

数据匹配与智能推送算法研究及其在教学训练系统中的应用

王秀华¹, 秦留洋², 沈颖¹

(1. 国防大学 联合勤务学院, 北京 100080; 2. 北京航天测控技术有限公司, 北京 100041)

摘要: 军事院校在校学员和任职培训人员需要结合自身军兵种特征、岗位职务特点, 学习个性化知识、开展针对性训练, 教学训练业务和辅助系统工具需要满足实战化、个性化要求; 研究基于基础数据、操作数据、任务数据、浏览数据、标记数据等多维个性化数据的匹配分析、处理挖掘和精准推送技术, 通过数据采集、行为分析、特征提取、机器学习、自然语言处理、数据挖掘、数据匹配、多维推送、效果评估等手段, 支撑教学训练过程中各类型学员的自主学习和自我训练工作; 研制教学训练系统软件, 以实战化应用验证数据匹配与智能推送算法的研究成果, 系统使用效果表明基于数据分析匹配的智能化知识资源推送方法可以有效提高学员学习效率、学习主动性, 提高学员的业务水平、扩展学员的知识领域、锻炼学员的自主能力。

关键词: 教学训练系统; 数据挖掘; 数据匹配; 智能推送; 课程设计

Research on Data Matching and Intelligent Push Algorithms and Its Application in the Teaching-Training System

WANG Xiuhua¹, QIN Liuyang², SHEN Ying¹

(1. Joint Logistics College of National Defense University of PLA, Beijing 100080, China;

2. Beijing Aerospace Measurement and Control Technology Co., Ltd., Beijing 100041, China)

Abstract: It is necessary for military academy students and on-the-job training personnel to combine their own characteristics of military service and job responsibilities, and to learn personalized knowledge and carry out targeted training. Teaching and training business and auxiliary system tools need to meet the requirements of practical and personalized training. Based on multi-dimensional personalized data such as basic data, operation data, task data, browsing data, and tagging data, research is conducted on the matching analysis, processing mining, and precise push technology. Through data collection, behavior analysis, feature extraction, machine learning, natural language processing, data mining, data matching, multi-dimensional push, and effect evaluation, it supports self-learning and self-training for various students in the teaching and training process, develops teaching and training system software, and verifies the research results of data matching and intelligent push algorithms through practical application. Application results show that based on data analysis and matching, the intelligent knowledge resource push method can effectively improve the learning efficiency and initiative of students, enhance their professional levels, expand their knowledge fields, and exercise their self-abilities.

Keywords: teaching-training system; data mining; data matching; intelligent push; course designing

0 引言

在数字技术浪潮席卷全球的当下, 教育领域正经历着前所未有的范式变革。传统教学模式受限于统一化教学流程与有限的资源供给, 难以满足新时代学习者个性化、差异化的学习诉求。学员普遍面临学习资源供需失

衡、知识获取路径模糊等现实困境^[1]。尤其在军事院校的教学训练体系中, 如何突破传统桎梏, 构建适应现代化战争人才培养需求的教育生态, 已成为亟待攻克的关键课题。

为破解这一难题, 基于数据深度分析的业务知识资源智能推送技术应运而生。该技术深度融合人工智能、

收稿日期:2025-06-13; 修回日期:2025-09-03。

作者简介:王秀华(1977-),女,博士,副教授。

引用格式:王秀华,秦留洋,沈颖. 数据匹配与智能推送算法研究及其在教学训练系统中的应用[J]. 计算机测量与控制, 2025, 33(10):289-297.

数据挖掘、智能匹配等前沿科技,通过对学员浏览行为、兴趣偏好、学业基础等多源数据的深度剖析,实现知识资源与业务资源的精准适配。系统通过构建多维度特征识别与提取网络^[2-3],利用数据差分算法实现特征数据的智能匹配,以学员日常学习数据与学业基础数据为支撑,构建动态差分计算模型,从而实现学习资源的精准靶向推送^[4-5]。

值得注意的是,不同用户群体的数据特征与服务需求呈现显著差异^[6]。智能推送教学训练系统通过基础信息和基本情况的消息推送注册^[7],利用大数据分析、机器学习和人工智能等先进技术^[8],实现计算机自动化系统模拟人类学习行为,以此获取新的技能^[9],实现学习资源的自动化、智能化推送。该系统能够精准捕捉学员的学习习惯、兴趣倾向与知识掌握程度,为其定制个性化的学习资源与训练任务,有效提升教学训练的针对性与实效性。

综上,本文提出的方法主要解决以下关键核心问题:

- 1) 如何精准匹配学员个性化学习需求与知识资源;
- 2) 如何通过智能推送提高学员学习效率和主动性;
- 3) 如何辅助教学管理决策科学化,优化教学安排与训练方案。

本文将从数据解析与特征挖掘、数据决策与智能推送等相关技术进行教学训练系统设计中的关键技术说明,并结合关键技术,详细阐述基于智能推送的教学训练系统实现架构、功能设计和结果验证过程。

1 数据解析与特征挖掘

智能推送教学训练系统的核心在于其智能化和个性化的特征^[10]。通过对学生学习数据的实时分析,系统能够识别出学生的薄弱环节和学习偏好,从而为其推荐最适合的学习内容。这种基于数据驱动的教学方法不仅提高了学习效率,也增强了学生的学习兴趣 and 参与感。

此外,智能推送教学训练系统还具有实时反馈和评估的功能。学生在学习过程中可以随时获得系统提供的反馈,帮助他们及时调整学习策略,优化学习效果。这一特性使得教学过程更加灵活和动态,能够适应不同学生的学习节奏^[11-12]。

数据匹配与智能推送方法的实现,数据的正确解析和高效的特征挖掘是关键和基础。基于学员的学习和浏览数据的整合分析,可以提高推送资源的匹配度和学员的点击率,从而提高教学训练的效率。

1.1 数据特征提取方法

学员浏览数据特征提取是关键数据挖掘和机器学习过程的重要任务,在学员行为分析、推荐内容投放等场景中具有重要作用。本文通过数据收集、预处理、特征生成和特征选择 4 个步骤实现对学员浏览数据的数据特

征提取,为后续数据聚类 and 关联提供特征要素支撑。

1.1.1 数据收集

浏览数据通常包括学员的浏览记录、点击行为、停留时间及页面交互等信息。浏览数据主要通过如下方式收集:

1) 日志文件:系统服务器会记录学员的访问日志,包括学员的 IP 地址、访问时间、请求的 URL、响应时间等信息。

2) 浏览器插件:通过浏览器插件进行学员浏览行为的捕捉,收集更为详细的信息。

3) 专门工具:设计学员用户行为追踪工具帮助系统管理员分析学员的行为模式。对接收到的网络数据进行数据异常检测,将数据按是否通过检测进行分类^[13];

1.1.2 数据预处理

数据可能包含错误重复、无效值和缺失值等不规范情况^[14],在特征提取之前,需要对收集到的数据进行预处理,根据系统的业务要求,预处理步骤主要包括去重、时间归一化、缺失值处理和数据清洗等,其中:

1) 去重:采用数据要素比对和差异性计算的方法,实现去除重复的浏览记录,确保每个事件的唯一性。基础比对要素包括学员 ID、访问时间戳、请求 URL 等核心字段,若核心字段完全一致,则直接判定为重复记录;若存在部分字段差异,则计算字段差异度(设置差异阈值),低于阈值的记录视为重复。通过合理设置时间戳精度和时间窗口实现准确提高重复识别度,确保无有效数据被误删。时间戳统一精确到秒级,避免因毫秒级差异误判非重复记录(如学员快速刷新页面的连续请求);采用滑动窗口法,对窗口内同一学员的重复 URL 访问记录仅保留首次或末次访问,保证访问记录的差异性。

2) 时间归一化:采用时间解析方法将时间戳转换成统一的格式,方便后续的分析。对于结构化时间戳,采用标准格式(YYYY-MM-DDTHH:MM:SS)进行转换。对于非结构化时间描述,解析相对时间,结合系统当前时间推算绝对时间戳。

3) 缺失值处理:采用同类数据分析和平滑插值等方法,实现对缺失的数据进行填补或删除,以保证数据的完整性。结合数据类型与业务场景选择适配方法。对于数值型数据(如停留时间、浏览次数),优先采用平滑插值法。若数据时序特征明显(如按小时/天的访问记录),选择线性插值或样条插值,通过相邻有效数据的趋势拟合缺失值;若数据分布较离散(如不同课程的点击量),则采用 K 近邻(KNN)插值,基于同类学员的相似行为数据推算缺失值。对于类别型数据,采用同类数据分析填充法。通过统计同岗位/同军兵种学员的高频类别值,或基于业务规则进行缺失值补充。

针对不同缺失比例数据差异化处理可以有效提高数

据处理和分析的效率与准确度,若缺失率低于预设阈值,直接填充缺失值;若缺失率过高,填充后通过数据分布检验(本文选取正态分布开展分布检验)验证填充效果,确保填充后数据与原始分布偏差满足要求。

4) 数据清洗:采用数据要素转换和数据识别等方式,实现去除无效数据,包括:

(1) 页面错误类。HTTP 状态码为 4xx、5xx 的访问记录,仅保留 2xx(成功)和 3xx(重定向)的有效记录;

(2) 异常行为类。超出合理范围的极端值,如异常挂机、恶意刷新等;

(3) 格式错误类。通过正则表达式校验学员 ID、URL 路径,剔除格式不符的脏数据。

采用“先过滤后校验”的流水线模式:第一步批量过滤错误状态码记录,第二步通过规则引擎校验异常行为,第三步通过格式校验剔除无效字段,最终保留的有效数据占比高于阈值数据预处理环节能够更精准地支撑后续特征提取与关联分析,为智能推送算法提供高质量的基础数据。

1.1.3 特征生成

特征生成是特征提取的核心步骤,通过数据特征的定义和归类,选择合适的技术手段对特征要素进行生成,从而提高数据的应用效率和准确率。本文针对学员浏览数据的特征生成过程主要包括以下几个方面:

1) 基本特征:

浏览次数(B_N):学员在一定时间内访问某个网页的次数。

停留时间(S_T):学员在某个页面停留的时间,可以反映学员用户对内容的兴趣。

跳出率(O_R):学员在访问某个页面后没有进行其他操作而直接离开的比例。

2) 行为特征:

序列特征(SE)^[15]:学员的浏览路径,可以用来分析学员的行为模式,比如“首页→课程页→详情页”。

频率特征(FQ):学员在特定时间段内的访问频率,比如日、周、月的访问量。

设备特征(EQ):学员使用的设备类型,不同设备的浏览行为可能存在差异,本文所述系统均在教学实验室 PC 机上,主要分析学员应用的设备属性和为主场所等,分析学员的登录习惯。

3) 社交特征:

社交媒体分享:学员用户通过社交媒体分享网页的次数,反映内容的传播性。本文所述系统建立在校园内部网络环境下,主要分析学员在内部论坛内的分享情况。

评论和反馈:学员用户对网页内容的评论数量和情感倾向,可以作为学员用户满意度的指标。

1.1.4 特征选择

特征选择是为了去除冗余和无关的特征,提高模型的性能。主要方法包括:

1) 过滤法:根据特征与目标变量的相关性进行筛选,计算各特征与目标变量(如学习效果评分)的皮尔逊相关系数,筛选系数绝对值高于阈值的特征。

2) 包裹法:通过训练模型来评估特征子集的效果,以决策树或支持向量机(SVM)为基模型,迭代剔除对模型性能贡献最低的特征,实现递归特征消除(RFE),直至保留最优特征子集。

3) 嵌入法:结合模型训练过程进行特征选择,比如使用 Lasso 回归中的 L1 正则化,通过正则化参数 λ 控制特征系数稀疏性,系数为 0 的特征被自动剔除。

数据特征选择的算法适用需要平衡筛选效率与模型适配性:包裹法对于数据规模适配较为繁琐,更加适用于数据规模变化较大的情况,不同数据规模的方法适用较为复杂,不易形成统一标准,且迭代次数配置难以固定,在教学训练系统这种数据规模变化不大的环境下,包裹法降低了筛选效率;嵌入法通过网格搜索结合交叉验证确定回归特征系数最优值,需要迭代更新初始搜索范围,计算复杂度和收敛效率难以有效提高,另外嵌入法通过特征系数的非零占比与模型解释性权衡,为确保保留特征能直观反映人员行为规律,对特征系数类型的选取要求较高,给实现过程增加了复杂度。

本文针对教学训练系统这一典型业务场景,选择过滤法实现数据特征选择,

灵活设置阈值,满足不同的业务场景需求,通过简单的阈值大小调节就可以有效的实现对数据保留特征的筛选,如教学训练系统中通常将阈值设为 0.3~0.5。若需保留更多潜在相关特征(如兴趣偏好类特征),阈值可下调至 0.2;若需精简特征维度(如减少模型计算量),阈值可提高至 0.6。通过特征剔除前后的模型准确率对比,若剔除低相关特征后准确率下降不超过 5%,则认为筛选有效。可以看出,过滤法既简化了业务逻辑,又提高了筛选效率,通过上述细节优化,数据特征提取过程更贴合教学训练场景的实际需求,既保证了数据质量,又为后续数据关联分析与智能推送提供了精准的特征支撑。

学员的浏览数据特征提取是一个系统性工程,涉及数据收集、预处理、特征生成和特征选择等多个环节。通过科学的特征提取方法,可以为后续的学员用户行为分析、个性化推荐和智能推送提供有力支持^[16],为提升学员用户体验、学习效率和知识价值做出贡献。

浏览数据特征提取的流程如图 1 所示。

1) 数据收集:从网页日志中收集学员用户的浏览数据,包括访问的网页、时间戳、学员用户 ID 等信息。

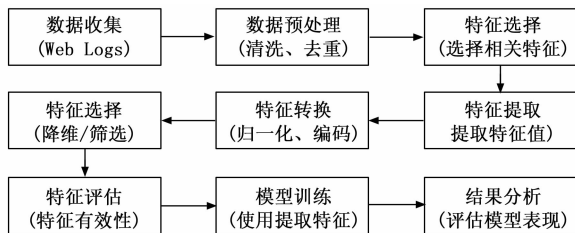


图 1 数据特征提取过程

2) 数据预处理: 对收集到的数据进行清洗, 去除重复项和无效数据, 确保数据的质量。

3) 特征选择: 根据业务需求选择与学员用户行为相关的特征, 例如访问频率、停留时间、点击率等。

4) 特征提取: 从原始数据中提取出选择的特征值, 例如将浏览时长、浏览深度等转化为数值型特征。

5) 特征转换: 对提取的特征进行归一化处理或编码, 以便于后续的模式训练。

6) 特征选择: 通过降维技术 (如 PCA) 或特征筛选方法, 进一步减少特征空间的维度, 保留最重要的特征。

7) 特征评估: 使用统计方法或模型评估特征的有效性, 确保所选特征对任务的贡献。

8) 模型训练: 使用提取和选择的特征训练机器学习模型, 以实现特定的目标 (如学员用户分类、推荐等)。

9) 结果分析: 分析模型的表现, 包括准确率、召回率等指标, 评估特征提取的有效性及其模型的实用性。

1.2 数据关联分析方法

常见的数据挖掘方法包括数据分类、回归分析、关联分析等。

数据分类方法用简单的规则表示, 易于理解和解释, 能够有效处理大规模数据集, 具有较好的可扩展性, 在特征选择和数据预处理得当的情况下, 分类模型能够提供较高的预测准确率, 且支持处理多类别问题。然而, 分类模型可能会在训练数据上表现良好, 但在测试数据上表现不佳, 尤其是在样本量不足或特征过多时, 存在过拟合风险。其次在类别分布不均的情况下, 分类模型可能对少数类的预测效果较差。以及分类算法通常依赖于特征的选择和质量, 不合适的特征可能导致性能下降^[17-19]。

回归分析能够提供定量的预测结果, 适用于连续变量的预测, 线性回归模型相对简单, 易于实现和解释, 提供了丰富的统计推断工具, 可以评估变量之间的关系和显著性。然而, 线性回归假设自变量与因变量之间存在线性关系, 若真实关系非线性, 则模型性能会下降。回归模型对异常值敏感, 可能会影响整体的拟合效果。自变量之间存在高度相关性时, 可能导致回归系数不稳定, 影响模型的解释能力^[20-21]。

关联分析能够揭示变量之间的潜在关系, 关联规则

挖掘是一种无监督学习方法, 不需要预先标注数据, 适合大规模数据集的分析。生成的关联规则通常直观易懂, 方便进行决策支持。然而, 常规的关联分析方法关联规则往往生成大量规则, 可能导致信息过载, 难以提取有效的知识。关联分析通常不考虑时间序列因素, 可能无法捕捉动态变化的关系。在大规模数据集上, 关联规则挖掘的计算复杂度较高, 可能需要大量的计算资源^[22-23]。

本文提出一种基于特征提取的数据关联分析方法进行数据的挖掘和规则的匹配, 解决关联分析方法的问题。首先, 特征提取后数据要素更加清晰, 弥补了信息规则过载的问题。其次特征提取的时序特征提取弥补了动态化不足的问题。最后, 经过特征处理后, 无用冗余的数据被预处理步骤提出, 减少了数据量, 降低了系统的计算复杂度。

基于特征提取的数据关联分析的主要目的是对数据的关联规则进行挖掘, 从而明确数据应用目的于关系数据集中数据之间的关联和相关性。在教学训练系统的设计中, 随着大量学员的系统应用数据的不断收集和累积, 通过对其数据库中挖掘此类模式, 交叉学员学习浏览行为分析, 查找学员放入“知识购物篮”中的不同知识项目之间的关联来分析学员的学习习惯和学习要点, 从而辅助设计教学训练业务决策过程, 例如学习资源设计、课程目录设计、考核练习设计、学习评价设计等。

数据关联分析的实现主要以关联规则的支持度和置信度两种度量为评价标准^[24]。通常, 如果关联规则同时满足最小支持度阈值和最小置信度阈值, 则被认为是强关联性的, 基于该规则进行关联的数据被学员认为是“感兴趣”的可能性便更高。

令 $I = \{i_1, i_2, i_3, \dots\}$ 是项目的集合, 事务数据库 D 是由一系列具有唯一标志 TID 的事务组成, 每个事务对应项目集上的一个子集, 即 $T \subseteq I$ 。关联规则可以表示为:

$$\{X \Rightarrow Y \mid X \subseteq I, Y \subseteq I, X \cap Y = \phi\}$$

则有:

支持度 S 指事务数据库中包含 $X \cup Y$ 的事务占事务数据库 D 的百分比, 置信度 C 是指事务数据库中包含 $X \cup Y$ 的事务数与包含 x 的事务数之比。

如图 2 所示, 数据规则挖掘的过程包括 5 个主要步骤: 提取目标的定义、数据选择、数据转换、数据挖掘以及最终的结果解释。

由于关联规则是基于概率的一种统计学意义上的分析, 因此存在关联规则不唯一的情况, 当算法给出的关联规则较多的时候, 决策过程就会出现无法确定重要且简洁的预测规则, 从而无法做出正确决策的问题。

对关联规则的有效评估是解决这个问题的方法, 对

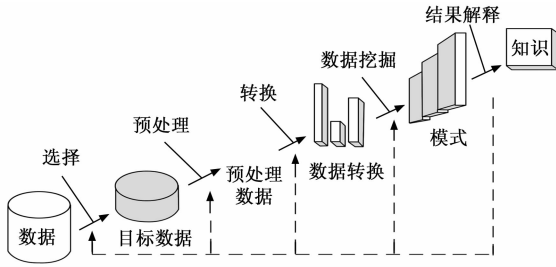


图 2 数据关联规则挖掘过程

给定目标进行规则过滤和排序的度量主要包括客观、主观及基于语义的度量。本文提出的基于置信度 C 的客观度量更加符合智能化的自动化系统进行实现, 且评估规则可量化。构造置信度评估模型如下:

$$CE(A \rightarrow C) = \frac{P(A \cap C)}{P(A)} = \frac{P(AC)}{P(A)P(C)}$$

即规则置信度 C 与结果发生概率 A 的比率, 反映规则的前提和结果的相关性。

2 数据决策与智能推送

数据决策与智能推送是基于数据关联分析结果的对关联性进行深度分析后对数据的匹配性进行决策并采用智能推送方法向学员用户进行推送, 并实时获取学员的点击反馈, 优化数据关联分析的规则和决策推送的结果^[25-26]。

本文提出一种基于数据深度分析的业务知识资源智能推送方法, 主要包括以下步骤:

步骤 1: 学员用户依据知识资源服务提供的相关业务流程, 通过学员用户接口访问知识资源服务并综合考虑自身职位特征与需求信息的基础上选择符合自身要求的学习任务;

步骤 2: 业务流程管理服务根据学员用户选择的任务信息获取相应的流程知识元描述数据, 同时将复杂任务进行分解, 形成若干业务活动, 进一步将分解后的任务描述信息发送到知识资源服务中心;

步骤 3: 知识资源服务中心的匹配推送管理服务在接收到学员用户选择的业务描述信息后, 将学员用户所选任务的描述信息与云端知识资源池中的海量知识资源进行匹配, 得到初步的知识匹配结果集。本步骤的实现就是基于本文提出的数据解析与特征挖掘的方法;

步骤 4: 智能决策管理模块在进一步在综合分析学员用户需求管理模块和学员用户行为管理模块得到的学员用户角色特征信息及学员用户行为特征信息的基础上, 基于模糊意见集中决策算法对初步的知识匹配结果集进行智能过滤与筛选, 以进一步提升所得结果与学员用户需求的匹配程度, 进而得到最终的知识资源匹配结果集;

模糊意见集中决策算法通过多维度评价指标对初步

知识匹配结果进行筛选, 通过 Borda 数与加权 Borda 数实现资源排序。

模糊意见集中决策算法具体如下:

1) 根据步骤 3 得到的初步知识资源匹配结果集, 将其设定为论域:

$$K = \{k_1, k_2, \dots, k_n\}$$

2) 根据论域中知识资源的特点, 采用学员用户角色需求、行为特征需求、同类型学员用户需求三个指标对论域中知识资源进行评价, 并分别给出知识资源之间的排序为:

$$S = \{s_1, s_2, s_3\}$$

其中: s_i 是第 i 种评价序列, $i=1, 2, 3$, 即 K 中元素的某一排序。

$$s_i = \{r_{i1}, r_{i2}, \dots, r_{in}\}$$

其中: r_{ij} 表示第 i 个指标下第 j 个资源的排序位置。

学员角色需求 s_1 : 反映资源与学员军兵种、岗位职责的匹配度;

行为特征需求 s_2 : 基于浏览记录、停留时间等数据反映资源与学习习惯的适配性;

同类型学员需求 s_3 : 参考同层次/同岗位学员的高频学习资源。

3) 知识资源 k 的 Borda 数 $B_i(k)$ 表示第 i 种评价序列 s_i 中排在 k 之后的元素个数, Borda 数直观反映资源在单一指标下的相对优势, 数值越大说明排序越靠前。若 k 在第 i 种评价序列 s_i 中排第 h 位, 则:

$$B_i(k) = n - h_i$$

4) 计算知识资源 k 的加权 Borda 数, 所采用的三个评价指标在知识推送过程中其重要性有所不同, 因此通过层次分析法确定各评价指标所占权重, 即: $W = \{\omega_1, \omega_2, \omega_3\}$, 其中 ω_i 是 i 种评价指标所占权重, 计算加权 Borda 数:

$$WB(k) = \sum_{i=1}^3 \omega_i \cdot B_i(k)$$

根据教学阶段动态调整评价指标权重, 入学初期侧重角色需求, 中期强化行为特征需求, 确保资源匹配与培养目标同步。权重 W 通过层次分析法确定, 包括:

(1) 层次构建:

目标层: 优化知识资源推送匹配度;

准则层: 学员角色需求 (C1)、行为特征需求 (C2)、同类型学员需求 (C3)。

(2) 判断矩阵构造:

通过专家打分 (结合教学经验) 对准则层指标两两比较, 采用 1~9 标度法 (1 表示同等重要, 9 表示极端重要) 构建判断矩阵。

$$A = \begin{bmatrix} 1 & a_{12} & a_{13} \\ a_{21} & 1 & a_{23} \\ a_{31} & a_{32} & 1 \end{bmatrix}$$

其中： $a_{ij} = 1/a_{ji}$ 。

(3) 一致性检验：

计算矩阵最大特征值 $\lambda_{\max}$ ；计算一致性指标 $CI = (\lambda_{\max} - 3)/(3 - 1)$ ；查表得平均随机一致性指标 $RI = 0.58$ (3 阶矩阵)；一致性比例 $CR = CI/RI$ ，若 $CR < 0.1$ ，则矩阵通过一致性检验。

(4) 权重计算：

采用特征向量法，对通过检验的矩阵求最大特征值对应的归一化特征向量，即为权重。示例中计算的。

5) 通过加权求和整合多指标评价结果，论域 K 中所有知识资源按加权 Borda 数的大小排成一个线性序列，最终 WB (k) 降序排列，截取排在序列前面的 N 个知识资源，根据学员日常资源负载调整，进而得到最终的知识资源匹配结果集。结合学员日均学习时长设定截取数量，军事院校学员每日推送 5~8 个核心资源，避免信息过载，同时保证学习深度。

6) 知识资源服务中心的匹配推送管理服务将符合学员用户业务需求的资源检索与匹配结果集依据学员用户在整个任务执行过程中所处的具体业务阶段进行分阶段推送。

3 基于智能推送的教学训练系统实现

教学训练系统的构建可以加快提升教研团队教战研战能力，提高实战化专业化教学育人质量水平，满足不同类型和水平的学员教学需求，支撑各级机关开展筹划和决策咨询。

在自主构建教学训练系统的过程中，考虑到学员的个体化差异 (包括能力差异、业务差异、课程差异、兴趣差异等)。

基于智能推送的教学训练系统基于学员的个体学习差异进行相关课程、学习材料等资源的智能编排和推送。智能推送功能的设计，有效的利用了机器学习和智能匹配等智能方法，实现“进度参差推进、学员个性学习、教员因材施教、知识针对性强、技能系统提高”的良性、高效学习体验。

3.1 系统体系架构

本系统建设于教研室等室内教学环境，为实现多学员的同时登录和同步使用，采用 B/S 架构 (Browser/Server，浏览器/服务器) 进行系统设计和开发。

如图 3 所示，系统基于分层技术体系架构，由基础支撑、资源要素、网络互联、服务供给和能力生成等 5 层构成。

1) 基础支撑层包括通信网络、计算存储、综合管理、时空基准和基础数据构成，即依托网络环境，构建系统基础设施，为上层应用提供硬件和环境基础。

2) 作为基于数据的智能化推送业务系统，资源要

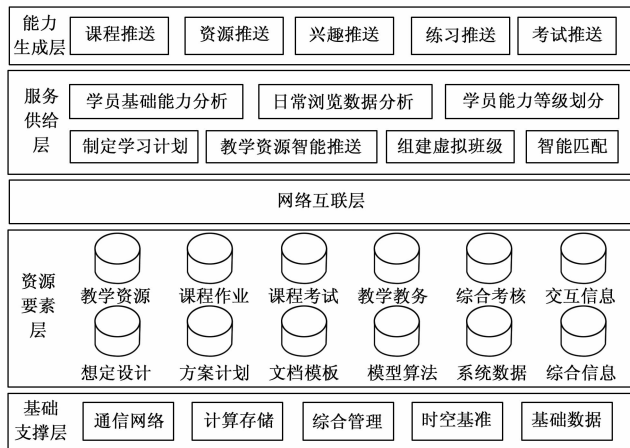


图 3 基于智能推送的教学训练系统体系结构图

素层是系统设计实现的重点内容。资源要素层作为系统架构中的存储层，存储了相关的历史数据和资源数据，并在系统交互过程中接受学员传输的数据来更新历史训练库，同时模型训练产生的日志以及模型参数等中间结果也是主要存储对象，便于后续的模式微调过程。原始的训练数据包括两类：业务预测数据集和数字表达学习数据集。

3) 网络互联层是整个网络和数据传输的传输架构，规定了网络传输环境和参数。军事院校教学训练系统需要根据实际要求搭建专门的网络环境和设施。

4) 服务供给层接收学员用户的请求输入，然后将系统的相关功能进行计算和反馈，在本系统中反馈给学员用户的功能内容主要包括模型训练服务 (包括数据分析、数据关联等训练计算过程)、功能加载服务 (包括智能匹配、智能推送等)，以及基础管理服务 (包括虚拟班级组建、学习计划指定生成等)。

5) 能力生成层是基于智能推送的教学训练系统实现的业务能力，在军事院校学员的学习中，主要实现学习课程、知识资源、兴趣内容、自主练习和考试考核的智能推送。

3.2 系统功能组成

根据教学训练业务的需求，基于智能推送的教学训练系统应具备如下能力：

- 1) 多维数据智能分析和多源教学资源匹配能力；
- 2) 学员能力等级划分、个性化学习计划的生成、制定能力；
- 3) 多维资源的智能推送能力；
- 4) 智能算法的调用、管理和更新的能力。

因此，基于智能推送的教学训练系统的主要功能包括智能分析模块、智能推送模块等，如图 4 所示。

如图 5 所示，通过学员基础能力分析和日常浏览数据分析等分析手段，获取推送依据。调用智能推送和数

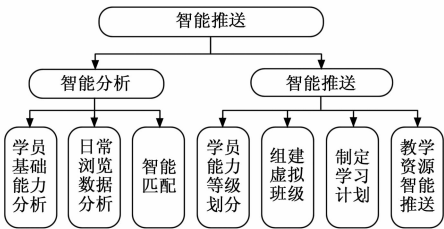


图 4 基于智能推送的教学训练系统功能组成图

据挖掘算法，完成对课程、资源和考试等的智能化和自动化推送。

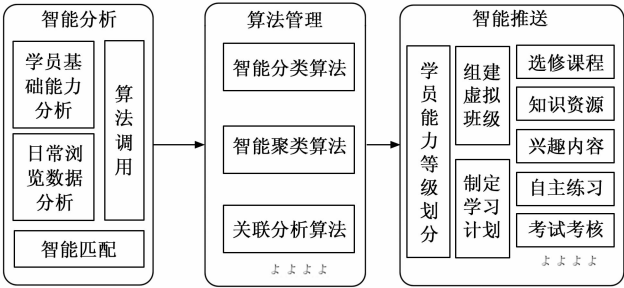


图 5 基于智能推送的教学训练系统业务逻辑图

3.3 系统功能实现

智能分析模块是相关学习内容智能推送的前提，系统通过分析学员能力从而匹配相关学习内容。有效提高了学员的学习效率，避免了重复学习和无效学习。有效填补了学员课外自我培训、自主学习和知识技能提升的空白。

1) 学员基础能力分析：学员基础能力是智能推送的重要依据，依据教学任务，系统会安排重点科目的等级考试，分析学员基础能力，对学员的已有知识水平有一个初步的掌握。

2) 日常浏览数据分析：系统会自动收集、分析学员的日常浏览数据，智能判定学员的学习浏览习惯和感兴趣内容，作为学员基础能力分析的补充，从而实现通过分析匹配进行定向推送。

3) 智能匹配：智能算法调用实现对算法的自动调用，实现智能聚类、分类和关联分析，从而为智能推送提供算法支撑。

智能匹配实现依据学员基础能力和日常浏览的数据分析结果，对课程资源进行智能匹配从而实现有针对性的智能推送和课程推荐。

智能推送模块通过智能分析和匹配的结果对学员进行课程资源的推送，从而引导学员完善自己的知识体系和学习计划。

1) 学员能力等级划分：对学员能力等级进行合理科学的划分是培训教员对学员的整体情况进行摸底的重要过程和辅助手段，系统通过学员的基础能力分析结

果，对各个学员在入学阶段进行学员能力级别的划分。从而作为等级班级和学习计划的依据。

2) 组建虚拟班级：根据学员的能力级别，知识学习分系统对同学科、同等级的学员进行智能的虚拟线上班级划分，打破行政班级壁垒，从而实现个性化学习。

3) 制定学习计划：根据学员的虚拟班级和感兴趣内容，系统辅助教员对不同的学员进行有效的个性化定制学习计划。

4) 教学资源智能推送：如表 1 所示，通过系统判定并规划的个性化学习内容，结合学员的学习进度、日常测评的表现和考核成绩、学习过程中对相关辅助资料的阅览情况等，有计划、有针对性的进行相关教学资源的自动推送。支持按时间、按学习阶段的定时推送，和根据数据分析结果的自主推送等多种推送模式。完善学员的课程学习体系。

表 1 智能推送资源表

推送资源	推送方式
学习课程	按照课程计划、学员学习情况等进行课外辅助课程、自学课程等课程资源的定时推送，匹配正课的学习阶段，为学校正课课程的学习提供辅助支撑
知识资源	按照课程计划、学员学习情况等进行参考资料、扩展阅读等知识资源的定时推送，匹配课程学习和扩展学习要求，为学员丰富知识储备、健全知识体系提供辅助支撑
兴趣内容	依照学员的日常浏览数据分析，主动推送系统判定的学员的感兴趣内容，包括阅读材料、案例、音视频资源等扩展兴趣内容
自主练习	按照课程计划、学员学习情况等进行扩展练习的定时推送，匹配课程学习和扩展学习要求，为学员正确应用知识、熟练掌握技巧提供扩展练习和自主应用平台
考试考核	按照课程计划、学员学习情况等进行考试考核的定时推送，匹配课程考核和自主评测要求，为学员自我评价、自我测试提供支撑

4 系统应用与方法验证

4.1 实验设计与实施

为全面、科学验证基于数据匹配与智能推送算法的教学训练系统有效性，通过校园局域网部署于教学实验室的相关计算机客户端上由学员进行使用。设计多维度、分层级的实验验证方案，通过分类群体分析、多指标数据采集、干扰因素控制，实现对系统性能的综合评估。

如表 2 所示，选取学院在校学员及任职培训人员共 32 人作为实验对象，依据“同质性分组”原则，结合军事教学场景核心差异维度，分为实验组（16 人）与对照组（16 人），两组在初始知识水平、学习动机、军兵种分布、岗位职务占比等关键指标上无显著差异（ $P > 0.05$ ）。

表 2 系统使用人员分布表

分组	人数	军兵种分布	岗位职务分布	初始知识水平 (平均得分, 满分 100)
实验组	16	陆军 4、海军 4、空军 4、火箭军 4	初级岗位 9、中级岗位 5、高级岗位 2	72.3±5.6
对照组	16	陆军 4、海军 4、空军 4、火箭军 4	初级岗位 9、中级岗位 5、高级岗位 2	71.8±5.8

注:初始知识水平通过军事专业基础测试获取。

实验周期覆盖“基础理论学习—应用实践训练—综合考核评估”完整教学阶段。

首先,完成实验对象基础数据采集,对实验组进行系统操作培训,确保学员掌握资源浏览、反馈提交等功能;对照组采用传统教学模式,由教员统一发放学习资料,无智能推送服务。实验组使用教学训练系统,系统按“每日推送 5~8 个核心资源”的频率,基于学员行为数据(浏览记录、停留时间)与角色特征(军兵种、岗位)动态推送课程、练习、扩展资料;对照组按传统教学计划学习,教员每周布置 1 次学习任务,提供固定参考资料。

学员根据自身兴趣,自主选择相关的兴趣内容进行详情浏览和查看,对不感兴趣的内容可以通过页面进行关闭,如图 6 所示。

在中期测试阶段,开展中期测试与主观反馈调查,中期测试侧重基础理论与应用技能,主观反馈采用结构化问卷(含系统易用性、资源匹配度等维度);对照组同步参与中期测试,不进行系统反馈调查。

最后在期末汇总考核中,完成综合考核,收集两组学员的最终成绩、学习行为数据(实验组浏览时长、点击率,对照组学习时长、资料查阅次数),并对实验组进行访谈调研。

4.2 实验结果分析

根据选择形成反馈,优化数据匹配分析算法和智能

推送的知识结构和知识内容,从而不断提高推送的准确率和匹配度。

为实现多维度评估,数据采集涵盖“行为数据—学习效果数据—主观反馈数据”三类指标,具体如表 3 所示。

表 3 指标数据表

数据类型	指标名称	采集方法
行为数据	实验组:资源浏览时长(分钟/周)、资源点击率(%) ; 对照组:学习时长(分钟/周)、资料查阅次数(次/周)	实验组通过系统日志自动统计(精确到秒),对照组通过《记录表》记录
学习效果数据	基础理论成绩(满分 100)、应用实践成绩(满分 100)、综合考核成绩(满分 100)、知识掌握稳定性(综合考核与初始成绩差值)	基础理论成绩通过线上闭卷测试获取;应用实践成绩通过专家独立评分评估;综合考核由院校统一组织考试
主观反馈数据	系统易用性、资源匹配、学习效率提升感知、改进建议	实验组填写《主观评价问卷》,并进行访谈

4.2.1 行为数据对比分析

整体行为趋势:实验组平均浏览时长从 25 分钟/天提升至第 42 分钟/天,点击率 38.5% 提升至 72.3%,呈现持续增长趋势;对照组平均学习时长稳定在 30~35 分钟/天,资料查阅次数维持在 2~3 次/周,无显著变化。分类群体(不同军兵种或岗位)行为差异不显著,只在具体学习内容上有业务层面的差距。

4.2.2 学习效果数据对比分析

实验组在综合考核中的平均成绩高于对照,且知识掌握稳定性更优,中期测试成绩差距不明显。分类群体(不同军兵种或岗位)学习效果差异不显著,如表 4 所示。

表 4 学习效果数据表

分组	中期测试	综合考核	知识掌握稳定性
实验组	81.5±4.2	83.2±4.5	10.9±3.2
对照组	80.7±4.8	76.5±5.1	7.7±3.5

4.2.3 主观反馈数据分析

85.6%的学员认为“系统推送的资源能直接支撑岗

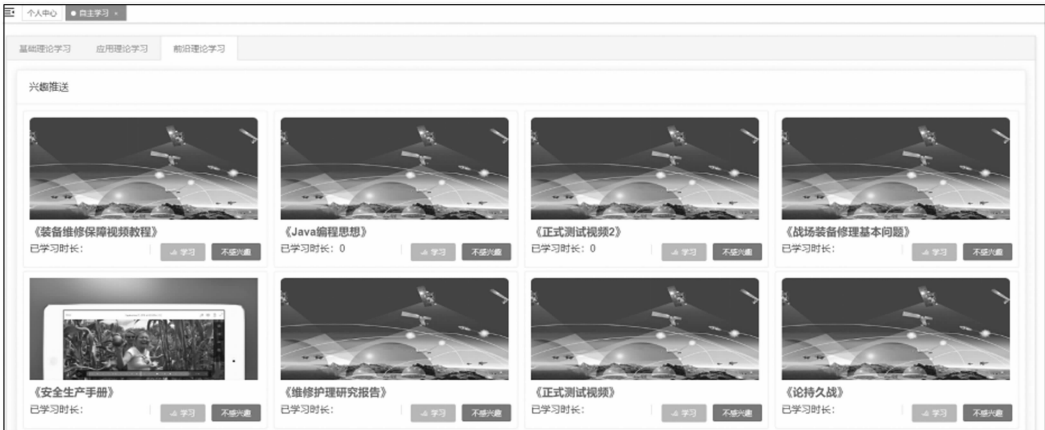


图 6 学员进行推送内容选择界面

位任务”，78.3%的学员表示“使用系统后学习时间更高效，减少了无效资料查找”。

4.3 实验结论

通过对学员的行为数据、学习效果数据、主观反馈数据分析可以看出，系统起始部署应用时，学员的推送内容的浏览时长较低，感兴趣程度并不高，学员的用户粘性较低。随着学员对系统使用的时间推迟延长，点击率和浏览时长均呈现明显增加的趋势，教学训练系统能显著提升学员的学习行为活跃度（浏览时长、点击率增长）与学习效果（理论、实践成绩提升），且知识掌握稳定性更优，相比传统教学模式具有明显优势。验证了本系统在设计实现中所应用的数据关联和决策算法的有效性。

5 结束语

基于智能推送的教学训练系统的设计实现，紧贴军事院校在校学员自主学习和训练的实际需求，采用基于数据挖掘、特征提取、关联分析匹配推送等智能化技术的智能推送方法，根据学员的浏览数据和系统使用情况，成功地为学员高效、准确地推送了相关的学习资源和知识资源，提高了学员的学习训练效率。

参考文献：

[1] 尹志. 专业教学与人工智能技术创新融合的路径探究[J]. 汽车维修与修理, 2022 (14): 16-18.

[2] 李达. 基于自适应特征提取网络的复杂环境下人脸识别[J]. 计算机测量与控制, 2024, 32 (8): 265-271.

[3] 秦垵忻, 王伟昕, 王砚生. 基于改进 MFCC 特征提取和 DNN 网络的机器人语音识别方法研究[J]. 计算机测量与控制, 2025, 33 (2): 246-253.

[4] 陈浩, 阎俏, 张桂青, 等. 基于模态分解与特征匹配的串联故障电弧识别方法研究[J]. 计算机测量与控制, 2021, 29 (11): 53-60.

[5] 陈方杰, 韩军, 王祖武. 基于改进网格划分统计的特征点快速匹配方法[J]. 计算机测量与控制, 2019, 27 (8): 231-235.

[6] 陆佳伟, 卢成炳, 王辰昊, 等. 基于 USDR 模型的云推荐方法研究[J]. 计算机测量与控制, 2018, 26 (8): 227-232.

[7] 刘练, 周凤星. 基于 APP 的智能家居环境监测系统的设计与实现[J]. 计算机测量与控制, 2014, 22 (7): 2018-2020.

[8] 何慧, 闫白洋. 智能系统常态化融入传统课堂教学的探索与实践[J]. 生物学教学, 2024, 49 (9): 39-42.

[9] 张馨予. 基于机器学习的车用主动防撞预警雷达信号识别系统研究[J]. 计算机测量与控制, 2020, 28 (7): 225-229.

[10] 戴天宇. 一种基于人机交互技术的新型教学系统的设计与实现[D]. 北京: 北京邮电大学, 2017.

[11] 彭艺, 宋浩, 苏黎犇, 等. 云环境下智能推送服务在数字化教学中的应用研究[C] //Information Engineering Research Institute, USA. Proceedings of 2013 2nd International Conference on Social Science and Education (ICSSE 2013), 昆明理工大学信息工程与自动化学院, 2013: 6.

[12] 林木辉, 张杰, 包正委. 智能教学系统中基于本体的知识表示及推送研究[J]. 福建师范大学学报(自然科学版), 2009, 25 (1): 120-124.

[13] 朱望纯, 曹健, 宋辉. 基于物联网的农场信息接收与发布平台设计与实现[J]. 计算机测量与控制, 2015, 23 (7): 2549-2552.

[14] 孙瑜. 基于云数据中心的多源异构数据治理技术研究[J]. 计算机测量与控制, 2024, 32 (3): 286-292.

[15] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17 (6): 734-749.

[16] GUYON I, ELISSEEFF A. An introduction to variable and feature selection[J]. Journal of Machine Learning Research, 2003 (3): 1157-1182.

[17] 杨吉才. 基于 ID3 决策树分类模型的自动评估学生技能水平的算法[J]. 无线互联科技, 2024, 21 (18): 17-22.

[18] 张璇, 左敏. 一种改进的朴素贝叶斯分类器在文本分类中的应用研究[J]. 北京工商大学学报(自然科学版), 2009, 27 (4): 52-55.

[19] 汪桂金. 随机森林算法的优化改进及其并行化研究[D]. 南昌: 南昌大学, 2019.

[20] 楚彬, 范东明. 改进的多元线性回归模型及其应用[J]. 测绘工程, 2014, 23 (6): 63-66.

[21] 杨静. 线性回归在经济预测中的应用[J]. 科技信息, 2012 (30): 276-277.

[22] 包勇. 基于遗传算法和改进兴趣度的关联规则挖掘算法[D]. 南宁: 广西师范学院, 2016.

[23] 于之虹, 黄彦浩, 鲁广明, 等. 基于时间序列关联分析的稳定运行规则提取方法[J]. 中国电机工程学报, 2015, 35 (3): 519-526.

[24] 才婷. 基于多源数据关联分析的关键技术应用研究[D]. 成都: 电子科技大学, 2020.

[25] 王克勤, 梁孟孟, 李靖, 等. 面向知识推送的设计情境建模及推理[J]. 机械科学与技术, 2019, 38 (11): 1654-1662.

[26] 胡嘉航. 基于用户画像的信息智能推送方法[J]. 无线互联科技, 2020, 17 (19): 161-162.