Semiconductor Optoelectronics
ISSN: 1001-5868

THE SPRING OF SEOUL: A BIG DATA ANALYSIS

Namkil Kang
Professor, College of Liberal Arts, Far East University, South Korea

ABSTRACT
The ultimate goal of this paper is to analyze 23 articles written in December, 2023 concerning the movie *spring of Seoul*. This research was carried out by python. To begin with, we elucidate the pretreatment of 23 articles. More specifically, we got rid of nominative case markers, topic markers, accusative case markers, genitive case markers, dative case markers, and several symbols for the big data analysis of 23 articles. A point to note is that there are no negative words in the two representative articles with respect to the spring of Seoul. A further point to note is that the word Seoul was the most occurred one, followed by the word movie, and the word audience, in that order. A major point of this paper is that in the word cloud, the keyword Seoul is the most noteworthy one, followed by the keyword movie, the keyword audience, and the keyword release, in that order. When it comes to the network analysis of 23 articles, the words director, military, sales, theater, etc. show up in the center and around the center. This in turn indicates that they are all the noteworthy ones in 23 articles. With respect to the sentiment analysis of 23 articles, it is worthwhile noting that we obtained +669 with respect to positive words, whereas we obtained -75 with respect to negative words. It can thus be inferred that the spring of Seoul has received many favorable reviews from viewers and the press.
Key Words: Big Data, Word Cloud, Network, Sentiment Analysis, Term Frequency

1. Introduction

The main purpose of this paper is to analyze 23 articles written in December, 2023 concerning the movie *spring of Seoul*. We collected 23 articles in the portal site Naver and we analyzed them in terms of python. First, we aim at going over the pretreatment of 23 articles. Second, we attempt to consider the two of 23 articles in which we can see the term frequency of verbs and nouns. By observing the term frequency of words, we can see what is happening in the two representative articles. Third, we try to observe the total term frequency of nouns and verbs. Note, however, that verbs, adjectives, and nouns play a pivotal role in the sentiment analysis of 23 articles. Fourth, this research aims to probe into the so-called word cloud in which key words appear in different fonts. This in turn indicates that important and central words are represented as bigger in size. Fifth, we attempt to provide the network analysis of 23 articles. Several pivotal words show up in the center of the map, whereas less pivotal words turn up in the corner. Finally, we attempt to provide the sentiment analysis of 23 articles in which we try to classify key words into the positive one and the negative one. After this, we assign –1 to the negative one, whereas we assign +1 to the positive one. By doing so, we can evaluate the spring of Seoul.

2. Results

2.1. The Pretreatment of 23 Articles

This section focuses on contemplating the pretreatment of 23 articles. In this paper, we obtained nouns and verbs from the original text. More specifically, the so-called nouns play an important role in the analysis of the word cloud. Also, nouns play a pivotal role in the network analysis of 23 articles. On the other hand, verbs and nouns play a big role in the sentiment analysis of 23 articles. More importantly, we obtained only nouns, adjectives, and verbs after processing the original text. To be more specific, we got rid of nominative case markers, which are not necessary in the big data analysis of 23 articles. Also, we eliminated accusative case markers, which are also not necessary for our analysis. Exactly the same can be said about genitive case markers and topic markers. We got rid of them since they are not necessary for our analysis. Also, we eliminated dative case markers and several symbols after processing the original text. To sum up, for the big data analysis of 23 articles, we excluded nominative case markers, topic markers, accusative case markers, genitive case markers, dative case markers, and several symbols. We obtained only nouns, adjectives, and verbs for the big data analysis of 23 articles.

2.2. The Term Frequency of Article 6 and Article 12

In the following, we aim at inquiring into the term frequency of article 6 (the newspaper Edaily) and article 12 (the newspaper Sports Kyeonghyang). The reason why we look into the term frequency of major words in the two representative articles is that we can guess whether the spring of Seoul is a good movie or not. By going over the term frequency of major words, we can guess that the spring of Seoul is well-made. Table 4 shows the term frequency of article 6 (the newspaper Edaily):

Table 4 The Term Frequency (the newspaper Edaily)

The Term Frequency of Words
'Do': 65, 'There's': 24, 'Spring': 23, 'Seoul': 23, 'Director': 18, 'Movie': 16, 'Kim Seong-su': 16, 'Actor': 15, 'Audience': 12, 10 Millions': 11, 'Hwang Jung-min': 9, 'Acting': 9, 'History': 8, 'Finished': 5, 'hot': 5, 'Rebellion': 5, 'Perfect': 5, 'Greetings': 5, 'Mind': 4, 'Meaning': 4, 'Mopping': 4, 'Photo': 4, 'Plus': 4, 'Entertainment': 4, 'Person': 4, 'Show': 4, 'Showing': 4, 'This year': 4, big': 3, 'It's excellent': 3, 'Record': 3, 'Right': 3, exist': 3, 'Assault': 3, 'Come out': 3, 'Theater': 2, 'Wind': 2, 'People': 2, 'Expectation': 2, 'The Center': 2, 'Best': 2, 'Praise': 2, 'Capital': 2, 'Korea': 2, 'Military': 2, 'Kim Seong-gyun': 2, 'Jeong Ui-seong': 2, 'Kim Ui-seong': 2, 'Guide': 2, 'Jung Man-sik': 2, 'Special Appearance': 2, 'Tension': 2, 'Story': 2, 'Emotional': 2, 'Excitement': 2, 'Last': 2, 'Parking': 2, 'This time': 2, ' sorry': 2, 'Fierce': 2, 'Forget': 2, 'Conclusion': 2, 'The day': 2, 'Choose': 2, 'Really': 2, 'Trust': 2, 'Dear': 2, 'Scent': 2, 'Story': 2, 'Real': 2

As alluded to in Table 4, key words that occurred in article 6 are largely positive. More specifically, the words spring and Seoul obtained 23 tokens, respectively. It is worthwhile pointing out that the word director obtained 18 tokens. It is particularly noteworthy that the word perfect obtained 5 tokens. It therefore seems appropriate to assume that the movie is

perfect. In this article, we cannot find out something negative, but we can find out something negative with respect to the event 12.12 military insurrection.

Now let us take a look at Table 5:

Table 5 The Term Frequency (Sports Kyunghyang)

The Term Frequency of words
'Do': 23, 'There is': 14, 'Seoul': 14, 'Spring': 13, 'Audience': 12, 'Movie': 9, 'Release': 8, 'See': 8, 'Theater': 5, 'Kim Seong-su': 5, 'Director': 5, 'Military': 5, 'History': 5, 'Actor': 5, 'Cumulative': 3, 'Excellent': 3, 'Korea': 3, 'Show': 3, 'Behind': 3, 'Generation': 2, 'Noryang': 2, 'Death': 2, 'Sea': 2, 'Shin Gun-bu': 2, Jung Woo-sung': 2, 'Hwang Jung-min': 2, 'Evaluation': 2, 'The same': 2, 'This year': 2, 'Power': 2, 'Move': 2, 'Act': 2, The Scene: 2, 'The Road': 2, 'Connect': 2, 'View': 2, 'Day': 1, 'Cold': 1, 'Temperature': 1, 'Summer': 1, 'Hot': 1, 'Young': 1

As exemplified in Table 5, the word Seoul obtained 14 tokens. What is interesting is that the word spring obtained 13 tokens. Note that nouns play a pivotal role in the keyword analysis. Particularly noteworthy is that the word history obtained 5 tokens. It must be stressed that the word hot obtained 1 token. It should be pointed out that the word excellent obtained 3 tokens. As illustrated in Table 5, we can find out something positive with respect to the spring of Seoul.

2.3. The Total Term Frequency of 25 Words

In the following, we aim to go over the total term frequency of 25 words that mainly occurred in 23 articles. Table 6 shows the term frequency of 25 words:

Table 6 The Total Term Frequency of 25 Words

Number	Word	Frequency
1	do	304
2	Seoul	178
3	movie	127
4	There is	113
5	audience	91
6	release	65
7	become	60
8	director	53
9	box office hit	50
10	actor	49
11	Ten million	40
12	Jung, Woo-seong	39
13	Kim, Seng-soo	36
14	record	31

15	rebellion	31
16	exceed ten million	30
17	watch	29
18	see a movie	24
19	do not	24
20	acting	24
21	accumulated	23
22	there is no	22
23	stage	21
24	this year	21
25	works	21

We ignore verbs such as hata 'do', issta 'there is', and toyta 'become' since they play a big role in the sentiment analysis of 23 articles. Note that only nouns play a role in the keyword analysis of 23 articles. Perhaps it is worthwhile noting that the word Seoul obtained 178 tokens (the highest frequency among nouns). This amounts to saying that the word is the most noteworthy one in 23 articles. It is worth noticing that the word is the most widely used topic in 23 articles. It must be stressed that the word movie obtained 127 tokens. It should be pointed out that the word Seoul is followed by the word movie. This in turn implies that the latter is one of the most widely used topics in 23 articles. It is appropriate to mention that the word audience obtained 91 tokens (the third highest among nouns). This keyword is also one of the most widely used topics among all words and counts as noteworthy. Additionally, the total number of viewers is about 10,000,000 people, which in turn indicates that the spring of Seoul has ever been one of the hit Korean movies. It is important to mention that the word box office hit obtained 50 tokens among all words that occurred in 23 articles. It therefore seems appropriate to conclude that the word Seoul was the most occurred one, followed by the word movie, and the word audience, in that order.

2.4. The Word Cloud of 23 Articles

This section is devoted to probing into the word cloud representing 23 articles. Figure 1 is the so-called word cloud that represents 23 articles:

Figure 1 The Word Cloud Representing 23 Articles

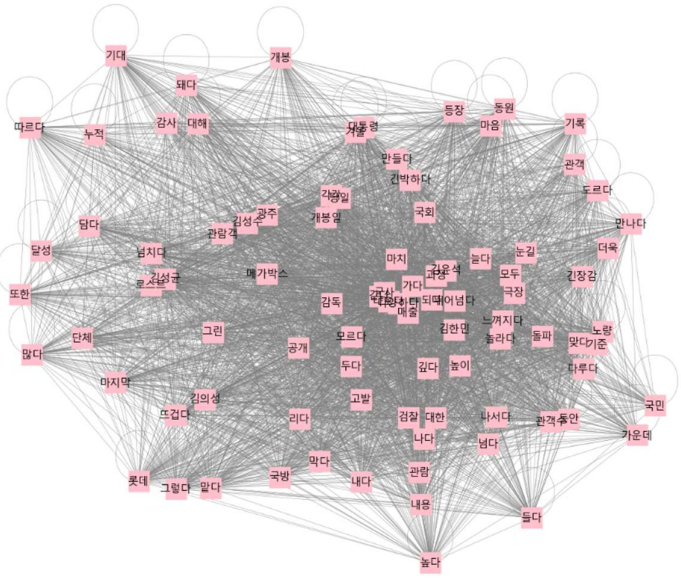
As exemplified in Figure 1, the keyword Seoul is the biggest. This in turn indicates that the word Seoul was the most noteworthy one as well as the most occurred one. When it comes to a word cloud, python produces the word cloud in accordance with the term frequency of central words. Note, however, that verbs play a role in the sentiment analysis of 23 articles and thus we ignore them again. With respect to the size of words, it is worthwhile pointing out that the word movie is the second biggest. It can thus be inferred that the keyword movie is the second most noteworthy one as well as the second most occurred one. It is worth mentioning that the keyword audience is the third biggest, which counts as the third most noteworthy one. It is also important to note that the word release 'kaypong' is the fourth biggest, which in turn suggests that the word was the fourth most frequently used one. It seems thus appropriate to conclude that the keyword Seoul is the most widely used, followed by the keyword movie, the keyword audience, and the keyword release, in that order.

2.5. The Network Analysis of 23 Articles

This section is focused on providing the network analysis of 23 articles. Figure 2 shows networks among nodes (central words) in 23 articles. The so-called networks refer to the relevance among main words. In other words, several central and noteworthy words show up in the central place, whereas less important words turn up in the corner. As illustrated in Figure 2, several main words such as the words director, Megabox, military, sales, etc. occur in the center, while less important words such as content, people, audience, etc. appear in the corner.

Now let us take look at Figure 2:

Figure 2 Networks



As indicated in Figure 1, the word military shows up in the center of the map, which in turn implies that this word is the most central one, thus becoming the theme of the spring of Seoul. Just as in the case of the word military, the word sales also turn up in the center of the map. From this, it is clear that this word is one of the most central words and one of the most noteworthy words. It is interesting to observe, on the other hand, that the word content, the word people, the word audience, etc. occur in the corner, which is deemed to be less significant in the map. We thus conclude that the words director, military, sales, theater, etc. show up in the center and around the center. This in turn shows that they are all the noteworthy ones in 23 articles. For big data and machine learning, see Kang (2023a, 2023b, 2023c, 2023d, 2023e, 2023f, 2024a, 2024b).

2.6. The Sentiment Analysis of 23 Articles

In what follows, this paper aims to evaluate the spring of Seoul. To begin with, we attempt to classify key words into the positive one and the negative one. After this, we assign -1 to something negative, while we assign +1 to something positive. After that, we multiple the frequency of each word by +1 or -1, depending on its property (on the basis of the negative one or positive one). Let us take a look at Table 7:

Table 7 The Sentiment Analysis of 23 Articles

Number	Word	Frequency	Positive or Negative	Score
1	audience	91	positive	91x1
2	director	53	positive	53x1
3	box office hit	50	positive	50x1
4	ten million	40	positive	40x1

5	Kim, Sungsoo	36	positive	36x1
6	Noryang	36	negative	36x-1
7	record	31	positive	31x1
8	exceed	30	positive	30x1
9	watch	24	positive	24x1
10	acting	24	positive	24x1
11	accumulated	23	positive	23x1
12	hot	19	positive	19x1
13	box office	19	positive	19x1
14	Yi, Sun-shin	16	negative	16x-1
15	collect	14	positive	14x1
16	sale	14	positive	14x1
17	achieved	11	positive	11x1
18	viewers	10	positive	10x1
19	direction	10	positive	10x1
20	reservation	10	positive	10x1
21	best	10	positive	10x1
22	perfect	9	positive	9x1
23	occupation	9	positive	9x1
24	mega box	8	positive	8x1
25	increase	7	positive	7x1
26	surpass	7	positive	7x1
27	sales	7	positive	7x1
28	hot	7	positive	7x1
29	conservative	6	negative	6x-1
30	make a new record	6	positive	6x1
31	Be ahead of breaking news	6	positive	6x1
32	continuing	6	positive	6x1
33	political circles	6	positive	6x-1
34	ratings	6	positive	6x1
35	expectation	5	positive	5x1
36	tension	5	positive	5x1
37	attention	5	positive	5x1
38	atmosphere	5	positive	5x1

39	overwhelming	5	positive	5x1
40	support	5	positive	5x1
41	favorable comment	5	positive	5x1
42	audience	4	positive	4x1
43	urgent	4	positive	4x1
44	content	4	positive	4x1
45	criticism	4	negative	4x-1
46	close	4	negative	4x-1
47	good	4	positive	4x1
48	attention	4	positive	4x1
49	excellent	4	positive	4x1
50	solid	4	positive	4x1
51	great	4	positive	4x1
52	dropping into a movie	3	positive	3x1
53	ensemble	3	positive	3x1
54	of all time	3	positive	3x1
55	excitement	3	positive	3x1
56	enthusiastic performance	3	positive	3x1
57	concern	3	negative	3x-1

As can be seen from Table 7, we obtained 52 positive words, whereas we obtained 5 negative words. More specifically, we have assigned +1 to positive words, while we have assigned -1 to negative words. After that, we multiple the frequency of each word by +1 or -1. By doing so, we obtained +669 with respect to positive words, while we obtained -75 with respect to negative words. It seems clear from our findings that the spring of Seoul is the hit Korean movie and worth watching it. Additionally, it is worthwhile noting that the movie has received many favorable reviews from viewers and the press. We thus conclude that the spring of Seoul got favorable reviews from the press.

3. Conclusion

To sum up, we have analyzed 23 articles written in December, 2023 concerning the movie *spring of Seoul*. In section 2.1, we have elucidated the pretreatment of 23 articles. More specifically, we got rid of nominative case markers, topic markers, accusative case markers, genitive case markers, dative case markers, and several symbols for the big data analysis of 23 articles. In section 2.2, we have argued that there are no negative words in the two representative articles. In section 2.3, we have maintained that the word Seoul was the most occurred one, followed by the word movie, and the word audience, in that order. In section 2.4, we have contended that in the word

cloud, the keyword Seoul is the most noteworthy one, followed by the keyword movie, the keyword audience, and the keyword release, in that order. In section 2.5, we have shown that the words director, military, sales, theater, etc. show up in the center and around the center. This in turn shows that they are all the noteworthy ones in 23 articles. Finally, we have provided the sentiment analysis of 23 articles. More specifically, we obtained +669 with respect to positive words, while we obtained -75 with respect to negative words. It seems clear from our findings that the spring of Seoul has received many favorable reviews from viewers and the press.

References

[1] Kang, N. (2023a). K-Pop in BBC News: A Big Data Analysis. *Advances in Social Sciences Research Journal 10*(2), 156-169.

[2] Kang, N. (2023b). K-Dramas in Google: A NetMiner Analysis. *Transaction on Engineering and Computing Sciences 11*(1), 193-216.

[3] Kang, N. (2023c). A Comparative Analysis of Tolerate and Put up with in the COCA. *Semiconductor and optoelectronics* 42(1): 1468-1476.

[4] Kang, N. (2023d). Sure of and Sure about in Corpora and ChatGPT. *Journal of Harbin Engineering University* 44(7): 1347-1351.

[5] Kang, N. (2023e). Turn out adj and Turn out to be adj in the Now Corpus and ChatGPT. *Journal of Harbin Engineering University* 44(8): 825-831.

[6] Kang, N. (2023f). Care for and Like in Corpora and ChatGPT. *Semiconductor and optoelectronics* 42(2): 188-198.

[7] Kang, N. (2024a). A Big Data Analysis of a Hot Political Issue. Studies in Linguistics 70: 149-165

[8] Kang, N. (2024b). A study of the Prediction of the Model Logistic Regression in Machine Learning : Focusing on a Survey.