AI为文学角色“测”人格

作者：胡珉琦 来源：中国科学报

本文原地址：https://www.iikx.com/news/progress/2257.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！



表1 王人公性格预测

	宜人性	尽责性	外向性	开放性	神经质
孙少安	67.91	84.23	94.13	81.97	75.61
孙少平	59.55	88.81	84.89	96.94	66.79

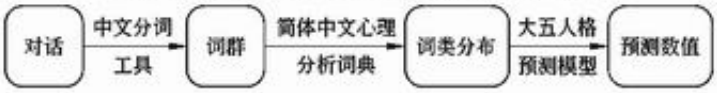
表2 结婚前后孙少安的性格变化

	宜人性	尽责性	外向性	开放性	神经质
结婚前	81.27	85.84	95.55	88.55	60.87
结婚后	60.69	83.38	93.36	78.38	83.59
差值	-20.58	-2.46	-2.19	-10.17	22.72
变化率（%）	-0.25	-0.03	-0.02	-0.12	0.37

表3 升学前后孙少平的性格变化

	宜人性	尽责性	外向性	开放性	神经质
升学前	50.89	82.70	53.25	89.94	77.85
升学后	62.18	90.67	94.47	99.03	63.44
差值	11.29	7.97	41.22	9.09	-14.41
变化率(%)	0.18	0.09	0.44	0.09	-0.23

用生态化识别的方法分析文学人物性格流程



AI为文学角色“测”人格。▲以《平凡的世界》中的主人公孙少安、孙少平兄弟为例，分析两位主人公的总体性格与经历重大生活事件后的性格变化。结果如表1、2、3所示。

在心理学中，要准确、客观测量一个人的性格，一般通过自我报告或填写性格测试来进行评价。可如果测量的对象是一个文学作品中的虚构人物呢?

传统文学分析方法只能基于定性研究，且与研究者的主观体验密切相关。如今，得益于生态化识别技术的人工智能分析法，科学家通过捕捉人物对话就能实现对文学角色的人格测量。

近期，中科院心理所和厦门大学科研团队的这一合作成果在线发表于《中文信息学报》和国际学术期刊Digital Scholarship in the Humanities。

文学人物定量分析难

1975年至1985年间，中国土地所有制改革开始实行，农村跟着进入了发展黄金期。作家路遥以一对兄弟——耿直、坚韧、质朴的孙少安和同样坚韧又更敏锐、灵活的孙少平的生存与生活故事，展示了中国西北农村一段荡气回肠的变迁史。

小说，是以刻画人物形象为中心的。《平凡的世界》作为当代文学史上的一部巨作，它所塑造的经典人物形象是许多文学批评家、评论家、作家研究和分析的对象。

而通过对小说人物性格的分析，可以帮助读者更准确、深刻地理解人物的语言、行动和内心世界，以及它所反映的现实背景。

论文通讯作者、中科院心理所研究员朱廷劭表示，单纯在文学界对小说人物的性格主要是定性分

析，不过，这样的心理分析主要依靠的是研究者在研究过程中的主观体验和文学素养。比如，通常研究者会从微观方面将小说中的一个或多个人物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。

另一种思路是从语言学的角度出发，通过对情态动词和高频词的统计，分析小说人物性格。这是一种定量方法，但并不主流。

说到对人的内在分析，心理学是有系统的、科学的研究方法的。只是过去心理学的研究对象都是真实存在的，而非虚构人物。但引入心理学方法研究文学人物完全是可行的。

研究显示，目前，小说人物的心理分析主要依据心理学家提出的人格三层次模型。其中，最基础的是作出对人格特质最基本的描述。人格特质是个体行为的稳定的、显著性倾向;第二层是个人关注，是描述个人奋斗、生活任务、防御机制、应对策略等大量有关人格动机和策略等方面的建构，从而了解一个人在不同情境下的动机、关切和策略;第三层是生活叙事，需要去探究一个人从出生到成年的发展历程，也就是人生故事。

但朱廷劭指出，相较于后两层，通过最初的人物性格或人格特质分析对文学角色进行的讨论是最少的。原因就在于，这类分析必须依据基于自我报告的心理测量，显然，这在虚构人物身上很难实现。

基于对话的智能分析法

不过，在有了生态化识别技术之后，自我报告的测量方法就不再是虚构人物人格分析的障碍了。这一技术是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。

在此之前，中科院心理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，开发了一款网络心理的研究工具——大五人格预测模型。简单说，研究人员可以利用社交媒体内容与大五人格量表的映射关系，对社交媒体使用者的人格进行自动识别，而无须通过量表进行测量。经过检验，这个模型预测值和量表测量结果之间的相关性已经达到了中等程度。

那么，小说人物能否使用大五人格预测模型进行分析?答案是肯定的。

社交媒体的内容一般是个人较为口语化的自我表达，也就是说，如果研究人员能把小说中与该人物相关的自我表达内容提取出来，录入模型，便可以得到预测结果。

在小说里，最符合人物自我表达属性的内容就是对话(直接引语)。以《平凡的世界》为例，研究人员首先按特定格式将小说文本中所有的对话提取出来，并以人物为分类条件拆分对话。每个人物的对话集就是系统分析的对象。

不过，角色有主次，分配到每个角色的对话体量也是不同的，体量过小必然会影响模型预测的有效性。孙少安、 孙少平、
田润叶和田晓霞是小说的4个主要人物，不仅具有人物代表性，得到作者的笔墨也是最多的。

接下去，研究人员需要对这4个人物各自的对话内容进行预处理。首先，他们利用中文分词工具，将所有自然语言进行拆分;然后，通过一种量化分析软件——中文心理分析词典，对分词得到的所有词汇进行词汇统计，得到完整的词类分布。该心理分析词典共有102个词类、6547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。

最后，根据词类统计的结果使用大五人格预测模型进行分析，得到该人物大五人格的预测分数，包括人物的宜人性、尽责性、外向性、开放性和神经质的分数。

该研究结果显示，从预测数值上看，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。

朱廷劭表示，在与现有的文艺学文献对这一小说人物性格的分析研究成果进行比较之后，文学智能分析的结果是与前者相符合的，从而证明了预测是有效的。

人工智能与文学的联合

这是一次客观的、量化的科学分析方法与文学的结合，给文学人物分析提供了一种全新的思路，可以屏蔽一些文学主观评价带来的偏差。

不过，仅将这种方法用于文学作品的个案分析意义有限。朱廷劭告诉《中国科学报》记者。

他认为，这种具有客观性、可重复性和处理大型语料库优势的技术，可以同时用于大样本量的分析，试图去寻找不同历史时期、不同时代背景下的文学作品中，个体人格特质、命运与所处环境之间可能存在的独特的关系。

此外，文学领域存在大量的纪实作品，比如人物传记、口述史、书信等，文学智能分析还可以为那些与我们相隔久远的历史人物进行一次科学的人格测试，作为史料分析的新的客观手段，也不失为一种有价值的应用方向。(来源：中国科学报 胡珉琦)

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发