

Characters and Events Analysis of Journey to the West by PCA and SPCA

Kang Lei

Department of Bioengineering

HKUST

lkangaa@connect.ust.hk

Abstract

In this paper, PCA and SPCA are used to evaluate the correlation between characters and events of Journey to the West. By principle components analysis of character-event data, we find that the characters in this novel can be mainly divided into two groups. The first group is composed of Tang seng, Monkey king, Pig and Sandy, who take the journey to the west, and the second group is composed of Jade emperor, Guanyin and Buddha, who set barriers and trials on the journey. In addition, by sparse principle components analysis, we find that Tang seng, Monkey king, Pig and Sandy are the most important characters in the novel and get similar conclusions to PCA.

1. Introduction

Journey to the West is a Chinese novel published in the 16th century during the Ming dynasty and attributed to Wu Cheng'en. It is one of the Four Great Classical Novels of Chinese literature. The novel is an extended account of the legendary pilgrimage of the Tang dynasty Buddhist Xuanzang who traveled to the "Western Regions", that is, Central Asia and India, to obtain Buddhist sacred texts and returned after many trials and much suffering. This article hopes to explain the plot and paving of the novel from the perspective of characters and events. Statistical method PCA and SPCA are used to analyze the characters and events matrix.

2. Datasets

Characters-scene occurrence matrices for 100 chapters of Journey to the West (CSIC 5011: Topological and Geometric Data Reduction and Visualization course website)

The matrix is 408*302, which means it has 408 events and 302 characters.

3. Methodology

3.1 PCA

For the character-scene occurrence matrix, 302 characters are regarded as p-dimension feature variables, 408 events are regarded 408 samples. We do PCA within python and get the following eigenvalues figure1 below. As shown in figure 1, top3 eigenvalues account for a large proportion of variance, so we extract the top3 principle components and the related events scores on them. The results are shown from table 1 to table 6.

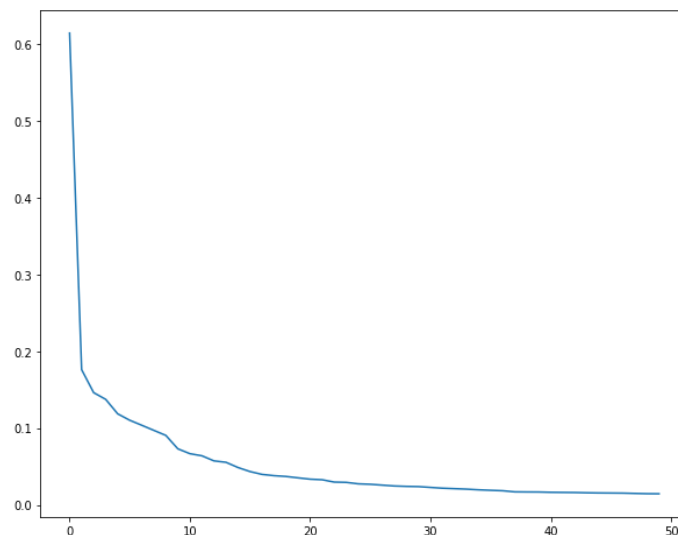


Figure 1: eigenvalues of PCA

3.2 SPCA

Similarly, SPCA is done in python, we extract the top3 principle components and the related events scores on them. The results are shown from table 7 to table 12.

4. Analysis

4.1 PCA analysis

The top3 principle components accounts for 26.8% variance. As shown in figure 1, the variance decrease slowly after the third principle components, so we analyze the top3 principle components.

4.1.1 The first principle component and the plots on its direction

Observing the loads on different feature variables of the first principle component, eight variables with the highest loads and eight with the lowest are respectively listed on table 1. One can notice that Pig, Sandy, Xuanzang, white horse and Monkey king occupy a very large proportion in the first principle components, and their importance are far greater than those characters with opposite loads, Such as Jade emperor, Guanyin, Tang emperor, Tota Li and Buddha. As a result, the first components can be regarded as the events involved with Pig, Sandy, Xuanzang, white horse and Monkey king. According to the reading experience, Pig, Sandy, Xuanzang, white horse and Monkey king are actually the main characters of the Journal to the west and the whole novel tells stories about their way to the west. What’s more, Jade emperor, Guanyin, Tang emperor, Tota Li and Buddha are also important in the novel, because the trials and sufferings that Xuanzang’s team meet in their journey is mainly set by Jade emperor, Guanyin and Buddha. Their importance is reflected by the loads on the positive directions of the first principle components.

Table 1: the first principle components

猪八戒	沙僧	唐僧	白龙马	孙悟空	天竺国小王子	天竺国二王子	天竺国三王子
-0.550497	-0.549392	-0.482363	-0.266515	-0.225097	-0.0264452	-0.0264452	-0.0264452
阿难	哪吒	如来佛	木吒,慧岸	托塔李天王	唐太宗李世民	观音菩萨	玉皇大帝
0.0304496	0.0447795	0.0455094	0.0457406	0.0458522	0.0469402	0.0507699	0.0804045

Considering the scores of different events on the first principle components, eight events with the lowest scores and eight with the highest are respectively listed on table2. After reading the eight chapters with the lowest scores, one can find that Pig, Sandy, Xuanzang, white horse and Monkey king appear in those

chapters simultaneously, and the frequency of occurrence is relatively high. For the chapters with the highest scores, Such as chapter 4, 5, 6 and 7. Jade emperor, Guanyin, Tang emperor appear in a high frequency. So the first principle components can successfully distinguish between the two clues of Tang Seng's team and Tian Ting's team (Jade emperor, Guanyin, Tang emperor).

Table 2: events scores on the direction of first principle components

chapter	89	73	64	27	64	99	73	53
0	-1.22	-1.15	-1.15	-1.14	-1.14	-1.14	-1.14	-1.13
6	4	8	6	6	6	5	7	
1.11	1.13	1.13	1.13	1.15	1.16	1.23	1.33	

4.1.2 The second principle component and the plots on its direction

Similar to the analysis method of the first principle components, observing the loads on different feature variables of the second principle component, eight variables with the lowest loads and eight with the highest are respectively listed on table 3. On the negative direction of the second principle component, Monkey king, Pig and Sandy account for a great proportion and they form as a team, while on the positive direction, Tang seng, Guan yin, Tang emperor take a very large proportion and form another group. What's more, most of the characters on the negative direction are monsters while many characters on the positive direction are immortals. So the second components can distinguish between two clues of monsters and immortals.

Table 3: the second principle components

孙悟空	猪八戒	沙僧	车迟国羊力大仙	车迟国鹿力大仙	车迟国虎力大仙	土地神	牛魔王
-0.499056	-0.329824	-0.122638	-0.0440782	-0.0440782	-0.0440782	-0.0429976	-0.0419057
地涌夫人	玉皇大帝	木吒慧岸	魏徵	唐太宗李世民	观音菩萨	白龙马	唐僧
0.0405248	0.0701416	0.0704422	0.0740929	0.133713	0.144314	0.411182	0.588335

Considering the scores of different events on the second principle components, eight events with the lowest scores and eight with the highest are respectively listed on table 4. By reading the eight chapters with the lowest scores, one can find that most events in these chapters are the fights between the team of Monkey king, Pig and Sandy and the groups of other monsters. The team of Monkey king, Pig and Sandy try to

protect Tang seng from those monsters. In addition, the fights happens on their journey to the west, so the chapters range from 24 to 90. In those chapters with the high scores, such as chapter 11, 12, 14, 15 and 83, most chapters are at the beginning at the whole novel. Tang seng, Tang emperor and Guan yin appears frequently in those chapters and they introduce in the whole story.

Table 4: events scores on the direction of second principle components

chapter	90	45	44	76	77	24	34	61
1	-0.927	-0.842	-0.842	-0.788	-0.788	-0.786	-0.775	-0.761
15	12	15	14	11	12	12	83	
0.949	0.964	0.973	0.974	1.08	1.11	1.22	1.28	

4.1.3 The third principle component and the plots on its direction

Similarly, table 5 shows eight lowest and eight highest loads on different feature variables of the third principle component and table 6 shows eight lowest and eight highest scores of different events on the third principle components. On the negative direction, Monkey king and Tang seng are important, by reading the chapters on this direction, such as 15, 16 and 98, one can find that Only Monkey king itself is trying to save Tang seng from monsters. In the chapters have high scores on positive direction, Pig and Sandy play an important rule to save Tang seng.

Table 5: the third principle components

孙悟空	白龙马	唐僧	方增长天王,四大天	方广目天王,四大天	哪吒	方持国天王,四大天	方多闻天王,四大天
-0.669133	-0.353222	-0.181275	-0.111403	-0.111298	-0.100076	-0.0943881	-0.0943881
木吒,慧岸	黄袍怪,奎星	通天河鱼怪	玉皇大帝	魏徵	唐太宗李世民	猪八戒	沙僧
0.0464485	0.0548757	0.060215	0.0884767	0.0995079	0.157584	0.257134	0.317676

Table 6: events scores on the direction of third principle components

chapter	15	16	98	76	77	83	15	16
2	-1.03	-1.03	-1.03	-0.882	-0.882	-0.881	-0.834	-0.802
10	10	30	11	29	10	29	49	

0.827	0.827	0.846	0.856	0.873	0.881	0.909	1.08
-------	-------	-------	-------	-------	-------	-------	------

4.2 SPCA analysis

SPCA is implemented in python with sklearn decomposition SPCA. The sparse parameter alpha is 0.25.

4.2.1 The first principle component and the plots on its direction

Similar to the analysis method in PCA, we can get the same tables in SPCA. For the first principle components in SPCA (Table6), the left side are Pig zhu, Sandy, Monkey King, they are major members of Tang seng's team. Another side are immortals who set barriers on the journey. Such as Taibaijinxing, Jade emperor. So the first principle component can separate two clues of Tang seng's team and immortals team. This conclusion is in agreement with the PCA analysis. In the chapters (Table7) with lowest scores, Tang seng's team appear frequently, while in the chapters with highest scores, immortals show up frequently.

Table 7: the first principle components for SPCA

猪八戒	沙僧	天竺国小王子	天竺国二王子	天竺国三王子	孙悟空	船上人	二十四路诸天
-0.716774	-0.697064	-0.00108069	-0.00108069	-0.00108069	0	0	0
黄风大圣	张道陵四大天师	十殿阎王	葛洪四大天师	玉皇大帝	敬之号许旄阳四	邱弘济四大天师	太白金星
0.000773429	0.00280959	0.00334566	0.0044558	0.00467796	0.00682479	0.00838738	0.0124821

Table 8: events scores on the direction of first principle components

chapter	26	22	58	49	55	84	99	98
0	-0.878238	-0.877913	-0.877793	-0.862382	-0.807505	-0.807385	-0.807384	-0.805818
15	15	15	16	16	21	87	92	
0.702302	0.702302	0.702302	0.702303	0.702303	0.702966	0.717695	0.734389	

4.2.2 The second principle component and the plots on its direction

In the second principle components, Tang seng is far more important than other people. The information carries by the whole axis is not obvious.

Table 9: the second principle components for SPCA

唐僧	沙僧	天竺国二王子	天竺国小王子	天竺国三王子	天竺国老王	朱紫国国王	地涌夫人
-0.993995	-0.0784473	-0.0144044	-0.0144044	-0.0144044	-0.0105154	-0.000615317	-8.10631e-05
木吒,慧岸	红孩儿	太上老君	通天河鱼怪	玉皇大帝	托塔李天王	土地神	哪吒
0.00314943	0.00384568	0.00628027	0.0085571	0.0249003	0.0342308	0.0383234	0.0410399

Table 10: events scores on the direction of second principle components

chapter	90	89	90	90	89	89	90	90
1	-0.609508	-0.575217	-0.571384	-0.571384	-0.571384	-0.571384	-0.566763	-0.560939
79	79	60	61	4	33	21	61	
-0.253009	-0.253009	-0.253011	-0.253011	0.924102	-0.0671136	-0.177032	-0.239712	

4.2.3 The third principle component and the plots on its direction

In the third principle components, On the negative direction of the third principle component, Monkey king and Pig account for a great proportion and they form as a team, while on the positive direction, Jade emperor, Wangmu, Tang emperor take a very large proportion and form another group. By reading the eight chapters with the lowest scores, one can find that most events in these chapters are the fights between the team of Monkey king, Pig and Sandy and the groups of other monsters. The team of Monkey king, Pig and Sandy try to protect Tang seng from those monsters. In those chapters with the high positive scores, such as chapter 9, 10, 11 and 12, most chapters are at the beginning. Tang emperor and Jade emperor appear frequently in those chapters and they introduce in the whole story.

Table 11: the third principle components for SPCA

孙悟空	猪八戒	土地神	车迟国众僧	车迟国道士	河神	船上人	二十四路诸天
-0.961358	-0.0622857	-0.0220939	0	0	0	0	0
徐茂功	东海龙王敖广	托塔李天王	黄袍怪,奎星	王母娘娘	魏徽	玉皇大帝	唐太宗李世民
0.0174509	0.0179375	0.0186223	0.0253478	0.0380967	0.095293	0.160988	0.180233

Table12: events scores on the direction of third principle components

chapter	60	61	79	79	98	61	76	77
2	-0.253011	-0.253011	-0.253009	-0.253009	-0.241368	-0.239712	-0.235193	-0.235193
12	9	11	11	10	10	10	11	
1.02265	1.0257	1.0257	1.02687	1.04229	1.07514	1.07573	1.0764	

5. Conclusion

From PCA, one can find that the novel has two clues of Tang Seng's team and Tian Ting's team (Jade emperor, Guanyin, Tang emperor). Pig, Sandy, Xuanzang, white horse and Monkey king are actually the main characters of the Journal to the west and the whole novel tells stories about their way to the west. What's more, Jade emperor, Guanyin, Tang emperor, Tota Li and Buddha are also important in the novel, because the trials and sufferings that Xuanzang's team meet in their journey is mainly set by Jade emperor, Guanyin and Buddha. They introduce in the whole novel and push the plots to develop. From SPCA, it verifies the conclusion that Pig, Sandy, Xuanzang and Monkey king are actually the main characters of the Journal to the west.

6. Acknowledgements

Thanks for the people who contribute to the datasets of Journal to west in Peking University.

7. Reference

- [1] Wan mengting. (2013) Character-Event Analysis and Network Analysis in Dream of the Red Chamber.
- [2] Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.