

引用格式:赵继春,孙素芬,郭建鑫,等. 基于本体的网络学习资源推荐算法设计[J]. 微电子学与计算机, 2021, 38(1): 64-69.
[ZHAO J C, SUN S F, GUO J X, et al. Recommendation algorithm design of network learning resource based on ontology[J].
Microelectronics & Computer, 2021, 38(1): 64-69.]

基于本体的网络学习资源推荐算法设计

赵继春^{1,2}, 孙素芬^{1,2}, 郭建鑫^{1,2}, 钟 瑶^{1,2}, 王 敏^{1,2}, 秦 莹^{1,2}

(1 北京市农林科学院, 北京 100097; 2 北京市农村远程信息服务工程技术研究中心, 北京 100097)

摘 要: 针对学习者在网络学习平台中面临的信息过载导致获取个性化信息不足问题, 应用领域本体构建方法, 对学习课件进行数据建模及表示, 构建了一种基于领域本体与学习者属性信息的特征模型, 模型结合领域本体主题及学习者基本信息要素具有语义支持并随时间推移进行更新. 以学习特征模型为基础设计一种融合相似度的协同过滤推荐方法, 有效缓解推荐算法面临的冷启动及其语义支持不足的问题. 实验结果表明, 所提的方法与传统的协同过滤推荐方法相比, 数据推荐的准确率 MAE 减小 12.29%, 有效提高数据推荐质量.

关键词: 本体; 特征模型; 资源推荐; 语义; 个性化; 相似度

中图分类号: TP311.5; G431

文献标识码: A

文章编号: 1000-7180(2021)01-0064-06

Recommendation algorithm design of network learning resource based on ontology

ZHAO Ji-chun^{1,2}, SUN Su-fen^{1,2}, GUO Jian-xin^{1,2}, ZHONG Yao^{1,2},
WANG min^{1,2}, QIN Ying^{1,2}

(1 Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China;

2 The Research Center of Beijing Engineering Technology for Rural Remote Information Services, Beijing 100097, China)

Abstract: In view of the problem that learners are faced with information overload in the network learning platform, which leads to insufficient access to personalized information, this paper applies the domain ontology construction method to model and represent the data of learning courseware, and constructs a learner feature model based on domain ontology and learner attribute information. The model has semantic support combined with domain ontology theme and basic information elements of learners and updates with time. Based on the learning feature model, a collaborative filtering recommendation method is designed to integrate similarity, which can effectively solve the problem of cold start and lack of semantic support. The experimental results show that the data recommendation precision of proposed method is better than the traditional collaborative filtering recommendation method, the accuracy of data recommendation decreased by 12.29%, effectively improving the quality of data recommendation.

Key words: ontology; feature model; resource recommendation; semantics; personalization; similarity

1 引言

随着网络与信息技术的快速发展, 网络在线学

习数据的体量呈现指数增长趋势, 网络学习资源从相对匮乏向过载转变导致数据利用效率不断降低, 学习者在数据体量大、结构复杂的资源前面临资源

迷向问题^[1]。因而,为提高学习资源利用效率,提升学习者的个性化学习支持,有必要构建高效、可操作性并能反映学习者特点的用户兴趣特征模型,为用户提供个性化的学习资源支持。

目前,网络在线个性化学习支持主要通过构建用户兴趣特征模型及数据推荐算法实现^[2],在用户兴趣特征模型研究方面,有学者依据用户学习计划的相似性构建兴趣特征模型^[3]、采用本体数方式表示学生特征行为概念^[4].在学习内容及路径规划推荐方面,通常采用基于项目推荐、基于内容推荐及其混合推荐^[5].为解决推荐系统项目矩阵稀疏及冷启动问题,有学者提出在传统协同过滤算法基础上提出基于实值的玻尔兹曼机模型^[6]、基于 SDAE (Stacked Denoising Auto Encoder) 方法混合推荐^[7]、组合深度学习与矩阵因子分解推荐方法^[8]、基于神经网络和改进相似度推荐^[9].多种个性化推荐技术的混合应用在一定程度上能缓解项目矩阵数据稀疏及冷启动问题.网络在线学习资源不同的内容有时具有相近或相同的含义,不同的内容在深层次上可能具有关联性.传统的相似度实现方法通常不能发现内容之间的语义关联关系,从而把相似的内容当做不同类别的内容,导致推荐效果不理想.因此,在用户特征模型的语义支持方面的研究还有于进一步补充.

为提升网络在线学习资源推荐的准确率,提升学习者特征模型的语义支持,缓解新加入平台学习者的冷启动问题,本文构建一种基于领域本体与学习者属性信息的特征模型,通过领域本体对应的学习者兴趣特征主题实现语义相似度计算,使学习者特征模型具有较好的语义支持. 在学习者特征模型基础上,通过融合相似度计算实现协同过滤推荐,推荐方法与传统的基于余弦相似度的协同过滤推荐方法相比,准确率大幅提升. 通过融合相似度调节参数设置,在一定程度上缓解了新加入平台学习者的冷启动问题,从而改进学习者在线学习体验.

2 网络学习资源领域本体构建

本文在北京农民网络在线学习平台资源基础上构建领域本体,平台经过十余年积累 16000 余部涉农视频学习课件,涉及内容包括农业种植技术、养殖技术、农产品加工、劳动技能、农村经济发展等。本文采用领域本体实例回馈检验概念方法构建本体^[10],应用 OWL 语言进行本体描述,编辑工具采用 Protégé。

首先确定视频学习课件 7 大主类,分别为课件作品、课件分类、来源单位、课件类型、课件人物、出镜角色、课件标签,构建的学习课件主要类及关系如图 1 所示,学习课件的主类、类层次和属性逻辑关系如图 2 所示。

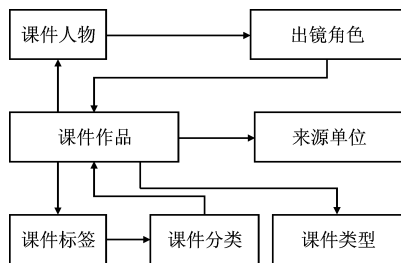


图 1 学习课件本体主要类关系

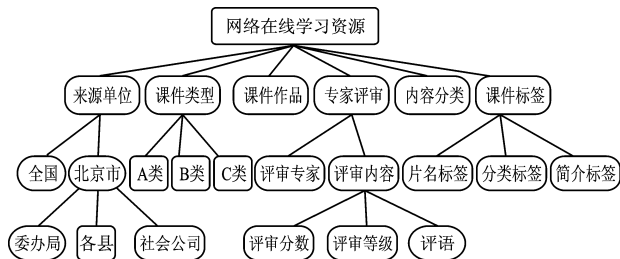


图2 学习课件类、类层次和属性逻辑关系

3 基于本体与用户属性信息特征模型

3.1 学习者特征模型构成

针对传统的用户兴趣特征模型存在语义支持不足问题,本文提出一种基于学习资源本体和用户属性信息的学习者特征模型,特征模型的实现方法如下:

学习者特征模型为:

$$\text{LCM}=(\text{LD},\text{LS},\text{LP},\text{T}) \quad (1)$$

由学习者属性信息 LD、特征信息 LS、行为轨迹信息 LP 及其时间变量 T 组成, LD 描述学习者基本特征属性信息, LS 特征信息的主题和学习课件本体信息对应, LP 为学习者学习日志集合.

学习者属性信息 LD 为:

$$LD=(ID,UI(S,A,P,T,F)) \quad (2)$$

学习者属性信息包含识别 ID 与属性信息要素 UI, 识别 ID 表述学习者特征模型类型, 属性信息 UI 包含学习者的性别 S、年龄 A、农业生产类型 P、农业经营类别 T、学习者好友 F, 采用向量空间方式表示.

使模型及时、有效、准确反映学习者的现状特征,需要对特征模型进行更新操作.综合考虑学习者在线学习体验和系统工作效率,对于学习者添加好友、添加收藏等操作对系统产生花销较小,特征模型参数采用及时在线更新方式.对于学习者产生的学习日志、添加课件标签等操作对系统产生的运算量较大,因而采用离线更新方式.

在对学习者特征模型查询更新操作时,如果新增兴趣主题能在模型中检索到,那么增加兴趣主题权重数值,如果检索结果返回为空值,则将特征主题插入到模型的兴趣主题集合.模型根据最近更新的时间 t_2 与当前时间 t_1 进行调整,将当前的主题兴趣度累加后作为最新的兴趣度数值,将模型的时间调整为 t_1 ,模型更新计算方法为:

$$UM(n) = BM(n) + RM(n) * \frac{m}{m + (t_1 - t_2)} \quad (11)$$

$UM(n)$ 为学习课件 n 更新后的兴趣度, $BM(n)$ 为当前计算得出学习者对学习课件 n 的兴趣度, $RM(n)$ 为上次记录的学习课件 n 兴趣度. m 为调节因子,数值越大表明学习者特征模型随时间衰减速度慢,反之表明随时间衰减速度快.

4 学习资源推荐算法设计

本文在结合学习资源本体与学习者属性信息特征模型构建基础上,提出一种融合相似度的协同过滤推荐方法.算法实现思路(图 4)为根据学习者特征模型的领域本体主题及其属性信息进行相似度计算,目的是获取学习者的近邻用户集合.应用协同过滤方法在近邻用户中,根据学习课件的评分数据获取学习者喜好的学习课程集合,同时应用相似度计算方法获得匹配度高的学习课件,预测学习者对匹配度高的课件打分数据,将预测打分数据排名靠前的课件推送给学习者.

算法相似度计算包括领域本体特征主题及学习者属性信息两个部分,将二者进行融合获得学习者近邻集合,根据学习者特征模型中学习课件打分数值获取推荐目标的偏好课件,并依据预测评分结果为用户推送学习资源.

4.1 学习者特征模型属性信息相似度计算

特征模型包含学习者的属性信息性别、年龄、农业生产类别、好友、农业经营类别等,将信息向量化处理,赋予权重并采用余弦相似度方法计算学习者之间的背景相似程度.

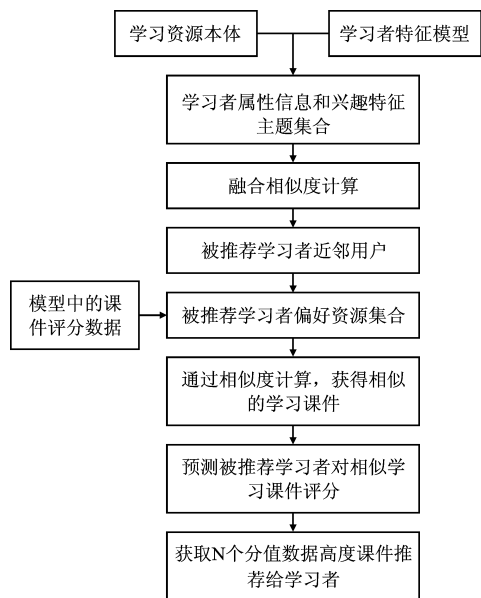


图 4 算法设计思路

$$UPsim(P_m, P_n) =$$

$$\frac{\sum_{i=1}^5 D_{m_i} * D_{n_i} * W_i^2}{\sqrt{\sum_{i=1}^5 (D_{m_i} * W_i)^2} \sqrt{\sum_{i=1}^5 (D_{n_i} * W_i)^2}} \quad (12)$$

$UPsim(P_m, P_n)$ 为学习者 m 和学习者 n 相似度, D_{m_i} 为学习者 m 的第 i 个向量化信息, D_{n_i} 为学习者 n 的第 i 个向量化信息, W_i 为第 i 个信息所占权重.

4.2 特征模型兴趣主题语义相似度计算

在领域本体相似度计算时,用户感兴趣的特定主题具有相应权重值,结合领域本体及其学习者对学习课件的评分信息计算相似度.用户感兴趣主题数据集 $ID = \{D_1, D_2, \dots, D_n\}$, $ND(D_m, D_n)$ 为两个感兴趣主题 D_m 与 D_n 在本体中关于同一类别结点最近一个,兴趣主题 D_m 与 D_n 在本体中的语义相似度计算公式如下:

$$IDsim(D_m, D_n) = \frac{2Depth(ND(D_m, D_n))}{Depth(D_m) + Depth(D_n)} \quad (13)$$

$IDsim(D_m, D_n)$ 为兴趣主题 D_m 和 D_n 的语义相似度, $Depth(D_m)$ 为从本体中同类资源根节点到用户标签 D_m 路径深度, $Depth(D_n)$ 为从本体中同类资源根节点到用户标签 D_n 路径深度, $Depth(ND(D_m, D_n))$ 是本体中同类资源根节点到 D_m 和 D_n 最临近共祖先结点路径长度. $IDsim(D_m, D_n)$ 数据范围 $[0, 1]$, $D_m = D_n$ 时,即两个主题相同, $IDsim(D_m, D_n) = 1$. 兴趣主题相似度数值随着共祖先结点深度增加而变大.

4.3 融合属性信息及兴趣主题相似度的近邻学习者集合

学习者属性信息及特征主题相似度计算完成后,依据二者融合结果获得相似度数值大小,从而得出目标推荐者的近邻学习者数据集,学习者 p 与 q 的融合相似度为:

LS(m,n) = r * UPsim(P_m,P_n) + (1 - r) * IDsim(D_m,D_n) (14)

LSsim(m,n) 为学习者 m 与 n 的融合相似度,UPsim(P_m,P_n) 为学习者 m 与 n 的属性信息相似度,IDsim(D_m,D_n) 为学习者 m 与 n 兴趣主题相似度, r 是调整参数,数值范围[0,1],学习者行为数据量较小时,将 r 参数值调大,主要依据用户属性信息计算相似度.当学习者行为数据积累到一定量值时,减小 r 参数数值,主要依据领域本体的特征主题计算相似度.因此新注册学习者的冷启动问题得到有效解决.

4.4 近邻学习者的协同推荐

推荐包括采集学习者打分数据集、预测打分数值及其数据推荐,设学习课件的数据集为{S_1,S_2,...,S_m},学习者数据集为{L_1,L_2,...,L_m},学习者对课件打分表示为{(S_1,F_1),(S_2,F_2),...,(S_m,F_m)},S_k 的评分数据集为 S_k = {(L_1,F_1),(L_2,F_2),...,(L_n,F_n)},L_K 感兴趣的学课件集合为 L_l = {S_x,...,S_y}. 预测 L_k 对课件 a = {L_i,...,L_j} 的打分数据,预测 L_k 对 S_h(S_h ∈ a) 打分数据方法为:

F_h = (sum_{i=m}^n (S_{hi} * F_{ki})) / (sum_{i=m}^n S_{hi}) (15)

式中,S_{hi} 为 S_h 与 S_i(S_i ∈ a) 相似度,F_{ki} 为 L_k 对 S_i 的评分.

预测分值排名靠前的课件推荐给学习者.

5 实验和分析

5.1 实验数据与设置

本研究团队研发构建了一套针对北京市农民远程培训的网络学习平台,从平台中采集 720 名学习者的 8235 条学习记录,包含 957 门学习课件,学习者对学习课件评分分为 5 个数值,评分数值反映学习者对课件的喜欢程度.

将实验数据划分为训练集和测试集,在数据集中随机抽取一定的比例作为实验训练集,按照数据总量的 20%、40%、60%和 80%抽取,剩下数据是测试集,采用学习者对学习资源的打分值和相似度分析预测测试数据的学习资源评分数据,分别采用基

于领域本体与学习者属性信息的协同过滤推荐(方法一)与传统基于余弦相似度的协同过滤方法(方法二)进行推荐.计算稀疏度 SPA、准确率 MAE 和覆盖率 COV.采用平均绝对误差 MAE 来衡量准确率优劣,数值越小说明算法的准确率越高^[11].采用覆盖率 COV 来衡量算法推荐范围好坏,数值越大说明推荐覆盖范围越广,算法挖掘推荐长尾效果越好.

SPA = M / (U * N) (16)

SPA 为稀疏度,M 为学习者已经评分资源数据,U 是学习者数量,N 是学习资源数量.

MAE = (sum_{i=1}^n |L_i - M_i|) / N (17)

{L_1,L_2,...,L_i} 为预测评分数据集,{M_1,M_2,...,M_i} 为学习者实际打分数据集,N 为集合的数据量.

COV = M_d / M (18)

M_d 是测试数据集中预测评分数据的数量,M 为测试数据集中的预测评分数据的总量.

5.2 实验结果及分析

通过对不同测试训练集划分,计算传统基于余弦相似度的协同过滤推荐方法(方法一)与基于融合相似度的协同过滤推荐方法(方法二)的稀疏度 SPA、准确率 MAE 和覆盖率 COV,数据见表 1.

表 1 不同训练集占比条件的 SPA、MAE 和 COV

训练集比例	SPA	方法一		方法二	
		MAE	COV	MAE	COV
0.2	0.024 1	0.921 3	0.853 5	0.801 6	0.841 3
0.4	0.036 9	0.905 5	0.872 3	0.784 7	0.871 5
0.6	0.048 3	0.863 3	0.891 3	0.721 3	0.883 4
0.8	0.060 6	0.812 6	0.903 4	0.703 5	0.897 2
平均值		0.875 7	0.880 1	0.752 8	0.873 4

分析表 1 的数据可知,训练集所占比例由 20%提升到 80%过程中,方法一和方法二的 MAE 都有所下降,也表明随着稀疏度 SPA 数值增加,推荐准确率 MAE 数值下降,推荐覆盖率 COV 数值在增加,在稀疏度 SPA 数值相对较大时,MAE 和 COV 数值变化趋向于平缓.方法一的平均准确率 MAE 为 87.57%大于方法二的平均准确率 MAE 75.28%,方法一的平均覆盖率 COV 88.01%接近于方法二的平均覆盖率 COV 87.34%.基于本体语义相似度的协同过滤推荐与基于传统协同过滤推荐方法相比,推荐的准确率 MAE 减小了 12.29%,推荐覆盖率 COV 基本保持一致.

6 结束语

本文提出一种基于领域本体与学习者属性信息

的学习者特征模型,结合领域本体信息计算语义相似度,并设计一种融合相似度的协同过滤推荐方法,推荐算法通过领域本体语义相似度和学习者属性信息实现.实验结果表明,与基于余弦相似度的协同过滤推荐方法项目,本文的推荐算法的准确率更优.通过设置融合相似度调节参数,在一定程度上缓解了新加入学习者的冷启动问题,从而改进学习者在线学习体验.然而在学习者数据和资源不断增加的条件下,推荐算法的时间复杂度也随之增加,后续研究中将进一步探索完善.

参考文献:

- [1] 王光,姜丽,董帅含,等.融合本体语义与用户属性的协同过滤算法[J].计算机工程,2019,45(10): 215-220. DOI: 10.19678/j.issn.1000-3428.0052499.
WANG G, JIANG L, DONG S H, et al. Collaborative filtering algorithm combining ontology semantics and user attribute[J]. Computer Engineering, 2019, 45(10): 215-220. DOI: 10.19678/j.issn.1000-3428.0052499.
- [2] 申云凤.基于多重智能算法的个性化学习路径推荐模型[J].中国电化教育,2019(11): 67-72.
SHEN Y F. Personalized learning path recommendation model based on multiple intelligent algorithms[J]. China Educational Technology, 2019(11): 67-72.
- [3] LAWSON C, BEER C, ROSSI D, et al. Identification of 'at Risk' students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution[J]. Educational Technology Research and Development, 2016, 64(5): 957-968. DOI: 10.1007/s11423-016-9459-0.
- [4] 杨桢,从传锋.基于大数据的网络教育方向选取方法研究[J].现代电子技术,2018,41(15): 87-91. DOI: 10.16652/j.issn.1004-373x.2018.15.020.
YANG Z, CONG C F. Research on network education direction selection method based on big data[J]. Modern Electronics Technique, 2018, 41(15): 87-91. DOI: 10.16652/j.issn.1004-373x.2018.15.020.
- [5] 黄立威,江碧涛,吕守业,等.基于深度学习的推荐系统研究综述[J].计算机学报,2018,41(7): 1619-1647. DOI: 10.11897/SP.J.1016.2018.01619.
HUANG L W, JIANG B T, LV S Y, et al. Survey on deep learning based recommender systems[J]. Chinese Journal of Computers, 2018, 41(7): 1619-1647. DOI: 10.11897/SP.J.1016.2018.01619.
- [6] 何洁月,马贝.利用社交关系的实值条件受限玻尔兹曼机协同过滤推荐算法[J].计算机学报,2016(1): 183-195. DOI: 10.11897/SP.J.1016.2016.00183.
HE J Y, MA B. Based on real-valued conditional restricted Boltzmann machine and social network for collaborative filtering[J]. Chinese Journal of Computers, 2016(1): 183-195. DOI: 10.11897/SP.J.1016.2016.00183.
- [7] WEI J, HE J H, CHEN K, et al. Collaborative filtering and deep learning based recommendation system for cold start items[J]. Expert Systems with Applications, 2017, 69: 29-39. DOI: 10.1016/j.eswa.2016.09.040.
- [8] ZHAO W X, LI S, HE Y L, et al. Connecting social media to e-commerce: cold-start product recommendation using microblogging information[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(5): 1147-1159. DOI: 10.1109/TKDE.2015.2508816.
- [9] 汪方胜,侯立文,蒋馥.领域本体建立的方法研究[J].情报科学,2005,23(2): 241-244. DOI: 10.3969/j.issn.1007-7634.2005.02.020.
WANG F S, HOU L W, JIANG F. Study on establishing domain ontology[J]. Information Science, 2005, 23(2): 241-244. DOI: 10.3969/j.issn.1007-7634.2005.02.020.
- [10] 邹锋.基于深度神经网络和改进相似性度量的推荐算法[J].计算机应用与软件,2019,36(11): 286-293.
ZOU F. A recommendation algorithm based on deep neural networks and improved similarity measurement[J]. Computer Applications and Software, 2019, 36(11): 286-293.
- [11] 王维,高岭,高全力.融合用户信任 and 用户兴趣漂移的协同过滤算法[J].微电子学与计算机,2019,36(7): 103-108. DOI: 10.19304/j.cnki.issn1000-7180.2019.07.020.
WANG W, GAO L, GAO Q L. Collaborative filtering algorithm based on user trust and interest drift detecting[J]. Microelectronics & Computer, 2019, 36(7): 103-108. DOI: 10.19304/j.cnki.issn1000-7180.2019.07.020.

作者简介:

赵继春 男,(1980-),博士,副研究员.研究方向为机器学习、数据挖掘. E-mail: zhaojichun123@163.com.

孙素芬 女,(1964-),硕士,研究员.研究方向为智能化系统.

郭建鑫 男,(1973-),博士,研究员.研究方向为智能化系统.

钟瑶 女,(1993-),硕士.研究方向为数据挖掘.

王敏 女,(1982-),硕士,助理研究员.研究方向为数据挖掘.

秦莹 女,(1982-),硕士,助理研究员.研究方向为数据挖掘.