

智能化时代背景下中小学校人工智能教育水平评价体系初探*

纪淑娟^{1*} 程艳林^{1,2} 左东晓¹

山东科技大学计算机科学与工程学院, 青岛 266000

山东省平度市蓼兰镇宅科小学, 青岛 266700

摘要 针对各中小学校的人工智能教育水平参差不齐、缺乏科学合理的评价标准问题, 建立一套指标体系, 评价各学校的人工智能教育水平。首先, 运用 LDA 主题识别方法从大量人工智能教育相关政策文件中提取主题词, 并结合文献研究法初选了中小学校级人工智能教育评价指标; 其次, 运用德尔菲法对评价指标进行筛选和修订; 第三, 运用一种改进的 AHP 方法对各指标进行权重赋值; 最后, 以青岛市四十二所学校作为实证研究对象, 验证该评价体系的有效性。

关键字 中小学人工智能教育, 评价体系, LDA 主题模型, 改进的 AHP 方法, 模糊综合评价

Evaluation System of Artificial Intelligence Education Level in Primary and Secondary Schools Under the Background of Intelligent Age

Ji Shujuan¹ Cheng Yanlin^{1,2} Zuo Dongxiao¹

1. College of Computer Science and Engineering, Shandong University of Science and Technology
Qingdao, 266590, China;

2. Zhaike primary school in Liaolan town, Pingdu, Shandong Province, 266700, China;

Abstract---In view of the uneven level of artificial intelligence education in primary and secondary schools and the lack of scientific and reasonable evaluation criteria, a set of index system is established to evaluate the level of artificial intelligence education in schools. Firstly, the LDA topic recognition method is used to extract the subject words from a large number of policy documents related to AI education, and the evaluation index of AI education at primary and secondary school level is selected based on literature research method. Secondly, the Delphi method is used to screen and revise the evaluation index. Thirdly, an improved AHP method is used to assign weights to each index. Finally, forty-two primary schools in Qingdao are taken as empirical research objects to verify the effectiveness of the evaluation system.

Keywords---Artificial intelligence education in primary and secondary schools; Evaluation system; LDA topic model; Improved AHP method; Fuzzy comprehensive evaluation

1 引言

近年来, 人工智能技术逐渐兴起, 对教育、医疗、互联网等多个行业产生了深远的影响。作为互联网技术之后的新一代“通用性技术”, 人工智能将对人们的思维方式和生活习惯带来巨大的改变。我国在 2017 年发布了《新一代人工智能发展规划》^[1], 并明确指出“要逐步开展全民智能教育项目, 在中小学阶段设置人工智能相关课程, 逐步推广编程教育, 建设人工智能学科, 形成我国人工智能人才高地”。

在中小学实施人工智能教育的过程中, 学校是实施人工智能教育的基本单元, 学校的设施配置、教学管理、师资配备等成为影响人工智能教育的重要因素。目前, 我国中小学人工智能教育处于起步阶段, 缺乏统一的评价标准, 各学校在课程体系、课程教学、教师素质、硬件配置等方面存在很大的差距。长远来看, 这种情况不利于国家层面人工智能教育水平的提升。

在这一背景下, 本文旨在构建一套科学、合理的评价体系, 从学校层面综合评估中小学的人工智能教育水平, 为精准把控各学校人工智能教育的优势与不足, 提供有针对性的反馈与建议, 从而推动人工智能教育的发展。论文其他部分组织如下。第 2 节详细介绍

* 基金资助: 本文得到青岛市社科规划研究项目(QDSKL2301131)支持。

介绍了评价指标体系的构建过程。第3节对评价指标体系的有效性进行了实证分析,第4节总结了全文。

2 中小学人工智能教育评价体系构建

2.1 评价指标体系初步构建

(1) 基于 LDA 模型的人工智能教育政策主题词提取

本文基于大量中小学人工智能教育的政策文件^[2,3]中获得人工智能教育的主题词,从中提取影响因素。在采集数据时,选取国务院、教育部、山东省教体局、济南市教体局、青岛市教体局、淄博市教体局和临沂市教体局七个单位的官方网站作为数据采集平台。

本文爬取了以上七个网站自“2017年1月1日”到“2022年12月30日”以“人工智能”“人工智能教育”“中小学人工智能教育”为关键词政策文件,随后对爬取的数据进行数据的清洗,去除无关及重复数据,共得到337条有效数据,运用Python的Jieba库对文本进行分词和去除停用词的处理,形成最终的语料库。之后运用LDA主题识别技术从以上大量文件中进行主题识别,依据困惑度曲线确定最优主题数,运行Python语言中的Latent Dirichlet Allocation

(LDA)库进行主题聚类,最终得到主题词如表1(左侧两列)所示。

(2) 一二级评价指标初步选取

通过LDA主题识别已经得到了中小学人工智能教育相关的关键词,基本了解了中小学人工智能教育有关的政策导向和要求,但要将这些关键词进一步纳入到评价体系内,就必须结合人工智能教育实际,将一系列关键词转换为可以量化的综合衡量指标。本文邀请多位一线的中小学人工智能教师进行讨论归纳,最终将LDA主题识别关键词归纳为一系列可以量化的衡量指标,对应关系见表1(第三列)。

依据以上分析结果,我们初步选取课程开设、课程体系、学科融合、教学资源、教学设计、教学过程、教学评价、学生能力提升、教师配备、教师基础素养、教师专业能力、教师教研与培训、教学硬件配置、场馆和设施建设、人工智能教育特色、人工智能教学管理共16个二级指标。

结合LDA主题识别的主题词,依据中小学人工智能教育实际,对以上二级指标进一步归纳总结,初步确定课程设置、课堂教学、师资水平、硬件设施建设、教育特色与管理等五个一级指标。

表1 LDA主题词与人工智能教育衡量指标对应关系

主题	主题词 (Top15)	人工智能教育量化衡量指标
Topic #0	评价 融合 内容 多元 机制 体系 导向 学科 能力 过程 资源 培养 评估 资源 模式	课程开设 (机制、导向) 课程体系 (内容、资源、培养) 学科融合 (融合、多元、体系) 教学资源 (资源) 教学设计 (目标)
Topic #1	课堂 过程 项目 核心 基础 目标 评价 合作 探究 思维 内容技能 核心 信息技术 方式	教学过程 (课堂、过程、技能、内容) 教学评价 (评价、项目) 学生能力提升 (基础、合作、探究、思维)
Topic #2	人才 素养 教学能力 体系 优质 研究 名师 评估 培训 态度 服务 教育 资源 平台 优质	教师配备 (人才、优质、名师) 教师基础素养 (素养、态度) 教师专业能力 (教学能力、服务) 教师教研与培训 (研究、培训、教育资源、平台)
Topic #3	系统 能力 个性化 创客 实验室 平台 信息化 技术 机器人 智能环境 模式 信息 方式 系统 数字化	教学硬件配置 (创客、个性化、机器人) 场馆和设施建设 (系统、实验室、平台、信息化、技术、数字化)
Topic #4	试点 体验 师资 个性化 教材 培育 中心 方式 研究 模式 课程体系 校本 人才 政策	人工智能教育特色 (试点、体验、个性化、校本) 人工智能教学管理 (培育、模式、政策)

(3) 三级评价指标初步选取

在中小学人工智能评价体系的构建过程中,希望对底层指标进行细化,选取可执行性强的具体指标,用LDA主题识别的政策文件关键词一般具有导向性、

引领性,但是不够具体。因此在三级指标的选取过程中,在主题识别归纳出的二级指标的基础上,需要运用文献分析法,参考相关的文献和具体政策文件,选取可执行性强的细化指标。在政策文件指标的选取过程中,采取对政策文本进行总结归纳的方式获得指标。

由此,在LDA主题识别的基础上,结合文献研究方法,完成了评价指标体系的初步构建,其中包含5个一级指标、16个二级指标,64个三级指标。

2.2 评价指标体系的筛选和修订

与文献[4]类似,本文也选择德尔菲法进行指标的筛选。这是因为德尔菲法是一种通过问卷形式征求专家意见的研究方式,具有匿名性、反馈性、可统计性的特点。本文选取17位专家组成专家小组运用德尔菲法对初选指标进行修订,专家小组成员包括高校从事相关研究的教授、长期开展人工智能教研活动的教研员、各学校人工智能教育的负责人和长期从事人工智能教育的一线教师,他们人工智能相关政策、理论能够较好地把握,且熟悉学校层面的人工智能教育环境,具有丰富的人工智能教育经验,对人工智能教育的评价指标的筛选和修订具有较强的现实意义。

在设计指标打分问卷时,考虑到初拟指标分为一级指标、二级指标和三级指标,指标较多,如果糅合在一份问卷中打分较为复杂,所以采用逐步打分的方式。每轮打分均由上而下设计了三份打分问卷,首先各位专家对一级指标进行打分,根据一级指标打分结果设计二级指标问卷并邀请专家进行打分,并依据打分结果设计三级指标打分问卷并邀请专家进行打分。完成一轮专家咨询后依据打分结果对指标进行修订,重新设计各级问卷并邀请专家打分,通过多轮打分直到所有专家对所有指标无异议。在打分过程中,将各级指标以问卷形式发送给各位专家,问卷通过电子邮件进行匿名回收。问卷中各项指标采用5分制进行评分,分值越高表示该项保留价值越大,最终通过统计每项指标的平均分和变异系数来判定专家的意见,如果某项指标平均分 M 低于3.0分,且变异系数 C 高于0.3,说明专家认为该指标重要程度较低,且专家分歧较大,该指标删除,否则保留;如果某专家认为某项指标非常不合理并给出理由,则在下一轮将该建议纳入问卷,直到多位专家达成共识。

(1) 第一轮专家意见咨询

第一轮咨询发放专家问卷17份,收回有效问卷16份,有效率为94.12%。经过整理各级指标专家的打分意见,对指标体系进行了以下修订:

① 五个一级指标的平均分均高于3.0分且变异系数均小于0.3,没有专家反馈某指标“非常不合理”,因此五个指标全部保留。

② 二级指标中的B5项“学生能力提升”平均分为2.625低于3分,变异系数为0.349高于0.3,且2号专家认为“B5项指标在评价过程中难以量化统计,

非常不合理”,因此将该指标删除,该指标下的三级指标不再纳入三级指标打分问卷。

③ 三级指标中的C37项“学习能力”、C44项“参与县级以上教研的人数比例”、D13项“物联网套件种类和数量”、D23项“是否连接超算中心”四项指标的打分平均值均低于3.0且变异系数大于0.3,故予以删除,不再列入下一轮打分问卷。

④ 5号专家提出“课堂教学”下的B35-B38“项目式学习效果”“小组合作学习效果”“是否鼓励学生发现问题”“是否进行探究性学习”四项作为学校级别的评价指标过于细化且难以量化,应将其归入B33项“课堂组织情况”,不应单独作为指标进行评价,将该建议纳入下一轮问卷再次征求其他专家意见。

(2) 第二轮专家意见咨询

根据第一轮修订后的中小学人工智能教育评价体系,本文编写了第二轮咨询问卷,以电子邮件的形式向16位专家发放问卷,回收有效问卷16份,有效率为100%。通过回收整理各专家意见,第二轮指标修订情况如下:

针对第一轮打分过程中5号专家提出“课堂教学”下的B35-B38“项目式学习效果”“小组合作学习效果”“是否鼓励学生发现问题”“是否进行探究性学习”四项指标应纳入B33项“课堂组织情况”的一个方面,不应单独作为学校评价指标问题,在问卷中进行描述,如果专家同意该建议则选择“通过”,不同意则选择“不通过”,最终有12位专家同意,通过率为75%较高,因此采纳该意见。将B35-B38“项目式学习效果”“小组合作学习效果”“是否鼓励学生发现问题”“是否进行探究性学习”四项指标删除,将其作为B33项“课堂组织情况”的评价标准。②其余各级指标的得分平均值均高于3.0且变异系数均低于0.3,没有专家就某项指标提出异议,故予以保留。

(3) 第三轮专家意见咨询

依据一二轮专家咨询结果设计问卷,进行第三轮问卷打分。统计结果发现,没有专家针对某项指标提出反对意见,所有指标的平均分都大于3.0分,且变异系数均低于0.3,所有专家的意见趋于一致,专家咨询结束。由此可以确立中小学人工智能教育评价指标体系。

中小学人工智能教育评价指标体系共包含“课程设置”“课堂教学”等共5项一级指标,“课程设置”“课程体系”“学科融合”等共15项二级指标,“人工智能课程开设年级”“人工智能课程月开课节数”等共50个三级指标,如表2所示。

表 2 中小学人工智能教学评价体系权重表

一级指标	权重	二级指标	权重	三级指标	权重				
A 课程设置	0.1713	A1 课程开设	0.0254	A11 人工智能课程开设年级	0.0068				
				A12 人工智能月课时安排	0.0186				
		A2 课程体系	0.0964	A3 学科融合	0.0495	A21 课程目标准确性	0.0127		
						A22 课程内容合理性	0.0127		
						A23 课程结构连续性	0.0442		
						A24 课程资源配置有效性	0.0268		
		B1 教学资源	0.0638	B2 教学设计	0.0387	A31 与人工智能融合的学科数目	0.0133		
						A32 学科融合方式是否合理	0.0362		
						B11 教材选择的合理性	0.0290		
						B12 是否配备教学平台	0.0065		
B13 教学硬件的配置是否完整	0.0107								
B14 多媒体资源是否丰富	0.0176								
B 课堂教学	0.3813	B3 教学过程	0.1735	B21 知识与技能	0.0112				
				B22 过程与方法	0.0057				
				B23 情感态度与价值观	0.0218				
				B31 内容安排合理性	0.0297				
		B4 教学评价	0.1052	C1 教师配备	0.0557	B32 教师讲解情况	0.0629		
						B33 课堂组织情况	0.0629		
		C 师资水平	0.2556	C2 教师基本素养	0.0263	B34 作品展示情况	0.0180		
						B41 评价方式	0.0304		
				C3 教师专业素养	0.1179	C4 教师培训与教研	0.0557	B42 评价内容	0.0592
								B43 阶段性评价情况	0.0156
C11 人工智能专职教师人数	0.0313								
C12 其他学科兼职教师人数	0.0161								
C13 人工智能教学负责人情况	0.0083								
C21 毕业专业与人工智能相关专业的教师比例	0.0039								
D 硬件设施建设	0.1148	D1 教学硬件配置	0.0839	C22 职称级别在中级以上的教师比例	0.0076				
				C23 人工智能教师平均教龄	0.0148				
				C31 专业知识和技能	0.0104				
				C32 教学态度	0.0333				
		D2 场馆和设施建设	0.0309	E1 人工智能教育特色	0.0207	C33 信息化技术运用能力	0.0333		
						C34 教学设计能力	0.0104		
		E 教育特色与管理	0.0770	E2 人工智能教学管理	0.0563	C35 课堂组织能力	0.0202		
						C36 教学评价能力	0.0104		
						C41 最近一年参加培训的教师比例	0.0253		
						C42 最近一年参加培训的总时长	0.0154		
E 教育特色与管理	0.0770	E2 人工智能教学管理	0.0563	C43 参加市级以上培训的教师比例	0.0057				
				C44 学校内部教研频率	0.0093				
				D11 开源硬件种类和数量	0.0226				
				D12 机器人种类和数量	0.0613				
				D21 实验室规模及数量	0.0089				
				D22 学生微机室规模及数量	0.0046				
				D23 校园网主干带宽值	0.0174				
				E11 社团开设情况	0.0242				
E12 相关竞赛成绩	0.0472								
E 教育特色与管理	0.0770	E2 人工智能教学管理	0.0563	E13 是否研发校本课程	0.0124				
				E21 人工智能教学理念	0.0113				
				E22 教学管理情况	0.0113				
				E23 硬件和设施使用管理情况	0.0042				
E 教育特色与管理	0.0770	E2 人工智能教学管理	0.0563	E24 教学评价机制	0.0042				

2.3 评价指标权重赋值

要提高评价实效,确定各指标权重是非常重要的。由于本文指标众多,为了保证权重赋值的科学性和准确性,与文献[4]不同,本文采用一种改进的AHP方法—基于最优一致性矩阵的层次分析法^[5-7]进行指标权重赋值。选择该方法是因为有利于专家打分时更准确地把握指标间的重要程度,省略了繁琐的一致性检验步骤,操作性强,有助于提高权重赋值过程中的科学性和准确性。应用该方法得到的各指标的权重系数如表2所示。

由该表可见,一级指标中课堂教学的权重系数最大($W=0.3813$),其次是师资水平($W=0.2556$),二级指标中教学过程($W=0.1735$)、教学评价($W=0.1052$)、教师专业素养($W=0.1179$)等的权重系数较大,说明以上方面在中小学人工智能教育过程中尤为重要。

3 中小学人工智能教育评价体系实证研究

为验证中小学人工智能教育情况评价体系的有效性和可行性,运用模糊综合评价的方法,对青岛市42所中小学的人工智能教育情况进行综合评价。这42所中小学校包括小学25所和初中17所,学校覆盖了青岛市的市南区、市北区、崂山区、李沧区、西海岸新区、城阳区、即墨区、胶州市、平度市、莱西市共10个区市。

3.1 实证研究过程

本文依据上文建立的中小学人工智能教育评价体系,每个学校邀请12-20名同行专家(包含人工智能教育教研员、学校人工智能教育负责人、经验丰富的人工智能教育一线教师等)对学校的人工智能教育情况进行评价,评定级别为优秀、良好、一般、较差、差五个等级,建立了包含五个要素的评语集,即 $V=\{V_1, V_2, V_3, V_4, V_5\}=\{\text{优秀, 良好, 一般, 较差, 差}\}$ 。依据专家评价结果,对各学校的数据进行分析计算。由于指标层级较多,采取自下而上的过程进行模糊综合计算,先依据打分结果计算各三级指标对于各二级指标的隶属度,通过模糊综合运算计算各二级指标的评价向量并据此形成二级指标对于一级指标的隶属度矩阵,再重复该过程计算各一级指标的评价向量及整体的综合评价向量,最后按照优秀(90分),良好(80分),一般(70分),较差(60分),差(50分)的标准对各评价向量进行赋分计算综合得分,可得出学校人工智能教育各一级指标得分和综合得分。

3.3 实证结果与分析

(1) 整体结果与分析

根据青岛市各学校模糊综合评价得分结果,绘制得分等级分布扇形图,如图1所示。由图1可以看出青岛市中小学人工智能教育整体处于中等偏上水平,其中处于“优秀”等级的学校有6所,约占学校总数的9%,处于“良好”等级的学校有18所,约占学校总数的43%,处于“一般”等级的学校有11所,约占学校总数的26%,处于“较差”等级的学校有5所,约占学校总数的12%,处于“差”等级的学校有4所,约占学校总数的10%。

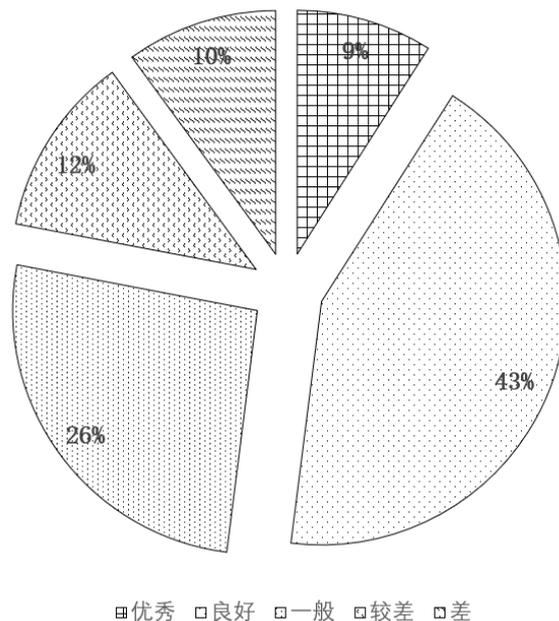


图1 青岛市各学校综合评价得分等级分布图

由此可见青岛市中小学人工智能教育水平整体上处于中等偏上水平,大部分学校综合评价结果处于“良好”和“一般”等级,整体发展趋势良好,但是处于“优秀”等级的学校较少,还有很多学校处于“较差”和“差”等级,因此青岛市中小学人工智能教育仍然存在一些问题,有一定的提升空间。

(2) 专项结果与分析

为了分析青岛市各学校在课程设置、课堂教学、师资水平、硬件设施配置、管理与特色等不同方面的得分等级情况,在一级指标层面将各学校等级分布情况绘制堆积柱状图,如图2所示。由图中可以看出各学校在课程设置和课堂教学方面整体表现较好,“优秀”和“良好”的学校相对较多,而“较差”和“差”的学校相对较少。管理与特色方面虽然“优秀”的学校最多有7所,但是处于“较差”和“差”的学校也较多,分别是4所和6所,因此管理与特色方面两级分化最严重,应重点进行“托底”。

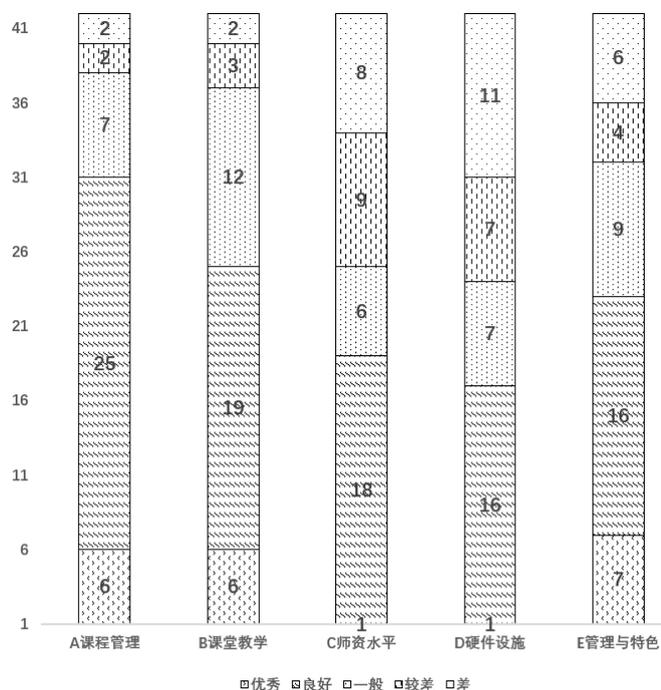


图2 一级指标得分等级分布对比图

由图3可以看出，青岛市各学校人工智能教育的师资水平和硬件设施配置整体情况较差，主要表现在处于“优秀”和“良好”等级的学校相对较少，而处于“较差”和“差”的学校众多，其中硬件设施方面处于“差”级的学校有11所，比重高达26.19%，足可以看出这一方面非常薄弱，因此在后面的研究中，师资水平和硬件设施是需要重点关注和改善的两个方面。

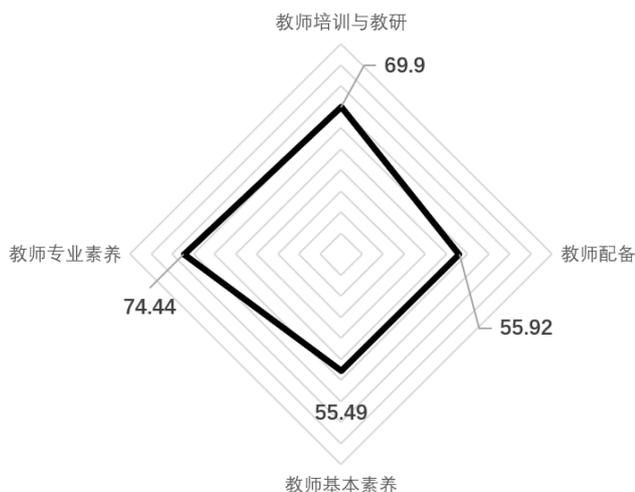


图3 师资水平下各指标得分雷达图

为了清晰地了解师资水平和硬件设施方面存在的具体问题，对二级指标得分求均值并进行对比分析，从而发现需要重点改进的短板和弱项。

在师资水平方面，将教师配备、教师基本素养、教师专业素养、教师培训与教研等四方面的得分绘制

成雷达图，如图3所示。由图可见，教师专业素养方面的得分最高，教师培训与教研方面的得分较高，说明在青岛市中小学的人工智能教育过程中，组织了一定数量的质量较高的教研培训活动，且效果较好，在一定程度上提升了教师的专业素养。但是由图中也不难发现，教师配备和教师基本素养得分明显低于另外两项，且得分差距很大，说明青岛市各中小学在人工智能教师配备方面存在较大问题，且教师基本素养不高，这两方面属于师资水平方面的短板，亟需重点加强。

通过对青岛市各学校模糊综合评价得分进行描述性统计分析，可以发现青岛市中小学人工智能教育在课堂教学、课程设置等方面得分较高，表现较好，但是在师资水平和硬件设施等方面存在短板，具体问题体现在以下方面：

- ① 人工智能教师配备情况不佳，各学校缺乏专业的人工智能师资配备。
- ② 人工智能教师基本素养不达标，难以满足人工智能教学的能力要求。
- ③ 各学校场馆和设施建设情况难以满足人工智能教育的要求，需加强配套设施建设。

4 结束语

在中小学普及人工智能教育的背景下，本文首先运用LDA主题识别的方法从大量的中小学人工智能教育相关的政策文件中提取主题词，从这些主题词中了解中小学人工智能教育的宏观导向。其次，在提取得到的主题词基础上结合文献研究法初步构建了中小学人工智能教育评价指标体系，并运用德尔菲法对指标体系进行筛选和修订；第三，运用一种改进的AHP方法对指标进行权重赋值，建立了一套中小学人工智能教育评价体系。第四，运用模糊综合评价的方法对评价体系的性能进行了实证分析。

该体系的建立，一方面可以为起步阶段各学校的人工智能教育实施提供指导，各学校可以依据此评价体系合理规划自身的人工智能教育，并通过量化的评价确保自身的人工智能教育平稳有序发展；另一方面丰富了人工智能教育相关的理论研究，为后续的研究提供了参考。

未来，可在本文的基础上进行以下几方面的研究：首先，对评价指标进一步完善，结合人工智能教育实际对指标进行进一步调整，使其更加科学合理；其次，可以扩大实证研究范围，选取更多学校进行人工智能教育评价，依据评价结果发现各学校在人工智能教育过程中存在的问题，并提出针对性的建议，以期促进各学校、各地区人工智能教育水平的提升；第三，进

行人工智能教育评价的分阶段研究,未来可以针对不同学段进行分层研究。例如,分别针对小学低年级、中年级、高年级和初中进行人工智能教育评价研究,建立个性化的评价体系。第四,未来可以进一步探索中小学人工智能教育与跟更高等教育的联系,使得出的评价体系更符合国家发展的需要。

参 考 文 献

- [1] 国务院关于印发《新一代人工智能发展规划》的通知[EB/OL]. (2017-07-20)[2023-01-30]. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.
- [2] 山东省教育厅关于印发《山东省教育信息化2.0行动计划(2019—2022)》的通知[EB/OL]. (2019-08-19)[2023-01-30]. http://edu.shandong.gov.cn/art/2019/8/19/art_11990_7739898.html.
- [3] 青岛市教育局关于印发《青岛市人工智能教育实施意见》的通知[EB/OL]. (2020-08-13)[2023-01-30]. http://edu.qingdao.gov.cn/zfxxgkml/ywflxx/zyjy/202203/t20220311_4715192.shtml.
- [4] 江祥奎,宁玉文.计算机类MOOC质量评价指标体系的设计与实证研究[J].计算机技术与教育学报,2023,11(2):1-6.
- [5] 赵中奇.带有自调节功能的层次分析[J].控制与决策,1992(01):53-57+71. DOI:10.13195/j.cd.1992.01.55.zhaozhq.010.
- [6] 任胜兵,冯迪,陈满男.基于最优一致性矩阵的灰色层次分析法研究[J].计算机工程与应用,2017,53(18):44-50.
- [7] 杜红珊,李洪杰,郑庆玉.不确定型AHP的最优传递矩阵法[J].曲阜师范大学学报(自然科学版),1997(04):39-42.