The Analysis of the 2018 New Year's Address of DPRK

Byungmarn Koh*, Minsoo Kim, Sunyoung Chang
Korea Institute of Nuclear Nonproliferation and Control, 1534 Yuseong-daero, Yuseong-gu, Daejeon, 34054
Corresponding author: marn@kinac.re.kr

1. Introduction

In 2017, North Korea announced in the new year's address that it was ready to launch an ICBM, and finally declared the completion of the nuclear capability after continuing military provocations including the sixth nuclear test and the testing of an ICBM. In series of process, the international community responded with sanctions on North Korea, the United States threatened North Korea with a strong military option and the Korean peninsula start the 2018 new year in a tense state.

First day on 2018, North Korean leader Kim Jong Eun announced the new year's address and some said it is a 'surprise message'. Therefore, it analyzed the new year message of North Korea more scientifically using the text mining method.

2. Methods of the Analysis

2.1 Text Mining

It is a technique that formalizes unstructured text data, extracts characteristic of formalized data and derives meaningful information from extracted characteristic by using natural language processing technology based on linguistics, statistics, and machine learning [1].

The analysis of the new year message of North Korea used the following techniques.

- Word Association Analysis: to analyze the association of words in a sentence and a document
- Keyword and Cluster Analysis: to extract frequently used words within a document and to classify it into similar groups

2.2 Association Rule Analysis

It analyzes the associations between individual data by extracting rules that occurred in large amount of data. Thus, in the text mining, it can be understood that the association rule is used to indicate the association between words used together when a particular word is used [2].

The association rule is the analysis techniques for finding the hidden patterns in data without reference values, and the dozens to thousands association rules are derived from the analysis [2]. Thus, the valuation standard for the association rule is very important and three measures are used:

- Support: The support for two words A and B can be understood as a percentage of sentences that contain both A and B throughout the document. It is used as a criterion for finding good rules that often appear.
- Confidence: It is the ratio (conditional probability) of a sentence that also contains the word B among the sentences containing the word A. The higher confidence rule, the more useful rule.
- Lift: If the association between word A and B is independent, the lift is 1. The closer the lift is to 1, the more likely it is that the associated rules are random chance.

2.3 Cluster Analysis

It is a method to group the similar or homogeneous data in the dataset into the cluster. It is often used in the initial analysis to confirm the meaningful data structure or to understand the structure of the data. It is also possible to identify hidden information or patterns via a cluster without the background knowledge of the dataset [2].

2.4 Data Pre-processing

An actual data may be in an incomplete state that includes the noise or lacks the consistency. Consequently, it is necessary to convert the actual data into the information data for the analyzing. There are techniques such as data refining, integration, conversion, and reduction [2].

3. Results of the Analysis

3.1 Results of the Pre-processing

In the text mining analysis for the new year's address, only the nouns were extracted and used. In the analysis process, the pronoun such as *'We'* and *'I'* was used very frequently, but it was removed so that it had no effect on the analysis results.

3.2 Results of the Association Rule

Figure 1 shows the result of association rule analysis for the new year's address of North Korea.

The 29 association rules were derived by applying the conditions of the support of 0.05 or higher and the confidence of 0.6 or higher. On the figure, the circle size indicates the support value and the circle color indicates the lift value. That is, the bigger the circle is the higher support value and the darker the circle color is the higher lift value.

In the message, it can be inferred that *'the nationality'* identity for *'the South'* and *'the North'* is emphasized, and *'the South Korean authority'* is required to play something role in the inter-Korea relationship. Also, it can be inferred showing antipathy against the US by using the negative word such as *'a machination'*.

It can be inferred that *'construction of the strong socialist country'* is promoted internally, and the relationship between *'the party'* and *'the people'*, and *'the struggle'* based on *'the people'* were emphasized.
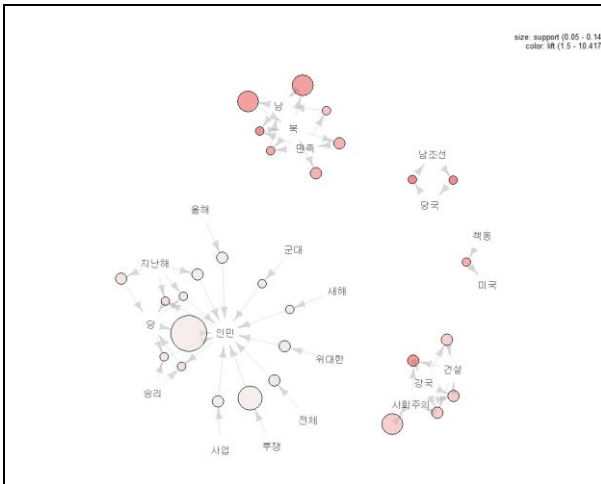

Fig. 1. The result of Association Rule Analysis.

3.3 Results of the Cluster Analysis

In order to analyze the word cluster, the 30 words which are frequently used in the new year's address are extracted and figure 2 shows the result that. The 30 words with high frequency were somewhat similar to the used words in the association rule analysis.
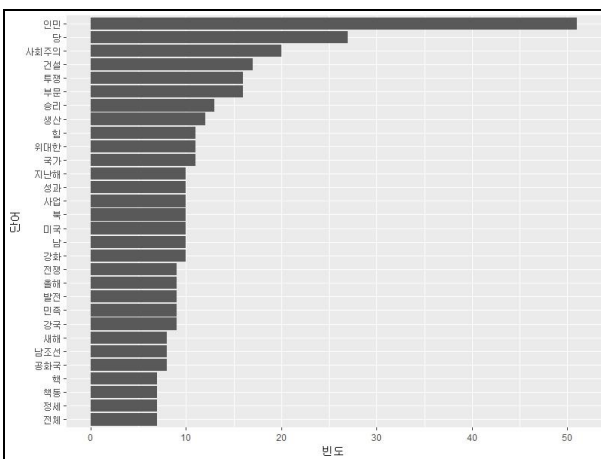

Fig. 2. The 30 words with high frequency

Figure 3 shows the results of a cluster analysis of the 30 words with high frequency. It forms a hierarchy from the bottom to the top according to the similarity of words, and the four red square are the classified clusters according to similar homogeneity.
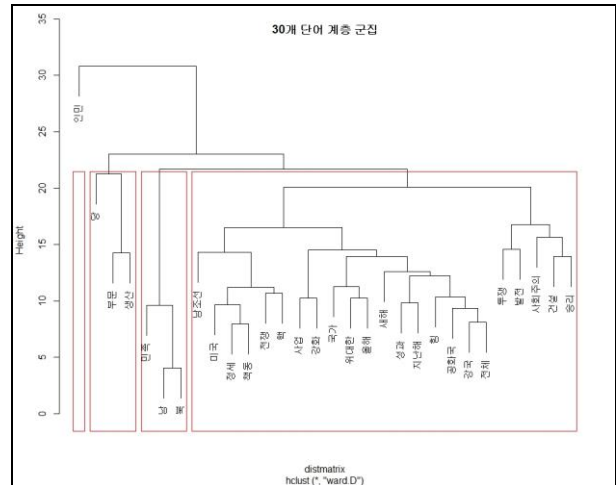

Fig. 3. The result of Cluster Analysis

It can be inferred that *'the party - the production · the sector'* is an economic issue and *'the nationality - the South · the North'* is the inter-Korean relationship.

Especially, the cluster including *'the socialist construction'* is higher located on the top than the cluster of *'the U.S.'* and *'the nuclear'*. It can be inferred that North Korea will be focusing on economic issues and inter-Korean relations after declaring of the completion of the nuclear capability at the end of 2017.

4. Conclusions

The analysis results for the 2018 new year's address of North Korea through the text mining analysis, there is a strong willingness to construct the socialist country internally and to improve the inter-Korea relationship externally. Also, it has been shown that priorities are given to various activities based on the party and the people rather than relations with the United States and the nuclear issue.

REFERENCES

[1] National Science Museum, Big Data, www.science.go.kr
[2] R Friend Blog, rfrend.tistory.com
[3] Text Mining with R, Julia Silge & David Robinson, www.tidytextmining.com
[4] Visualise Blog, visualize.tistory.com
[5] Dbrang Blog, dbrang.tistory.com