基于人工智能的教育质量评估系统研究

Research on Education Quality Assessment System Based on Artificial Intelligence

吴浩然[1]，袁　涛[2]，吴少平[3]，李亚梦[1]（1. 中国联通网络技术研究院，北京 100048；2. 中讯邮电咨询设计院有限公司广东分公司，广东 广州 510627；3. 华南师范大学，广东 广州 510631）
Wu Haoran[1]，Yuan Tao[2]，Wu Shaoping[3]，Li Yameng[1]（1. China Unicom Network Technology Research Institute，Beijing 100048，China；2. China Information Technology Designing & Consulting Institute Co.，Ltd. Guangdong Branch，Guangzhou 510627，China；3. South China Normal University，Guangzhou 510631，China）

摘　要：
科学客观的教育质量评估是当前教育行业的重要需求。人工智能赋能各行各业已成为未来社会发展的必然趋势。介绍了一种人工智能用于教育质量评估的方案。教育质量评估系统结合了大数据和人工智能技术，该系统利用各类人工智能算法对收集到的与教师教学质量有关的图像、声音、文字数据进行分析，并生成客观可视结果用于教育质量评估。

关键词：
人工智能；计算机视觉；教育质量评估
doi：10.12045/j.issn.1007-3043.2018.12.018
中图分类号：TP181
文献标识码：A
文章编号：1007-3043（2018）12-0083-06

Abstract：
Scientific and objective evaluation of educational quality is an important demand of the current education industry. Artificial intelligence enabling all walks of life has become an inevitable trend of future social development. The scheme of artificial intelligence for education quality assessment is introduced. The education quality assessment system combines big data with artificial intelligence technology. It uses various artificial intelligence algorithms to analyze the collected image，sound and text data related to the teacher's teaching quality，and it can generate objective visual results for educational quality assessment.

Keywords：
Artificial intelligence；Computer vision；Educational quality assessment

引用格式：吴浩然，袁涛，吴少平，等. 基于人工智能的教育质量评估系统研究[J]. 邮电设计技术，2018（12）：83-88.

1 研究背景

近年来，随着深度学习的兴起，人工智能又迎来了一个新的发展热潮。在这一新热潮的背景下，计算机视觉、语音识别、自然语言处理等领域技术日趋成熟，许多新产品不断出现并应用到了社会生产生活的各个领域。而人工智能与各生产生活领域结合产生的良好经济效益及社会效益，也引起了社会和政府层面的高度重视，国务院先后在2017年及2018年的政府工作报告中提及了人工智能发展战略规划，并在2017年印发了《新一代人工智能发展规划》[1]，工信部也在同年发布了《促进新一代人工智能产业发展三年行动计划（2018-2020年）》[2]。

随着我国高等教育的快速发展，教师教育质量已成为各高校及各级教育主管部门关注的重点。教育部也结合国家教育事业发展“十三五”规划工作要求及国家“互联网+”、大数据、新一代人工智能等重大战略任务安排印发了《教育信息化2.0行动计划》[3]。

基于以上背景，本文将讨论如何将人工智能技术应用于教育行业，推动教育行业信息化，使用人工智能的方法对教育质量进行监测，提升教师教学能力和水平。

收稿日期：2018-11-02

2 智慧教育使用的相关技术

随着现代信息技术及人工智能技术的不断发展，教育信息化程度也在逐渐提高。智慧教育就是借用新型的互联网及人工智能技术，使教育数字化、网络化、智能化和多媒体化，呈现出开放、共享、交互、协作的特点。

2.1 计算机视觉技术

计算机视觉是一门研究如何使计算机像人类一样“看”的科学。这项技术将摄像机等图像捕获设备当作“计算机的眼睛”获取图像或视频信息，并利用计算机视觉算法将获取到的图像、视频等信息进行处理。计算机视觉对图像的处理通常分为检测和识别两大类，在每大类下又有不同的分支应用。ImageNet[4]挑战赛对推动计算机视觉领域发展做出了突出贡献，2012年该比赛的冠军模型AlexNet[5]将深度学习应用在了计算机视觉中，取得了突破性的成果。深度学习由此引发了业界的高度重视，它也被应用到包括语音识别、自然语言处理在内的其他领域并取得了良好的效果。

计算机视觉领域使用深度学习方法对图像或视频进行处理的流程如图1所示。

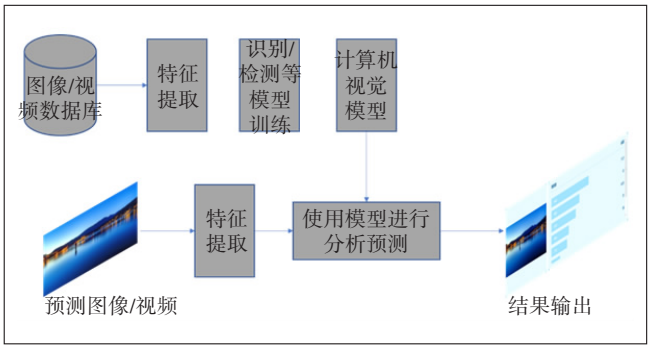

图1　图像处理的流程图

深度学习对图像或视频处理的流程主要包括特征提取、原始算法模型训练、预测图像特征提取和使用模型进行分析预测等步骤。

当前诸多人工智能产品均通过图形界面与用户进行交互，作为一项处理分析图像的技术，计算机视觉技术是人工智能领域必不可少的一项重要研究内容。在智慧教育应用中，计算机视觉技术能够对教室内教师授课图像进行分析处理，如行为分析、学生情感分析等。分析结果能够有效反映学生课堂注意力集中度，借此对教育质量进行评估。

2.2 语音识别技术

形象地讲，语音识别是一门让机器学会“听”的科学。这项技术将音频采集设备采集到的音频信号转为相应的文本或命令。

语音识别最早出现于20世纪50年代AT&T的贝尔实验室，他们的Audrey系统是第1个可以识别10个英文数字的语音识别系统。近年来随着人工智能尤其是深度学习技术的发展，语音识别技术也在不断实现突破。在特定场景下语音识别准确率已超过90%。目前语音识别技术也是人工智能领域落地最多、应用范围最广的技术。语音识别的主要流程如图2所示。

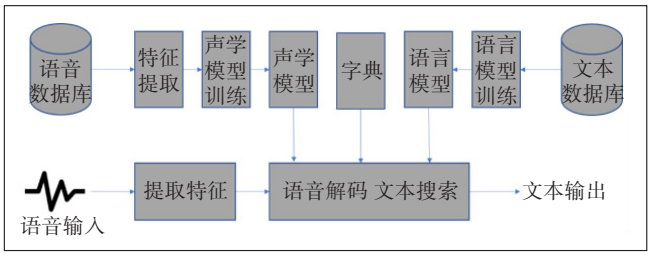

图2　语音识别流程图

语音识别流程相较图像处理多了字典及语言模型的训练，这2部分主要作用是在使用声学模型提取语音特征后，将语音特征解码转化为自然语言文本进行输出。

在智慧教育应用中，语音识别技术也大有可为。例如可以通过对教师授课时的语音进行识别分析判断教师授课过程中的文明教学程度或分析课堂上教师与学生的互动情况。

2.3 自然语言处理技术

自然语言处理是一门研究如何使用人类的自然语言与计算机进行有效交互的技术。最早的自然语言理解方面的工作是机器翻译。到了现代，自然语言处理主要基于机器学习，而深度学习出现后，一系列优秀的深度学习模型如RNN[6]、LSTM[7]被用于自然语言处理并取得了不错的效果。自然语言处理的主要流程如图3所示。

自然语言处理过程除了词向量编码外，其余与图像、语音处理流程类似。实际上，在处理不同形式数据时，最大的区别就在于数据预处理过程，即将不同类型数据处理为机器可进行特征提取、训练预测的数据形式。

自然语言处理技术不仅可以用于分析教师社交圈的同事或学生对教师的评价从而评估教师教学水

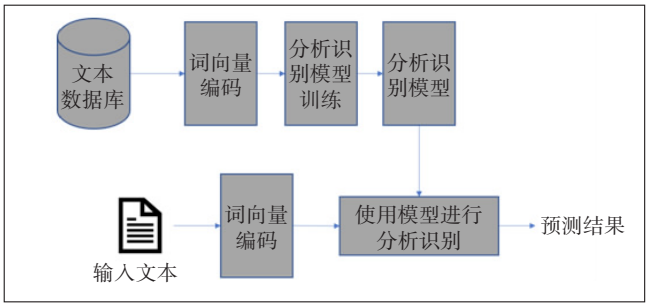

图3　自然语言处理流程图

平，还可用于各类辅助教师教学的应用，如作业辅助批改、辅助判卷等应用。

3 教育质量评估系统构建方案

3.1 教育质量评估系统构建背景

为培养高素质教师专业化创新型队伍，按照国家教育事业发展“十三五”规划工作要求，推进教师教育质量保障体系建设，提高师范类专业人才培养质量，教育部印发了《普通高等学校师范类专业认证实施办法（暂行）》，希望建立教师教育质量监测系统，并依托此系统，利用大数据信息及人工智能技术，进行师范类办学核心数据的监测、挖掘和分析，形成各类教育质量监测报告。最后将评价结果应用于教学改进，推动师范类专业人才培养质量的持续提升。

3.2 教育质量评估系统构建

教育质量评估系统架构如图4所示，整个平台分为3层，最下层的基础层将提供基础设备、数据资源及计算资源。基础设备包括网络设备、存储设备及终端设备。

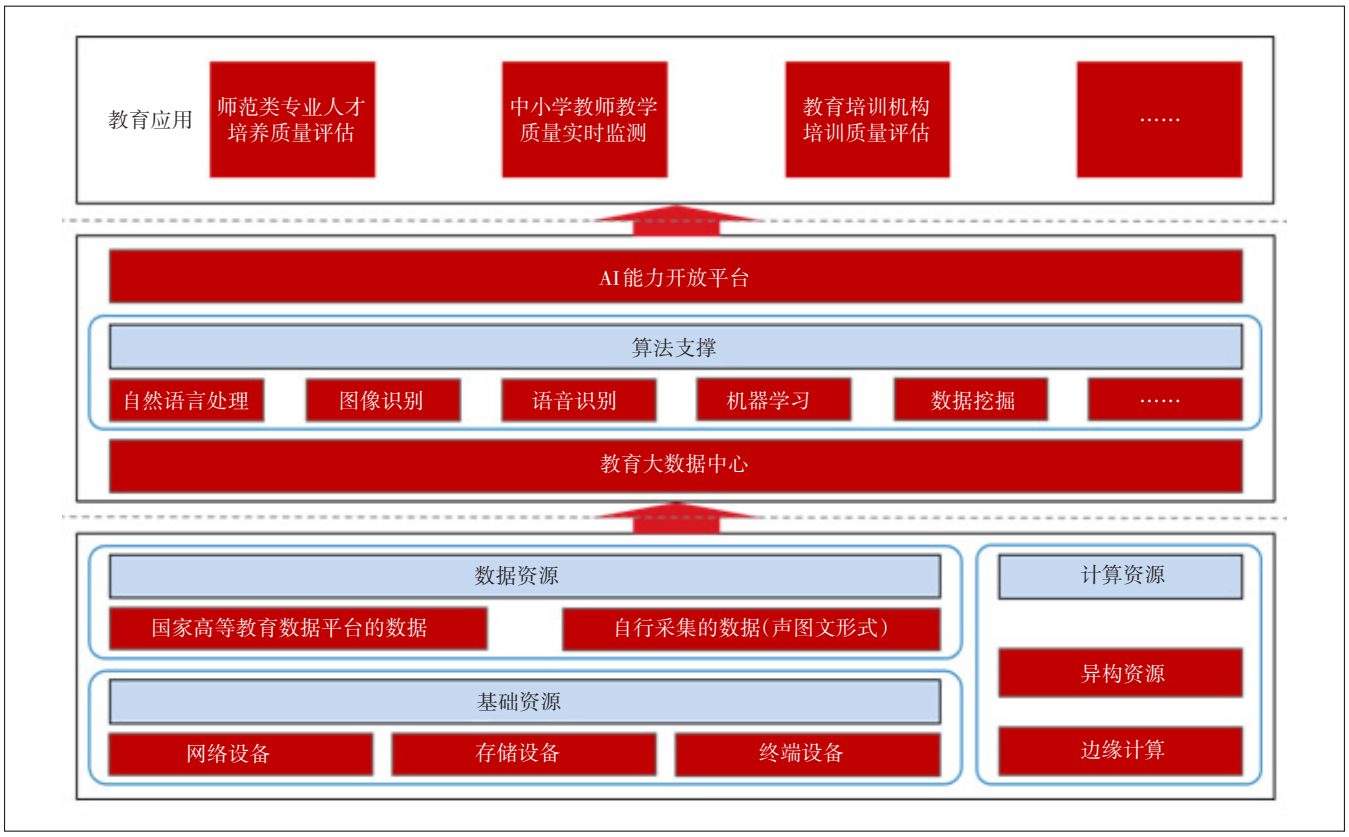

图4　教育质量评估系统架构

数据资源来源分为国家高等教育数据平台数据及自行采集数据2部分。这2类数据将存储于教育大数据中心内并用于人工智能算法的训练及教学质量的分析。其中自行采集的数据主要包括图像、声音、文字3类，如教师授课过程中的课程录像、教师授课过程中的语音信息、由学生或校内同事处通过问卷调查等方式采集的评价教师教学水平的文本信息。

异构计算和边缘计算等计算资源则可以解决人工智能算法应用所需要的计算力问题。

系统架构的核心层由教育大数据中心和AI能力开放平台2部分组成。大数据中心负责对数据进行清洗、整理并将数据用于上层AI算法训练及教师教学水

平分析。AI能力开放平台主要为最上层的应用提供自然语言处理、图像识别、语音识别、机器学习、数据挖掘等多种AI算法能力支持。

最上层的应用层将面向不同用户提供各类应用，如师范类专业人才培养质量评估应用、中小学教师教学质量实时监测应用、教育培训机构培训质量评估应用等。

3.3 教育大数据中心构建

教育大数据中心总体采用分层架构，其架构如图5所示。

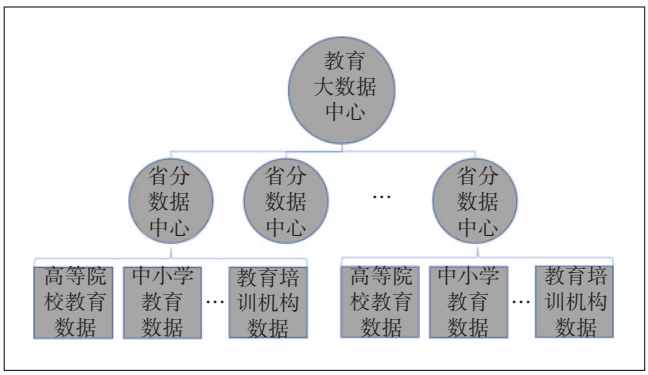

图5　教育大数据中心总体架构

大数据中心总体采用$M+N+1$的分层形式，教育大数据中心下辖N个省分数据中心，每个省都将建立1个省分数据中心用于收集存储本省的教育相关数据。每个省分数据中心将收集存储M类教育数据，包括中小学、高等院校、培训机构等不同类型的教育数据。所有数据由底层收集清洗[8]并不断向上整理汇总，最终存储于教育大数据中心。

数据中心可以通过API接口连接到国家高等教育数据平台，并从该平台获取高等教育数据。同时可以使用音频、视频采集设备对教师教学过程中的音频及视频进行离线或实时采集[9]。此外还可以使用网上调查问卷方式对教师教学水平进行调查，并通过调查问卷网站API接口获取到调查结果数据。同时还可采集教师的网络社交数据[10]。

数据采集后，平台会对不同类型的数据进行批处理[11]或实时处理[12]，批处理采用Hadoop或Spark框架，实时处理则使用Storm或Spark Streaming框架。

对数据进行整理清洗后，数据中心通过与AI开放能力平台之间的API接口将数据提供给AI开放能力平台用于AI算法的训练及教育质量的分析。

3.4 教育质量评估系统应用案例

教育质量评估系统可以利用图像、声音、文字等多种内容形式的数据评估教师的教学质量。本节将介绍一个以图像数据为基础的教育质量评估应用。

3.4.1 教室行为分析总体介绍

传统的教育质量评估方式多为专家进行课堂旁听再根据经验进行判断，这种方法缺少评估标准，导致不同评估人员做出的评估结果差异较大。并且专家仅能旁听到少量课程，难以全面了解教师的授课水平，教师若提前准备则评估结果可能与教师日常实际授课效果差异较大。此外，专家经验评估方式也存在流程繁琐、人工成本较大等问题。本文的方法利用人工智能算法，对教室内不同人员进行行为分析，并将行为分析结果进行统计分析，不仅能减少人工成本，且由于评测数据来源可以覆盖教师的每堂授课，其评测结果具有客观性。

由于教室内有教师及多名学生，因此教室行为分析可以分2个步骤：定位目标及对目标行为进行分析。这2个步骤均可通过计算机视觉及人工智能技术完成，分析过程完全自动无需人工干预。教室行为分析的流程图见图6。

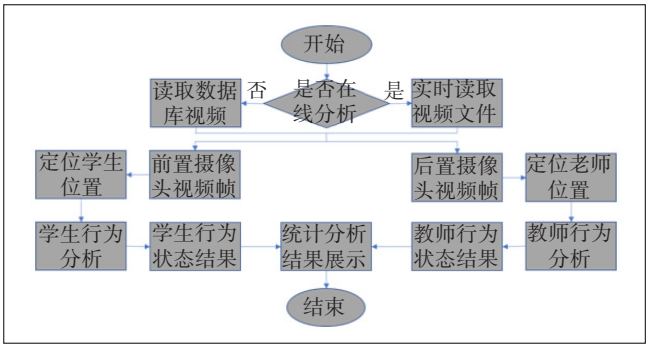

图6　教室行为分析流程图

教室行为分析可以使用在线视频文件或数据库中的视频文件。视频文件来自教室前后2个摄像头，前置摄像头用于采集学生数据，后置摄像头用于采集教师数据。通过目标检测算法，教室行为分析系统可以根据教师及学生特征差异自动识别人员身份，同时教室行为分析系统可以使用目标分析算法对不同人员的行为进行分析统计。教师的行为分为讲授、板书、指向、俯身、巡视5种，学生的行为包括举手、书写、阅读、讨论、听课、站立6种。且基于深度学习算法可迁移性强这一特点，如果专家认为有新的教师或学生行为对于分析教师教学质量有帮助，也可再训练微调模型，增加检测行为。

3.4.2 目标检测算法

要进行教室行为分析需要首先检测目标在图像中的位置。目前主流的深度学习目标检测算法有基于候选框及深度学习分类的方法和基于深度学习的回归方法2类，前者以Faster R-CNN[13]、R-FCN[14]为代表，后者以YOLO[15]、SSD[16]为代表。表1为目前主流目标检测算法在数据集COCO[17]上的平均精度均值（mAP）及检测速度。

表1　COCO数据集主流目标检测算法效果

模型	mAP/%	FPS
Faster R-CNN	42.7	5
R-FCN	51.9	12
YOLOv3 tiny	33.1	220
YOLOv3-320	51.5	45
YOLOv3-608	57.9	20
SSD300	41.2	46
SSD500	46.5	19

表1中YOLOv3及SSD后的数字为输入图像的大小，如SSD300表示使用长宽均为300像素的图像进行训练及测试。可以看到，随着使用图像大小的提升平均精度均值在逐步提升，但检测速度迅速下降。在COCO数据集上，基于候选框的目标检测算法Faster R-CNN及R-FCN在检测速度上都远小于其余基于深度学习回归的方法。监控视频帧数一般为20~25FPS，若以20FPS为标准，则YOLOv3-320，YOLOv3-608，YOLOv3-tiny，SSD300均达到了实时要求，且YOLOv3-608达到了最高的平均精度均值。若以25FPS为标准，则YOLOv3-608无法达到实时要求，此时能够进行实时检测且平均精度均值最高的算法为YOLOv3-320。表2为一些目标检测算法在VOC2007数据集的平均精度均值及检测速度。

表2　VOC2007数据集主流目标检测算法效果

模型	mAP/%	FPS
Faster R-CNN	73.2	7
Fast YOLO	52.7	155
YOLO（VGG16）	66.4	21
SSD300	74.3	46
SSD512	76.8	19

在VOC2007数据集上，若以20FPS作为实时检测的标准，则YOLO（VGG16）、SSD300、Fast YOLO达到了实时要求，其中Fast YOLO以155FPS的检测速度远远高于其他方法，但由于其过于追求速度，准确度下降明显，其平均精度均值仅为52.7%远低于其他模型。SSD300则在保证实时检测的前提下达到了最高的检测精度，其平均精度均值为74.3%，仅比最高的SSD512低2.5%。若不追求实时性，而是采用离线评估分析方式，则选用SSD512最佳。

从COCO及VOC2007 2个数据集的结果可以看出，不同算法在不同数据集上的效果也略有差异。因此，在选取目标检测算法时，应在实际数据基础上，根据需求（如是否有实时需求，平均精度均值需达到什么程度等）来选取。

3.4.3 行为识别算法

在准确定位教师及学生位置后需要使用行为识别算法分析教师及学生的行为。在2018年6月举办的计算机视觉顶级学术会议IEEE Conference on Computer Vision and Pattern Recognition（CVPR）上，Diogo C. Luvizon等人提出了一种多任务深度学习模型，该模型将姿态估计和行为识别2个任务放在了一个深度学习网络框架中，该模型的框架图如图7所示。

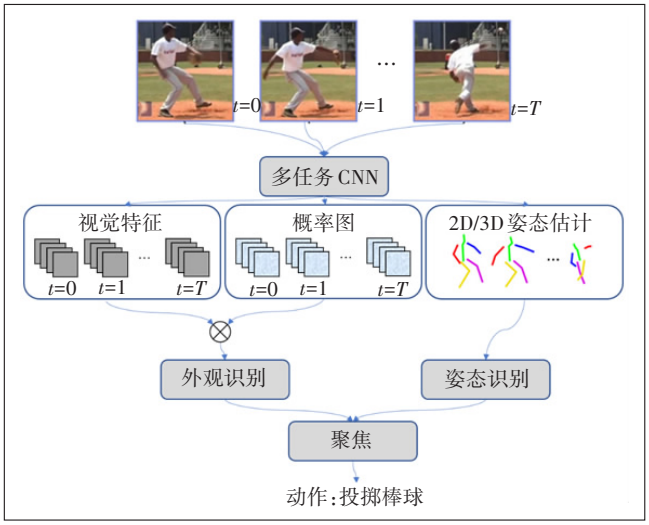

图7　多任务深度学习模型

将静态的RGB图像输入模型后，多任务CNN会进行姿态估计及动作识别。其中姿态估计使用了回归方法，利用了一个可微分的softmax来联合2D和3D的姿态估计。动作识别则分为基于姿态的识别和基于外观的识别2部分。最终2个任务的结果联合起来用于最终的行为识别。

该方法的一个重要思想是利用多任务之间相互促进的作用来提高最终任务的准确率。基于此思想及教室行为分析的场景，多任务结合的方法也可用于教室行为识别中。在对教师及学生行为进行识别时，

目标的姿态特征和外表特征都有利于行为的识别，可以将2种特征融合起来达到互利互补的效果。例如学生举手和书写的姿态特征会有明显的不同，此时再结合外表特征即可有效识别；学生书写和阅读的姿态特征则很近似，仅用姿态特征很难分辨，此时加入外表特征则可有效进行区分。基于此思想构建的教室行为识别深度神经网络模型如图8所示。

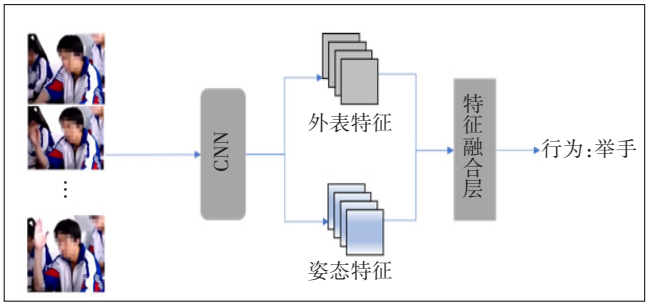

图8　教室行为识别深度神经网络模型

4 总结与展望

人工智能与各行各业深度融合是今后发展的必然趋势，利用人工智能的优势解决各行业的痛点问题、降低成本、提升效率是大势所趋。本文着眼于教育行业的痛点，结合大数据和人工智能技术，设计了一个基于人工智能的教育质量评估系统，用于解决教育行业中教师教育质量评估的问题。该系统相较传统的专家评估模式具有以下几点优势：

a）该系统从数据获取到教育质量分析评估基本无需人工干预。

b）相较传统的专家评估模式，该系统极大降低了人力成本。

c）该系统具有数据来源范围广、分析结果客观性强、全时性覆盖等特点，能够有效支撑专家进行教育质量评估工作。

目前该系统仍有数据量不足、分析方式单一等缺点。在未来，本系统将在以下几方面继续开展工作：持续迭代更新数据提升评估准确性；加入声音、文本等评估指标；除教学评估工作外，增加辅助教师教学的功能，丰富教师教学方式，提高学生学习效率。

参考文献：

[1] 新一代人工智能发展规划[M]. 北京：人民出版社，2017：2-6.
[2] 促进新一代人工智能产业发展三年行动计划（2018-2020年）[EB/OL].[2018-09-26]. http://xxgk.miit.gov.cn/gdnps/wjfbContent.jsp?id=5960820.
[3] 任友群. 走进新时代的中国教育信息化——《教育信息化2.0行动计划》解读之一[J]. 电化教育研究，2018(6)：27-28.
[4] DENG J，DONG W，SOCHER R，et al. ImageNet：A large-scale hierarchical image database[C]// Computer Vision and Pattern Recognition，2009. IEEE，2009：248-255.
[5] KRIZHEVSKY A，SUTSKEVER I，HINTON G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. 2012.
[6] ZAREMBA W，SUTSKEVER I，VINYALS O. Recurrent Neural Network Regularization[J]. Eprint Arxiv，2014.
[7] SAK H，SENIOR A，BEAUFAYS F. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition[C]// Conference of the International Speech Communication Association，2014. Interspeech，2014：338-342.
[8] 马晓亭. 基于大数据决策分析需求的图书馆大数据清洗系统设计[J]. 现代情报，2016，36(9)：107-111.
[9] 郭庆林. 基于实时数据采集的本科高校多媒体智能教室综合管理信息系统的分析与设计[J]. 电子测试，2016(11)：145-146.
[10] 徐雁飞，刘渊，吴文鹏. 社交网络数据采集技术研究与应用[J]. 计算机科学，2017，44(1)：277-282.
[11] 胡俊，胡贤德，程家兴. 基于Spark的大数据混合计算模型[J]. 计算机系统应用，2015，24(4)：214-218.
[12] 彭建华，李臣明，邱军林，等. 接收与处理分离的实时大数据处理模型[J]. 计算机科学与探索，2015，9(8)：906-913.
[13] REN S，HE K，GIRSHICK R，et al. Faster R-CNN：Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2017，39(6)：1137-1149.
[14] DAI J，LI Y，HE K，et al. R-FCN：Object Detection via Region-based Fully Convolutional Networks[C]// 30th Conference on Neural Information Processing System，2016.
[15] REDMON J，DIVVALA S，GIRSHICK R，et al. You Only Look Once：Unified，Real-Time Object Detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society，2016：779-788.
[16] LIU W，ANGUELOV D，ERHAN D，et al. SSD：Single Shot MultiBox Detector[C]// European Conference on Computer Vision. Springer International Publishing，2016：21-37.
[17] LIN T Y，MAIRE M，BELONGIE S，et al. Microsoft COCO：Common Objects in Context[C]// European Conference on Computer Vision. Springer International Publishing，2014：740-755.

作者简介：
吴浩然，工程师，硕士，主要从事人工智能应用、计算机视觉（图像检测、匹配）应用、教育信息化等研究工作；袁涛，工程师，硕士，主要从事可见光通信技术、人工智能应用、教育信息化研究工作；吴少平，研究实习员，硕士，主要从事教师教育、教育信息化、教育测评等研究工作；李亚梦，工程师，硕士，主要从事人工智能应用研究工作及知识图谱、自然语言处理相关研究工作。