

Five Centuries of Monarchy in Korea: Mining the Text of the Annals of the Joseon Dynasty

JinYeong Bak

Department of Computer Science
KAIST
Daejeon, South Korea
jy.bak@kaist.ac.kr

Alice Oh

Department of Computer Science
KAIST
Daejeon, South Korea
alice.oh@kaist.edu

Abstract

We present a quantitative study of the Annals of the Joseon Dynasty, the daily written records of the five hundred years of a monarchy in Korea. We first introduce the corpus, which is a series of books describing the historical facts during the Joseon dynasty. We then define three categories of the monarchical ruling styles based on the written records and compare the twenty-five kings in the monarchy. Finally, we investigate how kings show different ruling styles for various topics within the corpus. Through this study, we introduce a very unique corpus of monarchical records that span an entire monarchy of five hundred years and illustrate how text mining can be applied to answer important historical questions.

1 Introduction

Historical documents are usually studied qualitatively by researchers focusing on a close reading of a small number of documents. However, for a large corpus of historical texts, qualitative methods have limitations, thus quantitative approaches have been introduced recently (Moretti, 2005; Jockers, 2013). There is also research in applying text mining and natural language processing methods to identify patterns in a corpus of large and longitudinal documents (Mimno, 2012). In this paper, we introduce a unique corpus of historical documents from the written records that span almost five hundred years from the fourteenth century up to the late nineteenth century within the Korean peninsula. We apply text mining to this corpus to show the power of a computational approach in answering historical questions.

We first introduce *The Annals of the Joseon Dynasty* (Chunchugwan, 1863). Joseon is the last

monarchical nation in the Korean Peninsula from its founding in 1392 up to 1910. The Annals of the Joseon Dynasty are a series of books of historical facts, recorded almost daily during the Joseon dynasty. Whenever a king abdicated the throne, the Chunchugwan (office for annals compilation) updated the Annals for that king from all related official and unofficial documents. The Annals contain political, economic, social and cultural topics during the corresponding time periods.

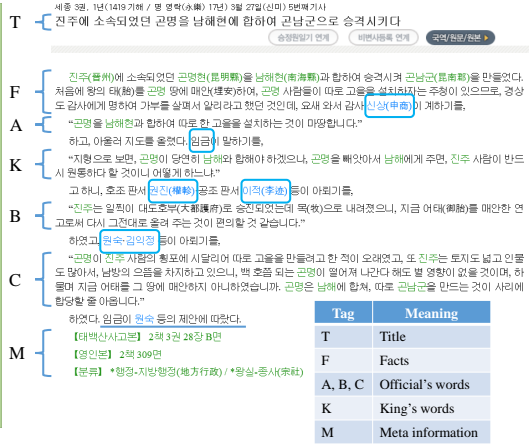
To illustrate the application of a text mining approach, we analyze each king's ruling style from the Annals of the Joseon dynasty. Being a monarchical system, almost all decisions within the government are confirmed by the king, where the king can make the decision on his own, or after discussing it with the government officials. We identify the patterns of each king's decision making and compare the patterns among the kings. The results show interesting patterns of the kings' ruling styles, including the tendency to make arbitrary decisions of the kings who were later dethroned because of tyranny. Additionally, we apply a topic model to the corpus and analyze the kings' ruling style for each topic.

2 The Annals of the Joseon Dynasty

In this section, we describe the details of *The Annals of the Joseon Dynasty* (from here referred to as the AJD) (Chunchugwan, 1863) and our process for building a corpus of the AJD. In its entirety, the AJD consists of records from twenty-seven kings over 519 years. However, the last two kings' (Gojong, Sunjong) books are usually excluded from research by historians because many facts are distorted. We follow that convention and use the books of the first twenty-five kings. These records, in their original Chinese text and in the Korean translations, are available publicly through



(a) Korean translation, Chinese original text and scanned image



(b) Structure of the article

Figure 1: Screenshot and structure of an article in the annals of the Joseon dynasty

King name	Period of reign	# months	# articles
Taejo	1392-1398	81	2,387
Jeongjong	1398-1400	24	624
Taejong	1400-1418	220	10,331
Sejong the Great	1418-1450	391	30,969
Munjong	1450-1452	27	2,670
Danjong	1452-1455	40	2,534
Sejo	1455-1468	165	10,832
Yejong	1468-1469	17	1,503
Seongjong	1469-1494	311	32,443
Yeonsangun	1494-1506	146	12,009
Jungjong	1506-1544	474	39,653
Injong	1544-1545	8	671
Myeongjong	1545-1567	272	15,044
Seonjo	1567-1608	438	26,712
Gwanghaegun	1608-1623	187	22,121
Injo	1623-1649	325	16,046
Hyojong	1649-1659	125	5,431
Hyeonjong	1659-1674	189	9,295
Sukjong	1674-1720	568	24,209
Gyeongjong	1720-1724	54	2,744
Yeongjo	1724-1776	639	36,731
Jeongjo	1776-1800	302	17,681
Sunjo	1800-1834	425	15,529
Heonjong	1834-1849	182	3,986
Cheoljong	1849-1863	180	5,771
Gojong	1863-1897	536	27,939
Sunjong	1907-1910	38	4,858

Table 1: Name, period of reign and the number of months and articles for 27 kings in Joseon dynasty

a website¹. We build our corpus by crawling all articles from that website², and this corpus comprise 1,893 books and 380,271 articles covering 472 years (1392 - 1863). Table 1 shows the basic statistics of our AJD corpus including the period of reign and the number of articles for each king.

Each *article* on the website consists of the tran-

¹<http://sillok.history.go.kr>

²We crawl and investigate the AJD from the site legally, because it is opened to the public by Korean government.

scription of the original Chinese text, the Korean translation, and the scanned images from the original books. Figure 1 shows an example article³. For this paper, we analyze the Korean translated text (Figure 1b), though we refer to the Chinese version to understand the meaning of some words that are not currently used in the modern Korean language. Each Korean article has a title (marked T in the figure) that is created by the translators, the body text (A, B, C, F and K in the figure) and the meta information (M) including the source, page, and tags of the article.

3 King's Ruling Style

Joseon was a monarchy, but a king could not make all decisions by himself. Instead, Joseon adopted a government system that most of the public issues are discussed with the government officials (Park, 1983; Kim, 2008) before the king made the decisions, which are all recorded in the AJD. Hence, by analyzing the decision making process in the AJD, we can understand each king's ruling style.

3.1 Categorizing ruling style

In Joseon dynasty, the king was the final decision maker. Even when the government officials discussed the public issues, a king's approval was needed. We can categorize each king's decision making process into three types. First, a king can order directly without discussion, which we call Arbitrary Decision (AD). Second, a king can discuss an issue with the officials and then direct his

³Article URL: http://sillok.history.go.kr/viewer/viewtype1.jsp?id=kda_10103027_005

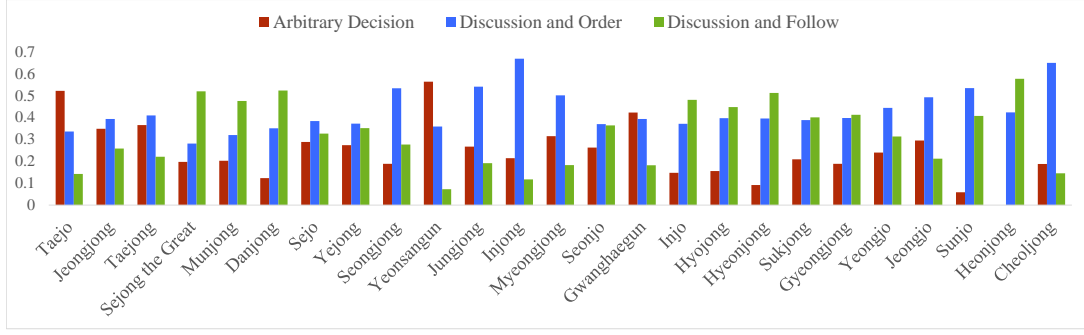


Figure 2: Joseon king’s ruling styles. Each king shows quite different ruling style ($p < 0.001$).

Decision	Words
Order	명하다, 하교하다, 전교하다, 命, 傳教, 下教
Approve	윤허하다, 허락하다, 允
Disapprove	불허하다, 허락하지 않았다, 不允
Reject	파르지 않았다, 듣지 않았다, 不從, 不聽
Follow	따르다, 따랐다, 從之, 依啓

Table 2: Example verbs for identifying king’s decision in the AJD. Words are written in Korean and Chinese alphabet.

order, which we call Discussion and Order (DO). Third, a king can discuss an issue with the officials and then decide to follow the officials’ suggestion, which we call Discussion and Follow (DF). The difference between DO and DF is that in DO, the king acts aggressively with his own opinion.

From these observations, we ask two research questions: 1) Can we identify and categorize kings with different ruling styles? 2) Do kings’ ruling styles differ depending on the topic?

3.2 Method

To understand each king’s ruling style, we first identify relevant articles that contain the king’s decision making because many of the articles describe non-governmental affairs, such as the weather, or simple status reports. These relevant articles contain direct quotations of the words of the king or the government official. The original texts do not contain any quotation marks, but translators added them to distinguish explicit quotations, which we can use to identify these relevant articles. Its size is 126K, 36% over all articles.

Each article contains who said what for an issue, and king’s final actions are written mostly in the last part of the article. For example, the underlined last sentence in Figure 1b says that the king followed the official’s suggestion. Hence, to identify king’s action for each issue, we focus on the

last sentence in each article.

First, we identify that the sentence subject is the king, because some issues are dealt by others. For example, Sunjo, Heonjong and Cheoljong’s mother or grandmother ruled as regent, so her decisions are recorded in the AJD. To identify the part of speech in Korean, we used HanNanum (Choi et al., 2012). And, we investigate the verbs that indicating decisions including order, follow, approval and reject. We use sixty verbs that describe king’s decision specifically, and table 2 shows example words. Finally, we classify these decisions into three types: 1) the king orders without discussions with the officials, and we label them as AD, 2) the king orders, approves, or rejects verbs in which their original Chinese characters show active decision making by the king, and we label them as DO, and 3) the king follows or discusses verbs which show passive submission by the king, and we label them as DF.

To identify topics, we use a Bayesian topic model, LDA (Blei et al., 2003). We implement it using Gibbs sampling (Griffiths and Steyvers, 2004), set 300 topics, and optimized hyperparameters after 100 iterations (Asuncion et al., 2009). We remove stopwords and words with document frequency of 30 or smaller.

3.3 Results and Discussions

We investigate the difference of ruling style between kings. We run multinomial test (Read and Cressie, 1988) between king’s ruling style distributions. Result shows that almost all kings are different significant from others ($p < 0.001$). It means that each king has his own ruling style.

Figure 2 shows the distribution of each category of ruling style. Overall, many kings do not act arbitrary. They discuss about many of the national affairs with officials. But, Taejo who is the

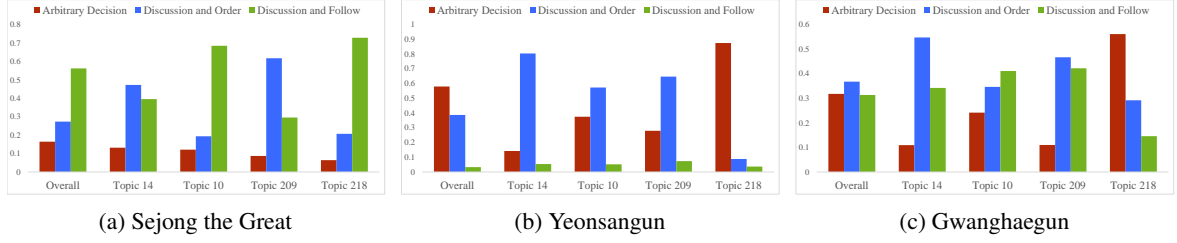


Figure 3: King’s ruling styles given a topic. It differs from overall ruling style (leftmost bars) ($p < 0.01$).

14	10	209	218
retirement	agriculture	remission	grants
vassal	grain	sin	a roll of cloth
retire	village	law	a piece
position	a piece	majesty	royal grant
person	storehouse	engage	saddle
capable	people	forgive	a piece
job	rice	favor	a part
duty	save	rebel	tiger skin
duties	bad year	person	epidermis

Table 3: Example of different topics ($p < 0.01$) from overall ruling style distribution

founder of the Joseon dynasty shows high value of AD. And Yeonsangun and Gwanghaegun who are evaluated as a tyrant also show high value of it. So we can imagine that tyrants tend to act arbitrarily.

We also identified those kings whose ruling style differed most from other kings. We use JS divergence which is the symmetric measure of the difference between two probability distributions. We compute JS divergence with each king pair’s ruling style distributions. Result shows that Heonjong (0.1220) and Yeonsangun (0.0998) have highest distance value. It means their ruling style are quite different from other kings. Because Heonjong’s grandmother governed the Joseon each year, so his actions are quite few. But, unlike Yeonsangun, Gwanghaegun (0.0454) who is known as a tyrant has similar value mean distance from other kings (0.0434). It means his ruling style is quite similar to other kings, and this result supports previous results in Korean historical study (Kye, 2008) that re-evaluate his reputation.

We investigate the difference of king’s ruling style based on the topic. We run multinomial test (Read and Cressie, 1988) between king’s overall ruling style distribution and specific distribution given a topic. Results show that some ruling styles given a topic are different significant from overall ($p < 0.01$). It means that the king’s ruling style when the topic is given is different from his usual style. Table 3 shows examples of topic. Figure 3

shows four kings’ overall ruling style and specific one given a topic. Comparing with the leftmost bars which is overall ruling style of the king, each ruling style given a topic is different from it. And, we can see that kings show similar/different ruling style for a topic. For example, kings tend to discuss and order (DO) to officials for retirement and remission topics. And, Sejong the Great and Gwanghaegun discuss and follow (DF) officials’ words for agricultural topic. But, for grants topic, Yeonsangun and Gwanghaegun act more arbitrarily (AD) than overall ruling style, and Sejong the Great follows more official’s opinions (DF).

4 Conclusion and Future Work

We introduced long and large historical documents, *The Annals of the Joseon Dynasty*. It contains lots of topics such as political, economic, social and cultural over 500 years. We looked at the ruling style of kings in Joseon dynasty and its difference by topics by computational methods.

This is ongoing work, and we are looking to find more hidden structures in the AJD. Currently, historians evaluate the king’s reputations (Park, 2004; Lee, 2010). This evaluation is done by many aspects, but one of the important feature is king’s ruling style (Kim, 2008). So we are looking to improve methods for analyzing ruling style more specifically. For example, we will look at the relationship with officials, especially who can make the king follows his opinion. This approach can be used to measure king’s leadership.

Acknowledgments

We would like to thank the anonymous reviewers for helpful comments, and National Institute of Korean History for checking legal issues. This work was supported by ICT R&D program of MSIP/IITP. [B0101-15-0307, Basic Software Research in Human-level Lifelong Machine Learning (Machine Learning Center)]

References

- Arthur Asuncion, Max Welling, Padhraic Smyth, and Yee Whye Teh. 2009. On smoothing and inference for topic models. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 27–34. AUAI Press.
- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.
- DongHyun Choi, Jungyeul Park, and Key-Sun Choi. 2012. Korean treebank transformation for parser training. In *Proceedings of the ACL 2012 Joint Workshop on Statistical Parsing and Semantic Processing of Morphologically Rich Languages*, pages 78–88. Citeseer.
- Chunchugwan. 1863. *Joseonwangjosillok (The Annals of the Joseon Dynasty)*. The name of the publisher.
- Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5228–5235.
- Matthew L Jockers. 2013. *Macroanalysis: Digital methods and literary history*. University of Illinois Press.
- Jeong Ki Kim. 2008. A study on the policy decision-making process of the chosun era: Focus on mandarins’ participation function. *Korean Public Administration History Review*, 23(1):161–186.
- Seung B. Kye. 2008. The causes for the coup of 1623 and their adjustment in choson korea. *NAMMYONG STUDY*, 26(1):439–478.
- Dukil Lee. 2010. *Say about Joseon King*, volume 1. WisdomHouse.
- David Mimno. 2012. Computational historiography: Data mining in a century of classics journals. *Journal on Computing and Cultural Heritage (JOCCH)*, 5(1):3.
- Franco Moretti. 2005. *Graphs, maps, trees: abstract models for a literary history*. Verso.
- Munok Park. 1983. *Discussion about new Korea government*. Shincheon.
- YoungKu Park. 2004. *Reading though a book about The annals of the Joseon dynasty*. Woongjin ThinkBig.
- Timothy RC Read and Noel AC Cressie. 1988. *Goodness-of-fit statistics for discrete multivariate data*. Springer New York.