

数据驱动的小学语文句式测评工具研究*



——以三余阅读 APP 为依托

吴娟^{1,2} 周建蓉² 李梦³ 骈扬¹ 黄云龙¹

(1. 北京师范大学 未来教育高精尖创新中心, 北京 100875;

2. 北京师范大学 教育学部, 北京 100875;

3. 清华大学 教育研究院, 北京 100084)

摘要: 随着大数据和人工智能等技术的快速发展, 数据驱动的学习诊断与能力测评成为教育领域的研究热点。针对小学生句式测评忽视阅读、写作等实际运用场景的现状, 文章以自然语言理解技术中的句式语料库构建和句式自动识别算法为技术基础, 设计了数据驱动的小学语文句式测评工具。之后, 文章以三余阅读 APP 为依托, 实现了此工具的数据采集、智能分析与诊断、结果可视化呈现等三大功能。最后, 文章通过算法实验验证了此工具对句式的识别基本达到了预期的效果, 并通过可用性调查发现此工具的接受度较高。文章设计的数据驱动的小学语文句式测评工具可对小学生的语文句式掌握情况进行智能测评, 不仅弥补了低频次、规范化测试的缺陷, 而且为语文知识与能力的动态、持续测评提供了理论和实践支持。

关键词: 数据驱动; 句式测评; 小学语文; 测评工具; 三余阅读 APP

【中图分类号】G40-057 【文献标识码】A 【论文编号】1009—8097 (2021) 12—0103—07 【DOI】10.3969/j.issn.1009-8097.2021.12.013

随着大数据、人工智能等新技术的飞速发展, 智慧学习环境不断优化, 学习过程中生成的各类数据信息被存储记录, 这为智能化学习评价提供了新的契机^[1]。基于数据对学生的学习情况进行分析, 进而对学生的能力和素养加以诊断测量, 正成为相关领域探讨的热点^[2]。在语言学中, 句子是能够表达一个相对完整意思的语言单位, 恰当地使用句式可将不同的语义信息连接起来组成更大的语言单位^[3]。在语文学习中, 句式是提高小学生阅读理解能力与写作能力的重要前提和基础, 现行的义务教育语文课程标准也对中小学生的句式习得提出了具体要求。基于此, 运用新兴技术对小学生的句式掌握水平进行动态的个性化测评研究便提上了日程。本研究尝试运用自然语言理解技术, 从数据驱动的技术视角设计并实现自动测评小学生句式掌握水平的工具, 以期智能测评小学生语文学习的过程状态提供借鉴。

一 数据驱动的小学语文句式自动测评的技术基础

当前, 小学语文句式测评大多采用规范化测试题目, 如辨认句式、改正句式、句式造句等^[4]。脱离语言运用情境的测试不仅枯燥而且片面, 无法评测出学生对句式的综合运用能力。数据驱动的句式测评是解决上述问题的一种途径, 即首先对小学生在语文学习中产生的数据信息加以分析和挖掘, 再以一定的算法程序对学生句式掌握状态进行个性化的自动测评诊断, 可同时考虑知识的运用情境和学生的过程状态。自然语言理解技术中的句式语料库构建和句式自动识别算法, 为实现数据驱动的句式测评提供了技术基础。

语料库是一种大规模的电子文本集合, 存放着在语言实际应用中出现的语言材料^[5], 构建小学生句式语料库是实现数据驱动的句式测评的基础。目前, 常用的句式语料库有汉语复句语料库^[6]和清华汉语树库的汉语句法树库^[7], 但因其语料源多为报纸、杂志等社会媒体材料, 故而与

小学生语文学习场景的适配度不高。借鉴已有句式语料库的语句标识规范,本研究将小学语文教材作为语料源,以义务教育语文课程标准的要求为依据,构建了小学语文句式语料库,其语料中标识的年级特征信息可为句式识别后的智能测评提供判断依据。

句式自动识别多采用基于规则的方法和基于统计的方法^[8]:①基于规则的方法是先识别检测对象的关系词,再检索相应的句式约束规则,最后根据约束规则与检测对象的匹配情况,自动识别检测对象所属的句式类型。该方法以句式的本体知识库和规则库作为计算基础^[9]。本体知识库是句式库,也是建立规则库的前提和基础。胡金柱等^[10]提出,可从“关系词的选取来源”“关系词的典型性分类”“关系词特征字段的设立”等方面建立句式库。规则库是为句式判定提供约束条件的文本合集,是计算机自动识别句式的主要依据。胡金柱等^[11]归纳出字面约束、关系标记位置约束、分句位置约束等12种约束规则;杨进才等^[12]将语法中的依存关系所形成的7类约束条件加入规则库,将字面特征与语法特征相结合来提高识别率。②基于统计的方法是用大量语料对模型(如贝叶斯网络^[13]、神经网络^{[14][15]}等)进行训练,再用训练后的模型对文本进行自动识别。目前,基于统计的方法尚没有针对小学语文的专用句式语料库,故本研究采用了较为成熟的基于规则的方法进行小学语文句式的自动识别。

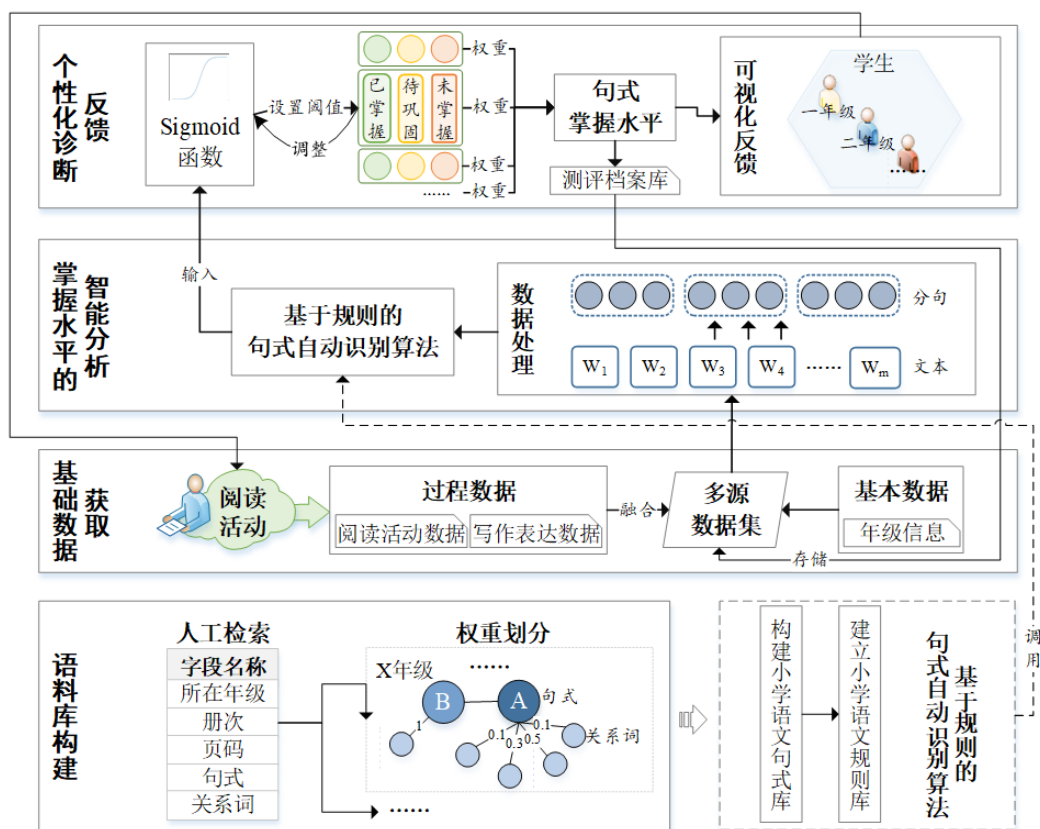


图1 数据驱动的小学语文句式测评工具设计框架

二 数据驱动的小学语文句式测评工具的设计

数据驱动的小学语文句式测评工具以小学语文各年级句式的掌握目标作为参照,对小学生

在阅读平台上留下的过程性句式运用数据进行动态采集、自动识别和分析挖掘,自动且个性化地诊断小学生句式的学习情况,并对测评结果予以可视化反馈。基于此,本研究构建了包含语料库构建、基础数据获取、掌握水平的智能分析、个性化诊断反馈四个环节的数据驱动的小学语文句式测评工具的设计框架,如图1所示。

1 小学语文句式语料库的构建

本研究选择部编版《小学语文》1~6年级、共12本教材作为语料来源,以人工检索句式的方式构建了小学语文句式语料库:先以人工方式对12本小学语文教材中出现的句式进行检索和整理,共获得1765条语料数据,语料信息包括年级、教材册次、页码、句式类型、具体关系词等5个字段。然后,依据语料库中的具体关系词字段对每种关系词出现的总次数进行统计,剔除总次数<3的关系词及其语料,构建出的小学语文句式语料库共收录59个关系词,分属于“并列”“递进”“选择”“转折”“因果”“假设”“条件”等7种句式。最后,以关系词在本年级此类句式中的出现的占比作为权重,来优化小学语文句式语料库,句式权重的计算如公式(1)所示。其中, g 表示某年级, W_g 为某关系词在 g 年级同类句式中的权重, $Count_item$ (某关系词)表示该关系词在 g 年级教材中出现的次数, $Count$ (该关系词所属句式的所有关系词)表示此类句式的所有关系词在 g 年级教材中出现的总次数。优化后得到的小学语文句式语料库(部分)如表1所示。

$$W_g = \frac{Count_item \text{ (某关系词)}}{Count \text{ (该关系词所属句式的所有关系词)}} \quad \text{公式(1)}$$

表1 小学语文句式语料库(部分)

年级	册次	页码	句式	关系词	权重	具体文本
一	上册	63	因果	因为	1	我最喜欢冬天,因为冬天可以堆雪人。
二	上册	104	假设	如果……就……	0.9	如果你愿意和好,就放一只风筝吧!
三	下册	41	递进	不但……而且……	0.3	这座桥不但精巧,而且美观。
四	上册	22	并列	一边……一边……	0.1	它一边飞,一边从嘴里发出一种声音。

2 句式测评基础数据的获取

学生在阅读平台上进行语文学习时会留下多种数据,如阅读过程中对相关句式进行批注、摘抄而形成的阅读活动数据,在写作和讨论等活动中使用各种句式进行语言表达而形成的写作表达数据等。其中,阅读数据代表学生处于知道或领会的层级,而写作数据表明学生达到运用层级,可见这两类数据能在不同程度上反映学生的句式知识掌握水平与运用能力。根据 Bloom 对认知领域的目标分类^[16],知识的运用比知识的领会所处的认知层级更高,而写作数据能在更大程度上体现学生的句式掌握程度,故对数据进行多源融合时,写作数据的占比应更高一些。此外,考虑到不同年级对于句式的掌握目标不同,故多源数据集中也会记录年级信息数据。

3 基于规则的句式掌握水平智能分析

在构建多源数据集的基础上,句式掌握水平智能分析可分为数据处理和基于规则的句式自动识别算法两个部分。由于句式自动识别算法需要分句维度的数据信息,因此需先对输入文本进行分句数据处理,再为基于规则的句式自动识别算法建立句式库和规则库。句式库是对句式的信息加以形式化表征,是句式自动识别的本体知识库,以关系类别、组配情况(单用、搭配)、

组配位置（前呼 f、后应 b）、组配对象、搭配形式（同形 1、异形 0）、能否易位（能 1、否 0）等 6 类特征字段对句式表征。规则库是明确判定句式的约束条件，是计算机自动识别不同句式的依据。以《汉语复句研究》为参考^[17]，本研究从以下四个方面对规则进行描述：①字面约束是对字面特征进行约束，如某句式前后能否出现特定字词；②分句位置约束，指对搭配关系词所在分句间距离的约束；③关系标记位置约束，指搭配关系词能否位于同一分句；④分句数目约束，指对分句的数量进行限制。

4 句式测评个性化诊断反馈

句式掌握水平智能分析的结果是获得某句式关系词在学生以往读写活动中出现的次数，其取值范围为 $[0, +\infty]$ 。由于无法通过此数据获悉学生对该句式的掌握水平，故采用 Sigmoid 函数来计算学生掌握某一关系词的概率，以此判断其句式掌握情况。函数表达式如公式（2）所示。其中， z 表示某一关系词的出现次数， e 为自然常数， $\sigma(z)$ 表示学生掌握这一关系词的概率。

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad \text{公式 (2)}$$

Sigmoid 函数能够计算特定输入值属于 1 的概率大小，当概率大于某一值时，就将其划分为一类。Sigmoid 函数值域为 $(0, 1)$ ，根据数据结果进行多次迭代和调整，将 0.6、0.8 设为划分阈值：值大于 0.8，认为学生已掌握此句式；值处于 0.6~0.8 之间，认为学生对此句式的掌握有待巩固；值小于 0.6，则认为学生未掌握此句式。为进一步呈现学生对某类句式的掌握程度，将学生的关系词掌握水平与该关系词的权重值相乘，得到学生对此类句式的掌握水平，并将其记录在学生测评档案库中。个性化诊断结果出来后，将以可视化的形式反馈给学生。

三 数据驱动的小学语文句式测评工具的功能

1 功能设计

对应图 1 所示设计框架中的基础数据获取、掌握水平智能分析、个性化诊断反馈三个环节，数据驱动的小学语文句式测评工具主要提供数据采集、智能分析与诊断、结果可视化呈现三大功能：①数据采集。数据驱动的小学语文句式测评工具首先需要记录学生的年级信息，同时采集学生在各类读写活动中的过程数据，如阅读活动数据、写作表达数据等。与此同时，学生的年级信息和句式测评结果也会被记录在多源数据集中。②智能分析与诊断。在采集各类数据的基础上，数据驱动的小学语文句式测评工具利用基于规则的句式掌握水平智能分析和个性化诊断方法，对学生的句式掌握水平予以测评。③结果可视化呈现。为使测评结果更加简洁、直观，学生的句式掌握水平采用进度条的形式予以呈现，学生进入工具首页即可查看。此外，学生点击各类句式进入某类句式界面后还可以查看不同关系词的具体掌握情况，其中“已掌握”“待巩固”“未掌握”三种情况分别以绿色、黄色、红色进行标识。以此掌握情况为依据，学生详情界面还进一步对测评结果予以分类呈现。为了进一步辅助教学，数据驱动的小学语文句式测评工具还为教师提供所有年级所需掌握句式的详情列表，用户以教师身份进入工具首页便可在教师详情界面进行查看。

2 功能实现

本研究团队在自主研发的三余阅读 APP 中，实现了数据驱动的小学语文句式测评工具的三大功能。三余阅读 APP 是北京师范大学未来教育高精尖创新中心研发的支持泛在学习环境下内

容与活动融合、课内和课外融合、线上和线下融合的一款移动阅读软件，能够对学生在语文学习和阅读活动中生成的各类读写活动数据进行采集、分析与反馈：①在数据采集部分，数据驱动的小学语文句式测评工具自动提取三余阅读 APP 中“批注”“摘抄”“小作家”“讨论区”等模块中的文本数据，同时采集学生的年级信息。②在智能分析与诊断部分，数据驱动的小学语文句式测评工具利用已开发的基于规则的句式掌握水平智能分析和个性化诊断方法，对学生的句式掌握水平予以测评。③在结果可视化呈现部分，数据驱动的小学语文句式测评工具依据学生或教师的不同用户身份，为学生呈现句式掌握整体情况的测评结果和各关系词掌握情况的详细信息，为教师呈现其所教年级的所有句式和句式关系词的具体信息。在三余阅读 APP 中，体现数据驱动的小学语文句式测评工具三大功能的部分句式测评界面如图 2 所示。其中，图 2(a)、(b)、(c) 为学生界面，图 2(d) 为教师界面。



图 2 三余阅读 APP 中的句式测评工具界面（部分）

四 数据驱动的小学语文句式测评工具的应用效果

为检验数据驱动的小学语文句式测评工具的应用效果，本研究同时进行了算法实验验证和可用性调查。

表 2 算法在 7 种句式上的实验结果

类别	并列	递进	选择	转折	因果	假设	条件	整体
精准率	84.6%	89.3%	90.0%	93.9%	89.4%	83.3%	95.0%	90.6%
召回率	97.1%	89.3%	90.0%	90.6%	88.6%	100.0%	95.0%	91.3%
F1 值	90.4%	89.3%	90.0%	92.2%	89.0%	90.9%	95.0%	90.9%

①算法实验验证：本研究采用精准率（Precision）、召回率（Recall）和 F1 值（F1-Measure）等计算机领域的常用指标^[18]，来测评数据驱动的小学语文句式测评工具中句式自动识别算法的质量。其中，精准率反映算法的查准率，召回率反映算法的查全率；而 F1 值是精准率和召回率的调和平均数，一般情况下如果 F1 值较高，就说明该实验方法比较有效。本研究选择三余阅读

APP 中的 439 条真实用户数据作为测试语料对算法进行检测, 得到算法在 7 种句式上的实验结果(如表 2 所示): 整体句式识别的精准率和召回率均在 90% 以上, 说明该工具对句式的识别基本达到了预期的效果。

②可用性调查: 为进一步优化数据驱动的小学语文句式测评工具的设计及其功能, 本研究对此工具的可用性进行了试用调查。作为目标用户, 北京市通州区 Z 小学的 13 名师生使用此工具完成指定的任务后填写调查问卷。问卷依据 Hwang 等^[19]设计的技术接受度量表修订而成, 包含认知有用性和认知易用性两个维度。问卷题项采用李克特五点量表进行计分(1 表示很不赞成, 5 表示非常赞成)。问卷调查结果显示, 此工具在认知有用性、认知易用性两个维度上的平均得分分别为 4.5 分、4.4 分, 表明用户对该工具的接受度较高。

五 结语

本研究从数据驱动的技术视角, 在构建小学语文句式语料库的基础上, 设计了数据驱动的小学语文句式测评工具。此工具嵌入句式自动识别算法, 基于学生句式运用读写活动中的过程数据, 对小学语文句式掌握水平进行了智能分析与诊断, 并对个性化诊断结果予以可视化呈现。以三余阅读 APP 为依托, 实现了数据驱动的小学语文句式测评工具的三大功能, 并通过算法实验验证和可用性调查, 验证了此工具的句式识别效果和接受度。

值得一提的是, 本研究采用的是基于规则的方法来进行小学语文句式的自动识别, 随着小学语文句式语料的不断丰富, 后期可采用深度学习算法实现句式的自动识别, 以进一步提高算法识别的智能性和正确率。此外, 受限于现有用户群体人数, 未来可面向更大范围的学生群体以获得更大量的应用数据, 再对工具加以完善和升级。

参考文献

- [1]姜强,赵蔚,李勇帆,等.基于大数据的学习分析仪表盘研究[J].中国电化教育,2017,(1):112-120.
- [2]王胜灵,石红薇,赵航,等.2019 年国家自然科学基金“教育信息科学与技术”的计量分析[J].现代教育技术,2020,(2):5-13.
- [3]黄伯荣,廖序东.现代汉语(增订六版 下册)[M].北京:高等教育出版社,2017:4.
- [4]潘建忠.图式与句式教学:一个转折复句的教学实验[J].心理科学,2000,(6):672-676.
- [5]邓纯华,胡金柱,李源.面向现代汉语复句信息工程的语料仓库构建研究[J].信息系统工程,2013,(9):146-148、151.
- [6]邢福义,姚双云.汉语复句语料库的建设与利用[A].第三届 HNC 与语言学研究学术研讨会论文集[C].北京:北京师范大学出版社,2005:432-439.
- [7]周强.汉语句法树库标注体系[J].中文信息学报,2004,(4):1-8.
- [8][12]杨进才,涂馨丹,沈显君,等.基于依存关系规则的汉语复句关系词自动识别[J].计算机应用研究,2018,(6):1756-1760.
- [9]沈威,姚双云.基于规则的复句中的关系词标注探讨[J].福建电脑,2007,(4):6-7.
- [10]胡金柱,吴锋文,李琼,等.汉语复句关系词库的建设及其利用[J].语言科学,2010,(2):133-142.
- [11]胡金柱,舒江波,胡泉,等.汉语复句关系词自动识别中规则的约束条件研究[J].语言文字应用,2015,(1):82-89.
- [13]杨进才,郭凯凯,沈显君,等.基于贝叶斯模型的复句关系词自动识别与规则挖掘[J].计算机科学,2015,(7):291-294、319.

- [14] 杨进才, 杨璐璐, 汪燕燕, 等. 基于神经网络的关系词非充盈态复句层次的自动识别[J]. 计算机科学, 2019, (S2): 103-107.
- [15] 孙凯丽, 邓沌华, 李源, 等. 基于句内注意力机制多路 CNN 的汉语复句关系识别方法[J]. 中文信息学报, 2020, (6): 9-17、26.
- [16] Krathwohl D R. A revision of Bloom's taxonomy: An overview[J]. Theory into Practice, 2002, (4): 212-218.
- [17] 邢福义. 汉语复句研究[M]. 北京: 商务印书馆, 2001: 177.
- [18] Powers D M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation[J]. Journal of Machine Learning Technologies, 2011, (1): 37-63.
- [19] Hwang G J, Yang L H, Wang S Y. A concept map-embedded educational computer game for improving students' learning performance in natural science courses[J]. Computers & Education, 2013, 69: 121-130.

Research on the Data-Driven Assessment Tool for Chinese Sentence Pattern in Primary School

——Relying on SanYu Reading APP

WU Juan^{1,2} ZHOU Jian-rong² LI Meng³ PIAN Yang¹ HUANG Yun-long¹

(1. Advanced Innovation Center for Future Education, Beijing Normal University, Beijing, China 100875;

2. Faculty of Education, Beijing Normal University, Beijing, China 100875;

3. Institute of Education, Tsinghua University, Beijing, China 100084)

Abstract: With the rapid development of technologies such as big data and artificial intelligence, data-driven learning diagnosis and ability assessment have become a research hotspot in the field of education. Aiming at the current situation that primary school students' sentence pattern assessment neglected reading, writing and other practical application scenarios, this paper designed a data-driven assessment tool for Chinese sentence pattern in primary school based on the construction of sentence pattern corpus and automatic sentence pattern recognition algorithm in natural language comprehension technology. After that, three major functions of this tool including data collection, intelligent analysis and diagnosis, and result visualization, were realized relying on SanYu Reading APP. Finally, the paper verified that this tool basically achieved the expected effect in the recognition of sentence patterns through algorithm experiments, and found that the acceptance degree of this tool was high through a usability survey. The data-driven assessment tool for Chinese sentence pattern in primary school could intelligently evaluate the primary school students' mastery situation of Chinese sentence patterns, which not only made up for the defects of low-frequency and standardized tests, but also provided theoretical and practical support for the dynamic and continuous evaluation of Chinese knowledge and ability.

Keywords: data-driven; sentence pattern assessment; Chinese in primary school; assessment tool; SanYu Reading APP

*基金项目: 本文为北京市教育科学“十三五”规划 2019 年度优先关注课题“数字媒体学习与传统学习效果的比较研究”(项目编号: CHEA19063)的阶段性研究成果。

作者简介: 吴娟, 副教授, 博士, 研究方向为技术增强语言学习, 邮箱为 wuj@bnu.edu.cn。

收稿日期: 2021 年 5 月 16 日

编辑: 小时

109