

幼儿语言能力评价大数据的可视化分析平台构建

王少能

杭州米越科技有限公司 浙江 杭州 310000

摘要：幼儿语言能力发展是早期教育重点关注内容，构建幼儿语言能力评价大数据的可视化分析平台意义重大。该平台基于幼儿语言能力评价指标体系，从幼儿园、家庭等多场景采集数据，运用大数据处理技术清洗、存储与管理数据，借助可视化技术将数据转化为直观图表，呈现幼儿语言发展趋势、能力对比等信息，还集成数据分析算法实现深度挖掘。平台能打破传统评价方式局限，为教育者、家长提供客观、精准的幼儿语言发展评估，助力制定个性化教育方案，提升幼儿语言教育质量。

关键词：幼儿语言能力；大数据；可视化分析平台；语言教育

1 引言

在幼儿成长过程中，语言能力是其认知、社交发展的重要基石。科学准确地评价幼儿语言能力，能为早期教育提供关键依据，助力幼儿更好地成长。然而，传统幼儿语言能力评价方式存在诸多局限，依赖人工观察和简单测试，主观性强且数据收集不全面，难以精准呈现幼儿语言发展全貌。随着大数据和可视化技术的迅猛发展，为幼儿语言能力评价带来新契机。通过构建可视化分析平台，可广泛收集多场景下幼儿语言相关数据，利用大数据技术深度挖掘分析，再以直观可视化形式呈现结果。这不仅能提升评价的客观性与科学性，还能帮助教育者和家长更好地了解幼儿语言发展状况，制定个性化教育策略，对推动幼儿语言教育进步意义非凡。

2 平台构建的理论基础

2.1 幼儿语言能力评价指标体系

幼儿语言能力评价指标涵盖多个维度。表达能力方面，考查幼儿词汇量的丰富程度，如能否精准运用不同场景词汇；观察语法运用正确性，像语句是否符合主谓宾等基本语法规则，以及表达流畅度，是否频繁卡顿、重复。理解能力也很关键，包括对简单指令的执行情况，比如能否按要求取物；对故事内容的理解，是否能回答相关问题。社交语言运用同样重要，看幼儿在与同伴交流时，能否礼貌表达想法、倾听他人观点，以及是否会主动发起对话。这些指标相互关联，共同构建起全面、科学的幼儿语言能力评价体系，为后续数据采集和分析提供方向，助力精准把握幼儿语言发展水平。

2.2 大数据处理技术原理

大数据处理技术在平台构建中发挥核心作用。数据采集阶段，运用传感器、日志记录等方式，从幼儿园课堂互动、家庭亲子交流、在线学习平台等多渠道收集

幼儿语言数据。收集后的数据繁杂，需清洗处理，借助数据去重、异常值检测技术，剔除重复、错误数据，保证数据质量^[1]。存储时，根据数据特点，选用关系型数据库存储结构化数据，如幼儿基本信息；用NoSQL数据库存储非结构化数据，像语音记录。分析阶段，借助MapReduce分布式计算框架并行处理海量数据，运用机器学习算法挖掘数据间潜在关系，比如分析幼儿语言发展与年龄、教育环境的关联，为幼儿语言能力评价提供数据支持。如图1所示。

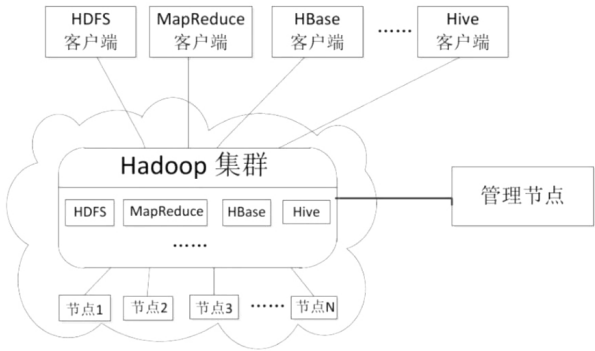


图1

2.3 可视化技术基础

可视化技术将复杂数据转化为直观图形，便于理解分析。柱状图适用于对比不同幼儿或同一幼儿不同阶段语言能力指标，如词汇量对比，柱子高低清晰展示差异。折线图可呈现幼儿语言能力随时间变化趋势，像表达流畅度提升曲线，让人直观看到发展进程。雷达图能综合展示幼儿在多个语言能力维度的表现，各维度指标围成多边形，清晰呈现优势与不足。可视化工具如Echarts，基于JavaScript开发，简单易用，可定制各类图表；Tableau功能强大，支持数据实时交互，通过拖拽操作快速生成可视化结果^[2]。这些技术和工具为幼儿语言能

力评价数据的可视化展示奠定基础。

3 平台需求分析与设计

3.1 用户需求调研与分析

针对幼儿教师,调研发现他们期望平台能实时呈现班级幼儿语言发展的整体情况,包括平均词汇量、表达能力分布等,方便掌握教学效果,据此调整教学计划。家长更关注自家孩子的语言成长细节,如发音准确性的进步、不同场景下语言运用的变化,以便在家庭中给予针对性陪伴和引导。教育研究者则希望平台提供多维度、大规模的幼儿语言数据,支持跨年龄段、跨教育环境的对比研究,挖掘幼儿语言发展规律。通过对这些用户需求的深入分析,明确平台应具备数据分类展示、个性化分析报告生成以及数据深度挖掘等功能,确保满足各方对幼儿语言能力评价数据的使用需求。

3.2 平台功能设计

平台核心功能围绕数据处理与分析展开。数据采集功能要支持多源接入,自动收集幼儿课堂发言、家庭对话录音等数据。数据处理功能对采集来的数据进行清

洗、标注,转化为可分析的结构化数据^[3]。可视化展示功能以多种图表呈现结果,如用饼图展示幼儿语言错误类型占比,便于直观发现问题。数据分析功能借助机器学习算法,实现幼儿语言能力的预测,例如预估下一阶段词汇量增长,以及对比分析不同教学方法对语言发展的影响,为教育者提供决策依据,助力制定更科学有效的幼儿语言教育策略。

3.3 平台架构设计

平台采用分层架构。前端展示层运用HTML、CSS和JavaScript技术,打造友好交互界面,用户可便捷操作,查看可视化图表、下载报告。中间逻辑层负责业务逻辑处理,接收前端请求,调用后端数据接口获取数据,根据用户需求进行数据处理和分析,如生成个性化语言发展报告。数据存储层选用合适数据库,关系型数据库MySQL存储幼儿基本信息、评价指标等结构化数据;非结构化数据如语音文件,存储在分布式文件系统HDFS中。各层之间通过RESTful API进行数据交互,保障系统高效稳定运行,支撑平台各项功能实现。如图2所示。

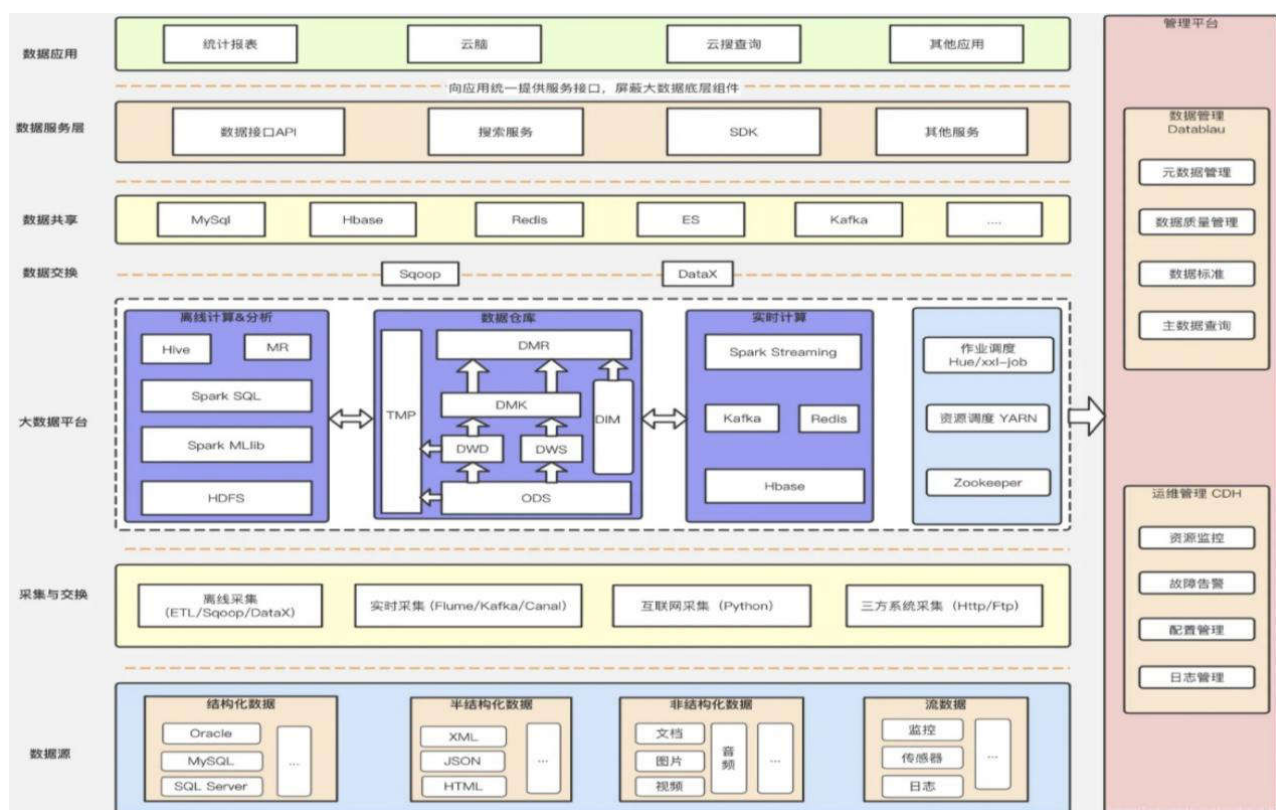


图2 平台分层架构的示意图

4 数据采集与处理

4.1 数据采集渠道与方法

数据采集渠道多元且各有方法。幼儿园是关键渠

道,教师通过课堂观察记录幼儿发言频次、内容等,借助智能设备录制课堂互动,利用语言学习软件记录幼儿学习轨迹;家庭环境也不容忽视,家长使用手机APP记录

亲子对话，分享幼儿日常语言表现，还可参与线上问卷调查，反馈幼儿在家语言环境和交流习惯；在线学习平台则自动记录幼儿参与课程的时长、完成作业情况、与虚拟角色互动数据等。多渠道采集可保证数据全面，从不同角度反映幼儿语言能力，避免单一渠道数据的局限性，为后续精准分析提供丰富素材，全面展现幼儿语言发展状况。

4.2 数据清洗与预处理

采集到的数据往往存在噪声和错误，需清洗与预处理。如图3所示。去重环节，通过比对幼儿ID、时间戳

等关键信息，去除重复记录，避免重复分析。异常值检测利用统计学方法，设定合理范围，如词汇量超出正常年龄段区间的视为异常，进行核实或修正。缺失值处理也很重要，若幼儿某阶段发言记录缺失，可根据前后数据趋势进行插值估算，或参考同年龄段幼儿平均水平填补。经过清洗与预处理，数据更加规范、准确，能有效提高后续数据分析的可靠性，确保基于这些数据得出的幼儿语言能力评价结果真实、可信，为教育决策提供有力支持。

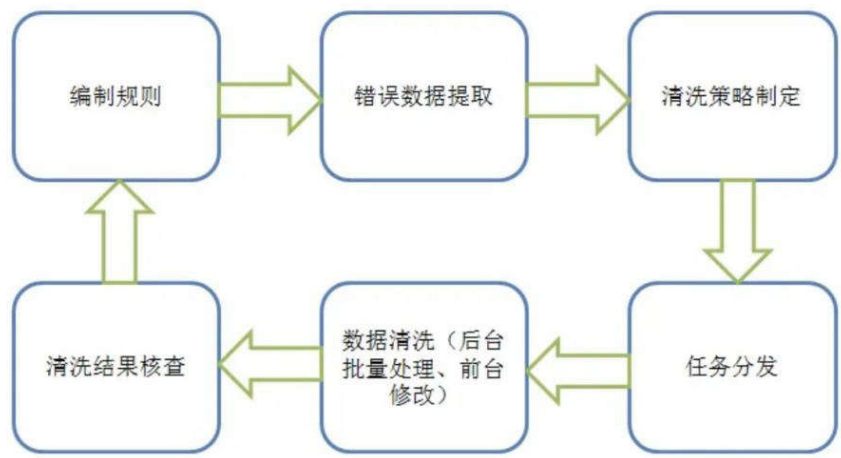


图3 数据清洗的流程示意图

4.3 数据存储与管理

数据存储与管理对平台稳定运行至关重要。存储方面，选用关系型数据库MySQL存储结构化数据，如幼儿基本信息、测评成绩，因其数据结构固定、查询精准；用非结构化数据库MongoDB存储语音、视频等非结构化数据，方便灵活存储与读取。管理上，建立严格的数据访问权限机制，教师只能查看所教班级幼儿数据，家长仅能访问自家孩子信息，保障隐私安全。定期备份数据，防止数据丢失，采用异地备份策略，应对突发灾难^[4]。同时，持续更新数据管理系统，优化存储结构，提高数据读写效率，满足平台不断增长的数据处理需求。

5 可视化分析功能实现

5.1 可视化图表设计与展示

可视化图表设计紧密围绕幼儿语言能力评价指标。词汇量增长用折线图展示，横坐标为时间，纵坐标为词汇量数值，能直观呈现幼儿词汇积累过程，如每月新增词汇量变化。语言表达错误类型占比用饼图呈现，不同错误类型如语法错误、用词不当等以扇形区域展示，占比一目了然，帮助教育者快速定位主要问题。幼儿在

不同场景下的语言表现，如课堂发言、家庭交流，用柱状图对比，柱子高度代表发言频次或表达时长，可清晰看出场景差异对语言运用的影响。通过精心设计并展示这些图表，让复杂的语言能力数据以直观易懂的形式呈现，为分析提供便利。

5.2 交互功能设计与实现

交互功能极大提升用户体验和数据分析效率。缩放功能允许用户在图表上通过鼠标滚轮或手势操作，放大感兴趣的局部数据，如查看某一时间段内幼儿词汇量骤增或骤减的细节。筛选功能可按幼儿年龄、性别、班级等条件筛选数据，对比不同群体幼儿语言能力，例如分析不同性别的幼儿在语法运用上的差异。数据查询功能支持用户输入关键词，快速获取特定幼儿的语言能力数据，如查询某个幼儿在过去一个月的表达流畅度评分。这些交互功能通过JavaScript、HTML5等前端技术实现，使用户能深入探索数据，挖掘更多有价值信息，满足不同用户对数据的个性化分析需求。

5.3 数据分析功能集成

数据分析功能集成让平台更具深度和价值。数据挖

掘算法用于发现数据中的潜在模式,如通过关联规则挖掘,找出幼儿语言发展与阅读习惯、家庭语言环境的关联,为教育干预提供参考。机器学习算法实现预测分析,以过往数据训练模型,预测幼儿未来语言能力发展趋势,帮助教育者提前规划教学方案^[5]。对比分析功能可将个体幼儿与同龄群体平均水平对比,或对比同一幼儿不同阶段表现,评估其发展进度。这些功能通过调用Python数据分析库如Scikit - learn、Pandas实现,结合平台数据存储和可视化展示,为幼儿语言能力评价提供全面、科学的分析结果,助力教育决策。

6 结语

构建幼儿语言能力评价大数据的可视化分析平台是早期教育领域的重要创新成果。在理论层面,科学完善的评价指标体系为精准评估幼儿语言能力筑牢根基,使评价过程有章可循;在实践方面,多源数据采集、高效处理以及可视化与深度分析功能的集成,有效克服了传统评价方式的弊端,为教育者和家长提供了更客观、精准的评价依据,显著提升了评价效率与精准度。未来,应积极拓展数据采集渠道,将更多元化的场景数据,如户外互动、兴趣班活动中的语言数据纳入平台;持续优化分析模型,考虑

更多影响幼儿语言发展的因素,提升其对不同幼儿群体的适配性。通过这些努力,进一步挖掘平台潜力,加强与教育实践深度融合,为幼儿语言教育提供更强大的支持,推动幼儿语言能力培养迈向新高度。

参考文献

- [1]谢庆斌,薛曼莉,陈昱玲,等.5岁幼儿执行功能对社会技能的影响:语言能力的中介作用[J].中国临床心理学杂志,2023,31(05):1194-1198.
- [2]姚俊,缪文淇,王玉微.大学生体质健康测试信息大数据可视化分析平台的构建及应用[C].第十三届全国体育科学大会论文摘要集——专题报告(体育统计分会).中国天津市天津市,2023:92-94.
- [3]金靛.大班幼儿语言表达能力评价研究[J].前卫,2024,(26):0062-0064.
- [4]朱学义.大数据财务分析创新体系的构建[J].会计之友,2024,(15):58-65.
- [5]王济农,孙帮涛,张驰.基于Three.js的地压监测可视化分析云服务平台构建与应用[J].有色金属(矿山部分),2023,75(06):21-25.