Transactions of the Korean Nuclear Society Autumn Meeting
Changwon, Korea, October 22-23, 2020

What Experts Care About in Satellite Observation : A Quantitative Analysis for Countering Nuclear Proliferation

Jae-Jun Han [a*] and Nam Kyung Kim [a]
[a] *Center for Nuclear Nonproliferation Strategy & Technology, Korea Institute of Nuclear Nonproliferation and Control, 1418 Yuseong-daero, Yuseong-gu, Daejeon 34101, Republic of Korea*
**Corresponding author: jjhan@kinac.re.kr*

1. Introduction

For countering nuclear proliferation, international society pays much attention to the potential or identified nuclear proliferant states. As providing advantages on accessibility to suspicious facilities in the rogue states, satellite observation has been playing an essential role in monitoring their potential nuclear activities.

Nuclear weapons as strategic arms have to consist of three components: fissile material, delivery system, and detonation devices. The weapons-grade fissile material can be acquired from uranium enrichment (about 90% U-235) or reprocessing used nuclear fuel (about 93% Pu-239). Hence, it is necessarily led to keep eyes on the relevant facilities and carriage of required materials by means of satellite surveillance. As for the delivery system, test sites and manufacturing factories are a matter of concern. Due to the size of a nuclear weapon, the area of interest for the delivery system is rather clearly to be revealed and investigated. On the other hand, there is no need to be special-exterior structures to develop detonation devices. Although any ambiguous activities at the border, e.g., trade traffic, could be captured on satellite imagery, a detailed investigation is under the implementation of strategic trade controls.

Identifying relevant activities not only for fissile material but also for delivery system involves expertise in the field of both nuclear fuel cycle and missile technology. Thus, the imagery analysts who are proficient in both areas, as well as remote sensing, are understandably small in number. They intermittently provide interpretation from satellite-observed information in the form of web-based articles via non-profit organisations such as 38 North and Center for Strategic and International Studies (CSIS). The articles inevitably include the logical basis for inferring nuclear proliferation, even though it is an external sign. As shown in Fig. 1, the annotations on satellite imagery represent what the experts practically capture, and the body contents of the article expand their interpretation.

This study quantitatively analyses the articles via web scraping to investigate what the experts care about in satellite observation. The following section presents the web scraping method for producing the article data set. Based on the data set, Section 3 discusses results from the article categorisation and the image annotation classification. The conclusions are drawn with future work in Section 4.

Fig. 1. Building construction continues in front of the ELWR (Excerpt from the 38 North article on 4 April 2018 [2])

2. Scope and Method

The web scraping is the process of automated gathering of data from the web pages which arrange objects in a more human-readable format [1]. Visiting every page and storing the hundreds of text contents and images are economically unfavourable. Thus, web scraping algorithms have been developed with the Python programming language [4] for the websites of interest: (1) 38 North [2] and (2) CSIS Beyond Parallel [3] which are the most active and being cited by the media. Both provide not only imagery-based articles but also text-only articles for various topics, and this study focused on the former only.

In terms of the entire platform of Python programming, the Selenium with Chrome web driver [5] is utilised to navigate webpages, and a parse tree for the pages that can be used to extract data from HTML is created via the Beautiful Soup library [6].

The scraping targets of an article are an article title, author(s), release date, body text, image(s), and image title(s). When scraping data from a website, targets are not always embedded in the identical HTML tag, so-called anomaly, which is the chronic challenge in the web scraping. For instance, the HTML tag content for the author(s) name occasionally is occupied as the organisation name instead, and the author(s) name is presented in the first line of the body text. Depending on the consistency of the first line, the automated gathering of the author(s)'s name may or may not be possible. If the number of the case is very few, filling

proliferation. The type of sign (direct or indirect) in the left column of Table 3 is decided for the classified activities based on connectivity to the facility for nuclear fuel cycle or weaponisation. For instance, class 'operation' stands for observation associated with fuel cycle facilities such as steam arise from the 5 MW_e reactor, snowmelt on the uranium enrichment plant, etc. On the other hand, class 'transport' including rail car/flat car, truck, etc., represents shipment or (at least) movement within the nuclear complex, and they are not directly connected to the facility for nuclear fuel cycle or weaponisation.

4. Conclusions

Due to the deficiency of the subject-matter experts in both the imagery interpretation and nuclear fuel cycle, this study has built the data set from satellite-observed information via the web scraping of the articles for countering nuclear proliferation. Through the 38 North and CSIS Beyond Parallel websites, the expert(s) has produced 389 articles of which topics have been categorised into the nuclear, missile, and others. In the nuclear category, specifically, the proportions of Yongbyon and Punngye-ri are the highest. The data set is expected to be a cornerstone for the in-depth analysis, including author network analysis, natural language process, etc.

Further, 510 annotated images for the Yongbyon nuclear complex have been extracted to classify what expert(s) care about in satellite observation. The indicators for inferring nuclear proliferation have been classified into nine classes: construction, pipeline, powerline, transport, dredge, object, fuel, operation, etc.

The results for classification are valuable in that they can be utilised by non-experts or novice interpreters in countering nuclear proliferation. Moreover, for automated change detection with algorithms, this provides a framework for satellite-observed objects that the system has to detect.

ACKNOWLEDGEMENT

This work was supported by the Nuclear Safety Research Program through the Korea Foundation Of Nuclear Safety (KoFONS), granted financial resource from the Nuclear Safety and Security Commission (NSSC), Republic of Korea (No. 1905009).

REFERENCES

[1] R. Mitchell, Web scraping with Python: Collecting more data from the modern web, O' Reilly Media, Inc., 2018.
[2] Articles – Topic: Satellite Imagery, Available at https://www.38north.org/topics/satellite-analysis/ (Accessed: 26 Aug 2020)
[3] Articles – Imagery, Available at https://beyondparallel.csis.org/imagery/ (Accessed: 26 Aug 2020)
[4] G.V. Rossum and F.L. Drake, The python language reference manual. Network Theory Ltd., 2011.
[5] Selenium WebDriver documentation, https://www.selenium.dev/documentation/en/ (Accessed: 26 Aug 2020)
[6] Beautiful Soup documentation, https://www.crummy.com/software/BeautifulSoup/bs4/doc/ (Accessed: 26 Aug 2020)
[7] Naver Clova Optical Character Recognition (OCR) platform, https://clova.ai/ocr (Accessed: 26 Aug 2020)