

文章编号: 1003-0077(2018)07-0128-09

## 小说人物性格的文学智能分析：以《平凡的世界》为例

吴育锋<sup>1,2,3</sup>, 吴胜涛<sup>1,2</sup>, 朱廷劭<sup>1</sup>, 刘洪飞<sup>4</sup>, 焦冬冬<sup>1</sup>

(1. 中国科学院 心理研究所, 北京 100101; 2. 厦门大学 传播研究所, 福建 厦门 361005;  
3. 厦门大学 人文学院, 福建 厦门 361005; 4. 北京联合大学 师范学院, 北京 100101)

**摘 要:** 以往小说人物心理分析主要是对人物性格的定性分析, 易受研究者个人主观经验影响; 而相比于描述繁杂的性格而言, 更加稳定系统的人格能够更好地描述并传达小说人物心理。该文采用基于数据挖掘的文学智能分析方法, 通过中文心理分析系统对《平凡的世界》人物对话进行处理, 得到人物的大五人格预测分数; 进而, 考察文学文献、小说剧情对预测分数的验证情况, 以确定这种方法的有效性。结果表明: 年轻的孙少平和田晓霞开放性相对较强, 而年长的孙少安和田润叶外向性较强; 此外, 孙少平和田润叶尽责性较强, 孙少安和田晓霞宜人性较好, 孙少安和孙少平情绪性较高。上述预测结果得到文献、剧情的支持, 这说明文学智能分析小说人物人格是有效的, 它为小说人物心理分析开辟出一条客观、体系化且智能化的道路。

**关键词:** 小说人物人格; 中文心理分析系统; 平凡的世界

中图分类号: TP391

文献标识码: A

## Identifying Novel Characters' Personality—An Example Study on *Ordinary World*

WU Yufeng<sup>1,2,3</sup>, WU Shengtao<sup>1,2</sup>, ZHU Tingshao<sup>1</sup>, LIU Hongfei<sup>4</sup>, JIAO Dongdong<sup>1</sup>

(1. Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China;  
2. Institute of Communication, Xiamen University, Xiamen, Fujian 361005, China;  
3. College of Humanities, Xiamen University, Xiamen, Fujian 361005, China;  
4. Teachers' College, Beijing Union University, Beijing 100101, China)

**Abstract:** The previous psychological investigations of novel mainly focus on novel characters' qualitatively, relying on researchers' personal experience subjectively, which is less stable or systematic as compared to personality analysis of the characters. Using the machine learning model based on Chinese psychology analysis system and characters' dialogues in *Ordinary World*, we got the four novel characters' big-five personality scores. Then, by comparing the predicted scores with documents related to characters' psychological analysis and novel descriptions, the validity of this method was established. The results revealed that: The younger characters (Shaoping Sun and Xiaoxia Tian) held relatively high Openness, while the older ones (Shaoan Sun and Runye Tian) held relatively high Extraversion. In addition, Shaoping Sun and Runye Tian had relatively high Conscientiousness, Shaoan Sun and Xiaoxia Tian had relatively high Agreeableness, and Shaoan Sun and Shaoping Sun had relatively high Neuroticism. The results demonstrate the applicability of personality analysis of novel characters, through an objective, systematic and intelligent approach.

**Key words:** novel characters' personality; Chinese psychology analysis system; *Ordinary World*

收稿日期: 2017-08-08 定稿日期: 2017-10-12

基金项目: 国家社会科学基金(17A2D041)

## 0 前言

通过对小说人物性格的分析,有助于我们更加准确、深刻地理解小说的人物心理及其所反映的现实生活。目前学界对小说人物的性格分析主要有两个思路。一个是文艺学的定性研究,依据研究者的人文素养和对小说文本的阅读体验,从微观方面将小说中的一个或多个人的性格概括为主要的几个方面,然后对这几个方面的性格特点用文本中的有关描写进行佐证<sup>[1]</sup>,在宏观层面进一步归纳同一作者笔下小说人物的性格特点<sup>[2]</sup>。另一个是从语言学的角度,通过选取若干段小说中的对话进行语用原则<sup>[3]</sup>、话轮转换<sup>[4]</sup>或其他会话分析,或者通过语料库对人物语言用词进行统计,得出频率较高的词作为关键词,通过关键词分析人物性格<sup>[5]</sup>。系统功能语言学的研究采用定量的方法,以某一种词汇语法框架/词分析理论或者相对稳定的词表<sup>[6]</sup>对小说中各种词的使用情况进行考察,以分析人物性格。

对小说人物性格进行定性分析是当前小说人物心理分析的主流,而对小说人物的心理分析也一般依赖于研究者在研究过程中的主观体验和文学素养。自刘再复“引入心理学方法研究文学人物”<sup>[7]</sup>之后,小说人物的心理分析主要聚焦于心理学家 McAdams 提出的人格三层次模型<sup>[8]</sup>之中的个人关注(personal concerns)(第二层)或生活叙事(life narratives)(第三层)——即角色在不同情境下的动机、关切和策略或探究角色人生的发展历程。很少有研究深入到最初的人物性格(personality)或人格特质(trait)(第一层)对角色进行讨论。

性格是一个人对现实稳定的态度和习惯化的行为方式,是个体行为的显著性倾向<sup>[9]</sup>。从心理测量学的角度看,性格较其他相对容易变化的心理状态相比,是相对稳定的人格特质,具有可比较、可描述和跨情境的优点。因此,性格或人格特质可以看作是个体心理活动的行为表现方式或模式,从性格的角度进行小说人物的心理分析,能够帮助对小说人物有更加准确、深刻的理解。通过性格来描述文学人物,可以直观地看出其最突出的心理特点,也可以比较不同人物的心理差异。人格特质研究领域包含多种理论模型,其中又以“大五”人格模型(big-five personality trait model,以下简称“大五人格”)最为当今学界所重视。

大五人格最早由词汇学研究者 Goldberg 在

1981 年提出<sup>[10]</sup>,其研究的核心假设是通过对自然语言中描述人格的语义相同或相近词汇的分析与缩减,能够得到人格更加基础的维度。之后的人格心理学家在此基础上进行了大量研究,虽然不同研究者之间就这五个因素的最佳解释有所争议,在一些因素的分子水平上也有不同的划分,但是在宏观上得到了跨文化普遍存在的五个最重要的人格特质模型<sup>[11]</sup>。本研究参照的大五模型<sup>[12-13]</sup>和建立大五人格预测模型参照的模型保持一致。下面简要介绍这五个维度的具体含义。

情绪性(neuroticism)与情绪的稳定相反,包括焦虑、愤怒、抑郁、自我、冲动、脆弱等负面情绪特征。神经质维度高的人可能常常表现得忧虑,表现得喜怒无常;神经质低的人常常是放松和情绪稳定,并且有能力应对压力。

外向性(extraversion)描述了一个人的活动水平、主导地位、社交性、表现力和积极情绪,包括热心、慷慨、自信、主动、积极、活跃等特征。外向性高的人表现出精力充沛、对人热情,可能喜欢和其他人逗趣,或是喜欢和他人一起聊天、玩耍;外向性低的人在社交之中比较含蓄,面对生人可能会紧张,表现出羞怯和拘谨,可能会给人以安静或者冷淡孤僻的印象。

开放性(openness)描述了个人精神和生活体验的广度、深度和复杂性,包括幻想、艺术、感觉、创新、兴趣、价值等特征。开放性高的人具有独创性,善于创造,有活跃的想象,易于接受新的思想,对许多不同的事物保持好奇,并且常常对艺术和美保持敏感与渴求;开放性低的人更加倾向传统,喜欢从事常规性工作,不喜欢不确定性。

宜人性(agreeableness)的主要特征是亲和力,包括利他、温柔、信任、坦诚、依从和谦虚等特征。宜人性高的人乐于助人、对人宽容、待人友善、体谅他人并喜欢合作;宜人性低的人可能有过强的自尊心,表现出喜欢挑人错误、对人粗鲁,并且常常与人发生争吵。

尽责性(conscientiousness)描述了坚持任务和对非目标行为冲动控制,包括竞争、有序、成就、义务、自律、缜密等特征。尽责性高的人做事计划性强、有效率,在他人眼中是个可信赖的人;尽责性比较低的人可能做事比较粗心、容易分心、缺乏条理。

近几年,随着计算机自然语言处理以及机器学习技术的日益成熟,研究者通过社交媒体内容与大五人格量表的映射关系,对 Facebook 用户<sup>[14]</sup>、新浪

微博用户<sup>[15]</sup>的人格进行自动识别。基于生态化的行为数据,利用机器学习实现个体心理特征的自动识别的过程,我们称之为生态化识别(ecological recognition, ER)。经典的基于自我报告的人格测量方法难以对文本构成的角色进行施测,生态化识别提供了一种行之有效的解决方法。近年来,通过运用统计学、模式识别、机器学习、数据抽象等数据分析工具研究文学的文学智能分析逐渐变得可行。

基于此,我们将生态化识别方法引入小说人物心理分析中,以文学智能分析的可重复的、客观性强的中文心理分析系统处理小说人物对话,以得到对应人物大五人格的预测分数。进而,再引用文艺学文献和关键情节来验证预测分数的有效性,从而说明文学智能分析方法的可行性。

## 1 研究方法

### 1.1 文本选取

我们选择《平凡的世界》<sup>[16]</sup>作为分析文本,主要基于以下两个标准:

(1) 对于小说人物性格的研究已经比较成熟。由于本研究属于对于方法的验证研究,因而我们需要有足够的文献结论来与我们得出的结论对比,这就要保证所选择的小说经过了一定时期的历史沉淀并且得到认可。因此我们从茅盾文学奖历年的获奖作品中选取,同时考察作品已有研究文献的数量。

(2) 小说人物语言和普通话的接近程度。小说人物的语言一般会受到其被设定的时空的影响,为了减少这种影响,我们选取文本的一个关注点就是小说人物语言和普通话的接近程度。因此,不仅小说所描写的年代应该尽可能晚。由于“北方方言是现代汉民族共同语的基础方言,以北京话为代表,内部一致性较强”<sup>[17]</sup>,入选小说的语言也应属于北方方言。

基于以上标准,我们选取了《平凡的世界》作为分析对象。《平凡的世界》“甫一出版,就得到了评论家们非常明确的肯定,很快又获得由政府颁发、评论界主持的权威奖项——茅盾文学奖,可以说是得到了中国文学评论界的最高荣誉;同样,读者对《平凡的世界》也有普遍的认可”<sup>[18]</sup>。笔者通过文献搜索,发现了足够数量的关于《平凡的世界》人物分析的文献。《平凡的世界》符合第一个标准。此外,《平凡的世界》描写了 1975—1985 年十年间,发生在我国黄

土高原某个省份的故事。有研究发现,“也许是因为路遥的精心提炼,《平凡的世界》里出现的方言土语词汇,非常明显地体现的是陕北的纯真和质朴方面,这就是《平凡的世界》要向我们展示的陕北特有的文化品格”<sup>[19]</sup>。《平凡的世界》中人物的语言以陕北方言为主,属于中原官话,隶属北方官话即北方方言。因此,《平凡的世界》也符合第二条标准。另外,“在中国文学的接受史上,《平凡的世界》具有某个方面的代表性意义:即以研究者和文学史所代表的学术界与评论者和读者大众之间存在着巨大的观点分歧。”<sup>[18]</sup>这使得《平凡的世界》具有特殊意义。

### 1.2 对话抽取

本研究采用网络电子版《平凡的世界》<sup>①</sup>进行对话抽取。为了尽量减少抽取过程中的主观影响,规定具有双引号的句子<sup>[20]</sup>为一个对话单位。具体抽取过程如下:

第一步,对文本进行预处理,去掉不属于文本的附加内容。将文本按照页码等分成 11 个部分,由 11 名研究生按特定格式将小说文本中所有的对话提取出来,格式如表 1 所示。

表 1 对话抽取格式

页码	行数	发言人	对话内容
354	44	孙少平	我已经习惯外面的这种生活
355	1	孙少安	这外面有个什么好处? 受死受活, 你能赚几个钱? 回去咱们合伙办砖厂, 用不了几年, 要什么有什么!
355	2	孙少平	钱当然很重要, 这我不是不知道; 我一天何尝不为钱而受熬苦! 可是, 我又觉得, 人活这一辈子, 还应该……

第二步, 汇总得到的 11 份对话抽取结果, 以人物为分类条件拆分对话, 并将每一个人物的全部对话集中至一个新的文件中, 最后得到以人物名字命名的多个文件。这些文件就是系统分析的对象。

最终我们得到了 220 名角色共 3 176 句话语, 包含 79 573 字和 18 719 个标点。小说中 220 名角色的话语句数的分布情况如图 1 所示。

① [http://down.txt99.cc/Zip\\_25023/%E5%B9%B3%E5%87%A1%E7%9A%84%E4%B8%96%E7%95%8C.zip](http://down.txt99.cc/Zip_25023/%E5%B9%B3%E5%87%A1%E7%9A%84%E4%B8%96%E7%95%8C.zip), 经核对与纸质版内容相同。

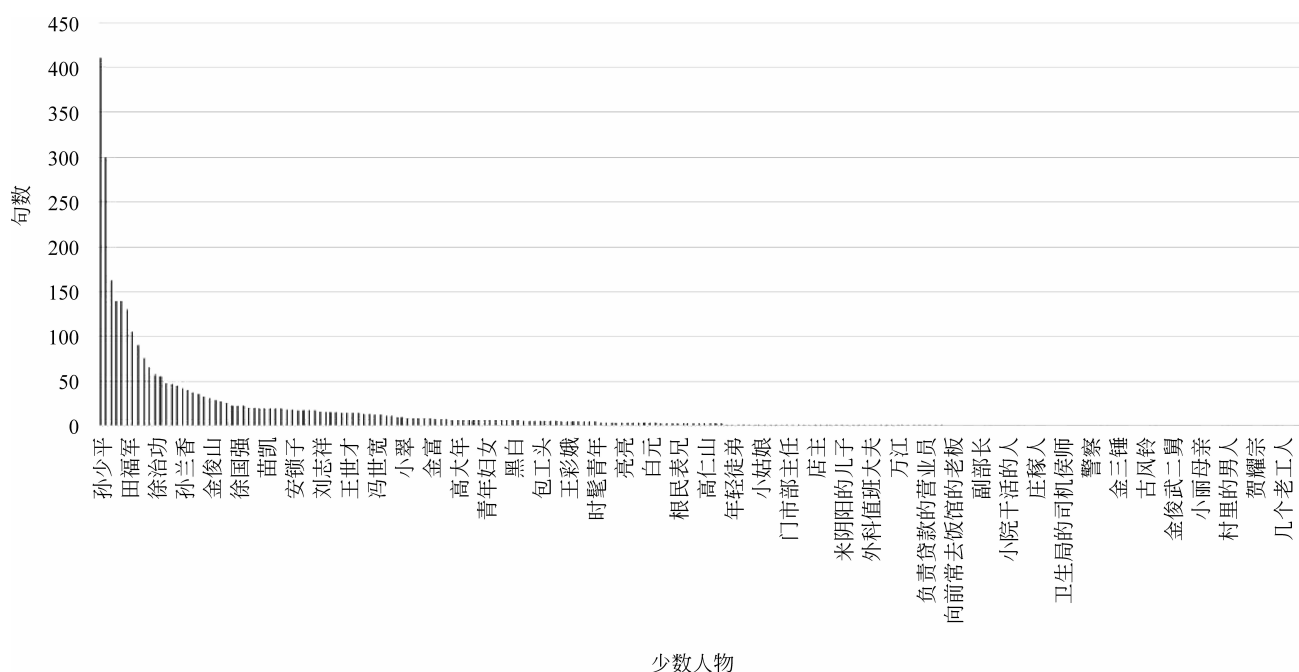


图1 对话统计情况

### 1.3 人物选取

语言艺术家叶圣陶在谈到文艺作品创作的时候认为：“写个作品总有个中心思想，跟中心思想有关系的就该用，而且非用不可，没关系的就不该用，用了就是累赘。”<sup>[21]</sup>叶先生指出了小说语言具有选择性的特点。小说《平凡的世界》的语言也具有明显的选择性。《平凡的世界》主要有三条线索，一是以孙少平为中心，“写了由于城乡差别和脑力劳动与体力劳动的差别，孙少平、金波、兰香、金秀等农村青年渴求知识，渴望现代社会城市的文明和丰富的精神生活与物质生活”；二是以孙少安为中心，“主要写了极左路线

给双水村人造成的贫困和三中全会后，双水村人奔富裕之路的艰难历程”；“第三条线索以田福军的升迁为序，展示了由村到县、地、省的政治斗争和路线斗争，这条斗争线索时明时暗，一直贯穿于作品的始终。”<sup>[22]</sup>结合我们对小说中人物对话句数的统计，对话句数超过100的有孙少平、孙少安、田润叶、田福堂、田晓霞、田福军和孙玉亭，句数分别为387、324、163、139、139和130，说明以上几个角色是本书的主要人物，跟中心思想的关系较密切。

我们将文献搜索得到的35篇对《平凡的世界》小说人物进行分析的文献进行统计，得到这些文献所分析人物及次数，如图2所示。

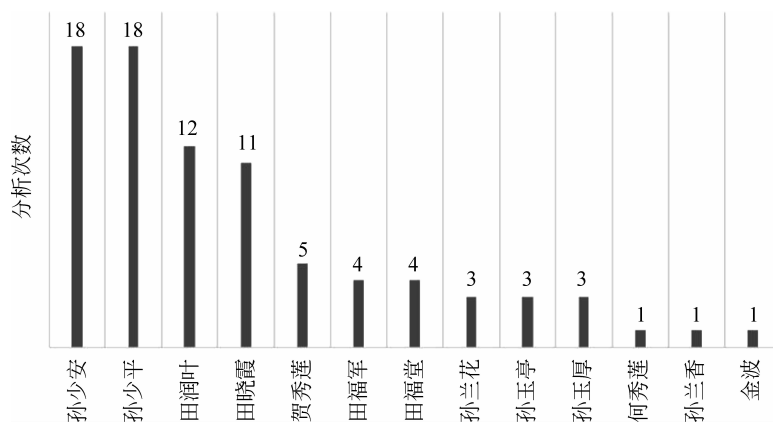


图2 《平凡的世界》小说人物分析频次统计

根据统计结果,我们认为《平凡的世界》中最受关注的人物为孙少安、孙少平、田润叶和田晓霞四个,而且四个人物相互之间既存在对比和差异,也蕴含着共性和统一。具体原因有三:①这四个人物是《平凡的世界》小说读者及研究者最为关注的对象,他们的性格引发了社会大众最强烈的心理共鸣,是社会大众的性格缩影;②这四个人物都是三中全会前后成长的年轻一代,避免了年龄因素的混淆;③孙氏两兄弟的对话代表了农家子弟的心路历程,而田氏两姐妹代表着城市干部子弟的心路历程,二者形成鲜明对比。因此,我们选择以孙少安、孙少平、田润叶和田晓霞四个人物为线索,评价中文心理分析系统用于小说人物性格分析的有效性。

#### 1.4 文本分析及人格预测

本文采用由中国科学院心理研究所计算网络心理实验室研发的“文心”中文心理分析系统(Text Mind)<sup>[23]</sup>进行对话分析和大五人格预测。“文心”系统主要由三部分组成,分别为中文分词工具、心理分析词典和大五人格预测模型。

中文分词工具基于哈尔滨工业大学社会计算与信息检索研究中心研制的语言技术平台(language technology platform, LTP)<sup>[24]</sup>, LTP 是一整套开放中文自然语言处理系统,通过切分词序列、词性标注、命名实体识别、句法分析、语义角色标注技术,最终达到通过计算机分析自然语言的目的。

“文心”系统的简体中文心理分析词典(simplified Chinese LIWC dictionary, SCLIWC)<sup>[25]</sup>参照 LIWC2007 和正体中文 C-LIWC 词库,由 21 名研究生分成三个翻译小组按照简体中文的使用习惯分别进行人工翻译,再由三名熟悉 LIWC 词典结构的研究生对三组翻译结果进行评议,最终采用得到最多研究生认可的翻译进入 SCLIWC。中文心理分析词典有 102 个词类、共 6 547 个词,包括与尽责性相关的成就词、情绪性相关的焦虑词、外向性相关的朋友词等词类,词类之间可相互重叠,也包括对标点符号和词长的统计(本文统称为词类,完整词类划分参

见“文心”系统说明)<sup>①</sup>。“语言探索与字词计数”(linguistic inquiry and word count, LIWC)是 20 世纪 80 年代 Pennebaker 等人在研究情绪书写的治疗效果时发明的基于计算机软件程序的文本分析工具,一种可以对文本内容的词语类别(尤其是心理学类词语)进行量化分析的软件<sup>[26]</sup>。经过 20 多年的发展,由于 LIWC 的良好信效度,该工具被心理学研究者广泛应用到各个方面,并被多国研究者引进。

大五人格预测模型<sup>[15]</sup>是中国科学院心理研究所计算网络心理实验室基于新媒体大数据和深度学习技术,为进行网络心理的研究而研发的工具。研究者邀请 563 名活跃微博用户在线填写大五人格量表(BFI-44),并将筛选后的 444 名合格用户作为训练数据集。然后利用量表得到的心理特征得分对微博用户的网络数据进行标注。我们随后利用 Pace-Regression 算法训练预测模型。PaceRegression 是一种基本的线性回归算法,比较适合高维度的回归模型建立。我们使用 Weka3.8 分类器中的 Pace-Regression 模块,使用系统的默认参数(debug 为 false, estimator 为 Empirical Bayes, threshold 为 2.0),训练得到用户大五人格的预测模型<sup>[15]</sup>。在大五人格五个维度,模型预测值和量表测量结果之间的相关性系数为 0.48~0.54,达到中等相关,预测精度好于其他经典算法。我们利用预测模型通过微博文本就可以自动计算出任意微博用户的大五人格得分,无需通过量表进行测量。微博内容一般是较口语化表达,与小说人物对话形式比较接近,所以本研究尝试将此预测模型用于小说人物性格分析。

“文心”系统对《平凡的世界》小说人物大五人格的预测主要分为三个步骤。首先,对该人物的全部对话进行分词。然后,根据中文心理分析词典的词类划分对分词得到的所有词汇进行词汇统计。最后,大五人格预测模型通过对词类统计的结果进行分析,得到该人物大五人格的预测分数。简要流程如图 3 所示。预测分数为标准化的分数,是以 50 为界的两极性分数,某个角色某维度的分数越接近 100,代表着该角色在该维度的某种倾向越强,越接近 0 则与之相反。

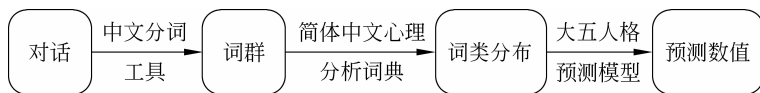


图3 文心系统工作流程简图

以上的分析方法实际上可以看作是语料库语言学领域词频统计研究的延伸,同时也是心理学理论

① 词表下载地址为: <http://ccpl.psych.ac.cn/textmind/>

分析小说人物性格的延伸。在语料库语言学领域,已有研究通过对情态动词(can、could)和高频词<sup>[27]</sup>的统计分析小说人物性格。心理学理论对小说人物性格的分析也已有先例<sup>[28]</sup>。而我们则在语料库语言学的词汇统计和心理学理论的性格分析两个方面都延伸了一步,以心理学大五人格理论为依托,利用文心系统对人物对话进行智能分析得出小说人物大五人格特质的预测分数,完成将文学智能分析引入小说人物心理分析的尝试。

## 2 结果与分析

### 2.1 大五人格预测分数

利用小说中抽取出来的主要人物的对话,我们调用“文心”中文心理分析系统中经过训练得到的中文分词工具、简体中文心理分析词典和大五人格预测模型,对每个人物的大五人格进行了预测,结果如表 2 所示。

表 2 四个主要人物大五人格的预测分数

姓名	孙少安	孙少平	田润叶	田晓霞	平均数 (标准差)
A 宜人性	67.91	59.55	62.7	72.5	65.65 (4.95)
C 尽责性	84.23	88.81	96	66.1	83.78 (11.05)
E 外向性	94.13	84.89	95	82.5	89.13 (5.50)
O 开放性	81.97	96.94	73.8	94.4	86.76 (9.39)
N 情绪性	75.61	66.79	63.8	61	66.80 (5.48)

预测结果所对应的该人物的大五人格分数,以及预测分数所显示出的该人物在各维度的倾向性,可以作为对该人物进行极化分析的参考。下面,我们结合文献对四个主要人物的大五人格预测得分进行分析讨论。

### 2.2 孙少安

从预测数值上看,孙少安具有极强的外向性,较强的尽责性和开放性。

极强的外向性表现为精力充沛、与人交往热情和慷慨。小说中的孙少安表现出精力充沛,“永远斗

志昂扬”<sup>[29]</sup>,与人交往热情<sup>[29-30]</sup>、慷慨<sup>[30]</sup>,这些都是极强外向性的典型表现。

较强的尽责性表现为较强的责任感、做事果决和可信赖。孙少安表现出对家庭的“强烈的责任感”<sup>[31-32]</sup>,做事果决“精明能干”<sup>[33]</sup>和可信赖“沉稳而又有主见”<sup>[34]</sup>,这些都符合较强尽责性的典型表现。

较强的开放性表现为具有独创性、善于创造和易于接受新思想。孙少安有“开创意识”<sup>[35]</sup>、善于“创新”<sup>[34,36]</sup>,易于接受新思想“接受新思想新事物比较快”<sup>[29,37-38]</sup>,这些特点说明他具有典型的较强开放性。

### 2.3 孙少平

孙少平具有极强的开放性,较强的尽责性和外向性。

极强的开放性表现为有活跃的梦想、对许多不同的事物保持好奇和超乎常人的精神深度。小说中的孙少平“极爱幻想”<sup>[36,39-40]</sup>,对双水村外面的世界“充满向往”<sup>[31,37,41]</sup>,热爱阅读,有着“更高的精神追求”<sup>[36-38]</sup>,这些特点反映出极强的开放性。

较强的尽责性表现为对目标的坚持、可依赖和负责任。孙少平面对工作中的困难时表现出“坚忍不拔”的奋斗精神<sup>[35,42-43]</sup>,让人觉得“踏实稳重”<sup>[40]</sup>、“独立自主”<sup>[37,39]</sup>可以依赖,同时“富有责任心”<sup>[37,44]</sup>,这些特点都来自于较强的尽责性。

较强的外向性表现出热心、积极和自信的特点。小说中的孙少平多次见义勇为、“热心”<sup>[36-37]</sup>,在学习和工作时都“积极”、“进取”<sup>[32,38]</sup>,在工作岗位上表现出“自信刚毅”<sup>[29]</sup>,这些特点都反映出孙少平典型的较强的外向性。

### 2.4 田润叶

田润叶有极强的尽责性和外向性。

极强的尽责性表现为对义务的认可和主动承担。小说中的田润叶表现出鲜明的自我牺牲的品质<sup>[34,45-47]</sup>。田润叶在小说中有两次重大的牺牲。第一次是为了二叔田福军的仕途牺牲自己的幸福而“违心地”答应嫁给李向前<sup>[48]</sup>;第二次是李向前出车祸而截肢以后,她认为自己应该尽到“妻子的义务和责任”<sup>[34]</sup>。传统道德观念的影响,使得她觉得对于自己的家族、家庭有必须承担的义务和责任,使她做出的选择“都是以牺牲自己的幸福或快乐为前提”<sup>[45]</sup>。这种自发的义务和责任感表明她具有典型

的极强的尽责性。

较强的外向性表现为主动和热情。田润叶在和孙少安的感情上一直处于“主动”追求的地位<sup>[34,37,49]</sup>,而且待人“热情”<sup>[48]</sup>,在社交性强的工作岗位上“做出了不俗的成绩”<sup>[29]</sup>,这些表现说明她具有较强的外向性。

## 2.5 田晓霞

田晓霞有极强的开放性,较强的外向性。

极强的开放性表现为易于接受新思想、喜爱冒险和有开放的心态。小说中的田晓霞的“思想敏锐”<sup>[45-46]</sup>、“善于思考”<sup>[30,50-51]</sup>,同时又“喜欢冒险”<sup>[37,45,47,49]</sup>,具有“开放的心态”<sup>[45]</sup>，“不安于平庸的生活”<sup>[37]</sup>,这些特点说明她具有典型的极强的开放性。

较强的外向性表现为活跃、积极和主动。田晓霞的性格“活泼大方”<sup>[46,48-49]</sup>,与人交往大胆、“洒脱”<sup>[45,48]</sup>,对待感情“敢爱敢恨”<sup>[45,47-48]</sup>,这些表现都属于典型的较强的外向性,因此田晓霞具有较强的外向性。

## 3 总结

通过对比文心系统对《平凡的世界》主要人物大五人格的预测分数和文艺学文献、小说关键剧情可以发现:开放性相对较强的孙少平和田晓霞是思想最超前的;尽责性较强的孙少平和田润叶是受传统道德观念影响较深的;外向性较强的孙少安和田润叶是年长且在社会上表现出良好交际性的;宜人性较强的是谦让、忍耐的孙少安和活泼大方的田晓霞;情绪性较强的孙少安、孙少平是因为贫穷而在生活中充满挫折和矛盾的。因此,文心系统的预测分数符合文艺学文献对人物性格的理解和作者对人物性格的塑造。通过文心系统完成的文学智能分析客观而全面地呈现了《平凡的世界》主要人物的大五人格特质,同时直观地反映出各人物间的心理差异,很好地反映了三中全会前后成长起来的一代人的心理特点。

因此,通过生态化识别将文学智能分析应用于小说人物心理分析是有效的。同时,文学智能分析的操作可重复性、客观性和对大型语料库的处理能力,使其更加具有实用潜能。

本研究也存在一定的不足。例如小说人物对话和微博内容的语言存在语用差异和语言演变带来的

差异。然而,虽然二者都可能限制人格预测分数的精确度,但对于通过大样本微博内容训练的文心系统而言,这种差异和微博内容内部不同群体的差异相似,在大样本的情况下对文学智能分析的有效性影响不大。因此对于这些问题的解决实际上属于模型优化的问题,可以按照对象语言的特点如年代(时间因素)、地域(空间因素)等选取尽可能相似的语料训练文心系统以提高精确度。另外,本研究可以更进一步地使用文学人物样本和人工一致性判定方法来说明方法本身的有效性。

本研究除了说明小说人物心理文学智能分析的可行性,也为文学领域的小说人物心理分析提供了一个新的研究方法,这种文学智能分析的方法简化、客观化了文艺学的心理批评,伴随着模型预测精确度的提高,将其应用于一些富有争议性的小说人物的研究,进而探求作者的创作心理,甚至解决一些文学公案,或者用以探索计算机对文学作品的学习,都是可能且相当有意义的。

## 参考文献

- [1] 包桂英. 小说《围城》中人物性格塑造及分析[J]. 作家, 2013(02): 15-16.
- [2] 曹萍. 巴金小说人物形象的性格矛盾分析[D]. 杭州: 浙江师范大学硕士学位论文, 2011.
- [3] 胡静. 从合作原则的违反剖析《傲慢与偏见》中的人物性格及人物关系[D]. 西安: 长安大学硕士学位论文, 2014.
- [4] 郭丽. 话轮转换框架下会话对文学作品中人物塑造的作用研究: 对《追风筝的人》的个案分析[D]. 天津: 天津大学硕士学位论文, 2014.
- [5] 黄银玲. 评斯坦贝克的《逃亡》: 他还尚未成人——基于语料库的小说人物语言分析[J]. 文学界(理论版), 2012(07): 103-105.
- [6] 隋晓蕾. 从评价理论的角度看小说《简·爱》女主人公的人物性格特征[D]. 长春: 东北师范大学硕士学位论文, 2007.
- [7] 刘智跃. 精神分析与新时期文学心理批评[J]. 当代文坛, 2008(06): 51-54.
- [8] 郭永玉, 张钊. 人格心理学的学科架构初探[J]. 心理科学进展, 2007, 15(02): 267-274.
- [9] 郭永玉, 胡小勇. 特质、动机和叙事: 人格研究的三种范式及其整合[J]. 心理科学, 2015, 38(06): 1489-1495.
- [10] Lewis R Goldberg. An alternative “description of personality”: the big-five factor structure[J]. Journal of Personality & Social Psychology, 1990, 59(6): 1216.

- [11] Veronica Benet-Martinez, Oliver P. John. Los cinco grandes across cultures and ethnic groups: Multitrait multimethod analyses of the big five in Spanish and English[J]. Journal of Personality & Social Psychology, 1998, 75(3): 729-50.
- [12] John O P, Donahue. et al. The big five inventory: Versions 4a and 54[R]. CA: University of California, Berkeley, Institute of Personality and Social Research. 1991.
- [13] John O P, Naumanm, et al. Handbook of personality: Theory and research[M]. New York: Guilford Press, 2008: 114-158.
- [14] Youyou W, Kosinski M, Stillwell D. Computer-based personality judgments are more accurate than those made by humans[J]. Proceedings of the National Academy of Sciences of the United States of America, 2015, 112(4): 1036-1040.
- [15] Li L, Li A, Hao B, et al. Predicting active users' personality based on micro-blogging behaviors[J]. PLOS ONE, 2014, 9(1): 1-11.
- [16] 路遥. 平凡的世界[M]. 北京: 十月文艺出版社, 2017.
- [17] 黄伯荣, 廖序东. 现代汉语(上)[M]增订五版. 北京: 高等教育出版社, 2011: 4.
- [18] 贺仲明. “《平凡的世界》现象”透析[J]. 文艺争鸣, 2005(04): 115-118.
- [19] 武杰. 《平凡的世界》文学语言研究[D]. 西安: 陕西师范大学硕士学位论文, 2013.
- [20] 黄伯荣, 廖序东. 现代汉语(下)[M]. [增订五版]. 北京: 高等教育出版社, 2011: 4.
- [21] 叶圣陶. 关于使用语言[J]. 人民文学, 1956(03): 121-125.
- [22] 孙悦. 简析《平凡的世界》的结构特点[J]. 贵州教育学院学报(社会科学版), 1998, 01: 42-44.
- [23] Gao R, Hao B, Li H, et al. Developing simplified Chinese psychological linguistic analysis dictionary for Microblog[C]//莫雷. 心理学与创新能力提升——第十六届全国心理学学术会议论文集. 南京: 中国心理学会, 2013: 364-366.
- [24] 刘挺, 车万翔, 李正华. 语言技术平台[J]. 中文信息学报, 2011, 25(06): 53-62.
- [25] Nan Zhao, Dongdong Jiao, Shuotian Bai, et al. Evaluating the validity of simplified Chinese version of LIWC in detecting psychological expressions in short texts on social network services[J]. PLOS ONE, 2016, 11(6): 1-15.
- [26] 张信勇. LIWC: 一种基于语词计量的文本分析工具[J]. 西南民族大学学报(人文社会科学版), 2015(04): 101-104.
- [27] 王俭. 基于语料库的海明威小说评论研究[D]. 上海: 上海外国语大学博士学位论文. 2012.
- [28] 范勤. 从依恋理论角度解析《呼啸山庄》中人物的悲剧性格[D]. 沈阳: 辽宁大学硕士学位论文, 2016.
- [29] 魏娟辉. 《平凡的世界》中人物形象的美学价值[J]. 西安文理学院学报(社会科学版), 2015, 18(03): 26-29.
- [30] 王倩. 论路遥的《平凡的世界》人物特色及形象层次性[J]. 鸭绿江(下半月版), 2014(06): 54-56.
- [31] 石欢欢. 《平凡的世界》: 平凡人物的“平凡世界”[J]. 名作欣赏, 2013, 09: 14-15.
- [32] 邵晶. 论《平凡的世界》孙少安的性格特征及其根源[J]. 中国西部, 2016, 11: 15.
- [33] 陈伟. 生活之树常青: 谈《平凡的世界》人物塑造[J]. 大庆高等专科学校学报, 1997, 17(03): 51-53.
- [34] 刘新生, 朱献贞. 拒绝蔑视: 论《平凡的世界》人物形象的精神探索[J]. 山东文学, 2000, 08: 61-63.
- [35] 张静. 《平凡的世界》主要人物形象分析: 以孙少安和田润叶为例[J]. 大众文艺, 2015, 21: 20.
- [36] 赵文煜. 论《平凡的世界》人物奋斗观[J]. 经营管理者, 2010(12): 324.
- [37] 许云生. 《平凡的世界》的人物造型[J]. 湖州师范学院学报, 1994(02): 42-47.
- [38] 汤思源. 《平凡的世界》人物形象分析[J]. 北方文学(下半月), 2011, 11: 36-37.
- [39] 密淑敏. 评《平凡的世界》中孙家两兄弟的性格差异[J]. 群文天地, 2012, 10: 46.
- [40] 金理. 在时代冲突和困顿深处: 回望孙少平[J]. 文学评论, 2012, 05: 136-142.
- [41] 葛有进. 路遥的《平凡的世界》人物形象分析[J]. 短篇小说(原创版), 2014, 30: 29-30.
- [42] 任桂娟. 浅析《平凡的世界》中孙少平、孙少安的形象内涵[J]. 黑龙江教育学院学报, 2008, 27(08): 131-132.
- [43] 周春雷. 路遥代表作《平凡的世界》人物思想分析[J]. 短篇小说(原创版), 2014, 12: 25-26.
- [44] 林国丽, 王天华. 小说《平凡的世界》中的人物情感建构[J]. 哈尔滨工业大学学报(社会科学版), 2016, 18(03): 101-105.
- [45] 牛夏. 从女性主义批评视角看路遥的《平凡的世界》[J]. 牡丹江师范学院学报(哲学社会科学版), 2013(05): 55-57.
- [46] 丁红梅, 王圣. 男权思想统照下的女性世界: 浅谈路遥《平凡的世界》中的几个女性形象[J]. 淄博学院学报(社会科学版), 2002, 18(01): 49-51.
- [47] 陈丽珍. 平凡的女性不平凡的人生: 浅论《平凡的世界》中三种女性的类型化形象[J]. 苏州职业大学学报, 2005, 02: 55-57.
- [48] 韩婷. 平凡女性的悲剧人生: 评《平凡的世界》中的人物田润叶[J]. 丹东师专学报, 2003, 25(01): 26-29.
- [49] 王季路. 《平凡的世界》中的三大女性角色研究[J]. 科教导刊(上旬刊), 2016, 04: 143-144.
- [50] 付慧颖. 有关《平凡的世界》三大女性角色的赏析[J]. 读书文摘, 2016, (08): 13.
- [51] 黄梓烜. 从《平凡的世界》中解读田晓霞的性格特点[J]. 环球人文地理, 2016, (16): 310.





吴育锋(1994—),本科,主要研究领域为计算语言学。

E-mail: 33120142201411@stu.xmu.edu.cn



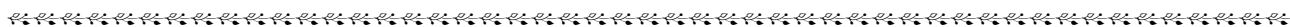
吴胜涛(1980—),博士,副教授,主要研究领域为文化价值观、社会情绪、大数据分析等。

E-mail: michaelstwu@xmu.edu.cn



朱廷劭(1971—),通信作者,博士,研究员,主要研究领域为机器学习、大数据心理。

E-mail: tszhu@psych.ac.cn



(上接第 108 页)

- [22] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [J]. arXiv preprint, 2014, arXiv: 1409.0473.

- [23] Srivastava N, Hinton G, Krizhevsky A, et al. Drop-

out: A simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.



宋皓宇(1994—),硕士研究生,主要研究领域为人机对话系统和自然语言处理。

E-mail: hysong@ir.hit.edu.cn



张伟男(1985—),博士,讲师,主要研究领域为人机对话系统、自然语言处理和信息检索。

E-mail: wnzhang@ir.hit.edu.cn



刘挺(1972—),博士,教授,主要研究领域为自然语言处理、文本挖掘和文本检索。

E-mail: tliu@ir.hit.edu.cn