

中图分类号：G434

论文编号：10006SY1524124

北京航空航天大學
硕 士 学 位 论 文

基于学习风格诊断的
自适应学习系统

作者姓名 李如寐

学科专业 计算机科学与技术

指导教师 殷传涛 副教授

培养院系 中法工程师学院

Adaptive Learning System based on Online Learning Style Identification

A Dissertation Submitted for the Degree of Master

Candidate: Li Rumei

Supervisor: Prof. Yin Chuantao

Sino-French Engineering School

Beihang University, Beijing, China

中图分类号：G434

论文编号：10006SY1524124

硕士论文

基于学习风格诊断的 自适应学习系统

作者姓名 李如寐 申请学位级别 工学硕士

指导教师姓名 殷传涛 职称 副教授

学科专业 计算机科学与技术 研究方向 在线教育，数据挖掘

学习时间自 2015 年 9 月 1 日 起至 2017 年 12 月 31 日止

论文提交日期 年 月 日 论文答辩日期 2018 年 3 月 7 日

学位授予单位 北京航空航天大学 学位授予日期 年 月 日

关于学位论文的独创性声明

本人郑重声明：所呈交的论文是本人在指导教师指导下独立进行研究工作所取得的成果，论文中有关资料和数据是实事求是的。尽我所知，除文中已经加以标注和致谢外，本论文不包含其他人已经发表或撰写的研究成果，也不包含本人或他人为获得北京航空航天大学或其它教育机构的学位或学历证书而使用过的材料。与我一同工作的同志对研究所做的任何贡献均已在论文中作出了明确的说明。

若有不实之处，本人愿意承担相关法律责任。

学位论文作者签名：_____

日期： 年 月 日

学位论文使用授权书

本人完全同意北京航空航天大学有权使用本学位论文（包括但不限于其印刷版和电子版），使用方式包括但不限于：保留学位论文，按规定向国家有关部门（机构）送交学位论文，以学术交流为目的赠送和交换学位论文，允许学位论文被查阅、借阅和复印，将学位论文的全部或部分内容编入有关数据库进行检索，采用影印、缩印或其他复制手段保存学位论文。

保密学位论文在解密后的使用授权同上。

学位论文作者签名：_____

日期： 年 月 日

指导教师签名：_____

日期： 年 月 日

摘要

随着计算机及网络技术在教育领域的飞速发展，越来越多的学习者开始利用互联网进行在线学习。在学习资源爆发式增长的情境下，如何帮助学习者提供合适的学习资源和学习指引成为了重要的研究方向。对于开发针对学习者知识背景、兴趣、目标和个人特质的自适应学习系统，普遍的解决方法就是对学习者和学习资源进行建模，并利用自适应推荐方法将二者结合。学习风格模型就是常用的学习者模型之一，主要对学习者情感、社会、生理和心理层面的特征进行描述。然而，由于在线学习与传统学习的明显不同，导致学习者行为特点乃至学习风格都产生了变化，传统的学习风格模型已不再适用。同时，如何将学习风格与自适应推荐方法所结合也是一大难题。

本研究针对当下在线学习者行为特点设计了在线学习风格模型（Online Learning Style, OLS），该模型包含了动机、交流、视觉、言语、感悟、直觉、序列和综合共八个维度，表示了在线学习者在情感、社会、生理和心理四个不同层面的个人特质。另外，本研究提出了基于学习风格的自适应推荐方法（Adaptive Recommendation based on Online Learning Style, AROLS）。AROLS 首先基于学习风格对学习者进行聚类分析，之后针对不同类别的学习者行为数据进行挖掘并提供学习资源推荐。本研究分别对传统推荐算法和深度学习方法进行了尝试，提出了基于协同过滤推荐的 CF-AROLS 和基于长短期记忆循环网络的 LSTM-AROLS。其中 CF-AROLS 使用关联规则对基于物品的协同过滤结果进行了筛选，实现了可变长度的动态推荐，提高推荐准确率。另外，考虑到学习资源的浏览存在序列关系，应用 LSTM 模型对学习资源浏览序列进行了建模。最后，本研究对两种方法分别进行了超参数配置的研究和推荐结果的对比，实验结果表明 LSTM-AROLS 由于考虑了学习资源浏览的序列关系，具有较好的表现。

在对学习者模型和自适应学习资源推荐方法在理论层面进行深入的研究后，本研究对自适应学习系统的架构和功能进行了设计，并使用基于 Javascript 语言的 Vue.js 前端框架和基于 Python 语言的 Django 框架对自适应学习系统的原型进行了实现。

关键词：在线学习，学习风格，个性化推荐，自适应学习系统

Abstract

Recent years, e-learning systems have been expanding its popularity and influences all over the world. Materials or what we call “tutorials” made by individuals around the internet are very popular resources preferable to learners. At the same time, popular e-learning websites, such as Moodle or Coursera, are also digitalizing materials incessantly for learners with different educational backgrounds and various needs. However, without proper guidance, it may be difficult for learners of different levels to choose suitable materials in their learning process facing the abundance of information.

Adaptive Learning System (ALS) has then become a significant stream of research and development in the field of e-learning. ALS gives instructions and learning resource recommendations based on different levels of expertise, interests, goals, educational backgrounds and personal traits of learners. This type of system usually consists of at least three parts: a content model that introduces the structure and contents of the learning resources, a learner model that represents features of the learner as an individual and an instructional model that dynamically matches a learner model with multiple content models and decides how they are presented.

The integration of learning style is the most popular direction of research considering learner model adaption. Learning styles are unique manners in which learners begin to concentrate on, process, absorb, and retain new and difficult information. Gaining insights into different learning styles can offer means to design and provide interventions that adapted to individual needs. However, as e-learning has vastly changed the way of knowledge acquisition, traditional learning style models based on systematic and linear teaching environment may not fit learners’ traits any more. Besides, it is of great importance to cooperate learning style with the instructional model in order to make personalized recommendations.

In this paper, we investigate the difference between the behaviors of learners in online and traditional environment, then we design a learning style model of eight features for online learners.

Based on our online learning style model, we propose a learning resource recommendation mechanism - AROLS (Adaptive Recommendation based on Online Learning Style). First, it generates learner clusters with different learning styles. Next, the behavioral patterns of each cluster are extracted by applying multiple techniques. Finally, AROLS creates a personalized recommendation set according to a learner's learning style and his browsing history. During our research, we apply two recommendation methods: CF-AROLS which uses association rules to filter the recommendation results of item-based collaborative filtering and generates a personalized recommendation set of variable size; LSTM-AROLS which applies long short-term memory model to analyze learners' browsing sequences. We also study different hyper-parameters to improve the performance of our methods. At last, we design and realize an adaptive learning system which focuses on analyzing learners' learning style and implements AROLS.

Key words: Online Learning, Learning Styles, Personalized Recommendation, Adaptive Learning System

目 录

第一章 绪论.....	1
1.1 研究背景.....	1
1.2 研究意义.....	2
1.3 研究现状.....	3
1.3.1 自适应学习系统相关研究.....	3
1.3.2 学习者模型相关研究.....	4
1.3.3 自适应推荐方法相关研究.....	5
1.4 研究内容及方法.....	6
1.4.1 研究内容.....	6
1.4.2 研究方法.....	7
1.5 论文组织结构.....	7
第二章 在线学习风格模型的研究.....	9
2.1 学习风格模型的发展及应用.....	9
2.1.1 学习风格模型的发展.....	9
2.1.2 学习风格的应用.....	10
2.2 学习风格的诊断.....	11
2.2.1 学习风格的显式诊断.....	11
2.2.2 学习风格的隐式诊断.....	12
2.3 在线学习风格模型（OLS）.....	13
2.3.1 传统学习风格模型的不足.....	13
2.3.2 在线学习与传统学习的对比.....	14
2.3.3 在线学习风格的相关研究.....	15
2.3.4 在线学习风格建模.....	16
2.4 在线学习风格的诊断方法.....	20
2.4.1 在线学习风格量表的设计与分析.....	20

2.4.2 在线学习风格的隐式诊断方法.....	23
2.5 本章小结.....	28
第三章 自适应学习资源推荐方法.....	29
3.1 以学习者为中心的自适应推荐的传统方法.....	29
3.1.1 基于学习者模型的自适应推荐方法.....	29
3.1.2 基于学习者模型的自适应推荐所面临的挑战.....	29
3.2 基于学习风格的自适应推荐方法（AROLS）	30
3.2.1 基于 BIRCH 的学习者聚类	31
3.2.2 基于协同过滤的自适应学习资源混合推荐方法（CF-AROLS）	33
3.2.3 基于 LSTM 的自适应学习资源推荐方法（LSTM-AROLS）	36
3.3 实验设计与结果分析.....	41
3.3.1 实验数据集.....	41
3.3.2 评价指标.....	41
3.3.3 数据预处理.....	43
3.3.4 学习者聚类数目的研究.....	44
3.3.5 CF-AROLS 超参数配置的研究	47
3.3.6 LSTM-AROLS 超参数配置的研究	50
3.3.7 实验结果与分析.....	52
3.4 本章小结.....	54
第四章 基于学习风格诊断的自适应推荐原型系统的实现.....	55
4.1 需求分析.....	55
4.1.1 业务需求分析.....	55
4.1.2 系统需求分析.....	56
4.2 系统设计.....	56
4.2.1 系统结构及功能模块设计.....	56
4.2.2 数据库设计.....	58
4.3 系统开发与实现.....	59
4.3.1 开发环境及关键技术.....	59

4.3.2 功能模块展示.....	60
4.4 本章小结.....	62
总结与展望.....	63
致谢.....	66
参考文献.....	67
攻读硕士学位期间取得的学术成果.....	72
附录.....	73
附录 1 在线学习风格（OLS）量表题目	73
附录 2 CF-AROLS 伪代码	74
附录 3 LSTM-AROLS 伪代码	75
附录 4 项目源代码.....	76

插图目录

图 1 学习风格诊断流程.....	11
图 2 学习风格要素、学习风格以及行为偏好的关系.....	16
图 3 OLS 量表题目的相关系数矩阵	22
图 4 聚类中心点学习风格分析图.....	22
图 5 OULAD 数据集结构.....	25
图 6 AROLS 的学习资源推荐流程	31
图 7 CF 树.....	33
图 8 Skip-gram 模型.....	37
图 9 RNN 网络（左）和单元结构（右）	38
图 10 LSTM 的单元结构.....	39
图 11 基于 LSTM 的学习资源路径预测模型框架	40
图 12 异常值检测.....	44
图 13 预处理后的学习者学习风格数据.....	44
图 14 BIRCH（左）和 K-means（右）的聚类结果对比.....	46
图 15 支持度阙值对频繁项集个数的影响.....	47
图 16 置信度阙值对推荐结果的影响.....	48
图 17 Kulc 度量阙值对推荐结果的影响.....	48
图 18 不平衡比对关联规则数目影响.....	49
图 19 浏览趋势阙值对推荐结果的影响.....	49
图 20 不同的学习资源向量长度对实验结果的影响.....	50
图 21 CF-AROLS 的对比实验结果	52
图 22 LSTM-AROLS 的对比实验结果	53
图 23 AROLS 自适应学习系统流程图	56
图 24 AROLS 自适应学习原型系统结构	57
图 25 AROLS 自适应学习系统模块	57

图 26 AROLS 自适应学习系统数据库表	58
图 27 AROLS 自适应学习系统主页	60
图 28 AROLS 自适应学习系统课程页面	60
图 29 AROLS 自适应学习系统个人资料页面	61
图 30 AROLS 自适应学习系统测试推荐结果	61

表格目录

表 1 Coffield 对 71 种学习风格模型的总结分类	9
表 2 基于学习风格要素对传统学习和在线学习特点的对比.....	14
表 3 经典的学习风格模型.....	17
表 4 OLS 模型的学习风格特征和对应行为	18
表 5 OLS 特征与学习资源属性的对应关系	24
表 6 OULAD 活动类型的描述和与 OLS 的对应关系	26
表 7 OULAD 数据集中学习者学习风格特征均值及标准差分析.....	27
表 8 BIRCH 聚类结果评估	45
表 9 针对阙值调参的 BIRCH 聚类结果评估	45
表 10 Birch 和 K-means 的不同类中学习者数目的对比	47
表 11 各类学习者训练数据和验证数据序列个数.....	50
表 12 LSTM 模型参数配置对实验结果的影响.....	51
表 13 AROLS 自适应学习系统数据类型及说明	59

第一章 绪论

1.1 研究背景

随着计算机技术的飞速发展，人们开始逐渐摆脱传统的、以教师为中心的面对面式课堂学习的束缚，提出新的学习方式并改善学习环境。从通过电视、广播等媒介的远程学习（D-learning），到通过计算机、手机等移动设备的在线学习（E-learning），人们不再受限于固定的学习资源，可以在任何时间或地点进行学习，自由度得到了很大的提升。然而，在线学习平台虽然具有灵活性、开放性和资源共享性等优点，但还存在着很多普遍问题，其中之一就是缺乏对学习者的引导。在线学习平台为学习者提供形式多样的文本、图像、视频等学习资源的同时，对学习资源的内容与质量缺乏控制，造成了平台资源组织的混乱，使得学习者容易在学习时出现信息的超载与迷失，降低学习效率。

为了对学习者进行适当的引导，产生了更加多样化、自由化、智能化的学习方式，我们称之为智慧学习（Smart learning）^[1]或自适应学习（Adaptive learning）^[2]。自适应学习通过与学习者交互，改变学习进程中学习资源的内容或表现形式，实现对不同学习者的个性化指导，以达到提高学习者效率的目标^[3]。与单纯的在线学习系统相比，自适应学习系统完全以学习者为中心，基于不同学习者的知识背景、认知水平、对信息展示的偏好等特征，帮助学习者在海量信息中筛选出适合自己的学习资源，并记录、挖掘和深入分析学习者的历史学习数据，在评估后进行个性化干预、指导，促进有效学习的发生。

自适应学习系统的建立通常包含三个核心要素：内容模型、学习者模型、教导模型。内容模型（也称学习资源模型）指对要学习的对象和内容进行建模，包括全面详细的学习成果和定义好的学习任务；学习者模型将学习者“模型化”，使得平台可以基于学生的表现来对学生的知识水平进行数据参考。学习者模型可以在不同的方面以数值化的方式来评估学习者的能力及水平，更进一步的可以实时跟踪学生的学习进度；教导模型决定了一个系统如何在特定的时间选择特定的内容呈现给特定的学习者。该模型提供了一个将学习者模型和内容模型结合在一起的自适应方法，根据之前收集的学习者相关数据推荐下一刻适合的学习内容或学习活动。

然而，尽管在此方面的研究已持续数年，自适应学习系统的技术还尚未成熟，至今

仍没有出现在世界范围内广为使用的系统。虽然一部分在线学习平台也拥有一定的推荐功能，但大多数只停留在学习资源的设计与实践层面，即“以学习资源为中心”，却缺乏对学习者特征的分析，不能针对学生的自身特征提供个性化指导，以至于无法达到预期的效果。因此，以学习者为中心的自适应系统无疑是未来教学模式的方向。设计开发一个能够根据学习者偏好进行动态调整的自适应学习系统显得尤为必要。

本研究的重点在于构建学习者模型和教导模型。通过挖掘学习者在线行为记录，并基于经典的学习风格模型，对在线学习者的认知能力和思维方式等特质进行建模和诊断。同时，结合在线学习者的学习风格模型，应用关联分析和协同过滤算法实现学习资源的自适应推荐。最终对基于在线学习风格诊断的自适应学习系统的原型进行实现。

1.2 研究意义

学习者模型是自适应学习系统的基础。近年来，越来越多的研究者们开始关注学习者在认知风格方面的差异，应用学习风格模型（Learning style model）作为个人特质来描述学习者的学习偏好与特点。然而，传统的学习风格模型大多是由来自生理学、心理学和社会学等不同领域的专家提出，在理论层面有些单一，难以涵盖学习过程中不同层面的内容。同时由于学习环境由传统课堂到在线课堂的变化，学习者相应的特质也难免有所更改。因此，对于在线学习者的重新建模十分必要。另外，经过对教导模型的调研发现，很少有研究考虑到学习者的学习风格，故基于学习风格对学习者进行自适应的教导模型是值得探索的领域。

为了针对学习者的各个特质进行学习资源自适应推荐，提高在线学习者的学习效率，本论文针对学习者模型和教导模型进行了深入研究，主要贡献有三个方面：

- 1) 基于传统学习风格模型，分析在线学习者的行为，提出在线学习风格模型(OLS)。并针对 OLS 模型设计了在线学习风格量表，对不同的学习者进行了调研。同时提出了通过对学习者行为日志记录的挖掘来诊断在线学习风格的方法。为在线学习者模型的构建和学习风格诊断提供参考价值。
- 2) 提出了基于 OSL 的自适应学习资源推荐方法 (AROLS)。AROLS 通过对不同学习者进行学习风格分类，挖掘相似学习风格的学习者群组的关联规则，结合协同过滤算法对学习者进行自适应推荐。为了给学习者提供合适的学习资源，AROLS 更注重推荐

准确率的提升，可为不同的学习者动态地调整推荐列表中资源的数目，在教导模型上有所创新。

3) 基于 OLS 学习者模型和 AROLS 教导模型，并设计实现了自适应学习系统的原型。该系统具有普适性和推广性，可承载不同的形式的学习资源，为之后的自适应学习系统提供参考和借鉴。

1.3 研究现状

1.3.1 自适应学习系统相关研究

目前，一些西方国家在自适应学习系统的概念、结构和参考模型的研究较为成熟，并在技术层面上尝试开发出提供个性化学习服务的系统。美国的 Perter Brusilovsky 是最早研究自适应学习系统的权威专家之一，他提出的 AEHS 模型^[4]一直是自适应学习系统中最通用的系统结构设计模型。Perter Brusilovsky 在 1996 年与他的同事开发出了全球首个自适应教学系统 InterBook^[5]，该系统能够实现自适应超媒体的制作和基于网络的超媒体传输。之后，Perter Brusilovsky 与 Gerhard Weber 共同设计出了 ELM-ART^[6]教学系统，该系统能实现一定的智能化自适应，可以进行实例问题解答并支持课程序列化选择。Knowledge Sea^[7]也是 Perter Brusilovsky 等人所开发的辅助教育系统，可以在用户的学习过程中依靠图的指引获取学习内容指导。此外，荷兰的 Debra 教授等人研究开发了 AHA!^[8]自适应超媒体系统，可以通过编辑自适应规则和更改用户模型设置来实现内容的个性化适应。澳大利亚的研究者 Wolf 基于 JAVA 设计了一款能够适应学习者需求的学习平台 iWeaver^[9]，该平台使用 Dunn and Dunn 学习风格模型作为学习者模型，并通过数据分析进行科学的推荐。

相比于国外，我国的对自适应系统的研究起步较晚。1999 年，北京师范大学的余胜泉教授首先进行了探索性的研究，他认为自适应学习系统的研究应包括学习背景的检测、学习内容的组合、学习策略的设置这三个重要部分^[10]。同时指出对学习背景的检测是自适应系统研究的重点，只有对学习者的学习水平和学习能力有整体上的了解，才能提供个性化的自适应反馈。华南师范大学的陈品德博士在 2003 年对自适应产生的因素等内容进行了分析，认为设计自适应学习系统应该从内容和导航两方面入手。他的团队设计研发了 Adaptive Tutor 原型系统^[11]。辽宁大学的丛春瑜等人在 2004 年基于 Web 设计了

一个自适应学习系统，该系统使得学习者可以利用导航功能进行分级、有次序的学习，也可以进行自由学习^[12]。南京师范大学的柏宏权博士设计实现了适应性智能教学系统 I-Tutor 的原型，该系统能对学习者的水平和偏好进行动态监测和分析，并提供适应性的学习材料^[13]。2012 年，东北师范大学的姜强博士研究了自适应学习系统中用户模型和知识模型的建模标准与参考规范^[14]。2014 年，山东师范大学的崔晓阳等人设计实现了基于 Felder-Silverman 的在线导学系统，可以根据学习者学习风格的差异性进行自适应推荐^[15]。近年来，随着数据挖掘等技术的发展，国内学者开始关注情感领域。辽宁师范大学的刘小丹等在 2015 年提出了基于情绪感知的学习者模型，利用表情识别技术来感知学习者的情绪，实时分析其状态与学习动机，以解决自适应学习系统无法提供与学习者双向情感交流的问题^[16]。

综上，国外对自适应学习系统的研究相对深入，而国内的研究主要集中在理论层面，在实际应用中相对不足。然而，不管是国外还是国内，对自适应学习系统的探索过程却是相似的。从最初只注重学习内容的适应性呈现和导航的设计，到更加注重学习者模型，将学习风格等心理学因素融入其中。此外，通过对国内外的研究现状进行分析后发现，大多数系统的不足主要集中在两个方面：学习者模型不完善，对学习者特征的分析太过片面；在学习者模型、教导模型和学习资源模型的设计相对随意，缺乏统一标准，无法在多系统间进行资源和数据的共享。

1.3.2 学习者模型相关研究

学习者模型是对个性特征的描述，是学习系统实现自适应机制的重要依据之一。学习者模型可以涵盖学习者的个人信息、学习兴趣、学习目标、背景知识和个人特质等特征。学习者模型中提供的信息使得自适应学习系统可以为学习者提供个性化的学习内容及导航^[17]。

学习者个性特征的选择是构建学习者模型最重要的一步。Self^[18]认为完整的学习者模型需要包含领域知识、目标、偏好、兴趣和学习进展等信息。Brusilovsky^[19]在 2007 年通过对已经存在对自适应超媒体系统中用户模型等总结归纳，分析出最常用的五个特征：知识状态(knowledge)，学习兴趣(interests)，学习目标(goals)，背景知识(background)和个人特质(individual traits)。其中个人特质包括认知风格和学习风格。2015 年，Nakic 等人^[20]对从 2001 年到 2013 年发表的论文进行统计，提取了自适应学习系统中典型的学

习者模型所包含的特征与发展趋势。根据文章的数量，最为普遍的是学习风格，其次是背景知识、认知风格、偏好和动机。

对于早期的自适应学习系统，学习者模型主要侧重于学习者的知识状态和能力差异，如分组模型和覆盖模型等。另外，还有一些学习者模型也考虑了认知方面的因素，例如前文提到的 AHA！ 和 ELM-ART 系统中的学习者模型就包含兴趣这一特征，也有些系统包含学习偏好的信息，如 Hyperbook、XAHM 和 HYNECOS 等系统^[21]。然而，这些参数指代的实际上都是学生的知识状况，其中的偏好和兴趣则反映的是学习者对特定知识点的兴趣。之后，随着研究的深入，研究者们开始把重点放到学习者的个性差异区分上，以更全面地反映学习者特征。这一阶段主要以认知风格作为学习者模型核心。在计算机辅助学习 SOLA 系统（Students' On-Line Advisor）^[22]的学习者模型中，就学生对序列和整体、探究和指导的方式进行偏好选择。后来，研究者们开始更多地倾向基于学习风格构建学习者模型。例如，学习系统 SAVER^[23]和自适应智能教学系统 EDUCE^[24]均采用了 Felder-Silverman 学习风格模型。Milosevi 和 Brkovic 等人^[25]在基于学习管理系统所提供的学习资源的自适应模型时，采用了 Kolb 学习风格模型。可见对于学习者建模而言，学习风格模型是不可忽视的一部分，是体现学习者个性差异的重要指标。

1.3.3 自适应推荐方法相关研究

教导模型也称自适应模型（Adaptive model）或自适应推荐方法，它将学习者模型和学习资源模型结合在一起，帮助学习者进行学习内容的导航，实现个性化学习。教导模型的自适应方式分为适应性导航支持（Adaptive navigation support）和适应性内容呈现（Adaptive content presentation）两种^[26]。适应性导航是指为不同特征的学习者提供学习路径建议，避免学习者在学习的过程中迷失方向，同时缩小学习者查找内容的范围和查找内容的时间，提高学习的效率。适应性内容呈现是指在相同的页面中为不同学习者呈现个性化的内容，以适应学习者的偏好，加快学习者对知识的吸收与理解。

Chedrawy 等人^[27]提出一种基于案例推理规则和物品的协同过滤的推荐策略。该方法首先利用协同过滤对各种信息项目进行推荐，之后基于案例推理选择最相关的信息，形成符合推荐。Cristea 等人^[28]研发了一种基于三层模型的自适应技术：自适应规则、自适应语言以及自适应策略。这种模式有利于标准化自适应技术，使系统之间的资源和技术互换成为可能。Koper 应用协同过滤算法，不使用学习者的个人信息，为学习者推荐课程^[29]。

Narimel 利用关联规则分析，实现了课程的推荐^[30]。方海光等人^[31]对移动学习资源的教导模型进行了探讨，使用教育建模语言提出了基于学习对象化的教导模型。该模型可以根据学习者的基本特点和学习环境为学习者推荐学习内容。赵学孔等人^[32]提出了基于 B/S 模式的个性化推荐模型，该模型以学习者模型和内容模型为基础，根据学习者的认知水平、学习风格和领域知识元对象来进行关联规则匹配，可进行学习路径与知识对象呈现的个性化。

总体来看，目前普遍的在线学习平台在教导模型设计上更加关注以系统为驱动的学习策略，而较少考虑对于学习者的个性化。在学习者学习的过程中，一方面需要学习者自主地进行实践和探索，另一方面也需要自适应模型针对学习者的学习方式进行分析和指导，并不断优化资源推进方式，这样不仅可以提高学习者的学习效率，同时可以借助丰富多样的学习资源推荐提高学习者的注意力和探索欲，并最大限度地利用系统内的学习资源。

1.4 研究内容及方法

1.4.1 研究内容

本研究首先对国内外自适应学习系统、学习者模型和自适应模型的研究进行分析总结，并在此基础上，以学习风格为学习者模型的核心，对在线学习者的行为数据进行具体分析，设计基于学习风格诊断的自适应导航方法并完成自适应系统原型的设计与实现，为学习者学习过程中提供学习资源选择建议，提升学习效率。

具体来讲，本文的研究内容主要包含以下三个方面：

- 1) 在线学习风格模型。传统的学习风格模型大多是单一模型，只从部分维度描述学习者的特征。而且由于从课堂环境到在线环境的转化，有些模型已不再适用在线学习者的特征描述。本研究通过对传统学习风格的总结和对在线学习行为特征的归纳，提出适合在线学习者的学习风格模型。
- 2) 基于在线学习风格模型的自适应学习资源推荐方法。在对学习者学习风格诊断的基础上设计自适应模型，主要实现适应性导航支持，辅助学习者的学习过程。
- 3) 自适应学习系统原型的开发，主要是在线学习风格识别模块和自适应模块的开发以及系统总体功能的设计与实现。

1.4.2 研究方法

本次研究的重点在于构建学习者模型和教导模型，通过挖掘学习者的在线学习行为记录，识别学习者的在线学习风格，并基于在线学习风格实现学习资源的自适应推荐。整个研究的过程分为以下几个阶段：

1) 提出问题阶段

通过对文献的总结分析，了解自适应学习系统中学习者模型与自适应模型的现状与发展趋势，为基于在线学习风格的自适应推荐提供理论基础。同时，在查阅文献的过程中进一步确定课题的研究方向和重点。

2) 理论分析阶段

该阶段主要目标是深入了解传统学习风格模型和寻找自适应模型解决方案。通过对传统学习风格的总结归纳和分析传统学习与在线学习的异同，寻求在线学习者建模的特征维度。同时，调研并学习机器学习的常见算法，挖掘开源数据集中的学习者行为数据，探索并研究各种学习资源推荐方法。

3) 实验评估阶段

通过反复的对比实验，最终确定在线学习风格模型和自适应推荐方法，并确定模型的相关参数设置。

4) 系统设计与开发阶段

在确定学习者模型和自适应模型后，以使用者的角度对系统进行需求分析，设计系统的功能模块、数据库模型以及框架结构，基于 Web 进行系统原型开发。

1.5 论文组织结构

本论文主要研究对于在线学习者学习风格的建模和基于该学习者模型的自适应学习推荐问题。全文共分为五个章节，组织结构如下：

第一章：绪论。该章节对本研究的研究背景和研究意义进行了阐述，并从自适应学习系统、学习者模型和自适应模型三方面进行了研究现状的总结和分析。接下来简要说明了本研究的主要研究内容与采用的研究方法，最后论文组织结构进行了梳理。

第二章：在线学习风格模型的研究。在阅读大量文献后，对学习风格模型的发展和应用进行了介绍和总结，同时解释了学习风格的显式和隐式诊断方法并进行对比。在对

传统学习风格模型的调研基础上，我们深入分析了在线学习和传统学习的区别，并基于 Dunn and Dunn、Felder-Silverman、Entwistle、Myers-Briggs 和 Kolb 学习风格模型设计了在线学习风格模型 OLS。同时，根据 OLS 设计了在线学习风格量表并进行调查研究，完善了量表的题目。最后提出了基于学习资源属性的学习风格隐式诊断规则，并在 OULAD 开源数据集上进行了实验和分析。

第三章：自适应学习资源推荐方法的研究。基于在线学习风格，提出了 AROLS 学习资源推荐方法。该方法首先对不同学习风格的学习者进行聚类分析，之后根据同一类学习者的学习资源浏览记录进行分析和建模。本研究先后提出了基于协同过滤的 CF-AROLS 和基于循环神经网络的 LSTM-AROLS，并对两种方法的超参数配置进行了大量的实验，选取表现更好的参数设置进行最终的评估分析。

第四章：基于学习风格诊断的自适应推荐原型系统的实现。对在线学习风格模型和 AROLS 自适应学习资源推荐方法进行深入研究和实验后，我们基于 Web 技术实现了基于学习风格诊断的自适应学习原型系统。首先对该系统进行需求分析和架构、功能、数据库设计，之后简要介绍了关键技术并对用户信息管理、学习资源管理和在线学习模块进行了界面展示。

第五章：总结与展望。该章节对本研究的主要工作进行了回顾和总结，并提出了一些研究不完善的问题，指出了下一步工作的方向。

第二章 在线学习风格模型的研究

2.1 学习风格模型的发展及应用

2.1.1 学习风格模型的发展

学习风格表示学习者在学习过程中经常采用的习惯性学习方式和倾向，是学习者在一定的生理特征基础上，在长期的学习活动中逐步形成的相对稳定的学习方式偏爱^[33]。它的概念最初于 1954 由 Herbert Thelen 在研究学习者个性化差异时提出^[34]，并且在之后的几十年中，心理学、社会学和教育学等不同领域的学者都建立了自己的学习风格模型。然而，因为每个学者的背景和出发点不同，导致每个模型都基于不同的侧重点对学习者进行了描述和分类，至今仍未有统一的学习风格定义和模型，但对学习风格特性的认识是大体一致的，及独特性、稳定性、兼有活动和个性两种功能。

从发展上来看，学习风格早期的研究重心是在不同学习方式、认知风格和学习策略的层面探讨学习风格特征。随后逐渐分化，注重使用实证的方法研究学习风格中的各个因素，借此分析它们对学习过程的影响。21 世纪以来，现代学者更加注重在教学实践中应用学习风格，提高教师对学生指导的表现。

从学习风格模型的侧重点来看，Coffield 对过去 30 年中 71 个学习风格模型进行了汇总，并将这些学习风格模型分为 5 类^[35]，如表 1 所示：

表 1 Coffield 对 71 种学习风格模型的总结分类

类别	学习风格模型或作者名称
第一类	Dunn and Dunn, Gregorc, Bartlett, Betts, Gordon, Marks, Paivio, Richardson, Sheehan
第二类	Riding, Broverman, Cooper, Gardner et al., Guilford, Holzman & Klein Hudson, Witkin
第三类	Apter, Jackson, Myers-Briggs, Epstein & Meier, Harrison-Branson, Miller
第四类	Allinson & Hayes, Herrmann, Honey & Mumford, Kolb, Felder & Silverman, Hermanussen Wierstra de Jong & Thijssen, Kaufmann, Kirton, McCarthy
第五类	Entwistle Sternberg Vermunt Biggs, Conti & Kolody, Grasha-Riechmann, Hill, Marton & Säljö, McKenney & Keen, Pask, Pintrich Smith Garcia & McCeachie, Schmeck, Weinstein Zimmerman & Palmer, Whetton & Cameron

在 Coffield 的分类中，第一类模型按照视觉、听觉、动觉和触觉等对学习风格进行了分类；第二类模型注重学习者的认知风格和能力，主要根据其认知模式进行分类；第三类模型认为学习风格是相对稳定的，是一组性格特征的组合；第四类模型认为学习风格不是一成不变的，而是相对稳定的；第五类模型涉及到学习方法、学习策略以及学习方向等。

学习风格是学习者模型的核心组件。学习风格体现了学习者的学习方式、学习策略、动机、偏好等不同方面的特征，对在线学习有十分重要的意义。同时，学习风格模型在理论和实践上都在不断完善，呈现出以下发展趋势：

- 1) 整合性。早期的学习风格模型较单一，随后一些学习风格模型也开始包含认知风格、动机等其他特征，在各个方面对学习者进行描述。
- 2) 动态性。相对于学习者基本信息等静态特征，学习者的学习风格是变化的。因此需要自适应学习系统实时对学习者的行为进行分析，动态更新学习风格。
- 3) 系统性。学习风格模型在将来会更多地考虑到与教导模型和学习资源模型的联系和相互作用，从系统的角度出发，将自适应系统的功能最大限度地发挥。

2.1.2 学习风格的应用

学习风格主要应用在教学活动中，在确定教学目标后，不仅要有选择地组织教学内容，更要借助于学习风格模型为学生设定有针对性的教学策略，有效的教学策略对达到教学目标十分重要。具体可基于以下方面制定教学策略：

1) 基于感知特征的教学策略

部分学习风格模型存在感知维度，并将学习者分为视觉型和言语型。视觉型的学习者对通过视觉感受到的信息接受度高，善于理解并记忆例如图表、图片或视频等呈现方式的内容。相反，言语型学习者善于记忆听到的和读到的内容。因此，根据学习者感知特征制定不同的教学策略有利于学习者对知识的记忆和理解。

2) 基于认知特征的教学策略

不同的学习风格模型在认知维度的划分不同，有的将学习者分为场独立型和场依存型，取决于学习者在学习过程中是否容易受环境影响，是否乐意与他人合作；有的将学习者分为分析型和综合型，取决于学习者是更善于从局部问题入手还是从全局的角度考虑问题；还有的将学习者分为谨慎型和冲动型，取决于学习者惯用的思考时间，对问题

的考虑是否更全面。所以，在不同的情境中，针对学习者不同的认知特征进行教学，可使学习者在学习过程中更为放松，提高对新知识的接受度。

3) 基于个性特征的教学策略

针对学习者外向或内向的个性特征制定教学策略，提高学习者的学习信心和积极性。

学习风格不仅可以应用于教学策略的制定，也可以应用于学生的自我评估。通过对学习者的学习风格进行定量测量和定性分析，使得学习者对自己的学习风格有所了解，清楚自己在各个特征上的长处和短处，在学习过程中主动选择合适自己的学习手段和方法。同时弥补自己的不足，以适应不同的学习环境，提高学习效率。

2.2 学习风格的诊断

学习风格的诊断根据是否需要学生额外回答相关问题可分为显式和隐式两种诊断方式（见图 1）。显式方法比较传统，通常要求学生填写针对该学习风格模型设计完善的测量量表。随着数据挖掘技术的成熟，人们逐渐倾向于学习风格诊断的隐式方法，即不需要学习者的额外操作，而在学习者的学习过程中对其学习风格进行诊断。

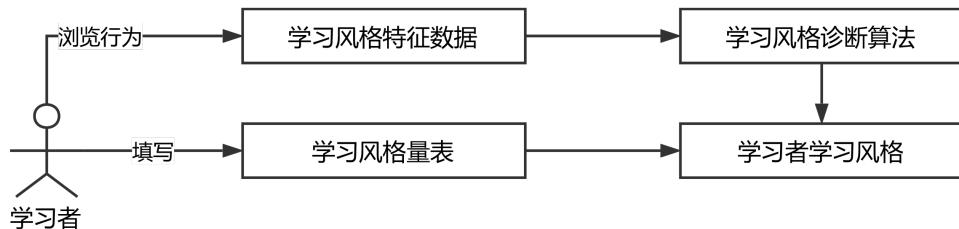


图 1 学习风格诊断流程

2.2.1 学习风格的显式诊断

由于学习风格模型的众多与分化，每个学习风格模型所对应的学习风格测量量表也各有不同。Kolb 经验学习理论针对学习过程进行建模，根据学习者对不同环节的偏好划分成聚敛型、发散型、同化型和顺应型四种。该理论较为成熟，存在不同修订版本的学习风格量表。Kolb^[36]针对成人的学习风格量表 LSI 包含 24 个题目，有不同的历史版本，最著名的包括 LSI-1976, LSI-1985 和 LSI IIA，在计算机教育、艺术、健康、职业培训和性别分析等领域应用广泛。Honey 和 Mumford^[37]对 LSI 进行了改进，于 1986 年设计了学习风格问卷 LSQ，在测量管理者的风格上更有效。Myers Briggs^[38]基于荣格的人格理论提出了人格类型模型，该模型从三个层面描述了人格类型特征。Myers Briggs Type

Indicator (MBTI) 测量量表基于此人格理论设计，从对应的三个层面，共四个角度对学习风格类型倾向进行判定，每个测验项目通常为两项选择题，被广泛应用于教育学和管理学等成人教育领域。除了从学习过程和人格特征方面对学习者的建模，也有一部分学习风格模型考虑到了认知偏好的维度，将学习者分为三类，即视觉型、动觉型和听觉型。该理论在描述和测量上更佳便捷，因此产生了大量基于此理论的学习风格量表，如神经语言程序疗法的两位创始人 Bandler 和 Grinder 制定的认知风格偏爱的学习风格量表^[39]；基于 Web 的 VARK 量表，包含 13 个生活具体场景并提供对应不同认知偏爱的选项，该量表可以自动进行计分统计和建议反馈；在线学习风格测验 Abitor^[40]，分为两个在线量表共计 78 个题目，可以根据测量结果提供反馈建议。

针对学习风格量表的研究主要集中在单一量表统计测量学特点的分析。Veres 等人对 LSI-1985 使用两个大样本对再测相关系数进行检验；Ruble 等人在检验 LSI-1985 测得的学习风格稳定性后发现前后的测验结果有正相关。Allison 和 Hayes^[41], Caple 和 Martin^[42]都对 LSQ 的统计学特征进行了检验；Swailies 和 Senior^[43]对 LSQ 的结构效度进行了检验。

然而，通过学习风格量表的测量方式较为传统，存在着很多问题。Koob 和 Funk^[44]认为学习风格很难被清晰地描述，也很难对相关概念进行准确界定，因为该理论的基础是在大脑发生的隐藏过程。同时，学习风格量表只使用了自我评价式的单一模式，过于主观，无法准确地测量。而且大部分学习风格模型认为学习者的特征是变化的，因此也需要学习风格量表不断修订和对学习者重新测试，保证结果的准确性。

2.2.2 学习风格的隐式诊断

由于学习风格量表过于主观和缺乏及时修订等普遍问题，越来越多的研究者开始尝试利用机器学习中普遍使用的分类算法对学习风格进行隐式诊断。隐式诊断方法分为学习风格特征数据的收集和学习风格诊断两个步骤。通过调研，在第一个环节中学习数据的获取源头可分为三种：

1) 日志信息

日志文件保存了学习者与系统交互的行为信息，包括学习者名称，页面地址，日期时间和在当前页面所产生的行为。通过挖掘学习者行为日志信息可以获取学习者的选择偏好和行为序列等数据，被广泛应用在各种情景。

2) 学习者背景信息

学习者背景信息是静态的，通常在学习者注册时录入，包括例如性别、年龄、文化背景、教育专业等信息。学习者背景信息在学习风格隐式诊断上应用较少，但研究表明它们对学习风格有潜在的决定性作用^[45]。

3) 定制化信息

除了学习者行为和基本个人信息外，还有其他较难获取的信息，如知识背景、学习目标等特征，这些信息需要在系统的前期设计时充分考虑，在学习者的学习过程中巧妙地进行提取。

Truong^[46]对学习风格在自适应学习系统的隐式诊断方法进行了调研分析，其中基于规则的学习风格和基于贝叶斯网络的学习风格最为常用。基于规则的诊断是指研究者们将不同的学习风格的诊断规则编码在自适应学习系统中，Graf 等人^[47]基于 Felder-Silverman^[48]学习风格的描述所设计的规则，例如某个学习者经常点击练习题，则该学习者具有积极的学习风格。通过对 127 名学习者进行测试，得到了可信的结果。贝叶斯网络是常见的分类算法，部分研究者们通过贝叶斯网络评估某学习者在当前行为下的学习风格分类概率。同时，也有部分研究使用了朴素贝叶斯方法，这些方法都得出了有说服力的结果。例如，Garcia 等人^[49]的评估结果展示了对于 Felder-Silverman 学习风格分类 58% 到 77% 的准确率。然而，现今大部分的学习风格诊断策略都是单一算法，很少有研究应用混合模型或集成学习方法，此方面的研究仍有待发展。

2.3 在线学习风格模型（OLS）

2.3.1 传统学习风格模型的不足

自适应学习系统中学习者的建模主要是为了量化个性化信息，以便更好地支持个性化学习。学习风格模型作为学习者模型中不可忽视的一部分，所体现的个体在学习策略、学习方式等方面的偏好在线学习中有重要意义。随着技术的发展，在线学习方式与传统学习方式在上述各个层面都有很大的不同，这也就使得学习者的学习方式和学习策略都发生变化，部分传统学习风格模型所描述的学习者特征已经变化、增加甚至减少，对在线学习者的描述不再准确。另外，传统的学习风格模型往往是单一模型，只关注学习者在部分特征上的差异。如 Kolb 模型只关注学习过程和信息处理方式，而 Myers-Briggs

模型只关注学习者的性格等，它们反应的学习风格概念从定义上看并不是全面的。同时，传统的模型维度往往较低，如常用的 Felder-Silverman 模型只有四个维度，在这个信息呈爆炸式增长的年代，很难将在线学习者进行有效的区分，降低了自适应系统的个性化程度。除此之外，也有学者对模型进行改进或整合，整合的模型相对完整，但应用较少。应注意的是，每个学习风格模型都有它们提出的背景和意义，因此在整合的过程中应充分理解分析各个模型，考虑每一个模型的结构效度，而不是将单一的模型进行简单拼凑。可见，针对在线学习者提出改进或谨慎整合的在线学习风格模型十分必要。

2.3.2 在线学习与传统学习的对比

在传统的学习环境中（特指固化的课堂教学形式、不变的教学方式、围绕书本的内容的课堂），知识只能通过课堂学习或教师教授才可以获得，因此传统教育采取的是以书本、课堂和教师为中心的教育方式，教师控制着教学的过程和节奏。相反，在线学习环境中的学习工具、学习资源、教学方式等都围绕着学习者提供服务，以学生为中心，学习者可以自主控制整个学习过程。邓恩夫妇（Dunn and Dunn）^[50]是较早研究学习风格的权威。他们认为学习风格的形成受多层面多种因素的影响，即学习风格要素。学习风格要素使学习者个体形成了习惯性的信息接受、加工和存储的偏爱方式。因此，可以借助学习风格要素的五大维度对传统学习和在线学习的特点进行对比，表 2 展示了对比结果。

表 2 基于学习风格要素对传统学习和在线学习特点的对比

要素	传统学习	在线学习
环境	时间地点受限，无自由发挥	时间地点不受限，自由度大大提高
情感	内容乏味、单调，容易厌学	学习者处于主导地位，发挥主观能动性
社会	只能从教师那里获取信息	可以与众多学习者互动
生理	教科书形式过于单一	富媒体技术丰富了接受知识的途径
心理	教学方式固化，学习者缺乏对知识的理解	自主探究问题，深刻理解知识

1) 环境

传统学习的教学时间、地点受限，学生不能根据自身需要或喜好开展学习活动。然而，在线学习者的学习时间、地点不再限于课堂和教师，他们随时随地都能通过电脑、手机等设备获取信息，对学习环境的选择更加自由。且学习资源丰富，有大量的网络资

源以及学习者在学习过程中自己生成的资源，用户随时可以获取最新的内容和资料。

2) 情感

传统学习中，学习计划、学习进度等都由教师安排规定，限制了学习者主观能动性的发挥；由于教学内容单调枯燥、教学方式固化，容易导致学习者产生厌学情绪。在线学习中，学习者处于主导地位，可以根据自身特点选择学习目标、学习策略、安排学习进程等。同时，很多在线学习者是在自身的学习需求下进行在线学习，具有主动的学习动机。

3) 社会

传统学习环境中，学习者通常需要询问教师以获得最为准确的回答。然而教师很难同时回答很多学生的问题，且教师的能力也是有限的，可能有些问题无法回答最准确的答案。因此这种一对多的单一方式限制了学生获得更多知识的可能。在线学习为学习者提供了交流信息等平台，学习者或教师们通过网络论坛中的协作，可以帮助学习者从多个角度分析问题，还可以使学习者有时间进行更多且全面的思考，扩展了知识的来源渠道，同时提高学习者自己解决问题的能力。

4) 生理

传统的学习内容较为呆板，教师也过于依赖教科书上的内容，导致学习者所学知识极度脱离应用情景，更使得学习变成了死记硬背的过程。网络的普及使得学习者可以选择不同的学习方式，提高学习者的积极性。

5) 心理

传统学习教学方式固化，无法适应学习者多元思考的认知特性，造成学习者对知识的产生、应用缺乏深刻的了解和掌握。在线学习的学习方式与教学方式逐渐多元化，由于学习者可以自己控制并重复查看，他们得以将更多的注意力放在具体的解决方法上，对于得到的知识记忆会更加深刻。

综上所述，在线学习在环境、情感、社会、生理和心理五大学习风格要素上都与传统学习不同，以至于两者所塑造的学习风格有着根本上的差异。可见对在线学习风格模型研究的重要性。

2.3.3 在线学习风格的相关研究

目前国内外针对在线学习风格的研究还不成熟。Jonassen^[51]对在线学习平台的数据

库中存储的学习者访问数据和日志文件进行了分析，从中归纳学习者行为。Lo 等人^[52]通过分析在线学习者浏览的页面，对学习者的学习风格进行研究。Lee 记录了在线学习者的学习需求和行为，并探究学习者的特征和学习效果评价方式。相比于国外的实践性研究，国内的研究更偏向于通过调查问卷等方式获取数据，并将分析出的特点与学习风格偏好相关联。郝丹等人在学习信息加工论、荣格的人格特征类型和 Kolb 经验学习理论将学习者按照生理维度、心理维度和经验维度进行划分，构建了中国远程学习风格特征的三维模型。黄海波^[53]针对在线学习者的特征展开了深入的分析和探讨，主要侧重于动机、学习需求、生理和心理等方面。李念^[54]基于在线学习行为分析提出了在线学习者学习风格三维模型。李素珍^[55]应用数据挖掘技术，基于网络学习行为分析并提出了包括环境、社会、心理、生理和情感的五维在线学习风格模型。

虽然部分研究者已经开始关注在线学习风格并有初步研究，但目前大部分自适应学习系统所应用的依旧是传统学习风格模型。在前文提到的 Truong 的综述文章中，使用 Felder-Silverman 学习风格模型的在线学习平台约占 50%，排名第二位为 Kolb 模型，占 8.6%，其次还有 Honey & Mumford 和 VARK 模型。因此，对于在线学习风格模型的设计需要取长补短，在尽量不破坏传统学习风格模型的情况下进行新的融合与创新。

2.3.4 在线学习风格建模

2.3.4.1 在线学习风格理论基础

在众多学习风格模型的基础上，本研究将综合邓恩夫妇的学习风格要素理论、Entwistle 学习风格模型^[56]、Felder-Silverman 模型、Myers-Briggs 和 Kolb 模型作为在线学习风格模型的理论基础。

如上文所述，邓恩夫妇将影响学习者学习风格的因素归结为学习风格要素。学习风格要素与学习风格的关系如图 2 所示。可见，学习风格要素影响并决定着学习者的学习风格，学习风格又表现为学习者在学习过程中的行为偏好，同时也可以通过学习者行为偏好来诊断其学习风格。



图 2 学习风格要素、学习风格以及行为偏好的关系

因此，本研究从学习风格要素出发，在学习风格要素的各个维度，对经典的学习风格模型（见表 3）进行整合，为在线学习者构建学习风格模型。

表 3 经典的学习风格模型

学习风格模型	简述	学习者分类
Entwistle	将学习风格按照学习途径来进行划分	深度 / 表面 / 策略型
Felder-Silverman	对前人模型的总结概括，包含四个维度，每个维度有两种类型的学习者	感知维度（感悟 / 直觉型） 信息输入维度（视觉 / 言语型） 内容理解维度（序列 / 综合型） 信息加工维度（活跃 / 沉思型）
Myers-Briggs	认为学习者的性格影响其学习方式	内倾 / 外倾型 感觉 / 直觉型 思维 / 情感型 判断 / 理解型
Kolb	强调学习过程的重要性	具体经验 / 抽象概括 反思观察 / 积极实践

在环境要素方面，较少有模型对其进行建模，因为这不仅包括地点，更涉及到温度、光线等多种物理因素，较难对此维度进行诊断，因此本研究假设在线学习者所处的环境都相同，即某一平台或系统的在线环境，该环境对于所有学习者而言是一样的。

对于情感要素，本研究借鉴 Entwistle 模型，使用“动机”这一特征来区分学习者。从 Entwistle 对于学习者的分类可以看出，学习者的学习程度有所不同，而造成这一差距的主要原因之一是学习者的内在动机。同时动机作为特征也易于量化，动机强的学习者在各方面的学习行为都会更加主动积极。

社会要素主要考察学习者是否喜欢独自学习，Myers-Briggs 模型就有对于学习者外倾或内倾的维度。本研究采用“交流”特征来区分学习者，通过分析学习者对于论坛的浏览和发言情况决定学习者交流能力的强弱。

生理要素可以从视觉、听觉和动觉等方面描述学习者的个人特质，考虑到动觉信息较难捕捉，本研究借鉴了 Felder-Silverman 模型，从视觉和言语两个特征区分学习者。

心理要素需要考虑的方面更多一些，该要素不仅反映了学习者对整个信息接受流程的偏好，从信息接收到理解，不同的学习者有不同的偏好。观察 Felder-Silverman 模型、Myers-Briggs 模型以及 Kolb 模型，它们都对学习者感知的方式有所描述，及通过具体

经验反思或理解高度概括的信息。同时，根据在线学习的经验，可以发现学习者的内容理解方式也有很大不同，有些学习者按顺序学习课程，而另外一些学习者则进行跳跃式的学习，最终把握知识，因此本研究加入了序列和综合两种不同的内容理解方式。

2.3.4.2 OLS 在线学习风格模型

通过 2.3.2 中对于传统学习模型的对比分析，为了在自适应学习系统中应用学习者的学习风格模型，本研究基于学习风格要素，同时借鉴其他现有经典模型，筛选出典型的“学习风格特征”，并将学习风格特征与学习行为一一对应，得到了如表 4 所示的 4 个要素下 8 个维度的在线学习风格 OLS（Online Learning Style）模型。模型的目的是针对学习者在学习过程中的个人特质进行建模，提供学习行为的解释和理论依据，并对在线学习者进行区分，最终基于在线学习风格为不同类型的学习者提供不同类型的资源、路径等，实现自适应学习机制。

表 4 OLS 模型的学习风格特征和对应行为

要素	学习风格特征	学习行为
情感	动机（Motivational）	在学习平台中停留总时间等
社会	交流（Communicational）	查看学习论坛、在论坛中发言等
生理	视觉（Visual）	浏览视频、图片类的材料等
	言语（Verbal）	浏览文字、音频类的材料等
心理	感悟（Sensory）	浏览故事性强、结合实际的材料等
	直觉（Intuitive）	浏览理论性强、高度概括的材料等
	序列（Sequential）	从前往后、按照规定好的顺序进行学习等
	综合（Global）	没有一定的逻辑顺序、分散式的学习等

模型中每个要素及特征的含义如下：

1) 情感

情感要素下的动机维度参考了 Entwistle 学习风格模型，该模型将学习者从学习途径进行划分，同时也表现了学习者的内在动机。情感维度主要表示学习者对学习的动机程度，对过程的坚持性和责任感，在 OLS 模型中主要通过动机程度进行描述。动机弱的学习者积极性低，遇到困难容易放弃。相反，动机强的学习者会积极主动完成学习，乐于查看其他资料丰富自己的课外知识。

2) 社会

该维度主要描述学习者在学习过程中与周围学习者的互动情况，用交流特征表示。交流性强的学习者会经常查看学习论坛，提出问题并发表自己的看法。而交流性弱的学习者更喜欢自己独立学习。

3) 生理

生理要素下的视觉、言语维度参考了 Felder-Silverman 模型。该维度基于信息加工理论，在生理感觉偏好的角度表现学习者的特征。视觉型学习者经常浏览图片、图表和视频等学习资源，而听觉型学习者更喜欢阅读文字或浏览音频等学习资源。

4) 心理

心理要素下的序列、综合维度参考了 Felder-Silverman 模型，这两个维度主要描述学习者在学习过程中对知识的处理偏好。序列型学习者通常按照规定好的章节顺序进行学习，很少跳过某些章节。而综合型学习者的学习过程缺乏规律，常常无条理地进行章节转换，日志信息缺乏条理，但这种类型的学习者常常有较强的全局观，善于从总体看问题。

另外，心理要素下的感悟、直觉同时参考了 Myers-Briggs 和 Kolb 模型。这两个维度基于荣格的人格类型理论，表现学习者的心理特征。直觉型学习者偏好浏览理论性强的内容，而感悟型学习者偏好浏览事实性材料，并从中进行知识的理解吸收。

2.3.4.3 OLS 模型的创新点

总体上，在线学习风格 OLS 模型有取舍地总结了影响在线学习风格的八个特征，解决了现有模型片面化、简单化和适用性低的问题。和上述经典的学习风格模型对比，主要有以下创新点：

- 1) 以在线学习者的行为方式及特征为出发点，对学习者的不同浏览偏好进行合理的抽象，同时对应部分经典的学习风格模型，具有坚实的理论基础。
- 2) 建立了相对完备的理论结构。从邓恩夫妇的学习风格要素出发，完整地对学习者学习风格进行描述，而不是只考虑某一层面的学习者偏好。
- 3) 特征互不对立。在前人的模型中，每一维度上的特征都是对立存在（如在 Felder-Silverman 模型中一个学习者只能是视觉型或言语型），但现实中学习者可能对于不同的学习内容采取不同的学习策略，故可能两者的特性都具有。如果某学习者对“序

列”、“综合”两个看似矛盾的特征选择了同样的分数，则恰巧说明此学习者在学习时善于变化，有能力去改变自己的习惯去学习不同的课程。所以，与其他学习风格模型都不同的一点是，OLS 模型中的八个特征是可以同时存在的，力求对在线学习者的特征进行描述，而不是将学习者绝对地判定到某个特定类型。

4) 特征易于量化。前人模型中的维度与特征过于抽象，难于量化，在传统教育模式下只能通过调查问卷的方式诊断学习风格，过于主观。而在线学习风格侧重于学习风格与行为偏好的紧密联系，如动机和交流维度，在自适应学习系统中易于衡量。

2.4 在线学习风格的诊断方法

2.4.1 在线学习风格量表的设计与分析

在改进传统模型并建立在线学习风格 OLS 模型后，需要考虑如何真正地对在线学习者对学习风格进行诊断，因此设计了 OLS 量表，以此显示地判定学习者的在线学习风格。为确定 OLS 模型及量表的设计和理性，本研究对 380 名学习者进行了在线学习风格诊断，下文主要介绍量表的设计过程和调查结果分析。

2.4.1.1 在线学习风格量表的设计

1) 初始设计

通过在线学习风格量表，可以有针对性地得到学习者的学习风格。本研究基于学习风格特征，结合在线学习在学习内容、学习活动等方面的特点，并参考已有经典学习风格模型的相关量表，设计了学习风格各维度的测量项目，对学习者在线学习行为进行提问，最终得到《在线学习风格测量》量表初版，共包含 23 个测量题目（见附录 1）。为避免题目繁多，造成测试者疲劳，对每个特征只设定了 2-3 道题目，每道题针对频率进行提问。为了避免无法将被测者区分的中性回答，同时便于之后的比较与计算，量表只设定了“0（从未）”、“1（很少）”、“2（一般）”、“3（经常）”四个选项。

2) 先行性研究

主要考察量表是否适合被试。在先行性研究阶段，我们随机小规模地选取了北京航空航天大学中法工程师学院的 10 名学生，请他们填写量表并对题目是否明确和题目顺序等提供意见，主要考察被试者是否可以完全理解量表的问题并在填写某些题目时不受其他题目影响。结果发现一些题目表述并不清晰，于是根据先行性研究的结果对量表

初版中的部分题目进行了修改。

3) 信度与效度检验

信度即可靠性，它是指采用同样的方法对同一对象重复测量时所得结果的一致性程度。本次研究采用重测信度法，随机抽取 10 名被调查者，请他们两次填写同一份量表，中间间隔 6 周。经过计算后，两次测试的平均皮尔逊相关系数为 0.713。依据再测信度检验标准，上述的检验结果是可接受的。

效度及有效性，指测量工具或手段能够准确测出所需测量的事物的程度。由于本次量表的制作是参考其他经典模型的量表，相同学习特征下的题目有极高相似性，只是为复合在线学习行为进行稍加修改，故可认为其内容具有一定效度。

2.4.1.2 在线学习风格量表的调查结果分析

在验证 OLS 量表的有效性后，本文在问卷星 (<https://www.sojump.com/>) 平台制作并发布了量表，一周时间内共收到 380 份填写结果。之后，本文对得到的在线学习风格数据进行统计分析，以此对量表题目进行筛选和改进。数据分析共分为以下步骤：

1) 数据预处理

在进行统计分析之前，通常会对数据进行预处理，以筛选不可靠或对结果有严重影响的数据。异常值处理的目标是去除偏离标准过大的值，减小某些过大或过小的值对结果的影响。另外，对于数据的归一化，由于每道题目只有 0、1、2、3 这 4 个取值，故对每个维度取平均值即可得到归一化的数据。

本次异常值处理主要是去除不可信的数据。筛选后，将不进行在线学习、填写时间在 100 秒之内（平均 220 秒）或对于两道对立测谎题给出相同答案的结果去除，剩余 338 条有效数据。

2) 题目相关性分析

在线学习风格模型共有 8 个特征，然而每个特征在量表中均对应 2-3 道题目。此步骤通过对题目之间的皮尔逊相关系数分析，去除相似度过高的题目。经计算的相关系数矩阵如图 3 所示：

	Motivation	Communication	Visual	Verbal	Sensing	Intuitive	Sequential	Global																	
	5	13	18	26	6	14	19	7	27	15	20	8	16	21	28	10	23	9	17	22	29	24	11		
Motivation	5	1																							
	13	0.22	1																						
Communication	18	0.30	0.25	1																					
	26	0.19	0.27	0.47	1																				
Visual	6	0.60	0.26	0.28	0.25	1																			
	14	0.32	0.33	0.34	0.22	0.41	1																		
Verbal	19	0.16	0.24	0.32	0.22	0.31	0.32	1																	
	7	0.23	0.21	0.29	0.29	0.27	0.28	0.22	1																
Sensing	27	0.31	0.22	0.19	0.21	0.25	0.21	0.07	0.27	1															
	15	0.15	0.20	0.28	0.19	0.15	0.30	0.18	0.08	0.17	1														
Intuitive	20	0.10	0.23	0.35	0.28	0.12	0.21	0.24	0.49	0.20	0.14	1													
	8	0.29	0.27	0.25	0.23	0.41	0.34	0.24	0.42	0.26	0.21	0.24	1												
Sequential	16	0.06	0.17	0.20	0.13	0.16	0.14	0.23	0.20	0.13	0.28	0.24	0.17	1											
	21	0.03	0.05	0.16	0.13	0.07	0.20	0.08	0.20	0.13	0.12	0.28	0.09	0.25	1										
Global	28	0.28	0.30	0.38	0.31	0.38	0.33	0.20	0.21	0.49	0.25	0.25	0.37	0.20	0.15	1									
	10	0.29	0.26	0.27	0.20	0.24	0.31	0.21	0.25	0.30	0.13	0.12	0.39	-0.07	-0.07	0.31	1								
Sequential	23	0.24	0.31	0.43	0.38	0.22	0.28	0.25	0.25	0.24	0.23	0.34	0.25	0.19	0.20	0.29	0.20	1							
	9	-0.09	-0.03	-0.01	0.07	-0.08	-0.04	0.03	0.08	0.01	-0.10	0.06	0.04	0.09	0.00	0.01	0.00	-0.03	1						
Sequential	17	0.05	0.19	0.30	0.18	0.07	0.22	0.27	0.22	0.17	0.20	0.36	0.18	0.30	0.31	0.23	0.09	0.29	0.06	1					
	22	0.10	0.09	0.35	0.29	0.10	0.08	0.20	0.27	0.15	0.08	0.42	0.11	0.17	0.35	0.25	0.16	0.28	0.00	0.46	1				
Global	29	0.25	0.25	0.27	0.21	0.22	0.25	0.19	0.12	0.34	0.20	0.20	0.38	0.11	0.05	0.44	0.32	0.27	0.00	0.12	0.12	1			
	24	0.28	0.14	0.28	0.18	0.29	0.20	0.14	0.30	0.29	0.09	0.27	0.28	0.11	0.21	0.41	0.18	0.28	-0.08	0.20	0.26	0.26	1		
	11	0.10	0.07	0.15	0.13	0.04	0.08	0.04	0.04	0.16	-0.01	0.12	0.17	0.07	0.03	0.12	0.14	0.08	0.15	0.11	0.04	0.16	0.01	0.07	1

图 3 OLS 量表题目的相关系数矩阵

可见，大部分题目之间是不相关的，但仍有小部分题目的相似度在 0.4 以上。其中：

- A. 6 题和 5 题、14 题、8 题的相关系数分别为 0.6、0.41 和 0.41；
- B. 18 题和 26 题、23 题的相关系数为 0.47 和 0.43；
- C. 20 题和 7 题语义相近，相关系数为 0.49，同时和 22 题相关系数为 0.42；
- D. 28 题和 27 题、29 题、24 题相关系数为 0.49，0.44、0.41；
- E. 17 题和 22 题语义想近，相关系数为 0.46，17 题的方差相对较小；

通过对方差和相关系数的比较，最终去掉了和两个以上的题目相关系数大于 0.4 的 6、18、20、28、17 五道题目，使所有题目都保持在弱相关的范围内。通过对量表结果的分析，最终筛选出 OLS 量表的题目。

为了考察在线学习风格模型对于学习者的区分能力，本研究对调查问卷的结果进行了 K-means 聚类分析，将 338 名学习者分为五组，每组中心点用雷达图表示，见图 4。

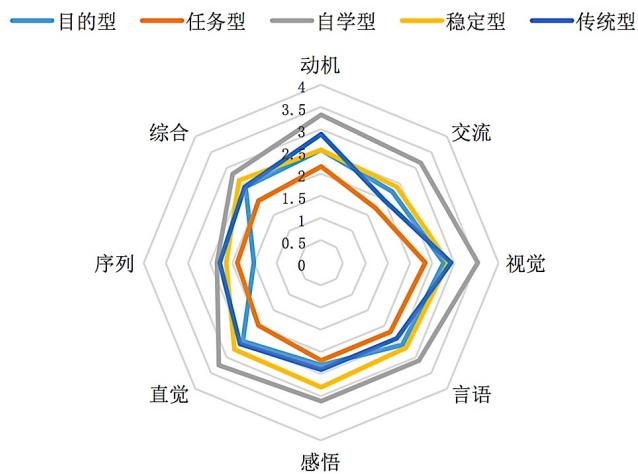


图 4 聚类中心点学习风格分析图

从图 4 中可以看出，八个维度很好地区分出了各种类别的学习者。尤其在动机、交流、序列这三个维度上，不同类别学习者的重合并不明显，可以很好地看出学习者之间的区别。同时，可以简要对这五种类别的人群进行分析：

- 1) 目的型学习者：在学习过程中会比较跳跃地学习，不太注重学习的递进顺序。此类学习者根据自己想要了解的知识点进行学习，在学习过程中有自己的节奏，很可能是一些已经具有某种领域背景的学习者，通过网络对自己已经有的知识进行查漏补缺。
- 2) 任务型学习者：在各个维度较一致，但都相对其他类型的数值小一些。此类学习者在网络平台的行为相对较少，可能是按要求进行网络学习的学生，他们的动机最弱，且由于所修在线课程的各种限制（如所有学习者都必须按老师的要求学习某节在线课程，学习资源缺乏多样性），因此在学习平台中并不活跃。
- 3) 自学型学习者：在各个维度的值都比其他类别要大，尤其是在交流维度和视觉维度。可见此类学习者热爱且擅长在线学习，同时积极与他人交流，对视频、文字等教程的理解能力都高于他人，接收信息能力强，可能是擅长自学的学习者。
- 4) 稳定型学习者：在各个维度都处在所有类别的平均值的位置，可能是长期进行在线学习，已形成稳定习惯的学习者。
- 5) 传统型学习者：在各个维度的值不太一致，在动机和视觉维度的值明显突出，同时在序列维度的值也高于其他类别。该类学习者可能经常在视频学习平台进行学习，在较规律的时间下按顺序学习视频课程，与传统学习的形式较为相似。

可见，在线学习风格模型不仅可以较好地区分在线学习者，同时可以对学习者的特点提供理论上的解释。

2.4.2 在线学习风格的隐式诊断方法

2.4.2.1 基于规则的 OLS 隐式诊断

由于缺乏大量且可靠的带有学习风格标签的数据，本研究使用基于规则的在线学习风格诊断方法，通过挖掘学习者日志信息，对学习者所访问的学习资源进行分类标签，通过对不同学习风格类型的学习资源点击量进行加权统计判定在线学习风格。该诊断方法分为两步：

- 1) 学习资源标记

本研究着重对学习者模型和自适应模型进行建模，因此简化了对于学习资源建模的

处理，假设学习资源与网页是一一对应的关系，即每个网页内只含一种学习资源和去往其他网页的链接。同时假设学习资源属性与在线学习风格特征是多对多关系，即浏览具有某些属性的学习资源可以表示该学习者具对应的多种学习风格特征。对于大多数在线学习平台来讲，与学习风格关联性强的学习资源和具体属性分类如表 5 所示。

表 5 OLS 特征与学习资源属性的对应关系

OLS 特征	学习资源属性
动机	课程主页类，课外补充知识类
交流	学习论坛类，师生互动类
视觉	视频类，图片类
言语	音频类，文献类
感悟	故事类，事实类
直觉	理论类
序列	章节类
综合	总结类，目录类

学习资源可以具有一种或多种属性，例如某个学习资源的展示方式是视频类，其中包含的内容是故事类，则该资源具有视频类和故事类两种属性。

2) 在线学习风格诊断

对学习资源进行标记后，可以基于学习者行为日志记录诊断其学习风格。令向量 (d_1, d_2, \dots, d_8) 表示每个学习者 x ， d_j 代表学习者 x 的第 j 个在线学习风格特征上的倾向，以数值方式表示。给定包含 M 个学习资源的集合 $I = \{i_1, i_2, \dots, i_M\}$ ，集合 I_j 是物品集合 I 的任意一个子集，有 $I_j \subseteq I$ ，表示包含属性 j 的学习资源集合。根据学习风格特征和学习资源属性的多对多关系，定义学习者 x 的学习风格特征 d 的计算公式为：

$$d = \sum_{j \in A(d)} w_j \cdot NClick(x, I_j) \quad (2.1)$$

其中 $A(d)$ 是学习风格特征 d 所对应的学习资源属性集合， w_j 是学习资源属性 j 对于学习风格特征 d 的权重，且 $\sum_{j \in A(d)} w_j = 1$ ， $NClick(x, I_j)$ 是学习者 x 对于具有属性 j 的学习资源的总点击量。

2.4.2.2 OLS 隐式诊断在教育数据集的应用

利用上述基于规则的 OLS 隐式诊断方法, 可以对记录了学习者行为日志和学习资源类别的在线学习平台进行学习者在线学习风格诊断。同时, 通过在真实数据集上的实践, 我们可以考察在线学习风格模型对学习者的区分程度。经过调查, 本研究决定对 Open University Learning Analytics dataset (OULAD) 数据集^[57]进行研究。该数据集来自英国的 Open University 在线学习平台, 包含 2013 年和 2014 年内 22 个课程模块、32593 个学习者的信息和 10655280 条学习者行为日志。在数据集中对所有学习者进行了匿名处理, 用唯一的标识符标注学习者和网页。OULAD 数据集中表的结构和关系如图 5 所示。该数据集共分为学习者信息、学习者活动和课程信息三部分, 对于隐式的学习风格诊断, 本研究主要对表 studentvle 和表 vle 中的信息进行挖掘。

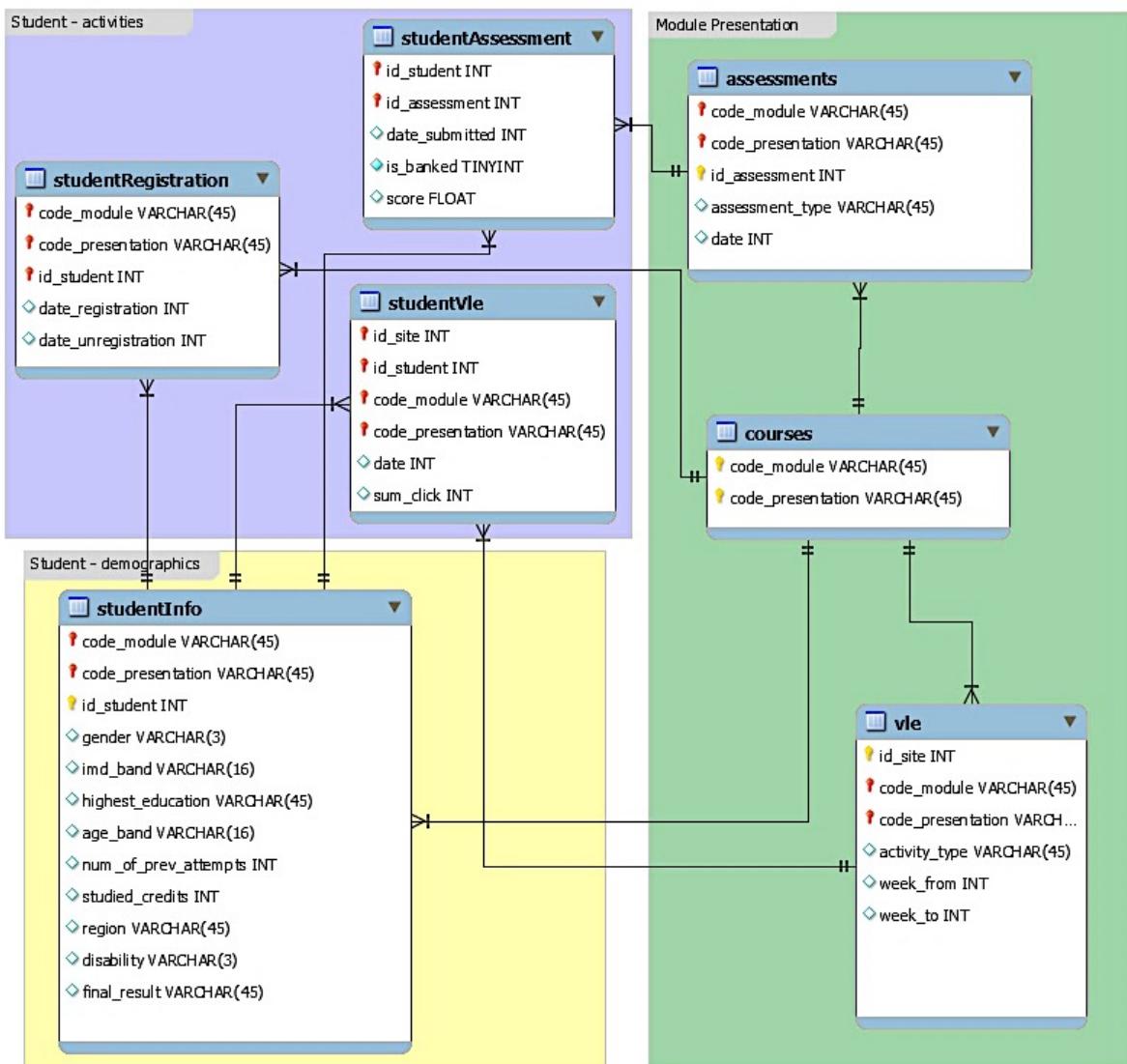


图 5 OULAD 数据集结构

表 studentvle 是学习者行为日志，记录了每个学习者所点击网页的模块代码、唯一标识符、日期和在该页产生的点击量总和。表 vle 是网页附加信息，表示了每个网页与对应课程模块的信息，其中活动类型 (activity_type) 字段表示了该网页的活动类型，共分为 20 类。经过对活动类型的分析，认为其中大部分类型可以与 OLS 隐式诊断方法中所需要的学习资源属性产生对应关系，并对在线学习风格产生影响。对 OULAD 数据集中活动类型的分析结果如表 6 所示。

表 6 OULAD 活动类型的描述和与 OLS 的对应关系

活动类型	描述	学习资源属性	OLS 特征
resource	提供 PDF 格式的书籍资料的网页	文字类	言语
oucontent	提供课程目标简介的网页	总结类	综合
url	提供视频的网页	视频类	视觉
glossary	提供对课程中术语的说明的网页	-	-
sharedsubpage	提供多个学院共用信息的网页	-	-
dualpane	提供一些信息并根据信息进行交互活动的网页	事实类	感悟
ouelluminate	提供课程直播服务的网页	视频类	视觉
ouwiki	百科网页	理论类, 文字类	直觉, 言语
homepage	课程主页	主页类	动机
forumng	论坛类网页	论坛类	交流
questionnaire	提供课程问卷的网页	-	-
repeatactivity	提供知识回顾的网页	总结类	综合类
subpage	提供具体章节内容的网页	章节类	序列
oucollaborate	提供助教与学生视频讨论服务的网页	师生互动类	交流
page	提供课程相关提示的网页	总结类	综合
folder	提供课程所需文件的网页	-	-
dataplus	提供音频、视频或 PDF 形式的附加资料的网页	-	-
externalquiz	外部网站测试题的网页	课外补充知识类	动机
htmlactivity	包含学习内容的可交互网页	-	-
quiz	小测验	课外补充知识类	动机

从表 6 可以看出，绝大多数的活动类型都可以与学习风格有对应关系。其中部分活动类型与学习风格无关，如提供课程术语说明的网页、提供学院信息的网页、提供课程问卷的网页等。另外还有部分网页等活动类型定义太过笼统，无法确定它们与某个学习特征的关系，例如提供课程所需文件的网页，该网页所提供的文件可以是文字或视频资源，也可以是总结类或具体章节类的学习内容，在没用具体学习资源数据的情况下无法进行深入分析。

基于公式 2.1 和表 6 的分析，在缺乏标记数据的情况下另学习资源属性对学习风格特征的权重相等且和为 1，挖掘学习行为日志，统计出具有学习行为的 26074 名学生的在线学习风格，其中每个学习特征的值表示该学习者对相关网页的点击总数。统计结果中每个学习风格特征的统计分析如表 7 所示。

表 7 OULAD 数据集中学习者学习风格特征均值及标准差分析

特征	动机	交流	视觉	言语	感悟	直觉	序列	综合
均值	121	119	9	26	1	20	96	98
标准差	179	248	16	39	2	60	130	161
最小值	0	0	0	0	0	0	0	0
25%	25	9	1	4	0	0	14	5
50%	64	45	4	13	0	0	44	31
75%	145	129	12	32	0	14	131	121
最大值	4465	4994	935	861	68	1708	1603	2273

可以看出，该数据集中的学习者在动机、交流、序列和综合这四个特征上的行为较多，且不同的学习者之间区分明显。然而，在视觉、感悟和直觉这三个维度上统计出的学习行为较少，原因主要是因为 OULAD 数据集来自普通的在线学习平台而不是专门基于学习风格打造的自适应学习平台，其中对于可以体现学习者心理要素的学习资源设计较少，且每个课程缺乏不同类型的学习资源，出现文字类的学习资源多于视频类的学习资源的情况。

在之后的章节中，本研究将会基于此次对 OULAD 中学习者学习风格的诊断，应用自适应推荐方法，实现基于学习风格的学习资源自适应。

2.5 本章小结

本章对经典的学习风格模型的发展进行了回顾，对学习风格的应用与诊断方式简要总结，归纳了传统学习风格的缺陷。之后对比了传统学习模式与在线学习模式的异同，基于 Dunn and Dunn 学习风格要素提出了适应当下在线学习者的在线学习风格模型 OLS。OLS 同时混合了 Myers-Briggs 学习风格模型、Kolb 学习风格模型、Felder-Silverman 学习风格模型以及 Entwistle 学习风格模型，从情感、社会、生理、心理共四个要素出发，在动机、交流、视觉、言语、感悟、直觉、序列、综合共八个互相独立的维度对在线学习者的特征以及偏好进行描述。

在建立了在线学习者学习风格 OLS 模型的同时，设计了相应的在线学习风格 OLS 量表，可以帮助自适应学习系统显式地确定某个学习者在各个维度的学习偏好。为了验证 OLS 量表的有效性和题目设置合理性，对 380 名学习者进行了调查，并根据调查结果对题目进行筛选。除了利用 OLS 量表设计了在线学习风格的显式诊断方法外，又提出了基于匹配规则和学习者行为日志的在线学习风格隐式诊断方法。最后，对 OULAD 开源数据集中的学习者行为日志和网页分类数据进行挖掘，诊断该数据集中学习者的学习风格并对各个学习风格特征的结果进行均值和方差统计。

通过进行在线学习风格模型的显式诊断和隐式诊断实验，验证了 OLS 模型对于在线学习者的可区分性，并提供了对于不同类别学习者行为的合理解释。在下一章节中，将通过学习风格模型与自适应学习资源推荐方法的融合进一步证实该模型在实际应用中的可行性及有效性。

第三章 自适应学习资源推荐方法

3.1 以学习者为中心的自适应推荐的传统方法

3.1.1 基于学习者模型的自适应推荐方法

学习者模型是自适应学习系统中的核心部分之一，学习风格又是构建学习者模型强有力的基础。基于学习者模型的自适应学习系统以学习者为中心，根据学习者在学习过程中呈现的偏好特征进行学习资源动态推荐。

国外对于以学习者为中心的自适应学习系统取得了不小的成果，主要的研究方向可归为两类：

1) 从学习者模型出发，利用机器学习中的分类算法，挖掘学习过程中的数据，分析学习者学习风格各个维度上的特征，对学习者进行个性化建模，并针对不同的学习者进行自适应推荐。该角度的重点是对学习者的完整建模，例如基于贝叶斯网络的智能导师系统，通过在导师的角度分析学习者的学习过程，从而对学习者的问题进行回答。前文中提到的 iWeaver 自适应学习平台则基于学习风格模型形成适合学习者认知风格的动态学习环境，同时提供导航帮助，以实现自适应。

2) 从内容模型出发，主要研究对相似学习风格群组进行学习内容推荐。研究者基于聚类算法挖掘相似学习者的浏览记录。这种方法可以帮助学习者找到相似的学习同伴，并快速获得学习资源推荐。在这方面的研究中，聚类算法是研究者们着重考虑的主要内容之一。Zakrzewska^[58]基于贝叶斯算法对学习者学习偏好分析后，比较分析了三种聚类算法的有效性；随后，又基于最近邻算法分析学习者的学习风格，以推荐相似的学习者群组。

国内的研究者们也开始重视个性化学习。东北师范大学的赵蔚^[59]等人基于学习风格搭建了自适应学习系统，并对学习风格进行了修订。该系统加入了推荐算法，可以根据学习者的学习风格和兴趣等特征进行学习方案的个性化定制，其中主要是个性化推荐学习资源和学习行为序列。熊玉珍^[60]的研究同样基于学习风格和学习行为，在学习者的汉语学习活动中进行相关推荐。

3.1.2 基于学习者模型的自适应推荐所面临的挑战

目前，虽然基于学习者模型的自适应学习系统已有较大的进展，但仍存在许多不足。

主要可以总结为对于学习风格的分析过于简单和缺乏两点。

首先，当前的一些自适应学习系统虽然在学习者模型中考虑到了学习风格，并将学习风格结合到学习资源的内容组织中，但由于模块的分离，使得学习过程与学习风格基本脱离。如山东师范大学的郭玲玲^[61]在研究中基于 Kolb 学习风格对学习者进行测试，并针对学习风格对在线学习行为和学习结果的影响进行了研究，证实了依据学习风格差异对学习者进行分组教学可以提高学习效果。然而此类研究多数只是将学习风格应用到前期的测试，并没有在后期针对学习风格的差异影响学习者的学习行为，无法根据学习者的学习风格进行偏好的自适应。

另外，大多数的系统虽然能够准确地分析学习者的兴趣并提供个性化的学习资源，但系统功能却相对零散。例如北京师范大学的余胜泉^[62]教授等人设计开发的学习元平台由学习资源模型的角度出发，基于元数据技术实现跨平台学习资源推荐和社交功能；东北师范大学的赵蔚等人在文献[59]中则以学习者特征模型为主，混合各种推荐算法，重点在于研究和更新自适应学习方案。上述系统虽然在教学中应用了计算机前沿技术，但系统并没有允许学习者参与到教学设计和活动方式的环节中，不能自主改变学习方式或学习环境等。此外，浙江师范大学的张家华等人^[63]在研究中也指出，当前自适应学习系统在学习理论和技术支持之间缺乏协调与沟通机制。

所以，基于学习者模型的自适应学习系统所面临的挑战不仅是最初对于学习者学习风格特征的完整建模和诊断，更在于之后如何将学习风格与学习资源和教导模型相结合，贯彻到系统的自适应功能中，形成完整的体系。

3.2 基于学习风格的自适应推荐方法（AROLS）

在对在线学习者学习风格进行建模和诊断后，为了解决基于学习风格的自适应推荐所面临的另一个问题：学习风格与教导模型和学习资源模型的结合，本研究提出了基于学习风格的自适应推荐方法 AROLS (Adaptive Recommendation based on Online Learning Style)。如图 6 所示，AROLS 主要包括两个部分，即基于学习风格的学习者聚类和针对学习者群组的学习资源推荐。

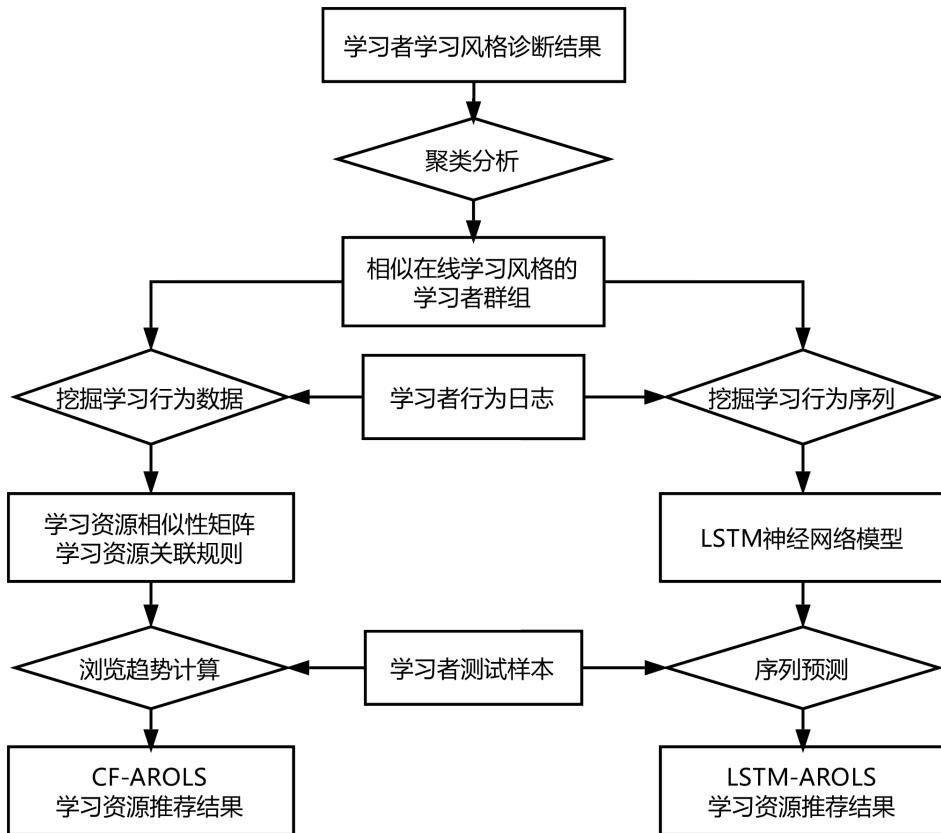


图 6 AROLS 的学习资源推荐流程

针对给定的训练集，即学习者学习风格诊断结果和学习者行为日志，AROLS 首先基于学习风格对学习者进行聚类，将相似学习风格的学习者分到同一群组中。之后对于不同的学习者群组，挖掘该类别学习风格学习者的行为模式和学习资源偏好，实现学习资源推荐。当训练过程结束后，一个学习者测试样本被输入到算法中，AROLS 通过找到该学习者所在群组以及历史行为，基于学习行为模型进行自适应推荐。

下文将分别介绍 AROLS 在各个部分的具体方法。对于学习资源推荐，我们分别研究了基于协同过滤的混合推荐 CF-AROLS 和基于长短期记忆 (long short-term memory, LSTM) 循环神经网络的学习资源推荐方法 LSTM-AROLS。

3.2.1 基于 BIRCH 的学习者聚类

聚类分析是把数据按照相似性归纳成若干类别，同一类中的数据彼此相似，不同类中的数据相异。其目标是在相似的基础上对收集来的数据分类。多项研究表明，学习风格相似的学习者具有相似的学习行为。因此，本研究采用聚类分析方法，以在线学习风格的八个维度为特征，将学习风格相似的学习者看为一个群组，对不同群组的学习者行为数据进行挖掘。

由于数据集中的数据适中，分布为凸，且维度较低，都为数值型数据，所以采用 BIRCH 聚类算法。BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) 是一种为大型数据集设计的混合聚类算法，它在初始阶段使用层次聚类算法，之后使用划分或其他算法，有效克服了层次法不可撤销的缺点。

BIRCH 算法通过聚类特征 (Clustering Feature, CF) 来表示一个簇。考虑包含 N 个学习者的集合 $X = \{x_1, x_2, \dots, x_N\}$ ，每个学习者 x_n 可由对应学习风格的 8 维向量 $(d_{n1}, d_{n2}, \dots, d_{n8})$ 表示。CF 是一个汇总了对象信息的三维向量：

$$CF = \langle n, LS, SS \rangle \quad (3.1)$$

其中 $LS = \sum_{i=1}^n x_i$ ，是 n 个学习者向量的线性和， $SS = \sum_{i=1}^n x_i^2$ 是 n 个学习者向量的平方和。聚类特征的本质是对簇的统计，通过 CF 可推导出簇的形心 x_0 、半径 R 和直径 D ：

$$x_0 = \frac{\sum_{i=1}^n x_i}{n} = \frac{LS}{n} \quad (3.2)$$

$$R = \sqrt{\frac{\sum_{i=1}^n (x_i - x_0)^2}{n^2}} = \sqrt{\frac{nSS - 2LS^2 + nLS}{n^2}} \quad (3.3)$$

$$D = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2}{n(n-1)}} = \sqrt{\frac{2nSS - 2LS^2}{n(n-1)}} \quad (3.4)$$

简化了对数据详细信息的存储，且具有可加性，即

$$CF_1 + CF_2 = \langle n_1 + n_2, LS_1 + LS_2, SS_1 + SS_2 \rangle \quad (3.5)$$

BIRCH 通过扫描数据集，建立一棵表示聚类层次结构的 CF 树（见图 7）。其中非叶结点存储了所有子节点 CF 的总和。构建 CF 树需要分支因子 (Branching factor) 和阙值两个参数，分支因子表示非叶结点最大子节点的数目，阙值表示叶结点中子簇的最大直径。CF 树在数据的插入中被动态的构造，如果插入新的数据后子簇的直径大于阙值，则进行分裂。同时 CF 树也支持在参数改变后进行重建。构建 CF 树后，可以针对形成的子簇应用不同的聚类算法得到最终的聚类结果。本研究采用凝聚的层次算法 (Agglomerative clustering)，它通过计算子簇形心的欧氏距离，将距离最近的子簇合并直至生成给定数目的簇。

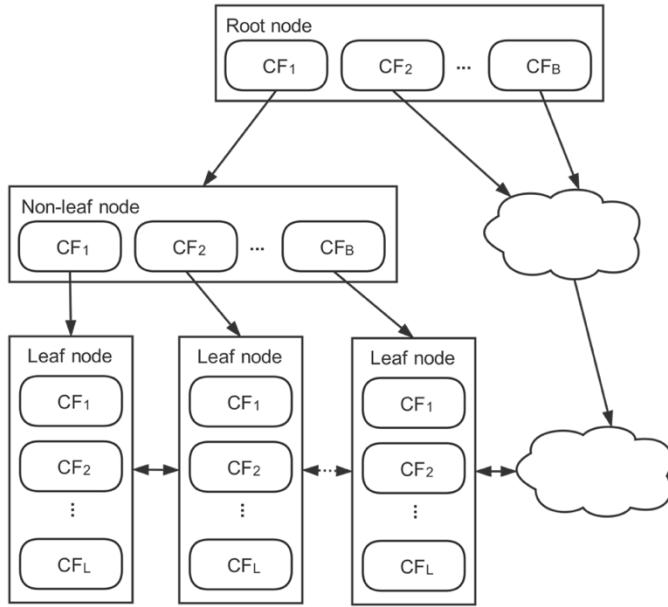


图 7 CF 树

BIRCH 算法具备很多优点。它将数据压缩，节约内存，同时只需要扫描一遍数据集，时间复杂度为 $O(n)$ ，CF 树的增删改操作也十分便捷。另外，BIRCH 算法受噪声的影响较小，因此不仅具有聚类功能，也可以常用于数据的预处理和简化。

3.2.2 基于协同过滤的自适应学习资源混合推荐方法（CF-AROLS）

基于学习风格对学习者进行分类后，为了挖掘相似学习风格学习者的信息，本研究对同一类别的学习者行为数据进行统计，主要分为学习资源相似性的计算和学习资源关联规则分析。

3.2.2.1 学习资源相似性矩阵的计算

相似性矩阵 (Similarity matrix) 也称为对象-对象结构，存放对象两两之间的相似度。对于 N 个学习者的集合 $X = \{x_1, x_2, \dots, x_N\}$ 以及 M 个物品的物品集合 $I = \{i_1, i_2, \dots, i_M\}$ ，可以用 $N \times M$ 的数据矩阵 (Data matrix) 表示学习者的学习资源浏览集合：

$$\begin{bmatrix} t(x_1, i_1) & \cdots & t(x_1, i_M) \\ \vdots & \ddots & \vdots \\ t(x_N, i_1) & \cdots & t(x_N, i_M) \end{bmatrix} \quad (3.6)$$

其中 $t(x_n, i_m)$ 表示学习者 x_n 浏览学习资源 i_m 的次数，如果未浏览该学习资源，则为 0。

通过数据矩阵，可以计算出 $M \times M$ 的学习资源相似性矩阵：

$$\begin{bmatrix} sim(i_1, i_1) & \cdots & sim(i_1, i_M) \\ \vdots & \ddots & \vdots \\ sim(i_M, i_1) & \cdots & sim(i_M, i_M) \end{bmatrix} \quad (3.7)$$

其中 $sim(i_n, i_m)$ 表示学习资源 i_n 和 i_m 的相似性，该值位于区间[0,1]，两个学习资源越相似该值越大，且有 $sim(i_m, i_m) = 1$ 。

关于相似度 $sim(i_n, i_m)$ 的计算方法有很多，应用较普遍的有皮尔森系数、闵可夫斯基距离的倒数等。由于上述算法主要针对有评分的记录（即用 $t(x_n, i_m)$ 表示学习者 x_n 对学习资源 i_m 的评分），不适用于本研究。我们使用余弦相似度（Cosine similarity）计算两个向量之间的夹角余弦值来评估相似度，夹角越小该值越接近1。

$$\cos(i_n, i_m) = \frac{|N(i_n) \cap N(i_m)|}{\sqrt{|N(i_n)||N(i_m)|}} \quad (3.8)$$

其中 $|N(i_n) \cap N(i_m)|$ 表示浏览过学习资源 i_n 和 i_m 的学习者数目， $|N(i_n)|$ 表示浏览过学习资源 i_n 的学习者的数目。

3.2.2.2 学习资源关联规则的挖掘

关联分析是一种在大规模数据集中寻找数据频繁模式（Frequent pattern）的方法。这些关系可以有多种形式，本研究主要挖掘频繁项集（Frequent item sets）和关联规则（Association rules）。频繁项集是经常一起出现的学习资源的集合，关联规则是两个学习资源之间可能存在的序列模式。

给定包含 M 个学习资源的集合 $I = \{i_1, i_2, \dots, i_M\}$ ，集合 I_j 表示集合 I 的一个子集，存在 $I_j \subseteq I$ 。对于 N 个学习者的集合 $X = \{x_1, x_2, \dots, x_N\}$ ，将每个学习者 x_n 浏览过的学习资源集合记为 I_{x_n} 。对于任意集合 I_j ，我们定义它的支持度为：

$$sup(I_j) = \frac{|N(I_j)|}{N} \quad (3.9)$$

其中 $|N(I_j)|$ 表示所浏览的集合中包含 I_j 的学习者数目。支持度表示关联规则或集合出现的频繁程度，给定一支持度阙值，支持度超过阙值的集合被称为频繁项集。

关联规则是形如 $I_{j1} \rightarrow I_{j2}$ 的表达式，表示学习者浏览集合 I_{j1} 的同时浏览集合 I_{j2} 的浏览模式。一条关联规则的支持度被定义为：

$$sup(I_{j1} \rightarrow I_{j2}) = sup(I_{j1} \cup I_{j2}) = \frac{|N(I_{j1} \cup I_{j2})|}{N} \quad (3.10)$$

其中 $|N(I_{j1} \cup I_{j2})|$ 表示同时浏览过学习资源集合 I_{j1} 和 I_{j2} 的用户数。

在寻找关联规则时，通常会计算规则的置信度（Confidence），置信度 $conf(I_{j1} \rightarrow I_{j2})$

表示当某个学习者浏览过集合 I_{j1} 时，也浏览集合 I_{j2} 的规则的确定性。如果某条关联规则的置信度和支持度都超过一定阈值，则认为该规则是强关联规则。利用支持度可以计算置信度，计算公式为：

$$conf(I_{j1} \rightarrow I_{j2}) = \frac{sup(I_{j1} \rightarrow I_{j2})}{sup(I_{j1})} \quad (3.11)$$

另外，由于支持度和置信度并不能充分表示关联规则的有效性，例如某规则 $I_{j1} \rightarrow I_{j2}$ 的置信度为 75%，而 $sup(I_{j2}) = 90\%$ ，即 I_{j2} 单独出现的概率还大于与 I_{j1} 同时出现的概率，则规则 $I_{j1} \rightarrow I_{j2}$ 是负相关的， I_{j1} 的出现实际上降低了 I_{j2} 出现的可能性。为了避免产生误导性的规则，本研究使用 Kulc 度量和不平衡比（Imbalance Ratio, IR）相结合来对规则 $I_{j1} \rightarrow I_{j2}$ 进行评估：

$$Kulc(I_{j1}, I_{j2}) = \frac{conf(I_{j1} \rightarrow I_{j2}) + conf(I_{j2} \rightarrow I_{j1})}{2} \quad (3.12)$$

Kulc 度量由波兰数学家 S.Kulczynski 提出，可以看作两个置信度的平均值，取值在 [0,1]，值越大表明关系越强。

$$IR(I_{j1}, I_{j2}) = \frac{|sup(I_{j1}) - sup(I_{j2})|}{sup(I_{j1}) + sup(I_{j2}) - sup(I_{j1} \cup I_{j2})} \quad (3.13)$$

IR 表示集合 I_{j1} 和 I_{j2} 的平衡程度， $IR(I_{j1}, I_{j2}) = 0$ 表示两者在数据集中的出现概率相同，该值越大表明集合的关系越不平衡，关联规则有效性越弱。

关联规则的挖掘分为两步，首先需要根据最小支持度阈值找出所有频繁项集，之后再由频繁项集产生满足最小置信度的关联规则。因为关联规则是由频繁项集的子集计算出的，故都是强关联规则。在大型数据集中，当支持度阈值设置较低时，常常会产生大量的频繁项集，故第一步的开销较大，决定了关联规则挖掘的总体性能。为减少计算量，本研究采用 FP-growth (Frequent-Pattern growth) 算法进行频集挖掘。该算法只对数据集进行两次扫描。该算法首先将代表频繁项集的数据压缩成一棵频繁模式树(FP Tree)，之后基于 FP 树为每个频繁项构造条件 FP 树，递归地发现频繁项。通过 FP-growth 挖掘频繁项集后，遍历数据集并计算频繁项集的支持度，最后根据频繁项集生成关联规则，并通过 Kulc 度量和 IR 比对关联规则进行筛选。

3.2.2.3 生成学习资源推荐列表

根据 3.2.1 将学习者分为 K 类后，类别集合用 $C = \{c_1, c_2, \dots, c_K\}$ 表示。我们利用 3.2.2.1 和 3.2.2.2 中提到的方法针对每一类别的学习者进行学习资源相似性计算和关联规则挖掘，分别用 $sim(i_a, i_b | c_k)$ 表示对于类别 c_k 的学习者来说物品 i_a 和物品 i_b 的相似度， $support(I_j | c_k)$ 和 $conf(I_{j1} \rightarrow I_{j2} | c_k)$ 分别表示对于类别 c_k 的学习者所挖掘到的学习资源集合的支持度和置信度。

根据基于物品的协同过滤 (Item-based collaborative filtering, Item-CF) 算法，我们可以根据相似度计算属于类别 c_k 的学习者 x_n 对学习资源 i_m 的兴趣：

$$P(x_n, i_m | c_k) = \sum_{i_k \in N(x_n) \cap S(i, N)} sim(i_m, i_k | c_k) \quad (3.14)$$

其中 $N(x_n)$ 是用户 x_n 浏览过的物品的集合， $S(i, N)$ 是和物品 i 最相似的 N 个物品的集合。可见，与学习者历史上浏览的学习资源越相似的学习资源，越有可能在推荐列表中获得比较高的排名。

接下来，我们利用关联规则对 Item-CF 所生成的推荐列表进行过滤。定义类别 c_k 中的学习者 x_n 对学习资源 i_m 的浏览趋势 (Tendency) 为：

$$T(x_n, i_m | c_k) = P(x_n, i_m | c_k) \times \sum_{I_j \in F(x_n)} conf(I_j \rightarrow \{i_m\} | c_k) \quad (3.15)$$

其中 $F(x_n)$ 是由学习者 x_n 的历史浏览记录所包含的频繁项集集合。

CF-AROLS 结合了协同过滤和关联规则的混合推荐算法通过生成可变长度的推荐列表，着重为学生提供当下学习资源的选择意见，而不是提供固定数目个选项。我们设定一浏览趋势的动态阈值，根据不同的学习者和其学习风格而变化，计算公式为：

$$T_{th}(x_n, L | c_k) = \mu \times \frac{1}{N} \sum_{i \in R_{CF}(x_n, L)} T(x_n, i_m | c_k) \quad (3.16)$$

其中 $R_{CF}(x_n)$ 表示 ItemCF 方法所为用户 x_n 生成的 TopN 推荐列表， $R_{AR}(x_n)$ 表示基于关联分析所为用户 x_n 生成的推荐列表。 μ 是一个可调节的参数，用来调节推荐列表的长度，实现更加个性化的推荐。CF-AROLS 方法具体流程参见附录 2。

3.2.3 基于 LSTM 的自适应学习资源推荐方法 (LSTM-AROLS)

协同过滤和关联规则虽然可以对学习资源进行推荐，但本质上只是对学习资源的共现概率进行统计。而学习则是一个过程，不同学习资源的浏览之间存在因果和顺序关系，

因此我们使用了深度学习中循环神经网络的方法对学习者行为进行了更深入的研究。LSTM-AROLS 主要通过训练 LSTM 模型对学习资源浏览序列进行预测，该方法分为两步，分别是分布式学习资源向量的构建和 LSTM 神经网络模型的训练。

3.2.3.1 分布式学习资源向量的构建

在通常情况下，自适应学习系统的学习资源数目是可变的，因此本研究参考 Mikolov 提出的词嵌入（Word embedding）方法 Skip-gram，该模型在众多研究中表现突出，可以将学习资源表示为固定维度的向量作为神经网络模型的输入。该方法不仅可以对模型的输入维度进行压缩，还可以表示学习资源之间的关系，使得相似的学习资源向量之间距离更小。

Skip-gram 模型将每个输入学习资源映射到一个对数线性投影层中，并预测该学习资源在浏览序列上下文中出现的概率。因此，另 $w(t)$ 表示 t 时刻所浏览的学习资源，Skip-gram 模型的最优化目标函数可表示为：

$$\sum_{-c \leq k \leq c, k \neq 0} \log p(w(t+k)|w(t)) \quad (3.17)$$

其中 c 表示浏览序列上下文窗口大小， $p(w(t+k)|w(t))$ 通过 softmax 函数计算。Skip-gram 模型包含三层，如图 8 所示，其中输入层是一个学习资源，它被影射到一个投影层，再输出层输出对当前学习上下文学习资源的预测。

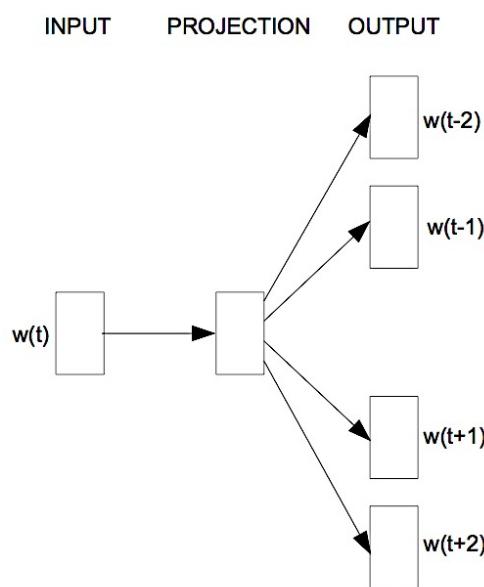


图 8 Skip-gram 模型

根据设置的维数和窗口大小，我们可以使用用户的学习资源浏览序列对 Skip-gram 模型进行训练，得到学习资源上下文模型和学习资源的向量矩阵。

3.2.3.2 栈式 LSTM 神经网络模型的训练

LSTM 模型是循环神经网络（recurrent neural network, RNN）变体的一种，RNN 模型由于在网络结构设计中考虑到了时序的概念，在时序数据分析中表现出了很强的适应性。图 9 展示了一般的 RNN 网络和神经元结构，其中 x_t 表示 t 时刻的输入单元， y_t 为 t 时刻的输出单元。输出序列可以通过如下公式计算：

$$h_t = f(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (3.18)$$

$$y_t = W_{hy}h_t + b_y \quad (3.19)$$

其中 h_t 表示 t 时刻的 RNN 隐层单元， W 表示层与层之间的权重系数矩阵， b 为偏置向量， f 为激活函数。

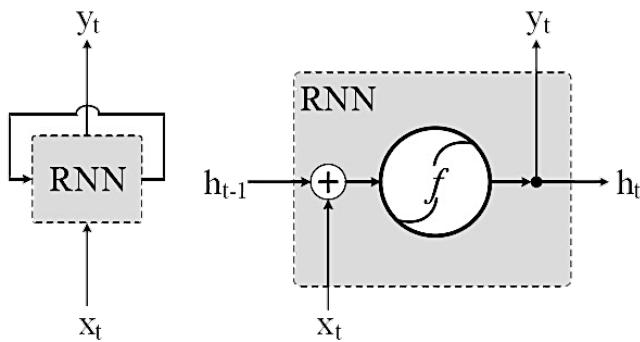


图 9 RNN 网络（左）和单元结构（右）

LSTM 模型对 RNN 进行了改进，具有长期记忆能力，弥补了 RNN 模型中梯度消失、梯度爆炸和长短期记忆能力不足等问题，在语言建模、语音识别和机器翻译领域应用广泛。LSTM 模型等单元结构^[64]如图 10 所示。在前向传播时，其计算方法为：

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (3.20)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (3.21)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3.22)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (3.23)$$

$$h_t = o_t \tanh(c_t) \quad (3.24)$$

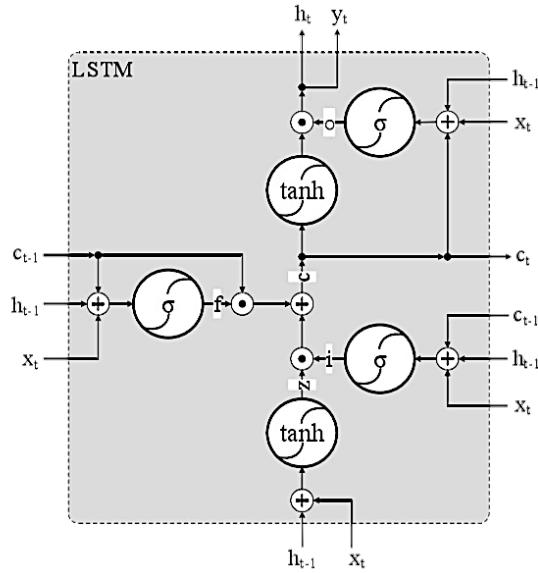


图 10 LSTM 的单元结构

上述公式中的 i, f, c, o 分别表示输入门、遗忘门、细胞状态和输出门， σ 为 sigmoid 激活函数， \tanh 为双曲正切激活函数。LSTM 的训练可以采用随时间的反向传播算法（back propagation through time, BPTT），首先根据上述前向传播公式计算输出值，之后按时间和网络层级两个传播方向反向计算 LSTM 单元的误差值，最后根据相应误差项计算权重的梯度并更新权重。

栈式 LSTM 神经网络可以看作为许多 LSTM 隐藏层堆叠在一起的结构，更深层的 LSTM 隐层可以对序列数据进行更深层的抽象，在研究上取得了很好的效果。本研究构建的栈式 LSTM 神经网络如图 11 所示。该框架包括输入层、Embedding 层、三个 LSTM 隐藏层和输出层。输入层的单元数目与学习资源向量的维度相同，接受学习资源序列，并由 Embedding 层将其转换为学习资源向量序列，之后，三个 LSTM 对序列数据进行特征提取和抽象，输出层的单元数目与输入层相等，提供下一个学习资源向量的预测。

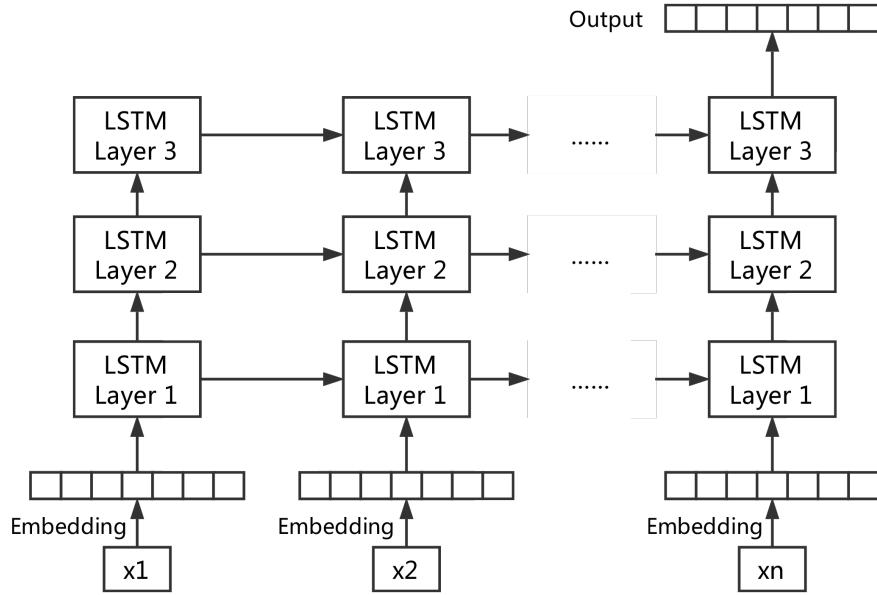


图 11 基于 LSTM 的学习资源路径预测模型框架

给定包含 M 个学习资源的集合 $I = \{i_1, i_2, \dots, i_M\}$, 对于 N 个学习者的集合 $X = \{x_1, x_2, \dots, x_N\}$, 将每个学习者 x_n 浏览过的学习资源序列记为 I_{x_n} 。给定任意窗口大小 k , 为生成训练数据, 我们可以根据 I_{x_n} 生成 $|I_{x_n}| - k$ 个长度为 $k + 1$ 的学习资源浏览序列 $\{i_{1+p}, i_{2+p}, \dots, i_{k+p}, i_{k+p+1}\}$, 其中前 k 个学习资源 $\{i_{1+p}, i_{2+p}, \dots, i_{k+p}\}$ 向量作为输入, 最后一个学习资源 i_{k+p+1} 的向量为标签。训练学习资源路径预测模型采用余弦距离平均值的相反数 (Cosine proximity) 作为损失函数, 对于输出向量 $\mathbf{o} = (o_1, o_2, \dots, o_D)$, 假设目标值为 $\mathbf{t} = (t_1, t_2, \dots, t_D)$, 它们的余弦距离可表示为:

$$\cos(\mathbf{o}, \mathbf{t}) = \frac{\sum_{d=1}^D (o_d \times t_d)}{\sqrt{\sum_{d=1}^D o_d^2} \times \sqrt{\sum_{d=1}^D t_d^2}} \quad (3.25)$$

对于 N 个训练数据, 损失函数为:

$$\mathcal{L} = -\frac{1}{N} \sum_{n=1}^N \cos(\mathbf{o}_n, \mathbf{t}_n) \quad (3.26)$$

为了最小化损失函数, 本研究采用基于梯度的随机优化算法 Adam (Adaptive moment estimation) 进行隐藏层权重的更新, 该算法结合了 AdaGrad 和 RMSProp 方法的优势, 根据不同参数改变学习学习率, 占用较少的存储资源, 在实际应用中表现优异。

对于不同学习风格类型的学习者, 利用上文中基于栈式 LSTM 的模型对该类学习者的学习资源浏览序列数据进行学习, 挖掘该类学习者的行为模式。之后, 对于新的学习者浏览序列输入, 将输出的向量以余弦相似度为衡量标准找到最相似的学习资源, 为该

学习者推荐下一个待浏览的学习资源。如果需要进行 TopN 推荐，则将预测结果放到输入序列的末尾进行再次推荐，直到推荐 N 个学习资源。LSTM-AROLS 方法的详细步骤参见附录 3。

3.3 实验设计与结果分析

3.3.1 实验数据集

对于 AROLS 算法的评估，本研究继续使用 2.4.3.2 中提到的 OULAD 数据集，重点在于对学习者日志中信息的挖掘。对于训练集和测试集的拆分策略，我们将学习者行为的前 80%作为训练集，后 20%作为测试集。这使我们可以根据学习者之前的行为记录对之后的行为进行预测。同时为了防止推荐算法中参数的过拟合，将测试集中的后 10%数据作为验证集，协助进行算法及参数的调整。

利用前文隐式诊断学习者学习风格特征的方法，对测试集中的学习者进行学习风格诊断，该数据包含学习者在动机、交流、视觉、言语、感悟、直觉、序列、综合这八个学习风格特征上的程度，均为数值型数据。我们使用此诊断结果和学习者行为日志作为 AROLS 模型的输入。

3.3.2 评价指标

1) 聚类趋势评估

在基于学习风格对学习者数据进行聚类分析前，需要对数据集的聚类趋势进行评估，即确定其中是否存在可以形成有意义的聚类的非随机结构。霍普金斯统计量（Hopkins Statistic）是一种检验变量分布的空间随机性的统计量。给定数据集D，霍普金斯统计量将它作为随机变量o的一个样本，考察o不同于数据空间中均匀分布的程度。其计算结果在 0 到 1 之间，0.5 表示数据集D是均匀分布的，不大可能有统计显著的簇。如果结果在 0.7 到 1 之间，则说明数据集D中具有有意义的非随机结构。

2) 聚类结果评价指标

对于部分聚类算法来说，最重要也是最难的部分的就是聚类数目K的选取，尤其是处理未标记的数据集时。本研究采用三种方式评定聚类结果的质量和稳定性（Stability），并辅助决定聚类数目。

Silhouette Coefficient (SC)是衡量聚类结果质量的一种指标，它的取值在-1 到 1 之间，

越接近 1 表明聚类效果越好。SC 系数的计算需要 a 和 b 两个参数，其中 a 表示一个学习者和其他同类型学习者距离的均值，b 表示一个学习者和最近的某个其他类内所有学习者距离的均值。某个学习者 x_n 的 SC 系数可以通过如下公式计算：

$$SC(x_n) = \frac{b - a}{\max(a, b)} \quad (3.27)$$

另一个可以用来评估非监督学习聚类效果的是 Calinski-Harabaz (CH) index，其取值没有一定范围，越大表明类内数据越密集，类间被分开得越好。对于 K 个类，CH 系数被定义为类间离散度与类内离散度的比值，可通过如下公式计算：

$$CH(K) = \frac{\text{Tr}(B_K)}{\text{Tr}(W_K)} \times \frac{N - K}{K - 1} \quad (3.28)$$

其中， B_K 表示类间离散度矩阵， W_K 表示类内离散度矩阵，他们如下公式定义：

$$W_K = \sum_{k=1}^K \sum_{x \in c_k} (x - \mu_k)(x - \mu_k)^T \quad (3.29)$$

$$B_K = \sum_k |c_k| (\mu_k - \mu)(\mu_k - \mu)^T \quad (3.30)$$

其中 $|c_k|$ 表示第 k 类中学习者的数量， μ 表示学习者集合 X 的质心。

结合 SC 系数和 CH 系数，我们可以在使 SC 系数和 CH 系数的评估结果尽可能优的情况下最大化聚类数目，因为对学习者的分类越详细，在之后所挖掘到每类学习者的信息就越具针对性。

聚类的稳定性同样可以帮助进行聚类数目的选择。应用稳定性的前提是所有数据都由同一隐含模型产生，即数据本身存在着某种分布，聚类的目标就是找到数据背后的分布模式。聚类结果越接近该模式，稳定性越高。在本研究中，给定某算法 A 和数据集 S ，对于聚类数目 k ，稳定性的计算分为以下几步：

- a) 根据原始数据集 S 产生 n 各不同的数据集 S_1, S_2, \dots, S_n 。
- b) 对数据集 S_1, S_2, \dots, S_n 分别应用算法 A 聚为 k 类，得到聚类结果 C_1, C_2, \dots, C_n
- c) 对于 $b, b' = 1, \dots, n$ ，计算不同聚类结果之间的最小匹配距离 $d(C_b, C_{b'})$ ：

$$d(C_b, C_{b'}) = \min_{\pi} \frac{1}{m} \sum_{i=1}^m 1_{\{C(X_i) \neq \pi(C'(X_i))\}} \quad (3.31)$$

其中 π 表示对 k 个标签的置换， m 表示 C_b 和 $C_{b'}$ 的公共数据点总数， $C(X_i)$ 表示对点 X_i 的标签。

d) 计算稳定性 $Stab(A, S, k)$:

$$Stab(A, S, k) = 1 - \frac{1}{n^2} \sum_{b, b'=1}^n d(C_b, C_{b'}) \quad (3.32)$$

稳定性表示了对于不同却有交集的数据集进行聚类，标签基本不变的点的数目占比。在对聚类的评估中，我们优先选取稳定性高的聚类数目。然而稳定性只是一种参考，显然类的数目越少稳定性越高，只分 1 类的稳定行为 100%。

3) 推荐算法评估指标

在本研究中，我们选取常用的准确率（Precision）、召回率（Recall）以及 F1 分数（F1-score）作为推荐算法的评价指标。准确率表示做出的正确推荐在推荐列表中所占比例，计算公式为：

$$Precision = \frac{\sum_{x \in X} |R(x) \cap T(x)|}{\sum_{x \in X} |R(x)|} \quad (3.33)$$

其中 X 表示测试集的学习者集合，对于学习者 $x \in X$ ， $R(x)$ 表示推荐算法生成的推荐列表， $T(x)$ 表示测试集中学习者真实浏览的学习资源。

召回率代表做出的正确推荐在学习者真实浏览记录中的比例，计算公式为：

$$Recall = \frac{\sum_{x \in X} |R(x) \cap T(x)|}{\sum_{x \in X} |T(x)|} \quad (3.34)$$

准确率和召回率的计算方式十分简单，但确实衡量众多推荐算法的重要指标。不过这两项评价指标经常会产生冲突，例如增加推荐列表的数量，可能会增加召回率，但同时也会降低准确率。为了寻找两者之间的均衡，我们可以使用综合指标 F1 分数，计算公式为：

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3.35)$$

F1 分数的取值在 0 到 1 之间，将准确率与召回率等权重混合在一起进行推荐效果的评估。

3.3.3 数据预处理

数据预处理主要分为异常值处理和标准化。本研究使用局部异常因子算法（Local Outlier Factor）去除了 1% 的数据点，该方法通过衡量每个数据点与周围数据点的相对密度，决定该数据点的异常程度。图 12 用深色标注出了异常点。

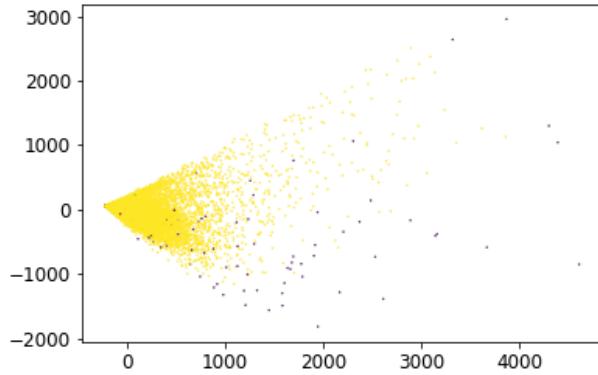


图 12 异常值检测

由于实际上 OULAD 数据集中课程页面可供点击点链接数目不同，以及某课程各个页面的 `activity_type` 设计不均衡，导致前文统计的学生在线学习风格数据在各个特征的值数量级不同。因此本研究采用 0-1 标准化方法，对原始数据的每个特征进行线性变换，使结果落到 [0, 1] 区间。

另外，考虑到学习风格的统计指标为学习者点击量，该值在实际应用中会不断增长，因此我们对每个学习者按照 L2 范数进行归一化，即对于 x 的学习风格向量 (d_1, d_2, \dots, d_8) 进行变换使得：

$$\sqrt{d_1^2 + d_2^2 + \dots + d_8^2} = 1 \quad (3.36)$$

经过异常值去除以及变换的数据如图 13 所示。其霍普金斯统计量为 0.90，具有很强的聚类趋势。同时也可通过观察看出，该数据分布不均匀，有多处的点聚集现象。此现象从侧面证明了基于在线学习风格模型的特征可以将不同特质的学习者进行区分。

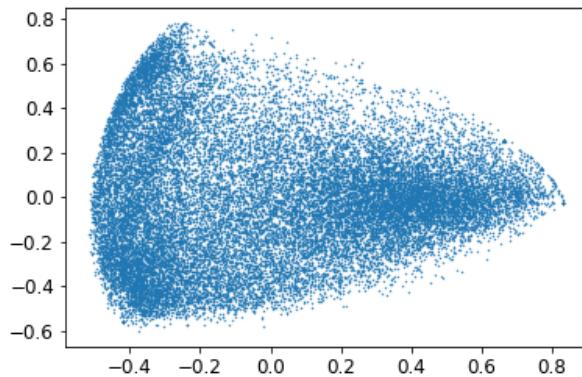


图 13 预处理后的学习者学习风格数据

3.3.4 学习者聚类数目的研究

BIRCH 聚类算法有三个参数：分支因子、阙值及聚类数目。其中分支因子和阙值决

定了 CF 树的结构，分枝因子越大，非叶结点能包含的子节点数目越多，因而 CF 树生成的子类越少。同理，阙值越大，非叶结点的直径越大，CF 树生成的子类越少。而聚类数目决定了最终的数目，该值必须大于 CF 树生成的子类数目。为获得最好的聚类效果，我们对分支因子区间[50, 1900]，阙值区间[0.1, 0.7]以及聚类数目区间[3, 10]的 BIRCH 聚类效果进行了评估，并对每个聚类数目选取了稳定性最高的参数配置，结果如表 8 所示。可见，分支因子的设置影响较小，因此对 CF 树的生成起主要作用的是阙值，且阙值在 0.4 到 0.5 左右时聚类结果最为稳定。考虑到分支因子对 CF 树的限制过于固定，如果学习者数量过于庞大，分支因子的限制会造成 CF 树高度的增加，因此将分支因子设置为 500。相反，阙值的限制更关注数据的分布，如果相当数量的点十分接近，半径较小，阙值的设置会使得这些学习者属于同一类，而分支因子会限制学习者的个数，可能会将十分相近的学习者分到其他类别。

表 8 BIRCH 聚类结果评估

分支因子	阙值	CF 树生成的子类数目	聚类数目	稳定性
50-1900	0.4	13	3	75.29%
50-1900	0.5	6	4	73.88%
50-1900	0.4	13	5	72.09%
50-1900	0.5	6	6	76.31%
50-1900	0.4	13	7	68.66%
50-1900	0.4	13	8	72.91%
50-1900	0.4	13	9	71.78%
50-1900	0.4	13	10	70.72%

针对初步调参的结果，我们针对阙值进行更进一步的调参，并计算结果的 SC 和 CH 系数，结果如表 9 所示。

表 9 针对阙值调参的 BIRCH 聚类结果评估

分支因子	阙值	CF 树生成的子类数目	聚类数目	CH 系数	SC 系数	稳定性
500	0.44	9	3	4212	0.22	79.48%
500	0.44	9	4	4848	0.22	78.87%
500	0.44	9	5	5189	0.24	78.78%
500	0.44	9	6	5740	0.26	78.65%
500	0.44	9	7	5556	0.26	78.51%
500	0.44	9	8	5697	0.27	78.39%
500	0.44	9	9	5733	0.28	78.12%
500	0.42	10	10	5265	0.28	76.60%

可知，当阙值为 0.44 时 BIRCH 对于学习者学习风格的聚类最为稳定，均在 78% 左右。但由于最终聚类数目的不同，结果的评估也存在一定差别，由稳定性的定义可知，聚类数目越少稳定性越可能高，因此对于聚类数目较大的情况，较难得到高的稳定性。由于本研究聚类的目的是使之后的学习资源数据挖掘更针对学习风格相似的学习者，且降低计算复杂度，如果聚类数目过低则失去了聚类的意义，而在学习者数量有限的情况下聚类数目也不宜过高。通过分析如上结果，发现当聚类数目为 6 时，CH 系数和 SC 系数都较高。虽然聚类数目为 9 时的结果也相当好，但 CF 树生成的子类数目也为 9，相当于 CF 树生成后并没有聚类，之后如果学习者数目增加导致子类数目增加后可能有较大的变化，较不稳定。因此最终选择将学习者依据学习风格聚为 6 类，应用 BIRCH 后的聚类效果如图 14 所示。

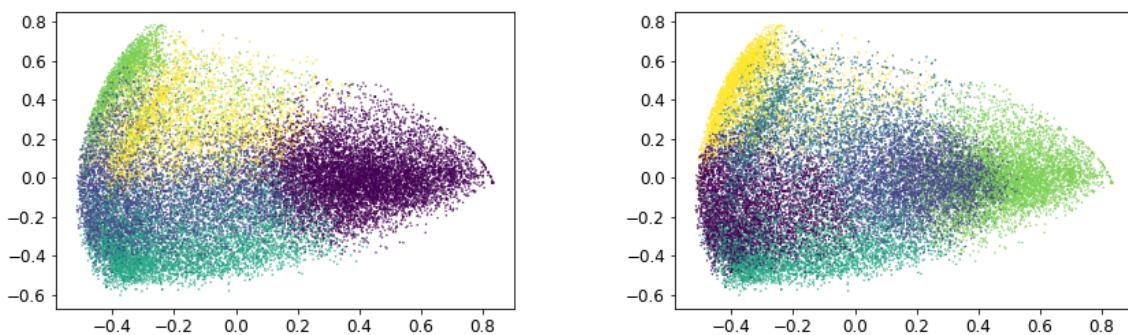


图 14 BIRCH（左）和 K-means（右）的聚类结果对比

为了更好地检验 BIRCH 的聚类效果，我们同时应用 K-means 对数据进行聚类。K-means 算法是最常用的分区聚类算法之一，主要基于欧氏距离对类进行划分，被广泛应用在各个领域。利用 K-means 将学习者根据学习风格聚为 6 类，如图 14 所示，聚类结果的 CH 系数为 6633, SC 系数为 0.22, 稳定性为 66.48%。虽然 CH 系数略高于 BIRCH，但聚类稳定性较低。从图 14 中可以看出，由于 K-means 是基于距离的聚类，因此聚类结果较均匀。但是，通过对不同类中数目的对比（见表 10）和数据点特征的查看，我们发现 K-means 只是简单的通过均值将数据均匀聚类，而 BIRCH 则更注重学习者特征取值。例如 BIRCH 所聚的第三类中只有 169 个学习者，通过查看数据发现该类学习者处理后的学习风格特征都为 0，而 K-means 则将这些行为数据较少的学习者和其他有行为的学习者归到了一起，不注重数据的特点而是均值。因此 BIRCH 的聚类结果显然更为合理。

表 10 Birch 和 K-means 的不同类中学习者数目的对比

	1	2	3	4	5	6
K-MEANS	5164	5570	3176	4058	5032	2813
BIRCH	8841	6399	169	5021	2339	3044

3.3.5 CF-AROLS 超参数配置的研究

CF-AROLS 中存在 5 个超参数，主要用于关联规则的挖掘，分别是频繁项集支持度阙值 minSup、频繁项集可信度阙值 minConf、Kulc 度量阙值 minKulc 和不平衡比阙值 minIR，同时还有浏览趋势阙值 minTend。我们分别对这些参数在验证集上进行实验，对协同过滤结果的 Top10 进行筛选，以研究它们对 CF-AROLS 表现的影响。

各类学习者群组在支持度阙值区间[0.12, 0.30]的频繁项集挖掘结果如图 15 所示。其中横坐标表示最小支持度阙值，纵坐标表示频繁项集数目。从图中可以看出，第 1 类学习者中的频繁项集绝大部分支持度偏小，当阙值为 0.13 时频繁项集数目大幅下降。而第 2 类和第 4 类学习者集合中的项集支持度都偏小。第 5 类和第 6 类的学习者中的频繁项集支持度最为理想，随着支持度阙值的升高而规律地下降。因此，为减少关联规则数目，提高计算效率，我们最终选择对学习者分别设置 0.13、0.11、0、0.13、0.17、0.21 作为频繁项集支持度阙值。

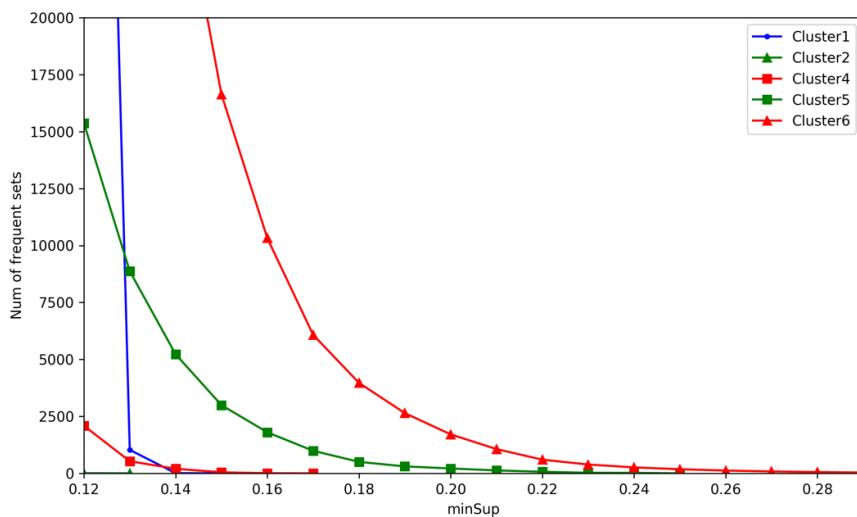


图 15 支持度阙值对频繁项集个数的影响

确定了支持度阙值后，我们对置信度阙值和 Kulc 阈值在区间[0.80, 0.98]进行了 Top10 对推荐实验，实验过程中将另一阙值设为 0，不平衡比阙值设为 1，浏览趋势阙值为 0.5。结果分别见图 16、图 17 所示。

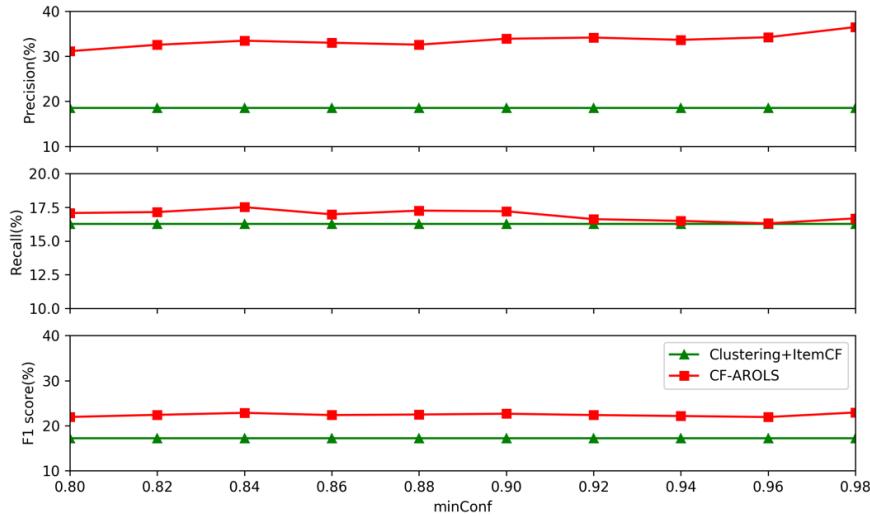


图 16 置信度阙值对推荐结果的影响

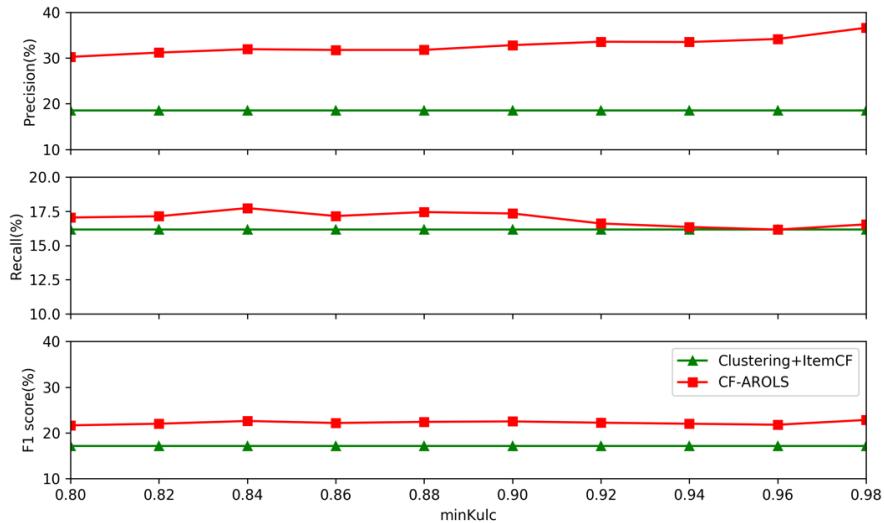


图 17 Kulc 度量阙值对推荐结果的影响

其中横坐标是 minConf 或者 minKulc 的取值，纵坐标分别是协同过滤和 CF-AROLS 方法的准确率、召回率和 F1 分数平均评估结果。从实验结果看出，置信度阙值和 Kulc 阈值对准确率的影响基本相同，因为置信度只是单向规则的条件概率，而 Kulc 度量则对规则的双向进行了衡量。同时可以看到，当 minConf 或 minKulc 增大时，CF-AROLS 的准确率逐渐上升，但并不稳定，原因是阙值较小时可用于过滤的关联规则多，而浏览趋势阙值较低，过滤力度不强。而较高的阙值增加了浏览趋势，但缺少大量的关联规则，因此过滤结果也有可能下降。对于召回率， minKulc 的影响比 minConf 更大一些，可见 minKulc 的对于关联规则的筛选较强于 minConf 。两者的 CF-AROLS 的召回率基本变化不大，可见 CF-AROLS 对于过滤掉不合适的学习资源很有效果。从 F1 分数的结果可以

看出，CF-AROLS 的推荐效果相比传统方法有一定的提升，且提升效果非常稳定。根据实验结果，最终选择 0.84 作为置信度阈值和 Kulc 阈值。

对于 maxIR 超参数的研究结果如图 18 所示。其中横坐标为 maxIR 阈值，纵坐标为该阙值设置所筛选的关联规则数目。实验中将 minConf 和 minKulc 都设置为 0.8，浏览趋势阙值为 0.5。从结果看出，数据集中的关联规则都较平衡，不规则比阙值的变动影响较小。因此为了不过多地减少关联规则数目，我们在最终测试时将 maxIR 的值设置为 0.15。

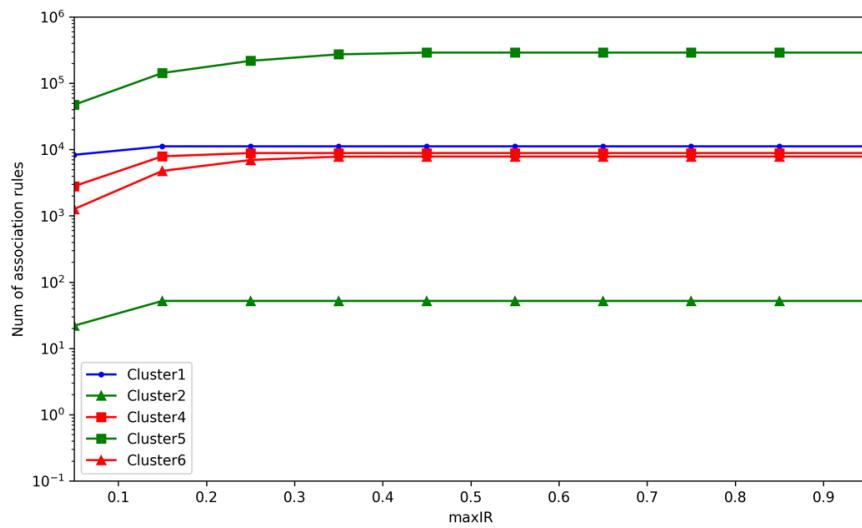


图 18 不平衡比对关联规则数目影响

最后，我们对浏览趋势阙值对推荐结果的影响进行研究。其他超参数的配置采用上文中的最优配置结论，对浏览趋势阙值 minTend 在 [0, 1] 的区间上进行了评估，实验结果如图 19 所示。

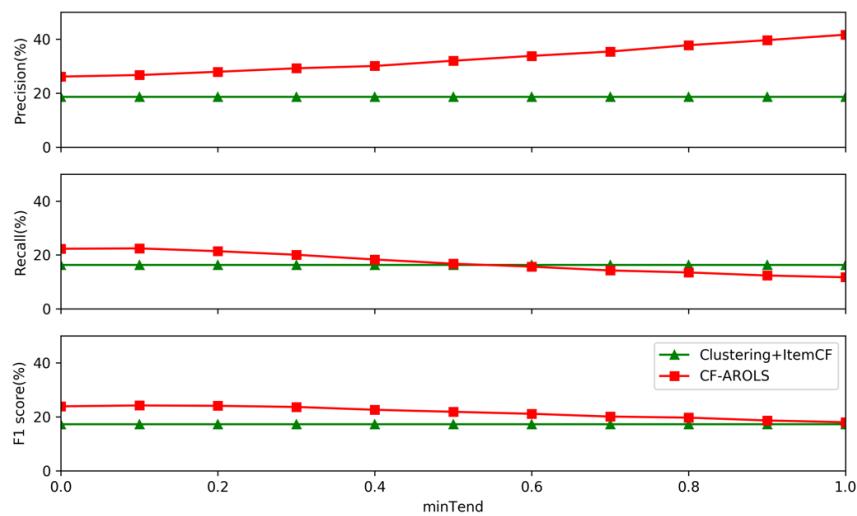


图 19 浏览趋势阙值对推荐结果的影响

从图中的结果可以看出，阙值越高，CF-AROLS 的推荐准确率越高，但同时也会缩短推荐列表的长度，导致召回率下降，使得 F1 分数同时下降。然而，在 CF-AROLS 的应用情境中，准确率的提升较为重要。因为召回率和测试集有很大的关系，很大程度上取决于测试数据的多少。所以我们在之后的实验中另 minTend 为 0.6，既提高了一定程度的准确率，又保证了召回率不会下降过多。

3.3.6 LSTM-AROLS 超参数配置的研究

学习资源的分布式向量可以表现学习资源之间的联系，该向量维度的不同会导致学习资源关系的表示能力不同，从而影响预测结果。理论上来讲，学习资源的数量越多，向量的维度应该越高。本研究对 6365 个学习资源的浏览数据分别训练了 128, 256, 400, 512 维学习资源向量，训练时上下文窗口大小为 5，并通过含有一个隐层的 LSTM 网络在验证集上进行训练和验证，序列窗口大小为 5，对于每一类学习者的训练数据和验证数据个数如表 11 所示。

表 11 各类学习者训练数据和验证数据序列个数

	1	2	3	4	5	6
TRAIN	3211196	1350191	0	1259660	792361	758013
VALIDATION	313572	121710	0	117122	76744	70517

对于不同学习者群组的平均验证结果如图 20 所示。其中横坐标表示训练次数，纵坐标表示 Cosine Proximity。学习者类别在图的上方显示。

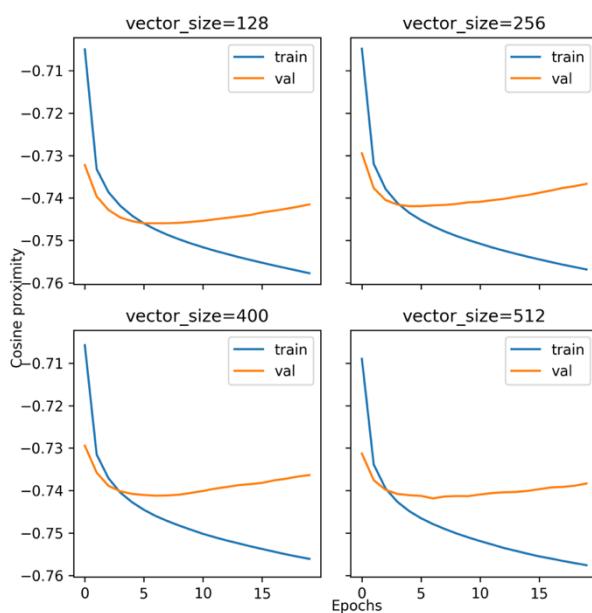


图 20 不同的学习资源向量长度对实验结果的影响

从实验结果看出，当向量长度为 128 时在验证集上的测试效果是最好的，余弦平均值相反数的最小值低于其他向量长度的最小值。并且，当向量长度过大时，测试结果出现了不稳定的情况。因此我们使用 128 维的词向量进行后续的实验。

同时，我们分别对 LSTM 隐层的层数 1、2、3 和隐层维度大小 64、128、256、512、1024 进行了研究。当 LSTM 层的隐层维度增加时，模型更容易过拟合。因此我们在之后对于 LSTM 隐含层数目的研究中设置 Dropout 为 0.2。模型参数配置及 Top10 推荐评估结果如表 12 所示。

表 12 LSTM 模型参数配置对实验结果的影响

LSTM 层数	模型参数配置	准确率 (%)	召回率 (%)	F1 分数 (%)
1	128-64-128	47.25	9.12	15.29
1	128-128-128	50.51	9.77	16.37
1	128-256-128	50.65	9.23	15.65
1	128-512-128	57.79	10.95	18.44
2	128-128-64-128	44.52	8.53	14.32
2	128-128-128-128	45.97	8.66	14.59
2	128-128-256-128	56.03	10.31	17.45
2	128-128-512-128	59.15	10.90	18.45
2	128-256-128-128	51.80	9.98	16.75
2	128-256-256-128	56.10	9.93	16.93
2	128-256-512-128	59.87	11.28	19.00
2	128-512-128-128	55.37	10.85	18.14
2	128-512-256-128	60.88	11.30	19.10
2	128-512-512-128	59.18	10.55	17.96
3	128-128-128-128-128	52.42	10.20	17.08
3	128-256-256-256-128	59.01	10.98	18.55
3	128-512-512-512-128	62.68	11.49	19.47
3	128-512-256-128-128	55.47	10.58	17.78
3	128-512-256-256-128	57.59	11.25	18.82
3	128-512-256-512-128	64.12	12.47	20.87
3	128-128-512-256-128	56.93	10.77	18.13
3	128-256-512-256-128	59.91	11.39	19.15
3	128-512-512-256-128	59.60	11.33	19.06
3	128-512-128-512-128	61.54	11.84	19.87
3	128-1024-1024-1024-128	63.06	11.90	20.04

从实验结果可以看出，推荐效果受栈式 LSTM 层数和隐层维度的共同影响。当隐层维度较小时，增加模型的层数可能会降低推荐效果，例如配置为 128-128-128 模型的 F1

分数为 16.37%，而 128-128-128-128 模型为 14.59%，但 128-128-128-128-128 的结果为 17.08%，可能是 20% 的 Dropout 使得两层 LSTM 欠拟合，但可以降低三层 LSTM 的过拟合。另外，模型层数的增加在普遍上可以提升推荐效果，可能是因为多层 LSTM 提取了历史浏览序列更加抽象的特征。当神经网络配置为 128-512-256-512-128 时在验证集上的效果最佳，因此我们选取该网络配置作为最终的 LSTM-AROLS 网络模型。

3.3.7 实验结果与分析

我们分别使用 3.3.5 和 3.3.6 中的推荐的组合为 CF-AROLS 和 LSTM-AROLS 的进行超参数的配置。为了验证 CF-AROLS 的有效性，我们设计三个实验在测试集上进行对比。采用传统的基于物品的协同过滤 ItemCF 作为基准（baseline），针对所有的学习者行为记录进行挖掘，计算学习资源相似度并进行推荐。第二个实验将 Item-CF 应用于基于学习风格进行聚类后的学习者行为数据上，针对同一类的学习者进行学习资源相似度挖掘和推荐。第三个实验使用本研究提出的 CF-AROLS 推荐方法，通过关联规则对协同过滤结果进行筛选。我们选用准确率、召回率和 F1 分数对结果进行对比，如图 21 所示。

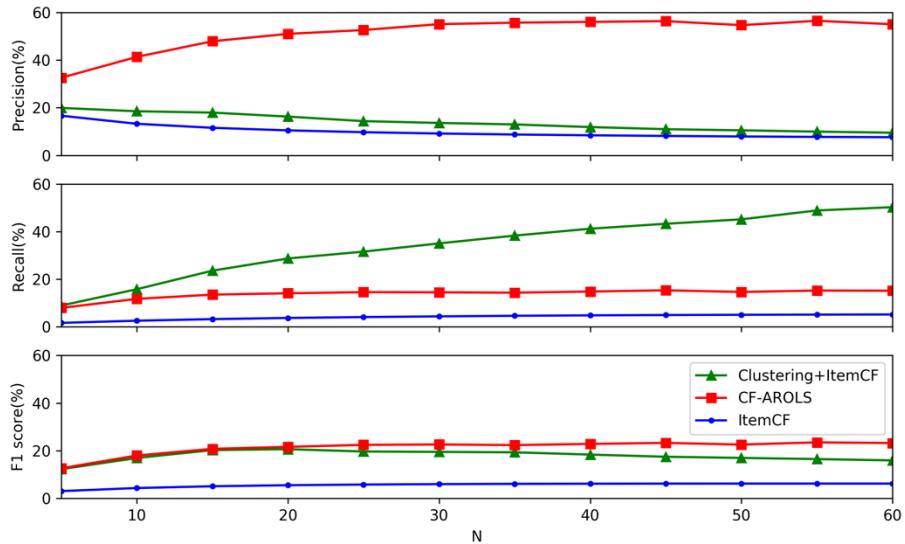


图 21 CF-AROLS 的对比实验结果

横坐标为协同过滤的推荐结果个数 N ，纵坐标为各个测量值的百分数结果。从图中可看出，在各项指标上，CF-AROLS 和基于学习风格聚类后的协同过滤推荐结果（Clustering+ItemCF）都优于传统的 ItemCF 推荐结果。另外，通过关联规则的过滤，CF-AROLS 在准确率上的表现大大优于其他两种方法，对比 Clustering+ItemCF 方法有

高达 50% 的提升，且总体表现（即 F1 分数）稳定。相反，Clustering+ItemCF 方法虽然在召回率上表现最好，但随着 N 的增大，准确率大幅下降。可见 CF-AROLS 在总体上得到了很好的推荐效果，且推荐个数越多，关联规则对结果的过滤效果越好。

CF-AROLS 能够获得较好的推荐结果主要有三点原因：首先，基于学习风格的学习者聚类将拥有相似个人特质和行为偏好的学习者聚集到一起，使得学习资源关系的挖掘更有针对性。其次，学习资源相似度计算和关联规则分析两种方法对学习资源的关系进行了有效的挖掘，选择出了大量共现概率高的学习资源。最后，通过将不同形式的学习资源关系进行结合，过滤了出现几率低的学习资源，并且动态地改变推荐列表长度，使得推荐准确率大幅提高。

为了验证 LSTM-AROLS 的有效性，我们设计两个实验进行对比。第一个实验使用所有的学习者数据训练一个 LSTM 模型，对测试集中的学习者进行学习资源推荐。第二个实验使用本研究提出的 LSTM-AROLS，在对学习者基于学习风格进行聚类分析（3.3.1）后针对不同类别的学习者训练多个 LSTM 模型。平均推荐结果如所示。

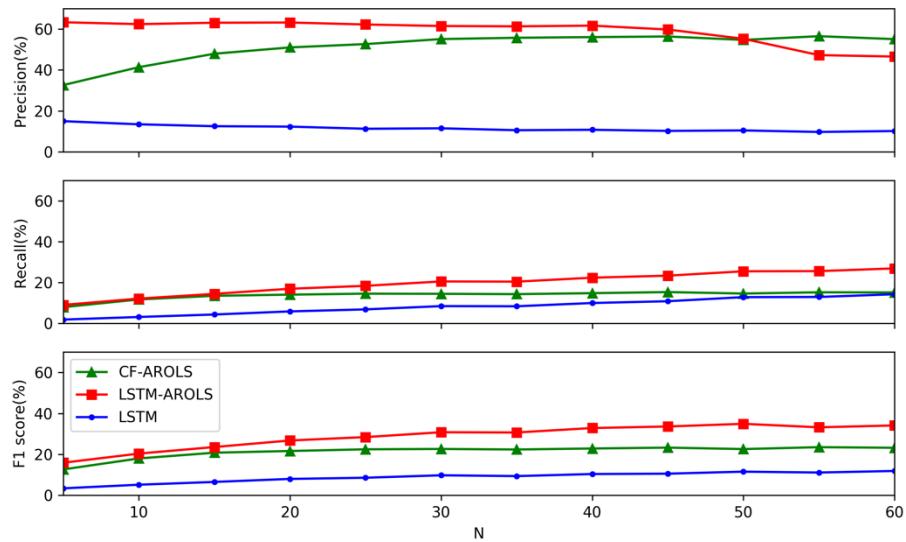


图 22 LSTM-AROLS 的对比实验结果

从结果可以看出，LSTM-AROLS 的表现大大优于单纯的 LSTM 网络的推荐结果，且在 N 值较小时比 CF-AROLS 的推荐结果更准确。因为最先开始的输入序列来自学习者的真实浏览记录，所以预测结果较准确，而之后逐渐偏差增大，导致准确率下降。但因为总体推荐数目的提升，召回率逐渐提高，这两者的综合使得 F1 分数比较稳定，且优于 CF-AROLS 的综合表现。

因为 CF-AROLS 的推荐考虑的主要还是学习资源的共现概率，较热门的学习资源，如

课程主页、论坛中热门帖子等的出现概率更大，所以在真实的应用中 CF-AROLS 可能无法达到和实验结果一样的效果。而 LSTM-AROLS 则对学习者的学习资源浏览序列进行建模，所得出的推荐更加符合真实情况和准确。

LSTM-AROLS 的突出表现首先归功于基于学习风格的学习者聚类，该步骤使得相似的浏览序列更加聚集，有助于 LSTM 模型的构建。其次基于 Skip-gram 的学习资源向量构建，不仅缩小了输入输出维度，还将学习资源上下文关系整合进模型中。最后是栈式 LSTM 循环网络模型，该模型有强大的序列建模功能，对学习资源浏览序列进行了逐步深层的抽象，使得推荐准确率大大提升。

3.4 本章小结

针对当前众多自适应学习系统的普遍问题，及通过学习风格与教导模型的紧密融合，实现真正以学习者个人特质为中心的自适应学习系统，本章分别介绍了两种可以实现自适应推荐机制的推荐方法：基于协同过滤的学习资源推荐 CF-AROLS 和基于长短期记忆循环神经网络的学习资源推荐 LSTM-AROLS。CF-AROLS 首先基于学习风格对学习者进行聚类分析，随后针对同一类学习者，挖掘他们对学习行为数据并计算学习资源相似度矩阵，查找符合条件的关联规则，最后利用关联规则对协同过滤结果进行过滤，筛选出有较高浏览趋势的学习资源列表。CF-AROLS 虽然有一定的推荐效果，但其根本思想是根据学习资源共现概率进行推荐，因此我们使用 LSTM 模型再度对聚类后的学习者浏览序列进行学习，训练出了符合不同学习风格的循环神经网络模型 LSTM-AROLS。随后，我们使用准确率、召回率以及 F1 分数作为评价指标，以传统的基于物品的协同过滤作为基准，对 CF-AROLS 和 LSTM-AROLS 进行了对比实验。实验结果表明，LSTM-AROLS 因考虑到了学习资源的序列关系，推荐结果显著高于 CF-AROLS。

第四章 基于学习风格诊断的自适应推荐原型系统的实现

基于前文中对学习风格隐式诊断和学习资源推荐方法 AROLS 的研究成果，我们设计并实现了基于学习风格诊断的自适应推荐原型系统。该系统主要对学习者模型、学习资源模型和教导模型进行了初步实现，并将前文中的研究成果进行了实施。下文将对系统的设计阶段和实现阶段进行详细介绍。

4.1 需求分析

4.1.1 业务需求分析

基于学习风格诊断的学习资源自适应推荐系统所提供的功能主要包含以下三个方面：

1) 用户管理功能

该系统中有三种基本用户角色：学习者、教师和管理员。用户管理功能可以对学习者的基本信息进行管理，如学习者注册和记录学习风格特征信息。同时，该功能可以对用户进行一定的权限控制，允许教师浏览部分学生信息，允许管理员浏览其他用户的信息。

2) 学习资源管理功能

包括拥有权限的教师及管理员对学习资源的创建、修改和发布，为学习者提供丰富的学习资源并提供针对课程或学习资源的讨论交流平台。同时允许管理员对学习资源进行学习资源属性标记，辅助系统对学习者学习风格进行诊断。

3) 在线学习功能

为在线学习者提供一个环境，使得学习者可以通过自主检索快速定位学习资源，收藏课程并进行学习。系统实时记录学习者的行为日志，并在后台进行分析计算，更新学习者的学习风格数据、学习资源的相似性和关联规则数据并提供个性化推荐。

系统的重点在于自适应教学功能的实现。对于学习者而言，系统的流程如图 23 所示：

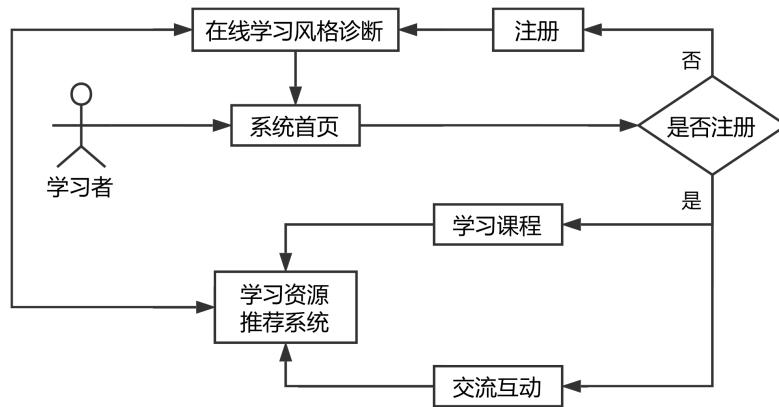


图 23 AROLS 自适应学习系统流程图

初次使用系统的学习者需要进行注册和选择性地填写在线学习风格量表，以初步诊断其在线学习风格。提交完相关信息后，可以查看编辑自己的信息和学习风格，并注册感兴趣的课程进行学习。学习者在学习过程中与学习资源的交互信息都会记录到学习资源推荐系统中，动态更新学习者的学习风格，并经过计算后根据其学习风格，在学习过程中对学习者进行自适应推荐。

4.1.2 系统需求分析

针对上述功能的实现，系统在性能方面应首先具备适应性，即保证提供给用户的推荐具有准确性和实时性，适应学习者的进度。其次，系统的可维护性也十分重要，这要求系统有一个好的框架模式，在面临未知故障时方便修复。对于学习者、教师等用户而言，系统应该具有有好的交互方式，在保证界面流畅的同时为用户提高使用体验。最后，系统还应具备灵活性，在学习资源和用户数量上升的同时可以灵活地进行扩展。

4.2 系统设计

本研究中，系统各个方面设计主要遵循“高内聚，低耦合”的设计原则，采用面向接口的编程思想，对系统结构和功能进行模块化，使得模块内部元素紧密结合，并且降低对其他模块的依赖，提高系统的流畅性和连贯性，并且利于后期的更改和维护。

4.2.1 系统结构及功能模块设计

在结构上，系统的基本结构主要分为三层（见图 24），用户通过客户层中的浏览器向系统后台发出 http 请求，接收到请求后，控制层与用户建立会话并解析请求，传递给业务逻辑层。业务逻辑层为系统的核心，包含了各个功能模块的主要业务逻辑。为回应

用户请求，业务逻辑层的相关功能模块调用基础服务层中的数据服务，通过数据库访问接口与数据库进行数据交互，获取数据后进行处理并回应用户请求。

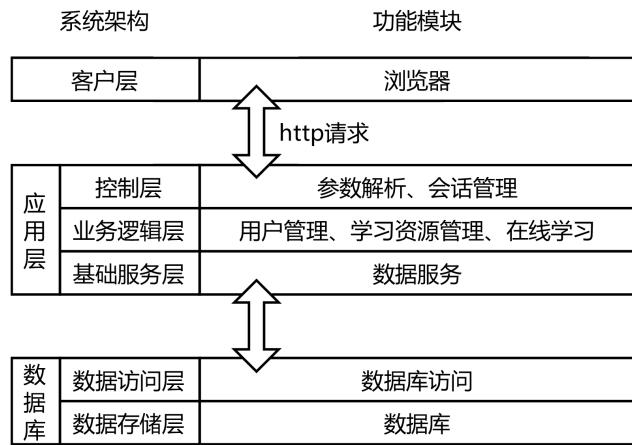


图 24 AROLS 自适应学习原型系统结构

在功能上，系统根据业务需求主要分为用户管理模块，学习资源管理模块和在线学习模块。每个模块根据不同的功能由若干子模块组成，如图 25 所示。

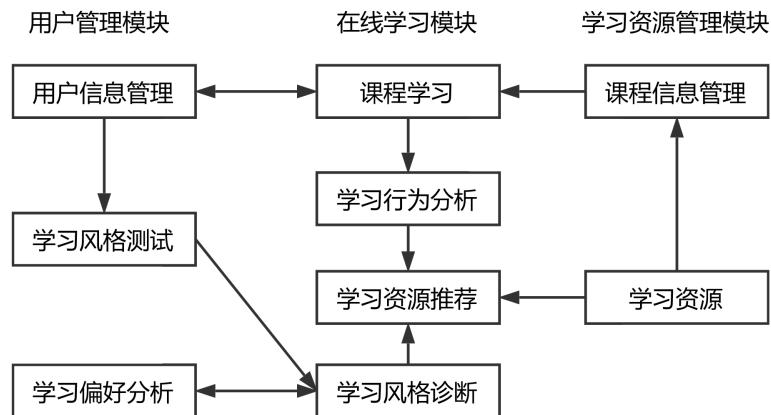


图 25 AROLS 自适应学习系统模块

用户管理模块主要负责管理用户信息，进行学习风格显式测试并生成学习者学习偏好分析报告。用户信息包括学习者或教师的用户名、邮箱、密码、权限等基本信息，该信息保存在数据库中，可进行更新和维护。学习风格测试采用 OLS 量表，用于检测初次使用系统的学习者学习风格。在学习者学习过程中，系统会收集并记录其他学习者行为数据，并提供给学习偏好子模块进行学习偏好分析。

学习资源管理模块主要由具有教师权限的用户进行操作，分为对课程基本信息的管理和学习资源管理。课程基本信息包括课程名称、负责人、概述、学生列表等信息，可以由创建人进行更新和维护。学习资源以网页为单位，该信息中包含各网页的学习资源

属性、生成时间、关键词等信息。

在线学习模块是系统最核心的模块，包含课程注册、课程学习、学习行为分析、学习风格诊断和学习资源推荐子模块。学习者首先需要进行课程的注册，之后才能进行课程的学习，系统会记录学习者对学习资源的浏览行为并进行隐式的学习风格诊断，更新学习者学习风格信息，根据更新后的学习风格和最新的浏览记录进行学习资源推荐。

4.2.2 数据库设计

数据是系统的核心，合理的数据库设计决定了系统后期维护的效率。在确定系统架构和功能模块的划分后，我们根据系统中将要存储和查询的数据进行数据库的设计。基于学习风格诊断的自适应学习系统主要包含用户、学习资源、在线学习行为的相关信息，我们根据这些信息设计了系统的数据表（见图 26）。

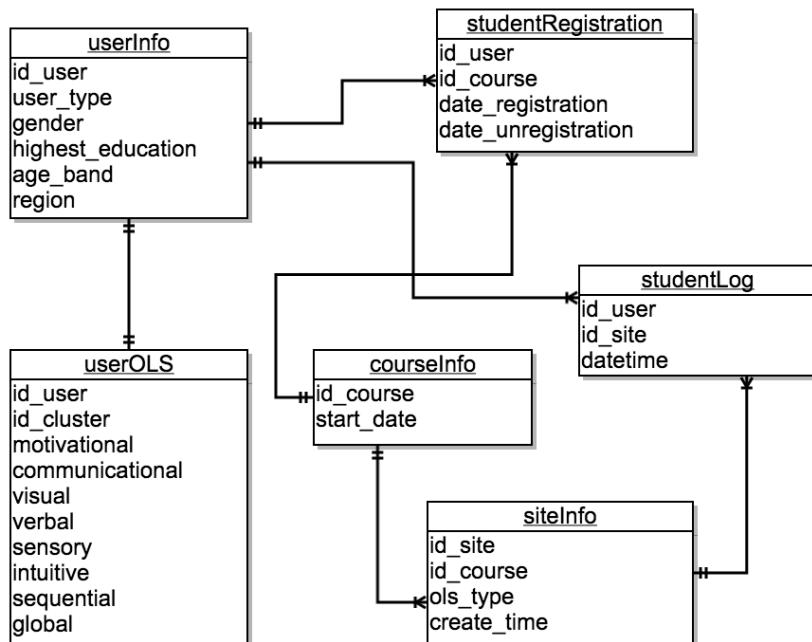


图 26 AROLS 自适应学习系统数据库表

用户基本信息和在线学习风格信息分别存储在 userInfo 和 userOLS 表中，课程信息和学习资源信息分别存储在 courseInfo 和 siteInfo 表中， studentRegistration 表和 studentLog 表分别记录了学习者对课程的注册和学习资源浏览日志。对于各表中字段具体的数据类型及说明见表 13。

表 13 AROLS 自适应学习系统数据类型及说明

字段名称	数据类型	说明
id_user	INTEGER	用户唯一标识符
user_type	VARCHAR	用户类型（教师、学习者、管理员等）
gender	VARCHAR	性别
highest_education	VARCHAR	最高学历
age_band	VARCHAR	年龄段
region	VARCHAR	地区
id_cluster	INTEGER	学习风格群组类别
motivational	DECIMAL	动机
communicational	DECIMAL	交流
visual	DECIMAL	视觉
verbal	DECIMAL	言语
sensory	DECIMAL	感悟
intuitive	DECIMAL	直觉
sequential	DECIMAL	序列
global	DECIMAL	综合
id_course	INTEGER	课程唯一标识符
start_date	DATETIME	课程开始日期
id_site	INTEGER	学习资源唯一标识符
ols_type	VARCHAR	学习资源属性
create_time	DATETIME	学习资源创建时间
date_registration	DATETIME	学习者课程注册时间
date_unregistration	DATETIME	学习者课程取消注册时间
datetime	DATETIME	学习资源点击时间

4.3 系统开发与实现

4.3.1 开发环境及关键技术

基于学习风格诊断的自适应学习系统需要实现前端和后端两个部分，前端系统是浏览器端的网页交互界面，后端系统实现学习风格隐式诊断、学习资源推荐等核心功能。对于前端，我们使用基于 Javascript 的 Vue.js 框架，对于后端，本原型系统采用 Python 作为开发语言，Django 作为 Web 框架。

1) Vue.js 框架

Vue.js 是一个易用、灵活且高效的渐进式 Javascript 框架，区别于其他框架，Vue.js 的设计为自底向上逐层应用，它的核心只关注视图层，容易上手和与第三方库整合。Vue.js 的设计基于 MVVM 设计模式，支持数据双向绑定和单文件组建，可以快速构建

界面元素。

2) Django 框架

Django 是 Python 写成的 Web 框架，采用了 MVC 模式，自带强大的数据库操作接口。同时，它提供自助后台管理，在不用编写的情况下即可获得完整的后台管理界面。对于 url 模块的设计，Django 采用正则表达式，使得网址的表达更为优雅。

4.3.2 功能模块展示

系统的名称为 AROLS，主页见图 27。打开网站后，可进行登陆和注册。



图 27 AROLS 自适应学习系统主页

学习者登陆后，可进入“公开课”栏目对课程进行学习。在课程页面中可选择自己感兴趣的课程并收藏（见图 28）。在已注册课程页面可取消对课程的收藏。也可点击授课教师或授课机构进行查看和收藏。

图 28 AROLS 自适应学习系统课程页面

在个人中心页面可查看当前登陆学习者的学习风格，系统会根据学习者的浏览数据或在线学习风格调查问卷的结果为学习者在线学习风格的各个维度打分（见图 29）。



图 29 AROLS 自适应学习系统个人资料页面

在学习者的学习过程中，AROLS 会根据前文中提到的 CF-AROLS 或 LSTM-AROLS 方法对学习者学习路径进行推荐。在本研究中，由于缺乏基于此系统的测试数据，只采用 OULAD 数据集提供的网页 id 和前文训练好的 LSTM-AROLS 进行推荐。图 30 为系统在测试时成功为当前学习者的推荐结果。

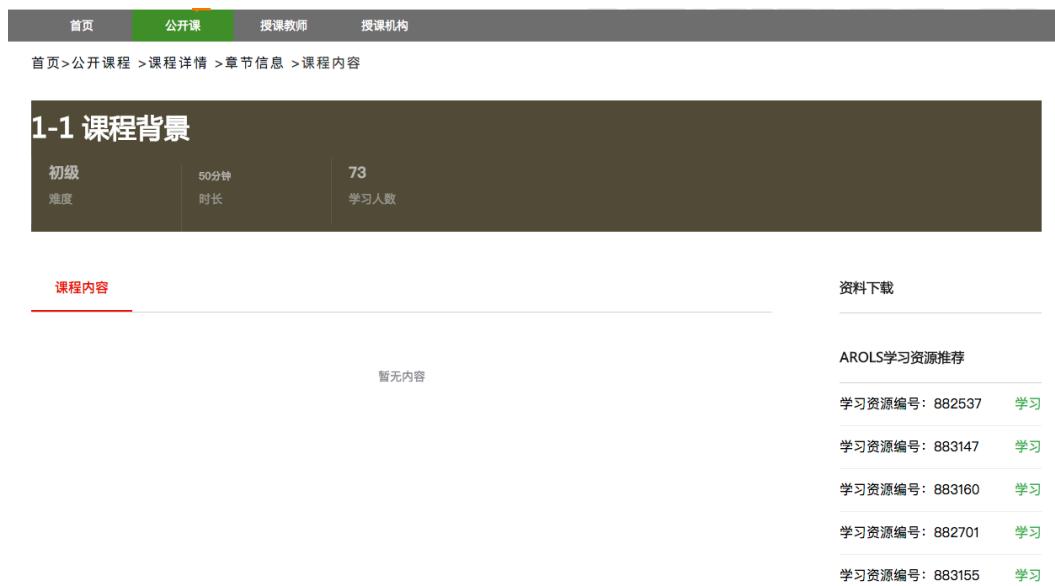


图 30 AROS 自适应学习系统测试推荐结果

4.4 本章小结

本章的主要内容为自适应学习系统的设计与开发。在设计初期，首先进行了需求分析，确定了系统的三个核心功能，即用户信息管理、学习资源信息管理和在线学习。接着，我们对系统的整体架构进行了设计，采用三层结构，分别为用户层、应用层和数据库，层与层之间通过接口进行数据交互，尽量降低模块之间的耦合性。同时，我们对三大核心模块进行了细分，对子模块进行了设计，使得子模块各司其职，实现“高内聚”的设计理念。最后，根据各个子模块涉及到的数据内容与结构，我们对底层的数据库表进行了设计。完成设计阶段后，本研究使用基于 Javascript 的 Vue.js 框架和基于 Python 的 Django 框架对系统进行了实现，展示了核心功能页面并对学习风格诊断功能和学习资源推荐功能进行了测试，结果显示系统可将前文中研究的 CF-AROLS 和 LSTM-AROLS 方法成功地应用于本章中搭建的系统。

总结与展望

本研究基于学习风格模型及前沿技术，完成了自适应学习系统中学习者模型和教导模型的研究和设计，并根据研究成果实现了基于学习风格诊断的自适应学习系统的原型。本研究的主要工作分为以下几个方面：

1) 学习者在线学习风格模型的构建

本研究对传统学习风格模型进行了大量调研，并在 Dunn and Dunn、Entwistle、Felder-Silverman、Myers-Briggs 记忆 Kolb 模型的基础上构建了在线学习风格模型 OLS，描述在线学习者的个人特质并对不同的学习者进行区分。在构建学习风格模型后，设计了对应的在线学习风格量表，用以显式地确定学习者在线学习风格。我们用该问卷对 380 人进行了调研，结果显示 OLS 模型中的学习风格特征可以对学习者进行显著的区分。之后我们通过调研结果对学习风格量表题目进行了分析和筛选，对参与调查问卷的学习者进行了聚类分析，验证在线学习风格模型对于学习者的分类效果。同时，本研究提出了基于规则的在线学习风格隐式诊断方法，通过分析学习资源的属性并赋予相关权重，根据学习者浏览记录自动诊断学习者的学习风格。最后，我们对开源数据集 OULAD 中的学习者学习风格进行了初步诊断和分析，进一步验证了在线学习风格模型在各个维度上对学习者的区分能力。

2) 学习资源推荐方法的研究

基于本研究构建的在线学习风格模型，我们提出了自适应学习资源推荐方法 AROLS。AROLS 首先对学习者进行聚类分析，将相似学习风格的学习者分为同一类别，之后针对每一类学习者的学习资源浏览记录进行挖掘，进行学习资源推荐。我们分别对两种方法进行了研究，分别为传统的协同过滤方法 CF-AROLS 和循环神经网络方法 LSTM-AROLS。CF-AROLS 对学习资源的相似度和关联规则进行计算，并利用关联规则对协同过滤的推荐结果进行筛选，提高了推荐的准确率。LSTM-AROLS 利用循环神经网络模型，对学习资源浏览序列进行建模，相比于 CF-AROLS 更多地考虑到了学习资源浏览的先后顺序，贴合真实情况，推荐效果有显著提升。同时，通过 AROLS 对比与传统推荐方法以及模型预测方法的优异表现，可以看出在线学习风格模型的引入大大增强了自适应学习系统的个性化推荐功能。

3) 基于学习风格诊断的自适应学习系统原型的实现

在对学习风格模型和学习资源推荐方法进行了充分的研究及验证后，我们分别使用基于 Javascript 的 Vue.js 框架和基于 Python 的 Django 框架对系统的前端和后端进行了实现。原型系统中主要包括学习者信息管理、学习资源信息管理和在线学习三个核心功能模块，分别对应自适应学习系统中的学习者模型、内容模型和教导模型。为了实现上述三部分功能，我们对系统架构、功能子模块以及数据库表进行了详细的设计，并使系统遵循“高内聚，低耦合”的设计理念，为以后更完善的系统实现提供了基础。

对比前人的研究，本研究主要有以下创新和贡献：

1) 在线学习者与传统学习者所处的环境差异巨大，因此经典的学习风格模型已不足以对在线学习者建模。本文通过对比在线学习特点与传统学习特点，构建了在线学习风格模型，对在线学习者进行了更为全面和细致的描述。同时设计了在线学习风格量表和隐式的在线学习风格诊断方法，简化了该模型的应用。

2) 本研究将学习者模型和教导模型紧密融合，对学习者聚类分析后对数据进行挖掘，并对不同的推荐方法进行了研究，其中 CF-AROLS 利用关联规则挖掘方法实现了推荐列表的动态变化，使得推荐结果更注重学习资源推荐的准确率。另外，本研究首次将深度学习方法应用到学习资源推荐中，基于 LSTM 模型生成了符合特定学习风格类别的学习者的学习资源浏览序列模型，并根据该模型进行学习资源推荐，为深度学习在自适应学习系统中的应用打下基础。

3) 基于学习风格诊断的自适应学习原型系统在帮助学习者学习的同时关注学习者的行为信息，动态更新其学习风格和系统后台的推荐数据，并为学习者提供学习风格分析。相比于当今各种在线学习平台，是完全以学习者特质为核心的个性化学习系统，使自适应学习系统的研究更进一步。

本文针对学习者模型和教导模型进行了深入的研究，并对自适应学习系统的原型进行了实现。但随着研究的深入，我们在研究中不断进行反思，也发现了很多不足之处。未来将在以下几方面进行更深入的探索：

1) 本研究提出的隐式学习风格诊断方法需要用到学习资源的信息，并为不同类型的学习资源分配权重，这两方面仍需要人工辅助，无法单纯依靠学习者的在线学习行为进行学习风格诊断。因此如何更好并隐式地获取学习者学习风格仍有待改进。

- 2) AROLS 学习资源推荐方法依赖于学习者的浏览记录，因此没能很好地解决“冷启动”问题，即针对新进入系统的学习者无法提供学习资源推荐。在以后的研究中，应扩充自适应系统的功能，比如加入学习者目标设置，根据学习者目标和学习风格对学习者进行学习资源推荐。
- 3) 本研究只初步实现了系统原型，对系统的很多方面缺乏考虑。比如如何更快地根据学习者行为信息更新其学习风格，如何在加入新的学习资源后对推荐算法进行更新等。

致谢

首先，感谢中法工程师学院为我提供了双学位的机会，忙碌的研究生生涯使我不仅在学业上和生活上都收获颇丰。在研究生第一年，我有幸在法国里尔中央理工大学交换，在体验法国文化的同时在纯法语环境下学习了众多课程，尤其是计算机方面的课程。另外，我有幸参与到学生项目中，负责应用程序的开发，在技术上有了很大成长，认识了很多有趣的朋友。研究生第二年，我回到北航进行计算机学硕的攻读，并同时进行了为期六个月的企业实习，收获颇丰。感谢学院和学校为我提供的平台，让我得以在自己感兴趣的专业方向上发展，并完成硕士学位。

其次，感谢我的导师殷传涛，在学业上一直对我进行悉心的指导，帮助我解决毕设上的困惑和研究方向选择。不仅如此，在事业上，殷老师也会给我提供有价值的建议，让我得以努力准备面试，获得理想的工作机会。

感谢我的同学丁瑞雪，利用她在自然语言处理方面上丰富的经验为我的学习资源推荐方法提供思路和建议，并在她百忙的实习和学习中抽空阅读我的代码，找出我方法中的问题，对我困惑的知识点进行解答。

感谢我的学弟张笑颜，对我的小论文进行审阅并提出修改建议，同时在熟悉我毕设工作的过程中不断与我讨论，使我进行更深入的思考和对方法的改进。

感谢我的众多朋友和同学，没有他们我无法快速完成学习资源量表数据的采集和分析。感谢他们花费宝贵的时间帮助我完成毕设。

需要感谢的人数之不尽，没有家人、老师和照顾我的朋友，我无法成为今天的我。随着毕业设计接近尾声，我的学生生涯也即将结束，对身边的老师、朋友和同学有太多的感恩和不舍。希望北航越来越好，成为中国乃至世界上理工科学校的佼佼者，希望中法工程师学院延续中国和法国教学的优点，培养出越来越多的优秀毕业生，也希望我的同学们都找到满意的工作，事业顺利。

参考文献

- [1] Hwang G J. Definition, framework and research issues of smart learning environments - a context-aware ubiquitous learning perspective [J]. Smart Learning Environments, 2014, 1(1):1-14.
- [2] Murray T. Authoring intelligent tutoring systems: An analysis of the state of the art[J]. International Journal of Artificial Intelligence in Education (IJAIED), 1999, 10: 98-129.
- [3] Oxman S, Wong W, Innovations D V X. White paper: Adaptive learning systems[J]. Integrated Education Solutions, 2014.
- [4] Brusilovsky P. Methods and techniques of adaptive hypermedia[M]. Adaptive hypertext and hypermedia. Springer Netherlands, 1998: 1-43.
- [5] Bursilovsky P, Eklund J, Schwarz E. Web-based education for all: a tool for development adaptive courseware[J]. Computer Networks and ISDN Systems, 1998, 30(1-7): 291-300.
- [6] Brusilovsky P, Schwarz E, Weber G. ELM-ART: An intelligent tutoring system on World Wide Web[C]. Intelligent tutoring systems. Springer Berlin/Heidelberg, 1996: 261-269.
- [7] Brusilovsky P, Farzan R, Ahn J. Layered evaluation of adaptive search[C]. Workshop on Evaluating Exploratory Search Systems at SIGIR06. 2006: 11-13.
- [8] Romero C, Ventura S, De Bra P. Knowledge discovery with genetic programming for providing feedback to courseware authors[J]. User Modeling and User-Adapted Interaction, 2004, 14(5): 425-464.
- [9] Wolf C. iWeaver: towards' learning style'-based e-learning in computer science education[C]. Proceedings of the fifth Australasian conference on Computing education-Volume 20. Australian Computer Society, Inc., 2003: 273-279.
- [10] 余胜泉. 适应性学习—远程教育发展的趋势[J]. 开放教育研究, 2000, 3: 14.
- [11] 陈品德, 李克东. 适应性教育超媒体系统—模型, 方法与技术[J]. 现代教育技术, 2002, 12(1): 11-17.
- [12] 丛春瑜, 刘家勋, 于滨. 一个基于 Web 的自适应学习系统[J]. 微型电脑应用, 2004, 20(9): 37-38.

- [13] 柏宏权. 适应性教学系统中个性化教学策略研究[D]. 南京: 南京师范大学, 2006.
- [14] 姜强. 自适应学习系统支持模型与实现机制研究[D]. 长春: 东北师范大学, 2012.
- [15] 崔晓阳. 基于 Felder-Silverman 学习风格模型的在线导学系统设计与开发[D]. 山东师范大学, 2014.
- [16] 刘小丹, 胡小红. 自适应学习系统中基于情绪感知的学生模型设计[J]. 中国教育信息化: 高教职教, 2015 (10): 85-88.
- [17] 王丽萍, 赵蔚, 魏久鸿, 等. 开放性社会学习者模型的研究进展视图[J]. 现代远程教育研究, 2015, 2: 012.
- [18] Student modelling: the key to individualized knowledge-based instruction[M]. Springer Science & Business Media, 2013.
- [19] Brusilovsky P, Millán E. User models for adaptive hypermedia and adaptive educational systems[M]. The adaptive web. Springer Berlin Heidelberg, 2007: 3-53.
- [20] Nakic J, Granic A, Glavinic V. Anatomy of student models in adaptive learning systems: A systematic literature review of individual differences from 2001 to 2013[J]. Journal of educational computing research, 2015, 51(4): 459-489.
- [21] 陈仕品, 张剑平. 适应性学习支持系统的学生成绩研究[J]. 中国电化教育, 2010 (5): 112-117.
- [22] Ortigosa A, Paredes P, Rodriguez P. AH-questionnaire: An adaptive hierarchical questionnaire for learning styles[J]. Computers & Education, 2010, 54(4): 999-1005.
- [23] Kelly D, Tangney B. Adapting to intelligence profile in an adaptive educational system[J]. Interacting with computers, 2006, 18(3): 385-409.
- [24] Cha H J, Kim Y S, Lee J H, et al. An adaptive learning system with learning style diagnosis based on interface behaviors[C]. Workshop Proceedings of International Conference on E-Learning and Games, Hangzhou, China. 2006: 513-524.
- [25] Milosevic D, Brkovic M, Debevc M, et al. Adaptive learning by using scos metadata[J]. Interdisciplinary Journal of E-Learning and Learning Objects, 2007, 3(1): 163-174.
- [26] 姜强, 赵蔚. 自适应学习系统述评及其优化机制研究[J]. 现代远距离教育, 2011 (6): 57-63.

- [27] Chedrawy Z, Abidi S S R. An adaptive personalized recommendation strategy featuring context sensitive content adaptation[C]. International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems. Springer Berlin Heidelberg, 2006: 61-70.
- [28] Cristea A, Calvi L. The three layers of adaptation granularity[C]. User Modeling. 2003, 2702: 4-14.
- [29] Koper R. An introduction to learning design[J]. Learning design, 2005: 3-20.
- [30] Bendakir N, Aïmeur E. Using association rules for course recommendation[C]. Proceedings of the AAAI Workshop on Educational Data Mining. 2006, 3.
- [31] 方海光, 刘静, 黄荣怀, 等. 基于学习对象化的移动学习资源自适应引擎的研究[J]. 中国电化教育, 2009 (7): 51-55.
- [32] 赵学孔, 徐晓东, 龙世荣. B/S 模式下自适应学习系统个性化推荐服务研究[J]. 中国远程教育, 2015 (10): 71-78.
- [33] 康淑敏. 学习风格理论——西方研究综述[J]. 山东外语教学, 2003 (3): 24-28.
- [34] Thelen H A. Dynamics of groups at work[J]. 1954.
- [35] Coffield F, Moseley D, Hall E, et al. Should we be using learning styles? What research has to say to practice[J]. 2004.
- [36] 沈博. 基于 Kolb 学习风格模型的调查和分析[J]. 中国地质大学学报: 社会科学版, 2006 (3): 66-69.
- [37] Honey P, Mumford A. Learning styles questionnaire: 80-item version[M]. Maidenhead, 2006.
- [38] Myers I B M, Quenk M H, Hammer A L. MBTI Manual: a guide to the development and use of the Myers-Briggs type indicator [M]. Manual: a guide to the development and use of the Myers-Briggs type indicator. Consulting Psychologists Press, 1998.
- [39] 张伟远, 欧阳丽红. 成人学习方式的理论, 分类及测量之述评[J]. 开放教育研究, 2002 (2): 30-34.
- [40] 郝丹, 张伟远, 陈丽. 远程学习者学习风格测量方法的研究[J]. 开放教育研究, 2005, 11(1): 51-54.

- [41] Allison C, Hayes J. Validity of the Learning Style Questionnaire[J]. Psychological Reports, 1990, 67: 859-866.
- [42] Caple J, Martin P. Reflections of Two Pragmatists: A Critique of Honey and Mumford's Learning Styles[J]. Industrial and Commercial Training, 1994, 26(1): 16-20.
- [43] Swailes S, Senior B. The dimensionality of Honey and Mumford's learning styles questionnaire[J]. International Journal of Selection and Assessment, 1999, 7(1): 1-11.
- [44] Koob J J, Funk J. Kolb's learning style inventory: Issues of reliability and validity[J]. Research on social work practice, 2002, 12(2): 293-308.
- [45] Vita G D. Learning styles, culture and inclusive instruction in the multicultural classroom: A business and management perspective[J]. Innovations in Education and Teaching International, 2001, 38(2): 165-174.
- [46] Truong H M. Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities[J]. Computers in Human Behavior, 2016, 55: 1185-1193.
- [47] Graf S, Liu T C. Supporting teachers in identifying students' learning styles in learning management systems: An automatic student modelling approach[J]. Journal of Educational Technology & Society, 2009, 12(4): 3.
- [48] Felder R M, Silverman L K. Learning and teaching styles in engineering education[J]. Engineering education, 1988, 78(7): 674-681.
- [49] García P, Amaldi A, Schiaffino S, et al. Evaluating Bayesian networks' precision for detecting students' learning styles[J]. Computers & Education, 2007, 49(3): 794-808.
- [50] Dunn R S, Dunn K J. Teaching students through their individual learning styles: A practical approach[M]. Prentice Hall, 1978.
- [51] Jonassen D H, Rohrer-Murphy L. Activity theory as a framework for designing constructivist learning environments[J]. Educational Technology Research and Development, 1999, 47(1): 61-79.
- [52] Lo J J, Shu P C. Identification of learning styles online by observing learners' browsing behaviour through a neural network[J]. British Journal of Educational Technology, 2005,

36(1): 43-55.

- [53] 黄海波. 网络环境下远程开放教育学习者特征分析[J]. 广东广播电视台学报, 2007, 16(2): 14-17.
- [54] 李念. 基于网络学习行为分析的评价模型研究[D]. 武汉: 华中师范大学, 2007.
- [55] 李素珍. 基于网络学习行为分析的网络学习风格与学习偏好挖掘模型研究[D]. 华中师范大学, 2009.
- [56] Marton, F, Hounsell D, Entwistle N J. The experience of learning: Implications for teaching and studying in higher education[M]. Scottish Academic Press, 1997.
- [57] Kuzilek, J. et al. Open University Learning Analytics dataset Sci. Data 4:170171 doi: 10.1038/sdata.2017.171 (2017).
- [58] Zakrzewska D. Cluster analysis in personalized e-learning systems[J]. Intelligent Systems for Knowledge Management, 2009: 229-250.
- [59] 姜强, 赵蔚, 王朋娇, 等. 基于大数据的个性化自适应在线学习分析模型及实现[J]. 中国电化教育, 2015, 1: 85-92.
- [60] 熊玉珍. 基于测评的汉语个性化学习环境的构建[J]. 电化教育研究, 2012, 3: 69-71.
- [61] 郭玲玲. 学习风格与在线学习行为之间的关系研究[D]. 济南: 山东师范大学, 2007.
- [62] 余胜泉, 杨现民, 程罡. 泛在学习环境中的学习资源设计与共享[J]. 开放教育研究, 2009, 1(5): 1.
- [63] 张家华, 张剑平. 适应性学习支持系统: 现状, 问题与趋势[J]. 现代教育技术, 2009, 19(2): 18-20.
- [64] Greff K, Srivastava R K, Koutník J, et al. LSTM: A search space odyssey[J]. IEEE transactions on neural networks and learning systems, 2017.

攻读硕士学位期间取得的学术成果

攻读硕士学位期间公开发表的学术论文：

1. Rumei Li, Chuantao Yin. Analysis of Online Learning Style Model Based on K-means Algorithm[C]. 3rd International Conference on Economics, Management, Law and Education, 2017: 692 -697. CPCI-S:源
2. Rumei Li, Chuantao Yin, Bertrand David. Online Learning Style Modeling for Course Recommendation[C]. 2017 International Conference on Intelligent Computing, Communication & Devices. EI:源

附录

附录 1 在线学习风格（OLS）量表题目

1. 生活中或课堂上，我碰到不懂的、不会的知识或技能会自己网上去查找
2. 学习某个知识或技能时，我会在网上逛相关论坛或网站
3. 通过网络学习时，我通常会选择视频、图片教程
4. 通过网络学习时，我喜欢学习事实性，启发我思考的内容
5. 通过网络学习时，目录对我的帮助不大
6. 我喜欢有创造性的工作
7. 解决数学问题时，我常常先找出答案，再思考如何得出解题步骤
8. 通过网络学习时，我会坚持学到自己满意为止，而不是觉得困难就放弃
9. 碰到问题，我会去看网络论坛上大家的讨论
10. 对于一份带有插图的网页，我通常会仔细阅读文字
11. 对于重复性工作，我始终会仔细地做每一个步骤
12. 通过网络学习时，我喜欢按照给定的章节顺序进行学习
13. 我喜欢通过网络进行自学，这使我有成就感
14. 我在问答论坛中提出自己的问题并解决他人的问题
15. 通过网络学习时，我喜欢听人讲解(如音频教程、类似传统课堂的视频教程)
16. 当我执行一项任务时，比起想出多种方法，我偏向熟练掌握一种方法
17. 通过网络学习时，我通常先学习完各个章节，才能对所学的技能 / 知识有整体把握
18. 通过网络学习时，我很容易理解某种理论的推论过程
19. 通过网络学习时，我喜欢看总结性的章节，它让我对知识的了解更深入
20. 通过网络学习时，(相比传统课堂)我的效率会更高
21. 回忆某件事情时，我的脑海中会出现一些情节或画面
22. 通过网络学习时，我喜欢能告诉我新事物和教我怎么做的学习材料
23. 通过网络学习新科目时，我试图建立该科目与其他科目的联系

附录 2 CF-AROLS 伪代码

算法：基于 TopN 协同过滤的自适应学习资源推荐算法 CF-AROLS

输入：学习风格诊断结果 \mathcal{D} , CF 树阙值 th , 聚类数目 k , 学习者浏览日志 log_data , 频繁项集支持度阙值 $minSup$ 、频繁项集可信度阙值 $minConf$ 、Kulc 度量阙值 $minKulc$, 不平衡比阙值 $minIR$, 浏览趋势阙值 μ , 学习者测试样本 x , 推荐个数 N

输出：为学习者 x 推荐的学习资源

遍历 \mathcal{D} , 根据 th 构建符合条件的 CF 树;

将 CF 树生成的相近子簇聚合, 直至类的数目缩小到 k , 将聚类结果在 \mathcal{D} 中标记;

遍历 log_data :

以浏览学习资源 id 为键, 浏览过该资源的学习者 id 列表为值构建 $item_user_log$ 字典;

以学习者 id 为键, 该学习者浏览过的学习资源 id 列表为值构建 $user_item_log$ 字典;

对于 $item_user_log$ 中的每一个学习资源 i_m :

计算 i_m 与之后所有学习资源 i_n 的余弦相似度 $cos(i_n, i_m)$ 并记录到学习资源相似度矩阵 M_{item} ;

通过 $user_item_log$ 中的浏览序列集合, 根据 $minSup$, $minConf$, $minKulc$, $minIR$ 生成关联规则集合 \mathcal{R} ;

根据 x 的 id, 在 $user_item_log$ 中找到 x 浏览过的物品列表 his_items ;

对于 his_items 中的每一个学习资源 i_m :

在 M_{item} 中找到与 i_m 最相似的 N 个学习资源, 对于其中每一个学习资源 i_n :

如果 $i_n \notin his_items$:

如果 $i_n \in P_{CF}$, 则 $P_{CF}[i_n] := P_{CF}[i_n] + cos(i_n, i_m)$, 否则将 $i_n: cos(i_n, i_m)$ 放入字典 P_{CF} ;

对于 \mathcal{R} 中的每一条关联规则 $l_{set} \rightarrow r_{set} (conf)$:

如果 $l_{set} \subset his_items$ 且 $r_{set} \cap P_{CF}.keys \neq \emptyset$:

将 $r_{set} \cap P_{CF}.keys$ 中的学习资源放入字典 P_{AS} ;

对 P_{CF} 中的每一个学习资源 i_m :

如果 $i_m \in P_{AS}$: $P_{CF}[i_m] := P_{CF}[i_m] * P_{AS}[i_m]$;

计算浏览趋势阙值 $T := \mu \times avg(P_{CF}.values)$;

对 P_{CF} 中的每一个学习资源 i_m :

如果 $P_{CF}[i_m] > T$: 输出 i_m ;

附录 3 LSTM-AROLS 伪代码

算法：基于 LSTM 的自适应学习资源 TopN 推荐算法 LSTM-AROLS

输入：学习风格诊断结果 \mathcal{D} , CF 树阙值 th , 聚类数目 k , 学习者浏览日志 log_data , 学习资源向量长度 k , 窗口大小 $window_size$, LSTM 模型, 学习者测试样本 x , 推荐个数 N

输出：为学习者 x 推荐的学习资源

遍历 \mathcal{D} , 根据 th 构建符合条件的 CF 树;

将 CF 树生成的相近子簇聚合, 直至类的数目缩小到 k , 将聚类结果在 \mathcal{D} 中标记;

遍历 log_data :

以学习者 id 为键, 该学习者浏览过的学习资源 id 列表为值构建 $user_item_log$ 字典;

根据 $user_item_log$ 中的浏览序列构建长度为 k 的学习资源向量矩阵 W , 窗口大小为 $window_size$;

将 $item_user_log$ 中的浏览序列变成长度为 $window_size + 1$ 的多个序列, 并将所有序列中的学习资源转换为向量, 构建学习资源向量序列 $seqs$;

$X_train := seqs[:, :window_size]$;

$Y_train := seqs[:, -1]$;

将 X_train 和 Y_train 作为 LSTM 模型的训练数据, 对模型进行训练;

根据 x 的 id, 在 $user_item_log$ 中找到 x 最近浏览过的 $window_size$ 个物品并转换为学习资源向量序列;

当 $N > 0$ 时:

将 x 的浏览序列 s 输入训练好的模型, 得到学习资源向量预测结果 $pred_y$, 并存入链表 $pred$;

$N := N - 1$;

$s := s[1:] + pred_y$;

将 $pred$ 中每个学习资源向量转换为 W 中与其最相似的向量的学习资源, 生成推荐列表;

附录 4 项目源代码

本研究中的算法以及系统源代码如下：

1. AROLS 研究代码: <https://github.com/leerumor/arols>
2. AROLS 系统原型: <https://github.com/leerumor/arolsweb>