

四大名著文本中的无标度规律

孙龙龙¹, 顾长贵¹, 冯靖², 吴果林³

(1. 上海理工大学 管理学院, 上海 200093; 2. 上海工程技术大学 高等职业技术学院,
上海 200437; 3. 桂林航天工业学院 理学院, 桂林 541004)

摘要: 从中国四大名著中分别提取每段句子数时间序列、每段字数时间序列和每句字数时间序列。采用去趋势波动分析法从以上各个层次分析每一名著, 发现在每个层次上每一名著均显示出无标度规律且标度指数都在 0.60 左右, 说明在各个层次上文本都表现出长程相关。进一步, 将每段字数时间序列划分成大约 10 个长度为 1 000 的时序片段, 使用去趋势波动分析发现《红楼梦》和《水浒传》这两本小说前后部分的标度指数都存在显著差异, 即这两本著作前部分的标度指数大约为 0.55, 而后部分大约为 0.65。这一显著变化, 佐证了《红楼梦》的作者为曹雪芹和高鹗的说法, 支持了《水浒传》可能是施耐庵和罗贯中合作完成的论点。

关键词: 四大名著; 去趋势波动分析法; 无标度规律

中图分类号: N 941 文献标志码: A

Scale-Free Characteristics of Four Great Classical Novels

SUN Longlong¹, GU Changgui¹, FENG Jing², WU Guolin³

(1. Business School, University of Shanghai for Science and Technology, Shanghai 200093, China;

2. Advanced Vocational Technical College, Shanghai University of Engineering and Technology,

Shanghai 200437, China; 3. Faculty of Science, Guilin University of Aerospace Technology, Guilin 541004, China)

Abstract: In the present study, three kinds of time series from Four Great Classical Novels were extracted, including the number of sentences in a paragraph (NSAP), the number of characters in a paragraph (NCAP) and the number of characters in a sentence (NCAS). The above time series were analyzed by the method of detrended fluctuation analysis (DFA). It is found that the scale-free characteristics exist at the paragraph level and the sentence level, and the scaling exponents are close to 0.60. Our finding indicates that long-correlations exist at different levels. Moreover, the time series of NCAP can be divided into ten over-capping segments with the length of 1 000. Based on the detrended fluctuation analysis, it is found that the scaling exponents for the time series of NCAP are significantly different between the first part and the second part of *A Dream of Red Mansions* as well as *All Men are Brothers*. To be specific, the scaling exponents of the first and second parts are about 0.55 and 0.65, respectively. This significant change in scaling exponents confirms that *A Dream of Red Mansions* was

收稿日期: 2018-01-04

基金项目: 国家自然科学基金资助项目(11505114); 上海市教委“青年东方学者”人才计划(QD2015016); 上海市高校青年教师资助计划(1016303806)

第一作者: 孙龙龙(1991-), 男, 硕士研究生. 研究方向: 非线性时间序列. E-mail: sll128245@163.com

通信作者: 顾长贵(1982-), 男, 副教授. 研究方向: 生物网络的动力学模型. E-mail: gu_changgui@163.com

finished by Cao Xueqin and Gao E . In addition, there is much chance that *All Men are Brothers* was written by Shi Naian and Luo Guanzhong together. Thus, our finding provides some suggestions for the controversy with respect to authors in both *A Dream of Red Mansions* and *All Men are Brothers*.

Keywords: *Four Great Classical Novels; detrended flutuation analysis; scale-free characteristics*

近年来,国内外许多学者都对文本语言内的无标度规律进行了详细的研究,并且取得了很多重要的成果。但前人大多数的研究对象都是英文文本,并且大多都是在单词层面,忽略了文本语言在其他层次上的无标度规律。本文的研究主要针对中文文本在句子、段落层次的无标度规律进行展开。

1 文字的无标度规律

人类语言学是非常复杂的社会系统^[1],是人类文化在社会以及生物层面长时间演化的结晶^[2]。在过去的一百年里,统计理论的日益完善,使得人类语言学的研究得到了长足的发展。其中,Zipf^[3]提出了著名的 Zipf’s 定律。它的主要思想是将单词按照其出现频率进行排序,频率最高的单词标记为等级 1(rank 1),频率第二高的单词标记为等级 2(rank 2).....依次排列,在双对数坐标系里单词频率与等级标号呈现负相关规律。由 Zipf’s 定律,文献[4]中将 *Holy Bible* 翻译成多种语言,研究了每一种语言的 Zipf 指数。文献[5]对日文文本中平假名文字频率分布进行研究,指出日文中平假名的出现频率服从 weibull 分布的论点。文献[6]中指出了中文汉字出现频率服从幂律分布的特点。

对于语言内部存在的无标度规律(幂律分布)现象,许多学者也作了深入的研究。Altmann 等^[7]解释了无标度规律在单词层次和单词字母层次之间的演化。Deng 等^[8]对中国近现代小说从汉字使用频率角度加以研究,发现汉字的使用具有无标度规律现象。Montemurro 等^[9]和 Bhan 等^[10]使用去趋势波动分析法(detrended flutuation analysis, DFA)对英文文本和韩文文本加以分析,发现英语和韩语内部存在的无标度规律。在文献[11-12]中,Ausloos 和 Gillet 将英文文本转化为世界语文本,同样也得到了在英文文本和世界语文本中都存在无标度规律的性质。以上研究大多使用去趋势波动分析法分析非汉语语言文本的无标度规律。此外,去趋势波动分析法还可以有效探测非平稳性

时间序列上的无标度规律,在 DNA 核苷酸序列、脉搏信号序列、金融时间序列、天气预测方面都有极为广泛的应用^[13-20]。

目前对于语言文本无标度规律的研究大多基于单词的层面,而没有从语句和段落的更高层次对文本加以研究^[19]。对语言文本而言,语句和段落是其重要的组成单元^[20]。一方面,语句为单词的使用提供了具体的语境,另一方面,语句的逻辑排列形成语句群、段落乃至整个文本,从而清楚地表达作者思想。此外,上述文献的研究对象大都是英文文本,而很少有研究者对中文小说文本进行研究。然而,中文小说文本与西方文学作品有很大差别,即中国的汉字是由绘画引申而来^[21],其后汉字经过几千年的演化使得中国人的思想表达方式与西方有很大不同。概括而言,汉字语言是世界上最为成熟的语言之一,而对于汉字在语句和段落层次上的无标度规律的研究,却很少有学者涉及。本文从语句层次以及段落层次使用去趋势波动分析法对中国古代小说文本四大名著加以研究。

2 四大名著文本数据

2.1 数据来源

本文所用到的数据均来自 4 大名著小说文本。四大名著是中国文学史中的经典作品,它们分别是《红楼梦》(*A Dream of Red Mansions*)简记为 ADRM,《三国演义》(*The Romence of Three Kingdoms*)简记为 TRTK,《水浒传》(*All Men are Brothers*)简记为 AMAB,《西游记》(*The Pilgrimage to the West*)简记为 TPTW。表 1 给出了四大名著相关文本数据。

表 1 四大名著相关文本数据
Tab.1 Related data of Four Great Classical Novels

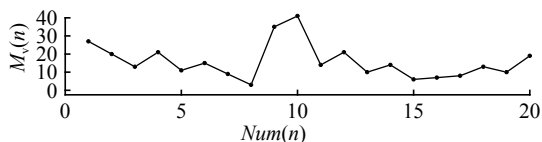
文本	章节数 N_t	段落数 N_v	句子数 N_w
《红楼梦》	120	1 856	34 467
《三国演义》	120	1 665	29 861
《水浒传》	100	2 434	35 859
《西游记》	100	1 947	26 705

2.2 时间序列的提取

本文分别从以上4本小说文本中提取每段字数时间序列、每段句子数时间序列和每句字数时间序列。忽略所有标点符号, 仅计算每段汉字的个数 M_t , 来构建每段字数时间序列; 以句号(‘.’)、问号(‘?’)、感叹号(‘!’)、省略号(‘……’)为一句话结束的标志并忽略其他标点符号, 来统计每句话汉字数 M_v 和每段的句子数 M_w , 并构建相应的时间序列。图1(a)为摘自《红楼梦》中的一段文字, 其中标记在红框内的标点符号为每句话结束的标志。图1(b)为与上述文本相对应的每句字数时间序列 $\xi_i = \{27, 20, 13, 21, 11, 15, \dots\}$, 其中 $Num(n)$ 表示句子数目, $M_v(n)$ 表示每句话的汉字数目。

兴儿又回道:“珍大爷那边给了张家不知多少银子, 那张家就不问了。”凤姐道:“这里头怎么又扯上什么张家李家咧呢?”兴儿回道:“奶奶不知道, 这二奶奶……”刚说到这里, 又自己打了个嘴巴, 把凤姐儿倒惹笑了。两边的丫头也都抿嘴儿笑。兴儿想了想, 说道:“那珍大爷的妹子……”凤姐儿接着道:“怎么样? 快说呀。”兴儿道:“那珍大爷的妹子原来从小儿有人家的, 姓张, 叫什么张华, 如今旁的待好讨饭, 珍大爷许了他银子, 他就退了亲了。”凤姐儿听到这里, 点了点头儿, 回头便望丫头们说道:“你们都听见了? 小忘八崽子, 头里他还说不知道呢。”兴儿又回道:“后来二爷才叫人搬了房子, 娶过来了。”凤姐道:“打那里娶过来的?”兴儿回道:“就在他老娘家抬过来的。”凤姐道:“好罢咧。”又问:“没人送亲么?”兴儿道:“就是蓉哥儿还有几个丫头老婆子们, 没别人。”凤姐道:“你大奶奶没来吗?”兴儿道:“过了两天, 大奶奶才拿了些东西来瞧的。”

(a) 《红楼梦》节选文字



(b) 相对应每句字数时间序列

图1 文本向时间序列的转化

Fig.1 Translation from text to time series

3 无标度规律分析方法

去趋势波动分析法是1994年由Peng等^[13]基于DNA机理提出的标度计算方法。该方法适用于分析非平稳性时间序列的长程相关性, 其优点是它可以有效滤去序列中的各阶趋势成分, 并能检测含有噪声且叠加有多项式趋势信号的长程相关^[13-20]。假设一时间序列为 $\{X_i\}$, $i = 1, 2, \dots, N$, 其中 N 为序列长度。

首先对该序列进行相空间重构, 可得到一系列时间序列片段 Y_n :

$$Y_n = \{x_n, x_{n+1}, \dots, x_{n+s-1}\}, n = 1, 2, \dots, N-s+1 \quad (1)$$

式中, s 为去趋势波动分析法中盒子的大小。

其次, 对每个时间序列片段用 q 阶多项式

$$F(i, n) = \sum_{\beta=0}^q \alpha_{\beta}(n) i^{\beta} \quad (2)$$

进行拟合, 其中 $\alpha_{\beta}(n)$, $\beta = 0, 1, \dots, q$ 均为拟合参数。

拟合曲线值作为每个时间序列片段的趋势。

从原序列片段中减去相应的趋势值, 可得到去趋势时间序列 Z_n :

$$Z_n = \{x_n - F(1, n), x_{n+1} - F(2, n), \dots, x_{n+s-1} - F(s, n)\}, n = 1, 2, \dots, N-s+1 \quad (3)$$

如果原时间序列存在长期相关性, 则有

$$DFA(s) = \sqrt{\frac{\sum_{n=1}^{N-s+1} \sum_{t=1}^s [x_{n+t-1} - F(t, n)]^2}{s(N-s+1)}} \sim s^H \quad (4)$$

式中, H 为标度指数(scaling exponent)。若 $H=0.5$, 则表明时间序列可用随机游走过程来描述; 若 $0.5 < H < 1$, 表明时间序列为长期持续的记忆性过程; 若 $0 < H < 0.5$, 表明时间序列为持续的反记忆性过程。

计算过程中, 多项式拟合函数中的阶数 q 取2, 即用最小二乘法拟合序列片段趋势。用 q 值取2的多项式对序列片段进行拟合, 使得每个标度范围内的数据点数目几乎相同, 提高了结果的准确性^[19]。

4 无标度规律分析结果

4.1 无标度规律分析

对四大名著每段句子数时间序列、每段字数时间序列、每句字数时间序列运用去趋势波动分析法分析, 结果如图2所示。其中: $\ln s$ 为对盒子大小取对数值; $\ln DFA(s)$ 为对涨落的大小取对数值。

每段句子数时间序列和每段字数时间序列的标度指数 H 几乎相同, 且接近于0.60。每句字数时间序列的标度指数 H 则偏小, 但其标度指数 H 值也接近于0.60。这说明中国四大名著小说无论在段落层次还是语句层次上均具有长程相关性, 且标度指数 H 与Holy Bible^[22]在单词长度层次得到的标度指数 H 几乎相同。这也进一步表明在中文小说文本中, 其语言内部无论是段落层次还是语句层次均存在固有的无标度规律。

4.2 时序片段的分析

为了更加细致地研究上述时间序列, 对以上时间序列设置滑动窗口 S , 把时间序列划分成相应的时序片段, 分别对每一个时序片段采用去趋势波动分析法分析。每段句子数时间序列和每段字数时间序列的序列长为 10^3 , 且滑动窗口长度 $S=1\,000$, 每句字数时间序列的序列长为 10^4 , 且滑动窗口 S 取10\,000。在每段字数时间序列结果中, 《红楼梦》和《水浒传》两本小说前后部分存在着标度指数的明显变化, 结果如图3所示。图

3 中(e), (f), (g), (h)分别为《红楼梦》、《三国演义》、《水浒传》、《西游记》在各个时序

片段上的拟合图。其中: $Num(n)$ 表示滑动窗口的数目; $H(n)$ 表示对应滑动窗口的标度指数。

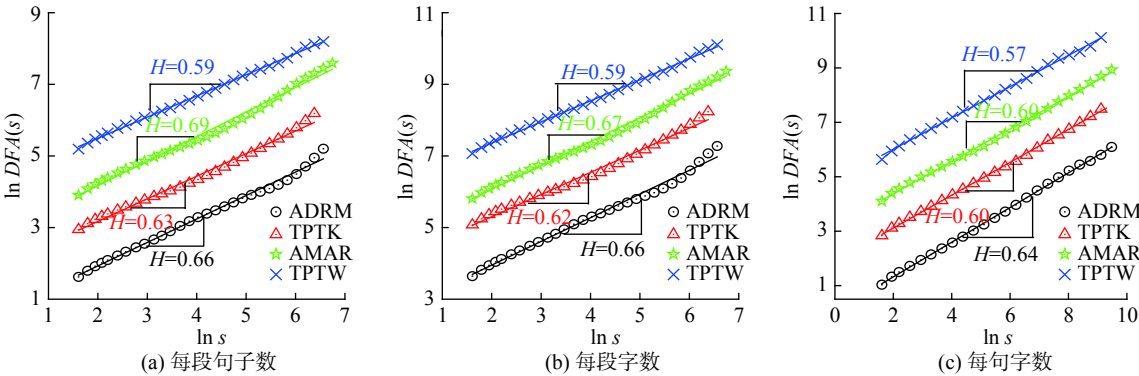


图 2 四大名著在各个层次的标度律
Fig.2 Scaling behaviors of Four Great Classical Novels across all levels

从图 3 可以看出,《红楼梦》和《水浒传》的标度指数变化相对明显,其变化区间分别为[0.55, 0.65]和[0.52, 0.71],且《红楼梦》和《水浒传》存在标度指数转变点 L_1 。《红楼梦》的标度指数转变点 L_1 在第 72 章节,且转变点前或后标度指数相对稳定。《水浒传》的标度指数转变点 L_1 在第 67 章

节,其转变点之前的标度指数呈递减趋势,转变点之后的标度指数呈递增趋势。《三国演义》和《西游记》的标度指数则无明显变化,其浮动区间分别为[0.57, 0.60]和[0.56, 0.59]。分别对《红楼梦》和《水浒传》转变点前后的标度指数做双样本 t 检验 (two-sample t test), 检验结果如表 2 所示。

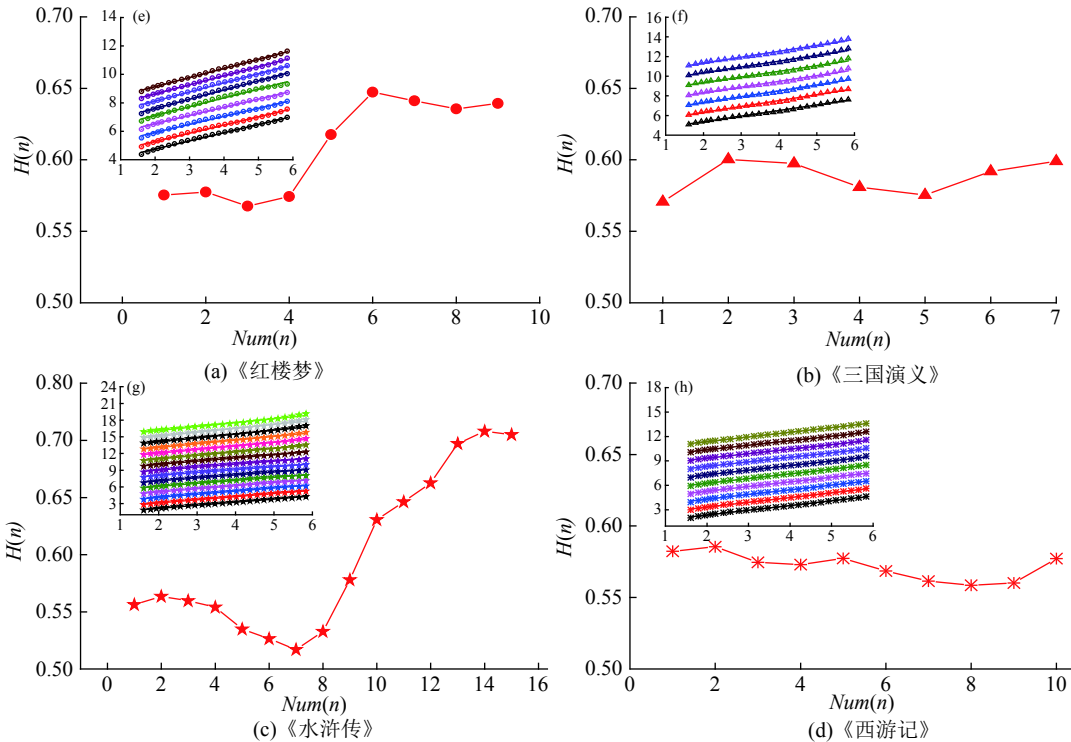


图 3 使用去趋势波动分析法分析每段字数时序片段的结果

Fig.3 DFA results of the time series' segments for the number of characters in a paragraph

对于《红楼梦》和《水浒传》,其标度指数转变点前后差异显著($P<0.001$)。事实上,对于《红楼梦》一书作者的争议一直存在,当下大众比较

认可的一种说法是《红楼梦》由曹雪芹和高鹗两人前后历经十几年时间创作完成,其中曹雪芹创作了前 80 章节,后 40 章节由高鹗在曹雪芹的思

想影响后续写完成^[23]。而对于《水浒传》一书的作者也存在很大争议, 普遍认为《水浒传》一书是由施耐庵一人完成^[24-26]。但也有人指出《水浒传》是由施耐庵和罗贯中共同完成的, 即“施耐庵的本, 罗贯中编次”^[27-29]。本文从数理统计的角度分析, 发现了《红楼梦》和《水浒传》书中每段字数时间序列存在标度指数的转变, 且转变点前后标度指数显著变化。为了验证上述发现, 本文从每段所含信息量的角度作了以下工作加以分析。

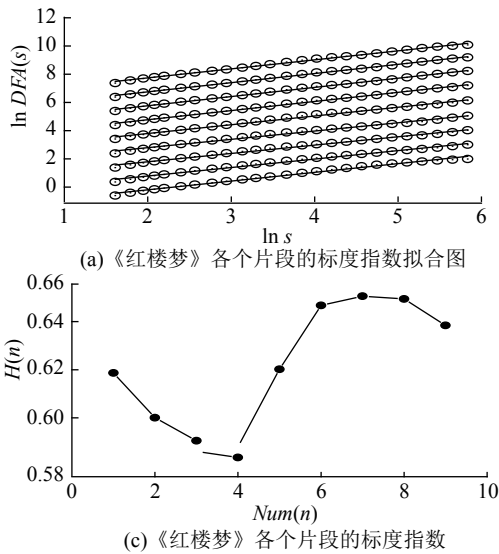
表 2 转变点 L_1 前后双样本 t 检验结果
Tab.2 Results of two-sample t test for the data before and after separation points L_1

文本	转变点 L_1 前		转变点 L_1 后		P 值
	均值	标准差	均值	标准差	
《红楼梦》	0.57	0.01	0.64	0.01	0.000
《水浒传》	0.54	0.02	0.66	0.05	0.001

令 $\{X_i\}, i = 1, 2, \dots, N_v$, 表示每段字数时间序列; $\{Y_i\}, i = 1, 2, \dots, N_v$, 表示每段句子数时间序列; $\{Z_i\}, i = 1, 2, \dots, N_w$, 表示每句字数时间序列。其中, N_v 表示段落数, N_w 表示句子数。则有:

$$X_i = \begin{cases} \sum_{t=1}^{Y_i} Z_t, & i = 1 \\ \sum_{t=H_{i-1}+1}^{H_i} Z_t, & H_i = \sum_{k=1}^i Y_k, \quad i \neq 1 \end{cases} \quad (5)$$

将每句字数时间序列 $\{Z_i\}$ 均分成 N_v 段, 每一段含有 w_v 个元素, 对每个片段运用香农熵 (Shannon entropy)^[30-32]:



$$H(Z) = - \sum_{i=1}^{w_v} P(Z_i) \log_2 [P(Z_i)] \quad (6)$$

由式(5), 每一个片段会得到对应的 $H(i)$ 的值。在此将 $H(i)$ 定义为每一段所包含的信息量, 对每段信息量时间序列作去趋势波动分析, 其结果如图 4 所示。

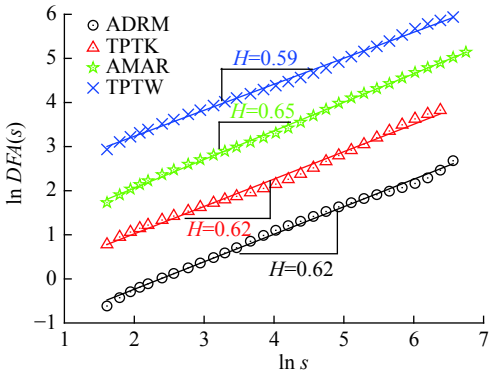


图 4 四大名著每段信息量的标度律
Fig.4 Scaling behaviors of each paragraph's information entropy for Four Great Classical Novels

图 4 给出了四大名著每段信息量时间序列的标度行为, 《红楼梦》、《三国演义》、《水浒传》、《西游记》的标度指数分别为 0.62, 0.62, 0.65, 0.59。由此可得, 每段信息量时间序列的标度指数和每段字数时间序列的标度指数几乎相同。

为了理解《红楼梦》和《水浒传》两本小说存在转变点的现象, 对每段信息量时间序列划分时序片段, 其序列长为 10^3 , 且滑动窗口 S 取 1 000。同样对每个时序片段运用去趋势波动分析法分析, 结果如图 5 所示。

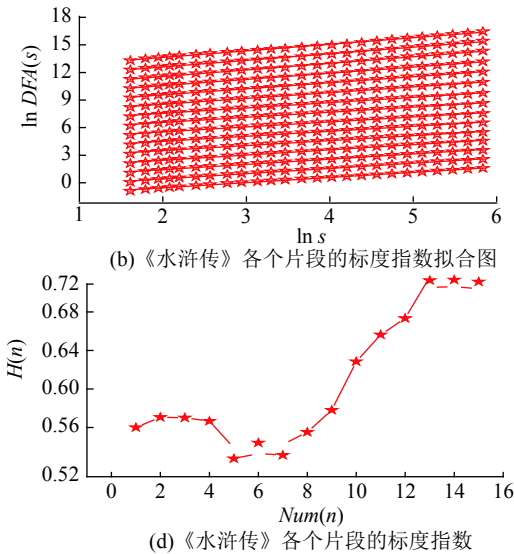


图 5 使用去趋势波动分析法分析每段信息量时序片段的結果
Fig.5 DFA results of the time series' segments for each paragraph's information entropy

由图 5 可知,《红楼梦》和《水浒传》的信息量时间序列也存在标度指数分段现象,《红楼梦》的每段信息量时间序列标度指数的转变点 L_2 与每段字数时间序列的标度指数转变点 L_1 一致,均在第 72 章节。《水浒传》的每段信息量时间序列标度指数的转变点 L_2 在第 62 章节。对《红楼梦》和《水浒传》每段信息量转变点前后进行差异性检验,表 3 列出了双样本 t 检验(two-sample t test)的检验结果。

表 3 转变点 L_2 前后双样本 t 检验结果

Tab.3 Results of two-sample t test for the data before and after separation points L_2

文本	转变点 L_2 前		转变点 L_2 后		P 值
	均值	标准差	均值	标准差	
《红楼梦》	0.59	0.01	0.65	0.01	0.008
《水浒传》	0.55	0.01	0.67	0.05	0.002

表 2 和表 3 结果表明,《红楼梦》和《水浒传》无论在每段字数时间序列,还是在每段所含信息量时间序列上,均存在标度指数的转变现象,且转变点前后的标度指数差异性显著($P<0.008$)。

5 结 论

前人在英文版的 *Holy Bible*^[22]中,使用去趋势波动分析研究了由单词长度构成的时间序列,发现了标度指数接近为 0.6 的无标度规律。本文在更高的层次即在段落层次和语句层次上使用去趋势波动分析我国的四大名著文本,即分析每一名著的每段句子数时间序列、每段字数时间序列和每句字数时间序列。研究发现,中国四大名著在各个层次上存在标度律,且每个层次上的标度指数也都接近于 0.60。这说明中文小说文本在各个层次上均满足固有的无标度规律,且具有相似的长程相关性,各个层次相似的无标度规律表明中文文本从微观层次到宏观层次具有相似的长程关联性,为重构语言形成与发展的理论模型提供帮助。

此外,本文还佐证了《红楼梦》的作者为曹雪芹和高鹗两人的说法,并支持《水浒传》一书有很大可能是施耐庵与罗贯中合作完成的观点。用去趋势波动分析法对每段字数时间序列的时序片段以及每段信息量时间序列的时序片段分析,得出了《红楼梦》和《水浒传》这两本小说在段落字数以及段落信息量上存在着标度指数的前后

显著差别。《红楼梦》的标度指数转变点为第 72 章节,即其前 72 章节和后 48 章节在段落结构和段落信息量上存在着前后变化,该转变点现象与学者们认为的《红楼梦》前 80 章由曹雪芹创作、后 40 章由高鹗完成的说法相接近。不同的标度指数表征不同作者的写作习惯以及表达方式,在一定程度上标度指数可以作为区分不同作者的依据^[20]。而对于《水浒传》,其标度指数也存在转变点现象,约为第 70 章节,从而支持了《水浒传》一书有很大可能是两人共同编写的结论,与当下许多学者认为的《水浒传》是“施耐庵的本,罗贯中编次”的说法不谋而合^[27-29]。

参考文献:

[1] KAPIEŃ J, DROŹDŹ S. Physical approach to complex systems[J]. *Physics Reports*, 2012, 515(3/4): 115–226.

[2] GRABSKA-GRADZIŃSKA I, KULIG A, KWAPIEŃ J, et al. Multifractal analysis of sentence lengths in English literary texts[C]//*Proceedings of 3rd World Conference on Information Technology (WCIT-2012)*. Barcelona, Spain: University of Barcelona, 2012.

[3] ZIPF G K. Human behavior and the principle of least effort[M]. Cambridge Mass: Addison-Wesley, 1949.

[4] MEHRI A, JAMAATI M. Variation of Zipf’s exponent in one hundred live languages: a study of the *Holy Bible* translations[J]. *Physics Letters A*, 2017, 381(31): 2470–2477.

[5] NABESHIMA T, GUNJI Y P. Zipf’s law in phonograms and weibull distribution in ideograms: comparison of English with Japanese[J]. *Biosystems*, 2004, 73(2): 131–139.

[6] LÜ L Y, ZHANG Z K, ZHOU T. Deviation of Zipf’s and Heaps’ laws in human languages with limited dictionary sizes[J]. *Scientific Reports*, 2013, 3: 1082.

[7] ALTMANN E G, CRISTADORO G, ESPOSTI M D. On the origin of long-range correlations in texts[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109(29): 11582–11587.

[8] DENG W B, WANG D J, LI W, et al. English and Chinese language frequency time series analysis[J]. *Chinese Science Bulletin*, 2011, 56(34): 3717–3722.

[9] MONTEMURRO M A, PURY P A. Long-range fractal correlations in literary corpora[J]. *Fractals*, 2002, 10(4): 451–461.

[10] BHAN J, KIM S, KIM J, et al. Long-range correlations in Korean literary corpora[J]. *Chaos, Solitons & Fractals*, 2006, 29(1): 69–81.

- [11] AUSLOOS M. Generalized Hurst exponent and multifractal function of original and translated texts mapped into frequency and length time series[J]. *Physical Review E*, 2012, 86(3): 031108.
- [12] GILLET J, AUSLOOS M. A comparison of natural (English) and artificial (Esperanto) languages. A multifractal method based analysis[J]. *CoRR*, 2008, abs/0801.2510.
- [13] PENG C K, BULDYREV S V, HAVLIN S, et al. Mosaic organization of DNA nucleotides[J]. *Physical Review E*, 1994, 49(2): 1685–1689.
- [14] GU C G, COOMANS C P, HU K, et al. Lack of exercise leads to significant and reversible loss of scale invariance in both aged and young mice[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(8): 2320–2324.
- [15] MUTUA S, GU C, YANG H J. Visibility graphlet approach to chaotic time series[J]. *Chaos*, 2016, 26(5): 053107.
- [16] STEPHEN M S, GU C G, YANG H J. Visibility graph based time series analysis[J]. *PLOS ONE*, 2015, 10(11): e0143015.
- [17] PAN X, HOU L, STEPHEN M, et al. Evaluation of scaling invariance embedded in short time series[J]. *PLoS One*, 2014, 9(12): e116128.
- [18] ZHANG W Q, QIU L, XIAO Q, et al. Evaluation of scale invariance in physiological signals by means of balanced estimation of diffusion entropy[J]. *Physical Review E*, 2012, 86(5): 056107.
- [19] YANG T G, GU C G, YANG H J. Long-range correlations in sentence series from *a story of the stone*[J]. *PLoS One*, 2016, 11(9): e0162423.
- [20] YANG Y, GU C G, YANG H J, et al. Evolution of scaling behaviors embedded in sentence series from *a story of the stone*[J]. *PLoS One*, 2017, 12(2): 0171776.
- [21] WEI B. The origin and evolvement of Chinese characters[M]. Poland: Wydawnictwo Uniwersytetu Gdańskiego, 2014.
- [22] DROŹDŹ S, OŚWIECIMKA P, KULIG A, et al. Quantifying origin and character of long-range correlations in narrative texts[J]. *Information Sciences*, 2016, 331: 32–44.
- [23] 刘上生. 曹雪芹的创作难题和高鹗的突破——试论《红楼梦》艺术构思的内在矛盾[J]. *红楼梦学刊*, 1989(2): 223–250.
- [24] 浦玉生. 《水浒传》作者施耐庵新证[J]. *菏泽学院学报*, 2011, 33(1): 46–54.
- [25] 任祖镛. 《水浒传》作者兴化施耐庵新证[J]. *东南大学学报(哲学社会科学版)*, 2014, 16(5): 105–109.
- [26] 汪吾金. 《水浒传》作者施耐庵“新证”中的几个“内证”问题——试评浦玉生先生的施耐庵研究[J]. *明清小说研究*, 2014(1): 86–97.
- [27] 许勇强, 邓雷. 近20年《水浒传》作者研究述评[J]. *东华理工大学学报(社会科学版)*, 2012, 31(4): 340–344.
- [28] 徐仲元. 施耐庵热与《水浒传》作者[J]. *内蒙古大学学报(哲学社会科学版)*, 1984(1): 88–94.
- [29] 黄霖. 宋末元初人施耐庵及“施耐庵的本”——兼析兴化、大丰“新发现”恰恰证明其地确无施耐庵[J]. *复旦学报(社会科学版)*, 1982(5): 50–56.
- [30] SHANNON C E. A mathematical theory of communication[J]. *The Bell System Technical Journal*, 1948, 27(3): 379–423.
- [31] 冯爱霞, 龚志强, 黄琰, 等. 全球温度场信息熵的时空特征分析[J]. *物理学报*, 2011, 60(9): 833–841.
- [32] 黄飞虎, 彭舰, 宁黎苗. 基于信息熵的社交网络观点演化模型[J]. *物理学报*, 2014, 63(16): 160501.

(编辑: 丁红艺)