

# 高校法语专业四级考试数据模块建设研究

方 敏

昆明文理学院 (原云南师范大学文理学院) 650000

**摘 要:** 高校法语专业四级考试, 即 TFS-4, 是检测大学法语专业学生法语水平的国家性考试。其数据模块建设的研究旨在利用特性化和区分、关联分析、分析和预测、聚类分析、孤立点分析、趋势和演变分析等几种数据挖掘的方法, 对该考试进行学生错题数据分析, 建设数据题库, 量化学生学习数据, 从而精准指导教学, 最终提高学生过级率。通过研究, 达到教育部合格评估要求, 进而提高学生就业率。

**关键词:** 法语专四; 教学改革; 数据挖掘; 模块研究; 精准指导

法语专四自 04 年首次开考以来, 经历了三年左右的探索, 题型自 07 年才基本固定下来, 共包括五大部分: 听写与听力、词汇、语法、阅读与理解、写作。大二法语专业学生, 要求从听说读写四方面全面发展法语能力。

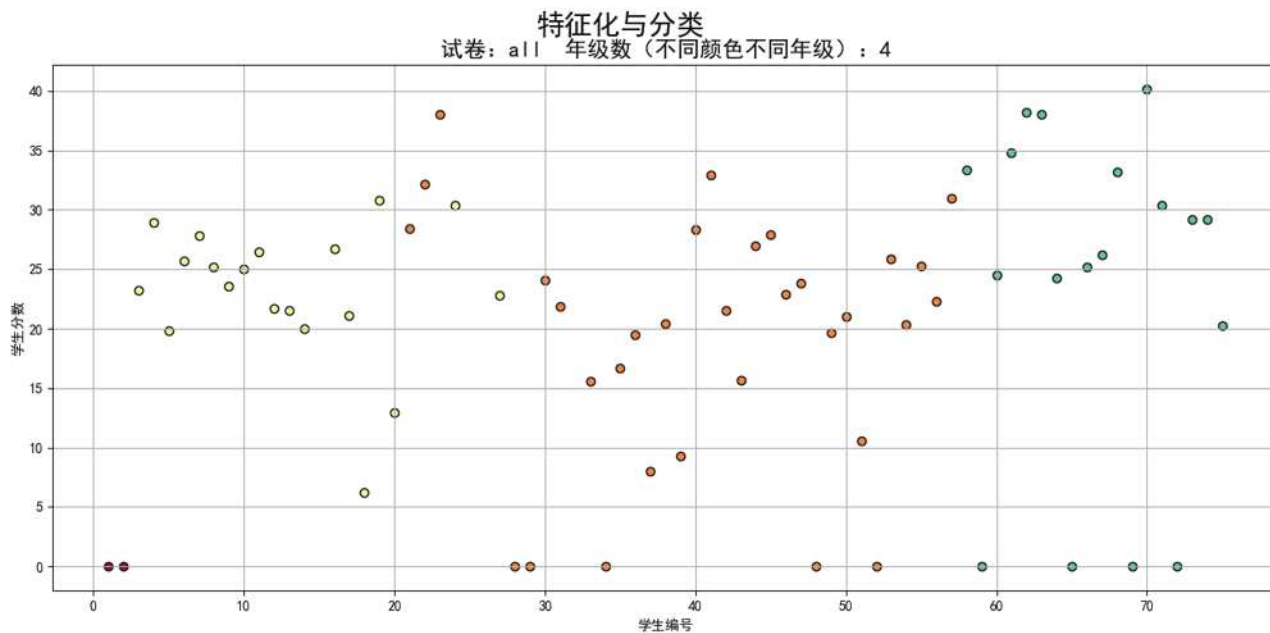
根据国家标准, 外语类人才培养目标旨在培养具有良好的综合素质、扎实的外语基本功和专业能力, 掌握相关专业只是, 适应我国对外交流、国家与地方经济社会发展、各类涉外行业、外语教育与学术需要的各外语语种专业人员和复合型外语人才。我们参照其制定了符合我们自身发展的人才培养方案。基于国标的高校法语专业四级考试过级率及备考训练是专业教学的

必要组成部分, 因此, 对该考试数据模块的建设研究也将与外语类学生高等教育有机结合。

在近两年的数据模块建设研究中, 阶段性成果如下:

一、特性化和区分: 为数据的特征化和比较产生描述 (当所描述的概念所指的是一类对象时, 也称为类描述); 统计不同 (同一) 学校不同层次学生学生, 区分不同学生对某个知识点同一学生对不同知识点的掌握情况, 通过数据筛查找到需要大范围集中讲解的知识点, 个别学生遇到的个别问题单独一对一解决;

基于试卷分数的学生分类, 可以按试卷或多个试卷搜索 (最新试卷)

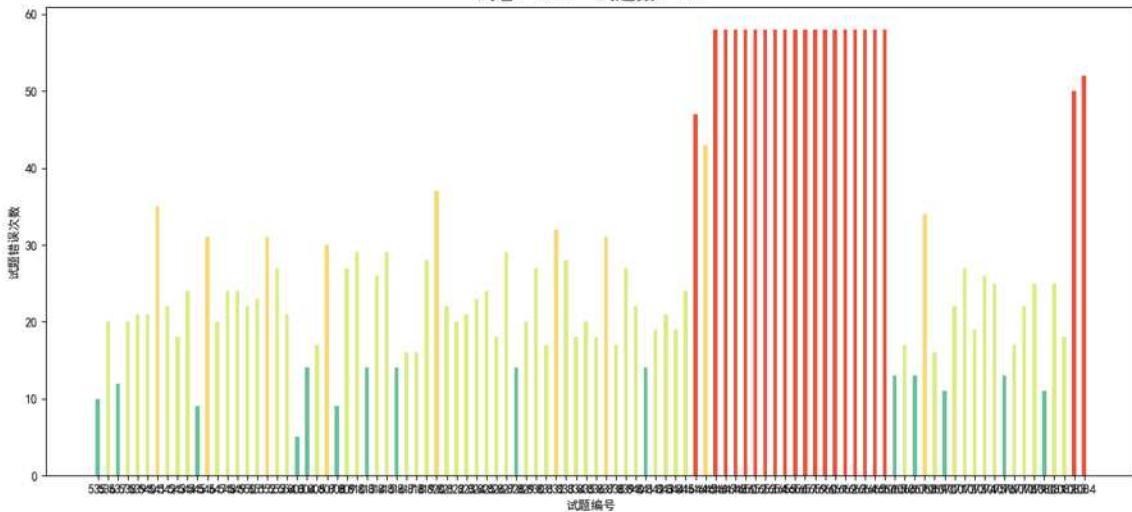


横轴为学生 ID 号 (共 75 个学生样本, 红色为 2016 级测试人员编号 2 人, 黄色为 2018 级学生 22 人, 橙色为 2017 级学生 33 人, 绿色为 2019 级学生 28 人), 纵轴为学生分数, 即历年真题试卷平均分。由此可见, 学生两极分

化严重, 最高分和做题较认真的学生主要集中在 2019 级。

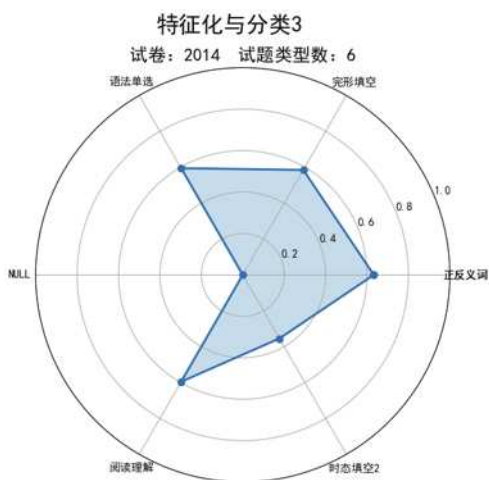
二、基于试卷错题的知识点分类 (试题分类, 人工知识点分析), 可以按单个学生或多个学生搜索 (单题柱状图)

特征化与分类2  
试卷: 2011 试题数: 100



横轴为题目ID号(2011年真题试卷共100题),纵轴为该题全部学生错题次数(共75个学生样本),红色为错误人数超过40人次的题目,橙色为错误人数超过30人次的题目,绿色为错误人数超过20人次的题目,蓝色为错误人数少于20人次的题目。由此可以看出,近78%的学生的错题集中在时态填空部分,阅读理解和语法填空题是第二个错题集中版块,28%的同学错误零星分布,每个版块都有薄弱环节。

三、基于试卷错题的知识点分类(试题题型分类,人工知识点分析),可以按单个学生或多个学生搜索(题型雷达图)(最新试卷)

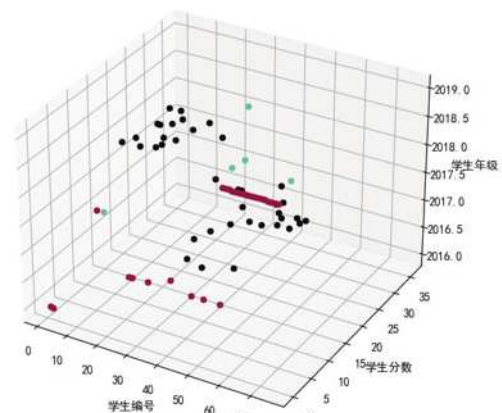


雷达图的每个轴为一种题目类型(共75个学生样本,2014年真题试卷共5种题型,该试卷没有时态填空1标记为NULL),每个顶点为该类型掌握情况,即归一化平均分。根据此图,正反义词模块超过60%,完成情况较好,完形填空、语法单选和阅读理解都呈现60%,水平均衡,时态填空情况稍差,不到40%的正确率。

三、聚类分析法:将物理或抽象对象的集合分组成为由类似的对象组成的多个类的过程;通过聚类推送学生薄弱知识点的考题着重训练,比如考生在时态学习和练习中错误多出现在过去时,系统题库会自动推送过去时的练习和答案解析;

按试卷题目对学生进行聚类,按学生聚类分析知识掌握情况和针对性题目辅导(三维图)(真题搜索为最新试卷,题型搜索为历次试卷)

DBSCAN聚类  
 $\epsilon = 15.0$   $m = 2$ , 聚类数目: 2

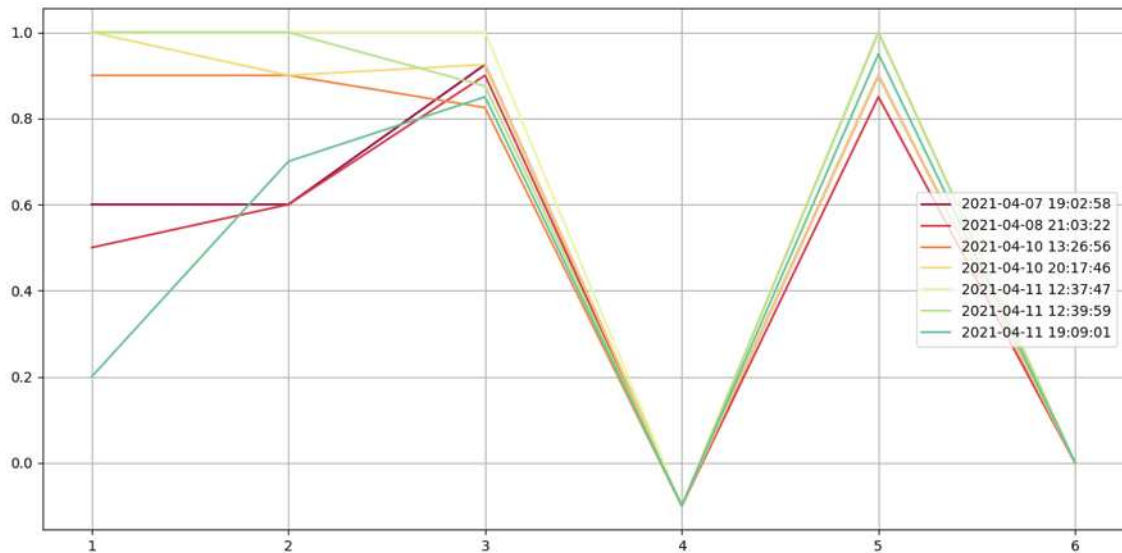


横轴为学生ID号(共75个学生样本),纵轴为学生年级(2016,2017,2018,2019级),Z轴为学生试卷分数(2015年真题试卷,去除听力部分的总分),根据错题交集的E领域值和核心对象值(20,4)进行聚类,得到红色和绿色两个学生集合,黑色为孤立点。此图由学生错题相似性将学生进行聚类,有效帮助教师对共性学生进行分类指导,组建专项学习小组,从题型的特点和学生的薄弱环节有的放矢,科学指导学生考前训练。

四、趋势和演变分析: 描述行为随时间变化的对象的发展规律或趋势(时序数据库); 通过题库数据挖掘分析学生易掌握和较难掌握的知识点, 对于浅显知识点简单复习即可, 根据数据分析难点如何循序渐进如何突破, 了解学生应该从哪个时期开始进入学习记忆理解高

峰, 或者某个学习时间应该加长, 哪个时期属于学生入门密集期, 应进入集中高强度训练阶段。

相同学生不同学习时间的试题得分(知识点掌握按题型分类)(真题搜索和题型搜索都为历次试卷单次统计)(时间按颜色区分)



横轴为题目类型(6种题型分类, 学生ID为63, 历年真题试卷得分), 纵轴为归一化试题得分, 颜色光谱(红橙黄绿蓝靛紫)根据做题时间先后顺序排列。由此可以看出, 该学生在考前运用数据题库持续训练, 学习状态趋于稳定, 并且成绩逐渐提高。

综上, 数据模块的建设研究科学助力日常教学。首先, 推动法语学科数据题库建设。对每道题进行多维度多层次精准标注, 上万道现有精选和我教师团队拓展题目共同构建智慧化学习资源, 基于法语教学知识体系和大数据分析, 实现知识点分析可视化, 智能化推送符合学生需求的重难点和薄弱题型, 提高教和学的效率, 有的放矢, 可以提升学生备考全国高校法语专业四级能力。其次, 信息化平台助力学生自测。教学资源可以随时调用, 学生自主测试, 可根据系统推送的标准答案和参考解析进行指导, 数据建设帮助教师适时调整教学策略。再者, 系统分析学生学习数据。借助大数据, 云计算, 学习过程全程跟踪记录, 自动存储学习数据, 学习成果

及时反馈, 学情报告可自动生成, 全面精准量化学习成果, 发现学生的潜在优势, 以有效数据为根据, 帮助教师由经验教学向精准教学过渡。

#### 参考文献:

- [1]王惠德, 曹德明.中国法语专业教学研究(基础阶段)[M].上海:上海外语教育出版社, 2005
- [2]曹德明, 王文新.全国高等学校法语专业四级考试指南[M].上海:上海外语教育出版社, 2014
- [3]胡庆, 张海月.法语专业四级真题精解[M].北京:北京时代华文书局, 2019
- [4]朱晔.法语专业四级考试答题技巧与模拟测试[M].北京:商务印书馆, 2015
- [5]王艳, 李双.全国高等学校法语专业四级考试样题集[M].北京:外研社, 2016
- [6]周志华.机器学习[M].北京:清华大学出版社, 2016
- [7]Jiawei Han和Micheline Kamber, 数据挖掘:概念与技术[M].北京:机械工业出版社, 2001