于多层次关系扩散的方法。通过对知识图谱中的实体-关系向量的学习，可以了解学生和课程的线性特性。某学员 t 选择过的所有课程集合为 $K_t=\{k_1, k_2,\cdots, k_n\}$，$K_t$ 中，课程名所包含的实体集合为 $E_t=\{e_1, e_2,\cdots, e_\mu\}$。将 K_t 作为知识图谱的种子集合。在知识图谱中，种子实体会沿着相关关系 r 向外传播，每个传播所包含的 3 个元素集，就是离种子集合的距离为 d 的三元组的集合为 $S_t^d(d=1,2,\cdots,h)$，S_t^d 将与课程分词实体进行多次交互，以获取学生对该课程的偏爱。将所有偏好值综合起来，得出学生的特征 $T_R=\sum_{d=1}^{h}O_t^d$，式中 O_t^d 表示学生 t 在第 d 层的兴趣特征。知识图谱特征学习模块的最终输出为：

$$\Phi_R=T_R\Theta K_R。\tag{6}$$

式中：Θ 表示内积计算；K_R 说明各课程的教学特征。

4) 连接预测。首先，通过对深层特征学习模块和知识图谱特性学习的向量进行线性运算，然后通过 Sigmoid 函数对其进行规格化，最后计算得出所需的预测结果：

$$y_{tc}=\sigma\left(\lambda^{\mathrm{T}}\begin{bmatrix}\Phi_R\\ \Phi_M\end{bmatrix}\right)。\tag{7}$$

式中：$\sigma(x)=1/(1+e^{-x})$；λ^{T} 为经过模型训练而得到的权重向量；y_{tc} 表示模型预测的学员 t 选择课程 c 的概率。

3.3 协同过滤与知识图谱推荐结果相似度融合

协同过滤方法虽然具有较高的推荐准确率，但在语义相似度方面缺乏支持。知识图谱增强了推荐结果的语义解释性，在特征学习得到语义推荐模型的基础上，再结合协同过滤思想对 2 个推荐模型进行线性加权融合，如图 5 所示。

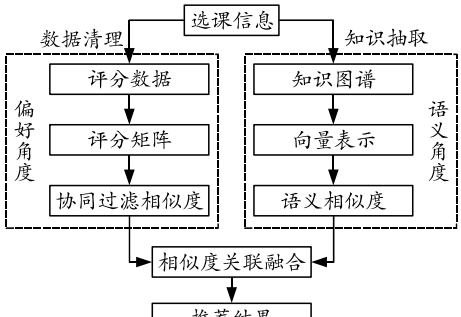

图 5 融合算法计算流程

融合方法为：

$$R(t,c)=\alpha\cdot t_{tc}+(1-\alpha)\cdot y_{tc}。\tag{8}$$

式中：α 为权重系数，且 $0<\alpha<1$，可以经过训练得到；$R(t, c)$为经融合算法后，向用户 t 推荐课程 c 的调优概率。

4 实验设计与分析

实验利用深度学习框架 TensorFlow 进行，训练数据集为中文通用百科知识图谱 CN-DBpedia Dump(包含 900 万+的百科实体以及 6 700 万+的三元组关系)[12]，实验数据集通过 Python 爬虫程序获取。鉴于现有军事职业教育平台中，学习者相关选课信息和评价信息过于稀少，评分机制尚未健全，

推荐精度大受影响，为验证实验设计模型的正确性，在试验中，选取互联网中国大学 MOOC 国家精品课程在线学习平台数据作为爬取对象。获取主要包括课程信息、用户与课程的交互信息及用户信息，然后利用 jieba 库工具对上述信息进行分词。之后，对爬取后的数据进行清洗、分层抽样等操作，得到学员数据集和课程知识图谱。在此基础上，通过模型来学习学员及课程特征，然后用训练好的模型预测学员选择课程的概率值；最后，评价推荐结果性能。

笔者采用融合算法与单一协同过滤算法对比，选取召回率(Recall)和 AUC(ROC 曲线下的面积值)2 个指标对模型性能进行评价。协同过滤的近邻值选为 30，实验中，融合算法的权重值 α 从 0 到 1 按照步长为 0.1 变化，进行 11 组×20 次实验，得到 α=0.6 时，召回效果最佳。

图 6 为实验数据中选取的某学员选课数据推荐结果的知识图谱。将该学员已选课程{Python 语言程序设计、机器学习、线性代数}作为训练集，将{神经网络与深度学习}作为测试集。图中，任意 2 门课程连线上的数值代表推荐权值，清晰描述了学员的兴趣关注点是以“Python 语言”和“线性代数”为知识基础的机器学习内容；因此，推荐系统中，围绕机器学习相关的课程推荐度较高。从共现实体的观点来看，这个学生很有可能会选择“神经网络和深度学习”，其学习效率会很高。

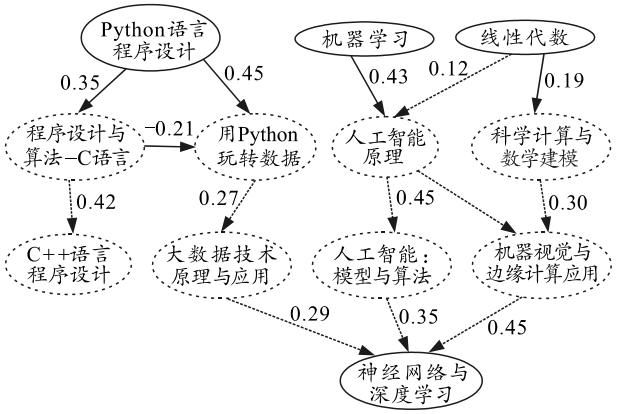

图 6　知识图谱表示下的选课推荐结果

5　结束语

军事职业教育背景下，选课推荐模式与普通推荐模式相比，实际上是一种具有高度相关的、有顺序的“商品”推荐模式。目前，大部分的知识都是从文字中抽取出来的关联和实体，然后再将其组合起来，忽视了许多人的常识；因此，提出一种基于神经网络和知识图谱的推荐模型，该模型兼顾了学习者的兴趣及与其自身相关的因素，为解决上述问题提供了借鉴。

参考文献：

[1] 中央军委. 关于加快推进三位一体新型军事人才培养体系建设的决定 [EB/OL], [2020-10-19]. https://baijiahao.baidu.com/s?id=1680980060303159354.

[2] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 589-606.

[3] 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, 53(3): 582-600.

[4] 李叶叶. 基于多种数据源的在线评论知识图谱构建[D]. 长春: 吉林大学, 2020: 14-18.

[5] 汤伟韬, 余敦辉, 魏世伟. 融合知识图谱与用户评论的商品推荐算法, 计算机工程, 2020, 46(8): 93-100.

[6] 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, 53(3): 582-600.

[7] 朱柳青. 基于深度学习的课程推荐与学习预测模型研究[D]. 杭州: 浙江工商大学, 2018: 6-20.

[8] 艾岩. 基于协同过滤推荐算法的选课系统的设计与实现[D]. 北京: 首都经济贸易大学, 2019: 8-18.

[9] 孔维梁. 协同过滤推荐系统关键问题研究[D]. 武汉: 华中师范大学, 2013: 12-26.

[10] 汤伟韬, 余敦辉, 魏世伟. 融合知识图谱与用户评论的商品推荐算法[J]. 计算机工程, 2020, 46(8): 96-97.

[11] 姚明臣. 机器学习和神经网络学习中的若干问题研究[D]. 大连: 大连理工大学, 2016: 2-5.

[12] XU B, XU Y, LIANG J Q, et al. CN-DBpedia: A Never-Ending Chinese Knowledge Extraction System.In International Conference on Industrial[C]// Engineering and Other Applications of Applied Intelligent Systems. Springer Cham, 2017.