Data Visualization of the Twelve Kingdoms

An Eastern Fantasy World

Alberto Cairo†
School of Communication
University of Miami
Miami Florida U.S.
alberto.cairo@gmail.com

Yutong Han
School of Communication
University of Miami
Miami Florida U.S.
yxh703@miami.edu

Clay Ewing
School of Communication
University of Miami
Miami Florida U.S.
c.ewing@miami.edu

ABSTRACT

Data visualization play an important role in today's data-centric world for discovering, getting familiar, communicating, and validating insights.We are familiar with the western fantasy world, like *Harry Potter*, *Game of Thrones*. But we know less about the eastern fantasy world. Some anime fans are familiar with *One Piece*, *Naruto*. However, even very loyal anime fans didn't hear this anime - *the Twelve Kingdoms*.

The Twelve Kingdoms originally is a series of fantasy novels wrote by Fuyumi Ono†. Because it's famous, the work was made into an anime in 2002. It's not the best-seller compared with Naruto or One Piece. But it has loyal audiences. In Bilibili (like Chinese Youtube), it has over 4,437,000 views. Over 106,000 screen bullets will stay with you when you watch the anime. In addition, the novels are available for online reading and download. It's convenient for me to reuse. I'd like to explore the story of *the Twelve Kingdoms* through data visualization.

After going through precedents *Fifty chapters of adorable cuteness Cardcaptor Sakura* and *Anne Sexton: Poetry Between Pain*, I decided to analyze both book series and anime. I used fan translation, English version book series of *The Twelve Kingdoms* on the website Eugene Woodbury. Due to the unique feature on Bilibili, I can drag and analyze the data of on-time bullet comments and bottom comments from the anime on the website.

After scratching data from the website, I used R to do proximity analysis to find out the relationships among words. From the result, I thought the bubble chart is a great choice to present the relationships among words. The center bubble should be the biggest bubble (often represent the main character like Youko, Kirin), other smaller bubbles are connected by it. All bubbles and the lines should be interactive, users can interact with them to get more info.

I know my article will be very long, since there are seven book series, bullet screen comments and bottom comments will be visualized and analyzed. So, I need to organize my website very convenient for users to go through. For the tooltip in bubble charts, I drew the pictures for different characters, even the sword. And I added detail descriptions for people who are not familiar with the characters to understand who they are. And you can also click the highlighted words in the main texts to see detail descriptions on the sidebar. If you are interested, you can click the button "Read the Book" to read English version book series. For books 6 and 7, they are composed of several short stories. So, I cut them into separate bubble charts.

CCS CONCEPTS

• Data Visualization • Program reasoning • N-gram

KEYWORDS

Python, Data Scraping, R, N-gram Chart, D3.js

ACM Reference format:

Yutong Han, Alberto Cairo and Clay Ewing. 2020. Data Visualization of the Twelve Kingdoms. In *Proceedings of ACM Woodstock conference (WOODSTOCK'18). ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/1234567890

1 INTRODUCTION

The anime *the Twelve Kingdoms* pushed me to read the original book series. It has interesting and dangerous adventures, with rich contents of awaking feminine consciousness and ancient Asian cultural value. Even today, it's still my favorite anime. But the original book series are Japanese, I can't understand. And the Chinese version book series I didn't find relative packages in R (a programming language, widely used among statisticians and data miners for developing statistical software and data analysis) to analyze them. Luckily, I found fan translation, English version book series of The Twelve Kingdoms on the website Eugene Woodbury.

For the anime, I know an interesting Chinese video-sharing website called Bilibili[1]. It has a special feature—a real-time commentary

†Fuyumi Ono: (小野 不由美, Ono Fuyumi, born December 24, 1960) is a Japanese novelist best known for writing The Twelve Kingdoms (十二国記, Jūni Kokuki), which was adapted into a popular anime series. She is married to Yukito Ayatsuji, the author of the horror novel Another.

1 Bilibili is like Chinese YouTube, which is based in Shanghai, themed around animation, comic, and game (ACG).

subtitle system that displays user comments as streams of moving subtitles overlaid on the video playback screen. Such subtitles are simultaneously broadcast to all viewers in real-time, creating a chat room experience in which users feel like watching and playing together with others.

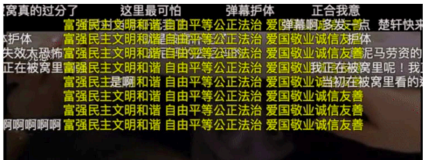

Figure 1: On-time bullet screen comments fly from right to left on the screen.

The Twelve Kingdoms has over 4,437,000 views on Bilibili. Over 106,000 screen bullets will stay with you when you watch the anime. Due to the unique feature on Bilibili, I can drag and analyze the data of on-time bullet comments and bottom comments from the anime on the website. Maybe I can even compare the difference between the comments from the website and the findings of the text analysis of the book series. Well, the ideal is plump, the reality is very skinny. Quickly, I met tons of troubles. But let's first look at the design.

8415	要不要点脸，打算给庆余年招黑吗	Or else face po
8416	我的天呐(ﾟДﾟ)此言此论真的是可歌可泣	My God (ﾟДﾟ) re
8417	庆余年还用阳子的初救呢……对庆余年没有任何好感甚至是很厌	Qingyu also use
8418	大半夜重温被楼主言论震惊到，类型不同如何比较，除了起点网文	Most of the nigh
8419	庆余年，，，	Qingyu years ,,,
8420	提庆余年我觉得层主你可能是反串黑吧	Mention Qingyu
8421	赞同楼上 除了起点网文，楼主最好多看看别的经典名著，起点yy	In addition to th
8422	楼上说得好，以上帝视角开金手指写得爽固然好，但也要认识到	Upstairs he said

Figure 2: The data I scraped from Bilibili.

After scratching data from the website, I used R to do proximity analysis to find out the relationships among words. The following image is what I got in R.

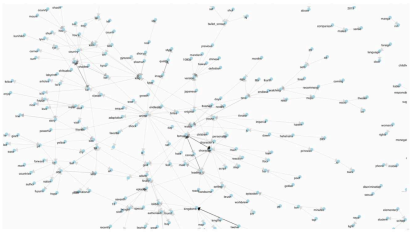

Figure 3: Rough n-gram chart I created through R

From the result, I thought the bubble chart is a great choice to present the relationships among words. The center bubble should be the biggest bubble (often represent the main character like Youko, Kirin), other smaller bubbles are connected by it. All bubbles and the lines should be interactive, users can interact with them to get more info.

2 OVERVIEW

In this case, I focused on the tree structure to visualize my data of the novels and the comments in Bilibili.

1. I introduced the work The Twelve Kingdoms. Because it's an unpopular work in the eastern anime market. I will shortly introduce the world view and special settings.
2. I cut tons of bubbles into different groups to help users to understand. I don't want to show everything at the beginning and make audiences feel lost. So, secondary groups will help a lot. At the same time, I will add detailed descriptions for these groups to help audiences further understanding. For example, the author didn't finish the story about Taiki and the king of Tai. So a lot of audiences complained that they fell into the pit, which is dug by Ono. And now they can't leave. It's pretty fun.

3 ROBLEM DEFINITION

I met tons of troubles I never imaged before. I overcame most of them, and some still need to do more work. I divided those troubles into three groups: scrape and clean the data, build the website, improve the website.

3.1 Scrape and Clean the Data

I spent nearly the whole of February to scrape and clean the data from the website Bilibili. And I met 3 problems:

1. Scraping Chinese is different from English. I learn how to use library pandas to scrape Chinese.
2. Because I'm in the US now, I can't load the comments unless use VPN. So, I asked my friend in China to run the code on his computer.
3. Another problem is I have no idea how to do n-gram in Python. Even though I watched some tutorials, it's still so hard for me to understand. Fortunately, Deb recommended the book Text Mining with R to me. Do n-gram by R is much easier. In this way, I successfully found out the relationships between characters, and words.

3.2 Build the Website

Then I focused on the visualization by d3 and website building during March. I tried different style charts, change the thickness of the line between bubbles, organize the context and charts.

I tried to create bubbles by D3 after solving the n-gram problem. The challenge at this step is to create force among bubbles and change the stroke between them into different thicknesses. Because I created treemap charts by D3 last semester, I wasn't stuck with the JSON file for a long time.

I know my article will be very long, since there are seven book series, bullet screen comments and bottom comments will be visualized and analyzed. So, I need to organize my website very convenient for users to go through.

3.2 Build the Website

For the tooltip in bubble charts, I drew the pictures for different characters, even the sword. And I added detail descriptions for

people who are not familiar with the characters to understand who they are. And you can also click the highlighted words in the main texts to see detail descriptions on the sidebar. If you are interested, you can click the button “Read the Book” to read English version book series. For books 6 and 7, they are composed of several short stories. So, I cut them into separate bubble charts.

Usually, more bullet comments mean the part is more important. The highest numbers of the episode often mean the climax of the story. So, I want to show both the distribution of the comments and the climax video.

Because I want to know which moment is the climax part per episode, I combined the chart and the video of the minute with the highest number of comments. You can use the button to change the episode to see the change.

When you click the button to change the episode, the chart, video, and the texts will also be changed.

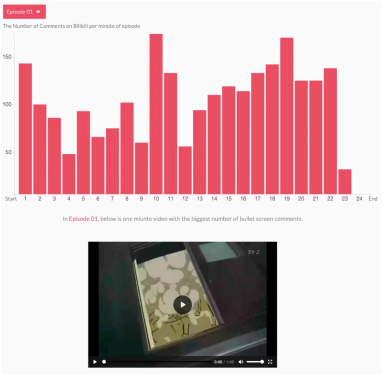

Figure 4: Add the button for users to jump among episodes.

Theorem/Proof/Lemma. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math

4 FINAL THOUGHTS

At the beginning of the semester, I found this image on the website. It’s about the process of the capstone. It’s funny so I saved it. In the beginning, I want to show the book series, bullet screen comments, and bottom comments, very ambitious.

Until the midterm, I felt like things are too overwhelming. I had no time to analyze the comments. So, my plan was changed. When the due date is getting close, I was more afraid my final work will become this horse. It’s cute but I don’t want to. It’s harder because of the home study. I need to keep myself away from computer games, TV dramas, and romantic novels. Luckily, I need to report my progress per week to Professor Cairo and Lenny, who pushed me to work hard. And you can see my work is not like the cute, unfinished horse. I can’t say my work is perfect. But it’s finished. And that makes me happy, and I hope it valuable for you.

1. In a Word 2013/2016 document, insert a picture.

ACKNOWLEDGMENTS

This work is supported by my Professor Alberto Cairo and Clay Ewing. They gave me a lot of valuable suggestions and detailed guidance. Without their help, I can't finish my capstone. Thank for them so much.

REFERENCES

[1] Julia Silge, David Robinson, 2007. *Text Mining with R: A Tidy Approach.* https://www.tidytextmining.com/index.html

[2] D Sarkar. 2008 *Lattice: multivariate data visualization with R.* In *Proceedings of the 6th. ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages* eBook ISBN 978-0-387-75969-2

[3] Chun-houh Chen, Wolfgang Karl Härdle, Antony Unwin, 2007. *Handbook of Data Visualization*. ISBN-13 : 978-3540330363

[4] IJ Ahrens, B Geveci, C Law 2005. *Paraview: An end-user tool for large data visualization*. Published in The Visualization Handbook 2005, Computer Science