Proceedings of the INMM & ESARDA Joint Virtual Annual Meeting
August 23-26 & August 30-September 1, 2021

MALICIOUS ACTION MONITORING TECHNOLOGY BY COUPLING DEEP LEANINGS OF IMAGE RECOGNITION AND NATURAL LANGUAGE PROCESSING

Kazuyuki Demachi
Nuclear Professional School,
School of Engineering,
The University of Tokyo

Shi Chen
Department of Nuclear
Engineering and Management,
School of Engineering,
The University of Tokyo

Masaki Sudo
Department of Nuclear
Engineering and Management,
School of Engineering,
The University of Tokyo

ABSTRACT

In recent years, deep learning has been very successful in various fields. In particular, since image data contains a lot of information, it is expected that deep learning will be applied to nuclear security technology for detecting the malicious action. However, most human action recognition using deep learning only sets OK / NO in advance for the combination of object identification result, so it is difficult to apply them to complicated nuclear security situation. On the other hand, in most nuclear facility, the nuclear security rules are determined with precise documents. So, it is desirable to detect malicious actions in captured images according to these rule documents. If it is possible to flexibly determine whether the captured image scene violates or matches the nuclear security rules, it is necessary to greatly improve its practicality. However, this requires different deep learning interfaces for image identification and natural language processing. Most of the deep learning research currently being conducted is closed only within each field, and there is no mutual access, that is, an interface. The human brain, on the other hand, usually links and processes all kinds of cognitive information, including language, sight, speech, and touch. Like the brain, deep learning abilities should develop anew when different types of deep learning are interconnected. One of the reasons this is difficult is that there is no common data format between different types of deep learning. In this study, it was determined that the logic representation is the most suitable as a common data format. The logic representation is a method of expressing the correlation between an object and a person by connecting nodes with edges. In this research, a deep learning interface method between image recognition and natural language processing was developed via a logic representation. This allows flexible judgment of malicious actions according to the rule document.

1. INTRODUCTION

An urgent lesson learned from Fukushima Daiichi accident is what can happen by natural disaster can also occur by malicious human's action. The accident raised a fear that terrorists could cause a similar accident by acts of sabotage against nuclear power plant (NPP) and it is noticeable that threats of terrorism for nuclear security are increased after the accident. When considering sabotage, the prime threat to nuclear power plants, attention should be paid to sabotage by insiders. Generally, insiders are the individuals with authorized access to nuclear facilities in transport who could attempt unauthorized sabotage. They could take advantage of their access authority and knowledge, to bypass dedicated physical protection elements or other provisions. Thus, we should value the catastrophic consequences of the attack or act of insider sabotage which may lead to loss of significant safety functions of NPP.

IAEA indicated that the physical protection system (PPS) of a nuclear facility should be integrated and effective against both sabotage and unauthorized removal [1]. The primary PPS functions are deterrence,

detection, delay and response. It is serious that if detection failed, delay and response would be invalid. Thus, detection of insiders' sabotage should be enhanced. Considering current countermeasures of PPS to insiders' sabotage, the most significant challenge is how to distinguish ordinary maintenance behaviors and malicious behaviors since some malicious behaviors may hidden in ordinary maintenance behaviors.

On the other hand, in recent years, deep learning has been very successful in various fields. In particular, since image data contains a lot of information, it is expected that deep learning will be applied to nuclear security technology for detecting the malicious action. However, most human action recognition using deep learning only sets OK / NO in advance for the combination of object identification result, so it is difficult to apply them to complicated nuclear security situation. On the other hand, in most nuclear facility, the nuclear security rules are determined as detailed text. So, it is desirable to detect malicious actions in captured images according to these rule texts. If it is possible to flexibly determine whether the captured image scene violates or matches the nuclear security rules, it is necessary to greatly improve its practicality. However, this requires different deep learning interfaces for image identification and natural language processing.

As mentioned above, a lot of AI research is currently being conducted. But most of them are closed only within each field, and there is no mutual access, that is, no interface. The human brain, on the other hand, usually links and processes all kinds of cognitive information, including language, sight, speech, and touch. Like the brain, deep learning abilities should develop anew when different types of deep learning are interconnected. One of the reasons this is difficult is that there is no common data format between different types of deep learning.

The logic representation is a method of expressing the correlation between an object and a person. In this research, a deep learning interface method between image recognition and natural language processing was developed via a logic representation. This allows flexible judgment of malicious actions according to the rule text.

So what are the possible common data formats? In our laboratory, after verification with several candidates, it was concluded that the logic representation [2] is optimal. In this research, three algorithms were developed for the purpose of creating an interface between image AI and natural language processing AI that use the logic representation as a common data form. This makes it possible to implement a very convenient and versatile technology, for example, the violation of rule captured in an image can be automatically determined simply by inputting a rule sentence as a determination standard.

2. REQUIRED FUNCTIONS

2.1 Conditions Required for Malicious Acts Judgment

All existing anomaly detection methods using image recognition AI require a large amount of data sets during normal times, but it is difficult to collect data during normal times because there is few illegal or malicious scene of nuclear security. So, the illegal and malicious acts detection AI must meet the following requirement-1;

Requirement-1: No need for training data during normal times.

Illegal or malicious acts related to nuclear security have a various definitions for situation and conditions such as when, where, who, and what are essential. Moreover, the rules that define such acts are very complex, too. However, it is not possible to identify what kind of incident is occurring only by detecting the deviation from the normal situation, and it is not suitable. Therefore, the following requirement-2 is required.

Requirement-2: Binary judgment based only on video features is not applied.

In the nuclear security rule text, various scenes are assumed using "texts" to prevent illegal and malicious acts, and acts that "must be done" and "must not be done" are defined for each scene. Therefore, the illegal and

malicious acts detection AI must be able to concretely judge what kind of action "must / must" be done depending on the place, situation, person, and object. Therefore, the following requirement-3 is required, too.

Requirement-3: Illegal & malicious acts expressed by nuclear security rule sentences can be specifically classified

2.2 Proposal of Malicious Insider Act Detection Method

Natural language is a language that humans use in their daily lives, such as Japanese and English, and the natural language processing is a technology that analyzes and processes these with a computer. Natural language processing is rapidly developing due to the increase in electronic text data due to the spread of the Internet and the establishment of data analysis methods by machine learning, and the highest accuracy record is set every year by a new model.

There are a rule-based approach and a machine learning approach as sentence classification methods. The rule-based approach has relatively high extraction accuracy because it manually develops important sentence extraction rules, but it takes a lot of time and effort to develop the rules. On the other hand, the machine learning approach can automatically classify by learning important sentences and non-important sentences in advance, but the accuracy decreases when the amount of training data available in a specific area is limited. For erasing unnecessary information, there is a classical method using dictionary data that extracts only necessary information by understanding the unique information of a word with a dictionary. As a semantic analysis, there is a method of individually matching the subject part, the verb part, and the object part by comparing with a predefined sentence pattern. Similar to this method, in the rule described in the style of "if-then", if the "if" condition is met, a warning message is presented if the "then" condition is also met.

2.3 Malicious Acts Detection System Currently under Development

Figure 1 shows the illegal / malicious acts detection AI system currently under development. First, the image recognition AI recognizes the face, place, object, whole body movement, and hand movement through a surveillance camera or smart glasses, and outputs 3W1H (who, where, what, how) as words. This method meets requirements 1 and 2 above. In this method, since the vocabulary output as image information is only the words learned in advance, the word list is saved as pre-registered data. In this method, YOLOv3 [3] is used for object recognition, FaceNet [4] is used for face recognition, OpenPose [5] is used for motion recognition, and Hand3D [6] is used for hand motion recognition. Then, we classify these by the method we developed using a convolutional neural network, and appropriately output 3W1H contained in the image information.

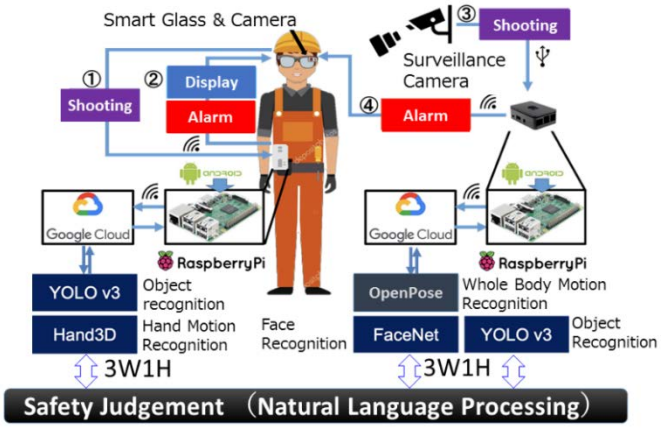


Figure 1. Malicious acts detection AI system currently under development

2.4 Image Recognition to Extract Logic representation of Captured Movie

Here, YOLOv3 [3] was used for object recognition, and OpenPose [5] was used for attitude recognition. Each architecture is shown in Figures 2 and 3.

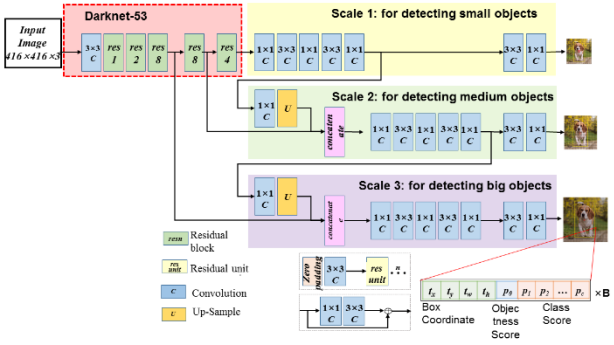


Figure 2. YOLO v3 architecture [3]

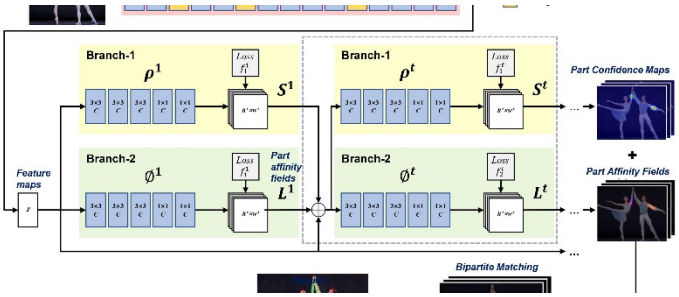


Figure 3. OpenPose architecture [5]

When YOLOv3 and OpenPose were detected, the object and attitude were set to S_i and T_j, and a threshold value was applied to the mutual Euclidean distance to determine the presence or absence of a correlation between S_i and T_j. (Figure 4)

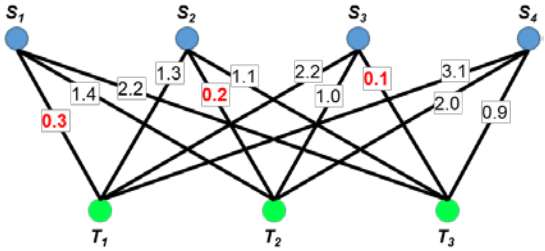


Figure 4. Relationship extraction based on Euclidean distance [7]

2.5 Natural Language Processing for Malicious Acts Detection

Next, the natural language processing related to the rule sentence shown in the lower part of Figure 1 will be explained. This also meets Requirement-3 above. In this research, the focus is on logically expressing this rule sentence after removing unnecessary information.

There is a previous study targeting English sentences as natural language processing for detecting malicious acts, but it is difficult to apply the same method to Japanese rule sentences because the grammars of English and Japanese are very different. Therefore, it is necessary to devise an original logic representation algorithm that is optimal for Japanese processing.

In general, sentences that define the same situation can take various expressions, which is the same in Japanese. The sentence obtained by converting the image recognition result and the sentence obtained by analyzing the rule sentence may have different expressions even if they describe the same situation. Therefore, in order to compare these correctly, it is necessary to remove the fluctuation between the words registered in advance for image recognition and the words in the rule sentence. Words in the rule sentence must be converted into words obtained by pre-registered image recognition. In addition, in order to compare the pre-registered word for image recognition with the word in the rule sentence, it is necessary to classify in advance whether the rule sentence is a prohibited sentence or a compliant one. Furthermore, a logic representation system that can correctly express the subject, predicate, and object of an action is required. In this research, a method for converting rule sentences

into logic representation s is developed for comparison with image information. Then, by comparing the two logic representation verbs, we aim to develop a method for detecting malicious acts.

i. An algorithm that classifies whether it is a prohibition rule or a compliance rule.
ii. Algorithm that correctly understands subjects, predicates, and objects
iii. Algorithm to remove word fluctuation
iv. Algorithm for creating comparable logic representation
v. Algorithm for comparing logic representation s of rule and image

are particularly important for natural language processing in a malicious activity detection system.

3. PROPOSED FOUR ALGORITHM BASED ON NATURAL LANGUAGE PROCESSING

3.1 Natural Language Processing for Malicious Acts Detection

First, as algorithm A, it is necessary to classify in advance whether the rule sentence is prohibited or compliant. In natural language processing, affirmation (to do) and negation (do not do) are likely to be interpreted in reverse, so it is effective to classify compliant sentences and prohibited sentences in order to improve the accuracy of sentence interpretation. In addition, unimportant sentences must be removed in advance so as not to make unnecessary judgments for situations that do not need to be dealt with. Algorithm A classifies rule sentences into three types: compliance sentences, prohibited sentences, and non-important sentences. Since the situation needs to be focused more than words, a natural language processing model that can capture the meaning of the sentence is needed. Moreover, since the data that can be used is limited, it is desirable that the required amount of training data is low.

Therefore, in this study, it was proposed to use BERT (Bidirectional Encoder Representations from Transformers) [8], which is a pre-learning model that can be applied to various tasks by additional learning. BERT is a natural language processing model announced by Google on October 11, 2018.

The BERT model is fundamentally different from existing task processing models. The natural language processing model using a conventional neural network can be applied only to specific tasks such as understanding sentences and analyzing emotions, but BERT is a pre-learning model that can be applied to various tasks by fine tuning. BERT consists of two steps: first pre-learning a large corpus and then fine-tuning according to each task. Only a small amount of fine tuning is required, so there is no need to collect large amounts of data for a particular task. Since pre-trained models corresponding to various languages are open to the public, users can perform highly accurate natural language processing by preparing a small amount of labeled data.

BERT pre-learning is a two-step process. The first is the Masked Language Model, which randomly masks or replaces 15% of the words in a sentence and predicts the correct word from the context.

The second is Next Sentence Prediction, which learns the task of guessing whether two sentences are "next to each other". Given adjacent sentences A and B, there is a 50% chance that B will be replaced by another sentence.

Figure 5. Task for masked language model

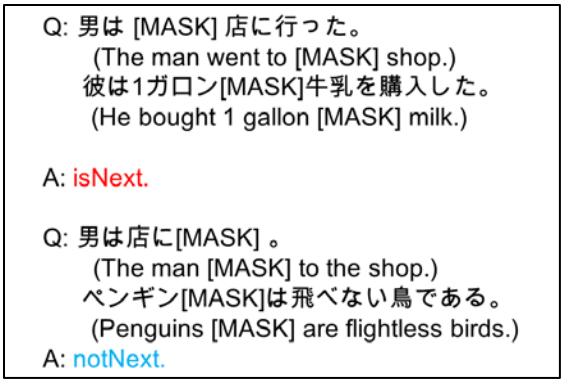


Figure 6. Task for next sentence prediction

The pre-learning of BERT is completed by applying this method to a large amount of text data. Once one trained model is completed, a wide variety of natural language processing will be possible after that with only fine tuning with a small amount of data.

Algorithm A needs to classify three types of sentences, but even with BERT it is difficult to do this with only a small data set. Therefore, in this study, a method using two-step binary classification was proposed to improve the judgment accuracy. BERT is better at binary classification than trivalent classification, and since fine tuning can be performed with different optimum data sets for first-stage classification and second-stage classification, this method is sufficient even with a small data set. Classification accuracy can be achieved. This makes it possible to remove not only compliance sentences and prohibited sentences but also non-important sentences with high accuracy in advance.

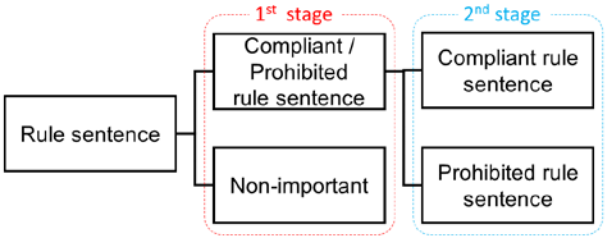


Figure 7. Two-step binary classification of rule sentences

In this study, an appropriate fine tuning dataset was prepared for each stage in order to obtain sufficient accuracy in the two-stage binary classification. In the first stage, there was no developed rule-based approach program that could be applied, and it took a lot of time and effort to set the rules, so additional learning data was created by myself. By manually labeling 200 rule sentences and additionally learning them, the rule sentences other than the training data were classified into two patterns, important or not. On the other hand, in the second stage classification, 5000 sentences of Livedoor news corpus [9] are labeled using a general rule-based approach algorithm for polarity judgment that can be used for binary judgment of compliance sentences and prohibited sentences. It was used as data for BERT fine tuning. Furthermore, we aimed to improve the accuracy by analyzing sentences whose classification was still incorrect even after learning the data of these 5000 sentences, and creating sentences that captured the characteristics by ourselves and adding them to the fine tuning data.

3.2 Parsing

In function B, it is necessary to clarify the subject, predicate, and object by analyzing the sentence structure of the rule sentence, and to grasp the sentence structure. As a result, important words and their part-speech information can be efficiently extracted in creating a logic representation, and the same part-speech can be

compared with each other when comparing with image information. Furthermore, if we look at the final logic representation and its comparison, the technique of defining a consistent syntactic structure and tag set is useful.

In morphological analysis in natural language processing, text data is divided into morphemes, which are the smallest units, based on language grammar and dictionary data, and part of speech information is determined for each. The parsing clarifies the dependency relationship while confirming the meaning in the sentence based on the part-speech information found by this morphological analysis. Especially in Japanese, where case grammar is established as the basic grammar, the dependency tree that clearly shows the role of each word is very important, and the meaning of the sentence can be correctly interpreted by using the syntax tree model. Here, if the word string $\boldsymbol{\omega}$ and the dependency tree are $\boldsymbol{d}$, and the edge score $\sigma(\langle i,d_i \rangle,\boldsymbol{\omega})$ is defined by machine learning, the dependency tree $\widehat{\boldsymbol{d}}$ that maximizes the sum of the edge scores is as follows. It becomes possible to search from all the trees for the input sentence.

$$\widehat{\boldsymbol{d}} = argmax \sum_{i=1}^{n} score\,, \quad score = \sigma(\langle i, d_i \rangle, \boldsymbol{\omega}), \quad \boldsymbol{d} = \langle d_1, d_2, d_3, \cdots \rangle \tag{1}$$

Universal Dependencies (UD) [10] aims to enable standardization of post-parsing processing methods, application of models learned in corpora of other languages, etc., and consistent tags in multiple languages and defines sets and syntactic structures. In addition to simplifying expressions, UD expresses all syntactic structures with inter-word dependencies and their labels, and does not consider phrase structures, in order to be robust against simple sentences and special structures.

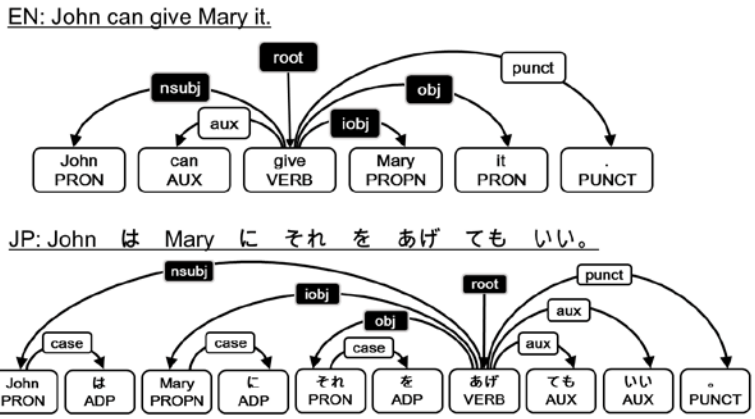


Figure 8. Analysis examples of English UD and Japanese UD [10]

For the dependency analysis in this study, GiNZA [11], which is a model that can convert Japanese into UD, was used for the next logic representation. Since UD defines a consistent tag set and syntactic structure in multiple languages, it has made it possible to apply logic representation creation methods in other languages to Japanese as well.

3.3 Synonym Correction

Next, in function C, it is necessary to solve the fluctuation of the word in the rule sentence and the word in the image information. Since the information output from the image is only the pre-programmed words, the words for creating a logic representation from the rule sentence should also be unified with the vocabulary of the image information. That is, words that are not used in the image information must be replaced with the words that are used in the image information by referring to the synonym database.

There are roughly two methods for searching for similar words: similarity calculation and thesaurus. Similarity calculation is an algorithm that uses a neural network to learn the vector representation of words. To learn the vector representation of a word, a distribution hypothesis is used that assumes that the meaning of the word is determined by the words around the word. However, in general, the distribution hypothesis not only learns synonyms, but also antonyms tend to co-occur near the word. Therefore, the similarity of the pair of antonyms becomes very high.

On the other hand, Wordnet [12] is a dictionary that manually defines the relationships between words, and was used as a synonym database for this study. By recording manually, the antonym pair is not misjudged, and it can be said that the reliability and completeness are sufficient in that it has been published for many years while repeating corrections and additions. Therefore, all words in the rule sentence can be replaced one by one with the same expression as the image word by searching on Wordnet.

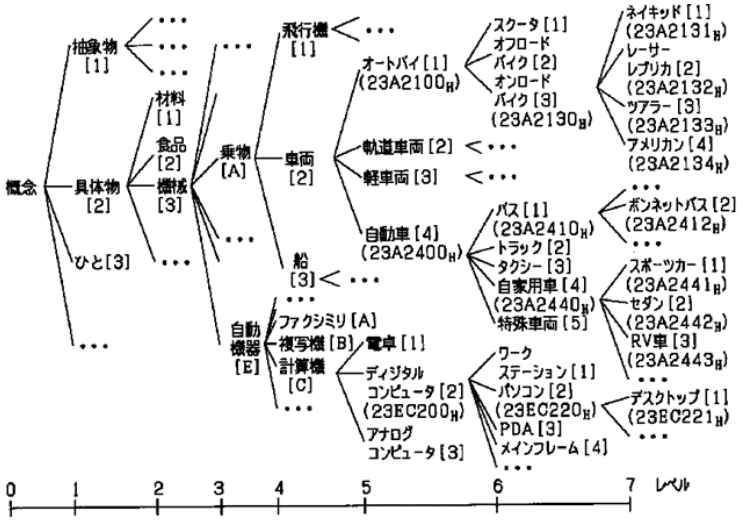


Figure 9. WordNet, a synonym database [12]

3.4 Logic representation

In function D, the rule sentence must be converted into a format that can be compared with the image information while leaving only the necessary information. That is, a logic representation method is required in which symbols representing contextual information are added to words. In image recognition AI, registered words are limited and the subject, predicate, and object are clear, so image information can be easily logically expressed by programming certain rules. The goal of this study is the logic representation of rule sentences, which must be more complex than image information. Since the logic representation can be converted into relational algebra operations, it has logical properties that hold in any database. Therefore, it can be determined by logical reasoning whether the two meanings are the same.

However, it is said that it is very difficult to logically express Japanese with existing methods. A typical example of a logic representation method is the Combinatory Categorial Grammar (CCG), which is mainly used for English. In CCG, the sentence structure is described by the CCG category, which expresses restrictions on the combination with surrounding words, and a small number of combination rules. Since it is a very simple model, it can be derived by full search, but it is prone to ambiguity when multiple his CCG trees can be derived. An existing study [13, 14] has shown that the use of dependent trees makes the CCG category a stronger constraint and achieves higher accuracy in English. In this research, we try to express Japanese logically by using Universal Dependencies, which is a common form of dependency analysis in English and Japanese.

3.5 Comparison with Image Information

In comparing logic representation s, it is very important to judge the match / mismatch of contents regardless of the order of words. In the logic representation creation method of this research, important words in a sentence are numbered in order from the front, so even the same word will have different numbers depending on its position in the sentence. Therefore, in this study, we search for matching predicates from the observance sentence and the logic representation of the image, and define the case where the words of the recipients match exactly as safe.

Regarding the logic representation of the prohibited sentence and the image, the case where the predicate matches, only one of them is a negative expression, and the dependency is an exact match is defined as BAD. The logic representation is represented by the relationship between the numbered x-word and e-word, and their corresponding case particles. Therefore, in order to correctly judge the meaning regardless of the word order, we created an algorithm that performs matrix operations so that the order of common words is aligned. The upper part of Figure 10(a) is the comparison of a logic representation for a compliant rule and Figure 10(b) is that for a prohibited rule.

In the compliant rule sentence in Figure 10(a), "A" is x0, "B" is x1, "C" is e0, "D" is x2, "E" is e1, and in the image output, "D" is x0, "E" is e0, "B" is x1, "F" is x2, "A" is x3, and "¬C" is e1. Because it is a compliance rule, the logic representation of the rule included in the logic representation of the image are extracted and compared. C matches on rule and image, but A and B do not match, so judge is BAD.

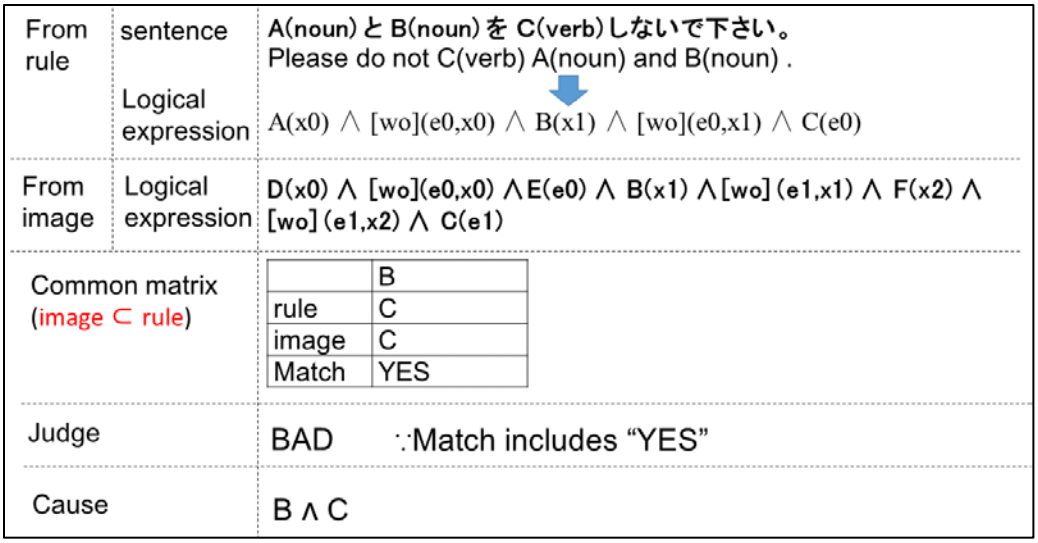
In the prohibited rule sentence in Figure 10(b), "A" is x0, "B" is x1, and "C" is e0, and in the image output, "D" is x0, "E" is e0, "B" is x1, "F" is x2, and "C" is e1. Because it is a prohibited rule, the logic representation of the image included in the logic representation of the rule are extracted and compared. C matches on rule and image, so judge is BAD.

From rule	sentence	A(noun)と B(noun)を C(verb)、D(noun)を E(verb)して下さい。 Please C(verb) A(noun) and B(noun), and E(verb) D(noun) .
	Logical expression	A(x0) ∧ [wo](e0,x0) ∧ B(x1) ∧ [wo](e0,x1) ∧ C(e0) ∧ D(x2) ∧ [wo](e1,x2) ∧ E(e1)
From image	Logical expression	D(x0) ∧ [wo](e0,x0) ∧E(e0) ∧ B(x1) ∧[wo](e1,x1) ∧ F(x2) ∧ [wo](e1,x2) ∧A(x3) ∧ [wo](e1,x3) ∧ ¬C(e1)
Common matrix (rule ⊂image)		

	A	B	D
rule	C	C	E
image	¬C	¬C	E
Match	NO	NO	YES

Judge	BAD ∵Match includes “NO”
Cause	A ∧ ¬C, B ∧ ¬C

(a) For compliant rule



(b) For prohibited rule

Figure 10. Logic representation comparison algorithm using matrices

4. RESULTS AND DISCUSSION

4.1 Calculation Result of Two-Step Classification of Rule Sentence

Since the nuclear security rule text cannot be published, the algorithm A is used to classify the occupational safety rule text at a general construction site as a substitute. In the first stage, important sentences and non-important sentences were classified, and in the second stage, the classified important sentences were further classified into compliance sentences and prohibited sentences. Since the purpose of this study is to determine unnecessary work risk for large-scale datasets, only 235 rule statements manually entered and labeled were used for the dataset.

In the first stage, transfer learning was performed using 212 of the 235 safety rule texts that were manually labeled as important and irrelevant as training data and 23 as test data. As shown in Table 1, there were 15 true positive (TP) sentences, 0 false positive (FP), 1 false negative (FN), and 7 true negative (TN). The precision and recall were 100% and 93.8%, respectively.

Table 1. First stage classification result

		Correct	
		TRUE(=Important)	FALSE(Not-important)
Judgement	TRUE(=extracted)	15	0
	FALSE(=not extracted)	1	7

$$\text{Precision} = TP/(TP+FP)$$
$$\text{Recall} = TP/(TP+FN) \quad (2)$$

In the second stage, 5000 general sentences of the Livedoor news corpus were labeled by a rule-based approach and used as training data. Then, all 235 sentences in the occupational safety manual, which were manually labeled as compliance / prohibition, were used as test data. Furthermore, we considered the sentences that were incorrect from the classification results of the test data, and aimed to improve the accuracy by creating the sentences that captured those characteristics and adding them to the training data. As shown in Table2, the judgment accuracy when using only the Livedoor news corpus as training data were, TP = 69, FP = 10, FN = 51,

TN = 105, and Precision and Recall are 87.3% and 57.5%, respectively. Next, after adding 6 self-made sentences to the training data, TP = 101, FP = 10, FN = 19, TN = 105, and the Precision and Recall rose to 90.9% and 84.1%, respectively. After adding more 5 self-made sentences, TP = 114, FP = 4, FN = 6, TN = 111, and Precision and Recall improved to 96.6% and 95.0%, respectively.

Table 2. Second stage classification results

(a) Only the Livedoor news corpus as training data

		Correct	
		TRUE	FALSE
Judgement	TRUE	69	10
	FALSE	51	105

(b) 6 self-made sentences added to the training data

		Correct	
		TRUE	FALSE
Judgement	TRUE	101	10
	FALSE	19	105

(c) After adding more 5 self-made sentences to the training data

		Correct	
		TRUE	FALSE
Judgement	TRUE	114	4
	FALSE	6	111

4.2 Calculation Result of Syntactic Analysis

In Algorithm B, syntatic analysis was performed on compliance rule sentences and prohibited rule sentences classified from the occupational safety rule text by Algorithm A. Figure 11 is an example of created dependency tree from a compliance rule sentence, “使用後はスイッチを切ってください= Please turn off the switch after use,” and Table 3 shows its extracted information. Figure 12 is also an example of dependency tree from a prohibited rule sentence, “側面や上面では研削しないで下さい。= Please do not grind on the sides or top.” The most important thing here is to clarify the subject, predicate, and object, and to understand the structure of the sentence, but it can be seen that each is extracted correctly.

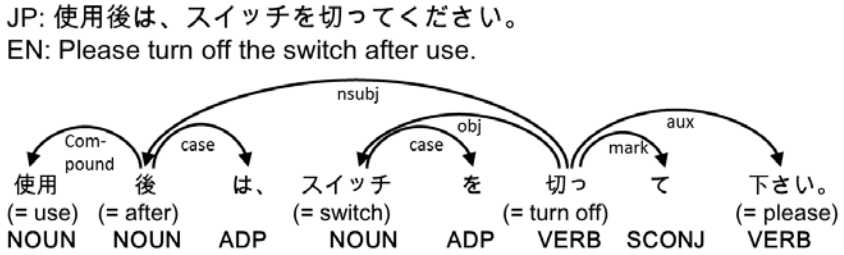


Figure 11. A dependency tree of a compliance rule sentence

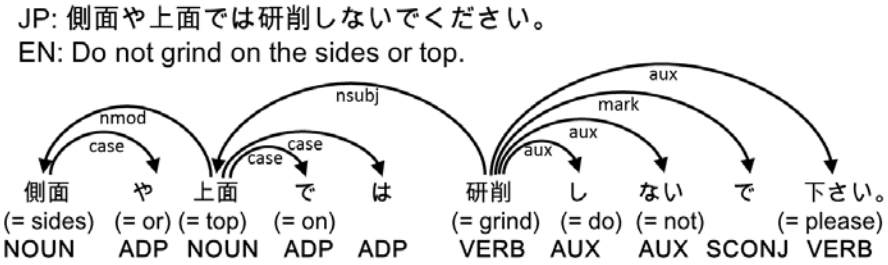


Figure 12. A dependency tree of a prohibited rule sentece

4.3 Results of Synonym Correction by Database

An algorithm that replaces each word with a synonym was created using a synonym database included in the publicly available Japanese Wordnet. At present, the pre-registered image word dataset has not been defined yet, so a image word was prepared virtually and verified. The algorithm convert the rule sentence word (B) to the registered image word (A) if synonym database of (B) includes (A) as shown in Figure 13. Here the synonym for "検査(= inspection) " has been corrected to "点検(= inspection)." This algorithm does not convert the rule sentence word (B) , if its synonym database is not including the registered image words (A). Figure 14 shows an example that the rule sentence word "切る (= cut)" has not been corrected to "着る (= wear)" , because its synonym database does not include "着る (= wear)."

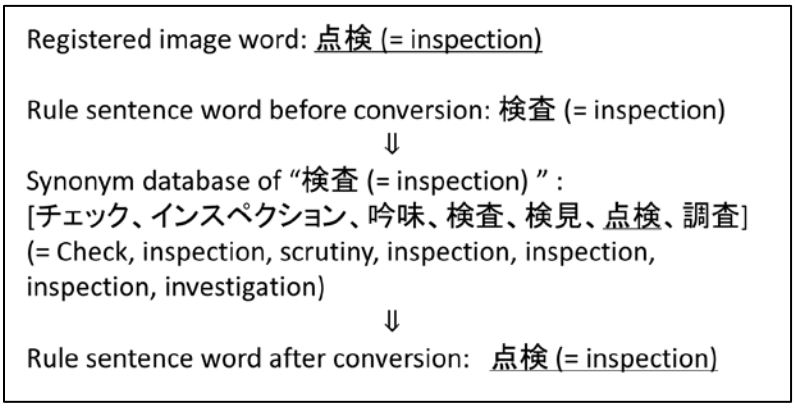


Figure 13. An example of a rule sentence word converted to an image word contained in the synonym database of rule sentence word.

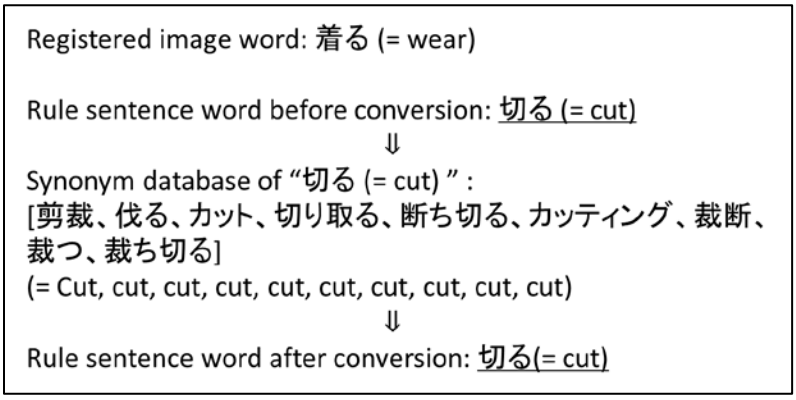


Figure 14. An example in which the image word is note included in the synonym database of the rule sentence word and the rule sentence word is not converted.

4.4 Results of Logic representation

Next, a logic representation was created that extracts only the main words and their relationships from the manually prepared occupational safety rule sentences. In Table 3(1) , the nominal word "めがね (= glasses)" and the inflectable word "利用する下さる (= please use)" are extracted from "保護めがねを利用してください。(= Please use protective goggles.)" They are linked by the case particle "を(wo)". Here, the expression "~してください(= please do)" is output as "~ する下さる(= please do)", but in the later logic representation comparison, the match is judged without considering "下さる(= please)".

In other results, logic representation s were created in the same way even if the number of nominal words, infectable words, and case particles increased.

For prohibited rule sentences in Table 4, "利用しない (= do not use)" could be correctly extracted as a negative expression in the form of "¬利用 (= ¬ use)". This makes it possible to determine whether or not the situation in the captured image violates the prohibited rule sentence.

Table 3. Example of logic representation of compliant rule sentences

(1)	保護めがねを利用してください。 Please use protective goggles.
	めがね(x0) ∧ を(e0,x0) ∧ 利用する下さる(e0) goggles (x0) ∧ [wo] (e0,x0) ∧ please use (e0)
(2)	作業場ではヘルメットを利用してください。 Please use a helmet in the workplace.
	ヘルメット(x0) ∧ を(e0,x0) ∧ 場 (x1) ∧ で(e0,x1) ∧ 利用する下さる(e0) helmet (x0) ∧ [wo] (e0,x0) ∧ place (x1) ∧ in (e0,x1) ∧ please use (e0)
(3)	ヘルメットと保護メガネを利用してください。 Please use a helmet and protective goggles.
	ヘルメット(x0) ∧ を(e0,x0) ∧ メガネ(x1) ∧ を(e0,x1) ∧ 利用する下さる(e0) helmet (x0) ∧ [wo] (e0,x0) ∧ goggles (x1) ∧ [wo] (e0,x1) ∧ please use (e0)
(4)	安全帯とヘルメットを利用し、また足元を確認してください Please use a safety belt and helmet, and check your feet.
	帯(x0) ∧ を(e0,x0) ∧ ヘルメット(x1) ∧ を(e0,x1) ∧ 利用(e0) ∧ 足元(x2) ∧ を(e1,x2) ∧ 確認する下さる(e1) belt (x0) ∧ [wo] (e0,x0) ∧ helmet (x1) ∧ [wo] (e0,x1) ∧ use (e0) ∧ feet (x2) ∧ [wo] (e1,x2) ∧ please check (e1)

Table 4. Example of logic representation of prohibition rule sentences

(1)	イヤホンを利用しないでください。 Please do not use earphones.
	イヤホン(x0) ∧ を (e0,x0) ∧ 利用(e0) earphones (x0) ∧ [wo] (e0,x0) ∧ ¬ use(e0)
(2)	イヤホンやヘッドホンを利用しないでください。 Please do not use earphones or headphones..
	イヤホン(x0) ∧ を(e0,x0) ∧ ヘッドホン(x1) ∧ を (e0,x1) ∧ 利用(e0) earphones(x0) ∧ [wo] (e0,x0) ∧ headphones(x1) ∧ [wo](e0,x1) ∧ ¬ use(e0)
(3)	作業場ではイヤホンを利用しないでください。 Please do not use earphones in the workspace.
	イヤホン(x0) ∧ を(e0,x0) ∧ 場(x1) ∧ で(e0,x1) ∧ 利用(e0) earphones(x0) ∧ [wo](e0,x0) ∧ place(x1) ∧ in(e0,x1) ∧ ¬ use(e0)

4.5 Logic representation Comparison

Here at present, the pre-registered image word dataset has not been defined yet, so the image output is virtual.

A logic representation was created from the occupational safety rule text and compared with the virtual image output. The rule sentence at Table 5 (1) is a logic representation output of "please use a helmet", while the image output is "using a helmet in workspace". Therefore, this is matching to the compliance rule sentence, and the judgment of "GOOD" is appropriate. The rule sentence at Table 5 (2) is a logic representation output of "please use a helmet and goggles", while the image output is "not using a helmet and using goggles". Therefore, this is contrary to the compliance rule sentence, and the judgment of "BAD" is appropriate. In addition, by outputting a common word for what judged to be "bad", it is possible to immediately understand what is the cause.

Similarly, the rule sentence at Table 6 (1) is a logic representation output of "please do not use an earphone", while the image output is "using a headphone ". Therefore, there is no common, and the judgment of "GOOD" is appropriate. The rule sentence at Table 6 (2) is "please do not use an earphone in workspace" while the image output is "using a headphone and an earphone". Therefore, this is contrary to the prohibited rule sentence, and the judgment of "BAD" is appropriate.

For compliance sentence, it is judged whether the image output includes the instruction item of the rule sentence, and conversely, in the prohibited sentence, it is judged whether the image output is included in the prohibited item of the rule sentence.

In this way, it was found that this method can accurately compare the meaning of logic representation s regardless of the number of words or word order.

Table 5. Results of logic representation comparison for compliance rule

<table>
<tr><td rowspan="5">(1)</td><td>From rule</td><td>ヘルメット(x0) ∧ を(e0,x0) ∧ 場 (x1) ∧ で(e0,x1) ∧ 利用する下さる(e0)
helmet (x0) ∧ [wo] (e0,x0) ∧ place (x1) ∧ in (e0,x1) ∧ please use (e0)</td></tr>
<tr><td>From image</td><td>ヘルメット(x0) ∧ を(e0,x0) ∧ 場 (x1) ∧ で(e0,x1) ∧ 利用 (e0)
helmet (x0) ∧ [wo] (e0,x0) ∧ place (x1) ∧ in (e0,x1) ∧ use (e0)</td></tr>
<tr><td>Matrix
(rule⊂image)</td><td><table><tr><td></td><td>ヘルメット(=helmet)</td></tr><tr><td>rule</td><td>利用 (= use)</td></tr><tr><td>image</td><td>利用 (= use)</td></tr><tr><td>Match</td><td>YES</td></tr></table></td></tr>
<tr><td>Judge</td><td>GOOD ∵Match = YES</td></tr>
<tr><td>Cause</td><td>-----</td></tr>
<tr><td rowspan="5">(2)</td><td>From rule</td><td>ヘルメット(x0) ∧ を(e0,x0) ∧ めがね (x1) ∧ を(e0,x1) ∧ 利用する下さる(e0)
helmet (x0) ∧ [wo] (e0,x0) ∧ goggle (x1) ∧ [wo] (e0,x1) ∧ please use (e0)</td></tr>
<tr><td>From image</td><td>ヘルメット(x0) ∧ を(e0,x0) ∧ ¬ 利用(e0) ∧ メガネ (x1) ∧を(e1,x1) ∧利用(e1)
helmet(x0) ∧ [wo](e0,x0) ∧ ¬ use(e0)∧ ∧[wo](e0,x1) ∧ use(e0)</td></tr>
<tr><td>Matrix
(rule⊂image)</td><td><table><tr><td></td><td>ヘルメット(= helmet)</td><td>めがね(= goggles)</td></tr><tr><td>rule</td><td>利用 (= use)</td><td>利用 (= use)</td></tr><tr><td>image</td><td>¬ 利用(= ¬ use)</td><td>利用 (= use)</td></tr><tr><td>Match</td><td>NO</td><td>YES</td></tr></table></td></tr>
<tr><td>Judge</td><td>BAD ∵Match = NO</td></tr>
<tr><td>Cause</td><td>ヘルメット(= helmet), 利用(= use)</td></tr>
</table>

Table 6. Results of logic representation comparison for prohibition rule

<table>
<tr><td rowspan="5">(1)</td><td>From rule</td><td>イヤホン(x0) ∧ を(e0,x0) ∧ 利用(e0)
earphone(x0) ∧ [wo] (e0,x0) ∧ use(e0)</td></tr>
<tr><td>From image</td><td>ヘッドホン(x0) ∧ を(e0,x0) ∧ 利用(e0)
headphone(x0) ∧ [wo] (e0,x0) ∧ use(e0)</td></tr>
<tr><td>Matrix
(image⊂rule)</td><td>NONE</td></tr>
<tr><td>Judge</td><td>GOOD ∵Match =NONE</td></tr>
<tr><td>Cause</td><td>----</td></tr>
<tr><td rowspan="4">(2)</td><td>From rule</td><td>イヤホン(x0) ∧ を(e0,x0) ∧ 場(x1) ∧で(e0,x1) ∧利用(e0)
earphone(x0) ∧ [wo] (e0,x0) ∧ place(x1) ∧ in (e0,x1)∧ use(e0)</td></tr>
<tr><td>From image</td><td>ヘッドホン(x0) ∧ を(e0,x0) ∧ イヤホン(x1) ∧ を(e0,x1) ∧ 利用(e0)
headphone(x0) ∧ [wo] (e0,x0) ∧earphone(x1) ∧ [wo] (e0,x1) ∧ use(e0)</td></tr>
<tr><td>matrix
(image⊂rule)</td><td><table><tr><td></td><td>イヤホン (= earphone)</td></tr><tr><td>rule</td><td>利用(= use)</td></tr><tr><td>image</td><td>利用(= use)</td></tr><tr><td>Match</td><td>YES</td></tr></table></td></tr>
<tr><td>judge</td><td>BAD ∵Match = YES</td></tr>
<tr><td></td><td>cause</td><td>イヤホン(= earphone), 利用(= use)</td></tr>
</table>

4.6 Discussion

Despite the small amount of data, Algorithm A was finally able to classify sentences into three categories: compliance, prohibition, and non-importance with a high accuracy of over 95%. Especially in the second stage classification, the accuracy increased sharply just by increasing the number of training data sentences. Therefore, it is expected that the accuracy will be further improved by adding the training data sentence, and the risk of misjudgment due to the sentence classification will be very low.

Algorithm B succeeded in Japanese dependency analysis using a method different from the conventional method. This has increased the possibility of natural language processing common to both Japanese and English, which has been considered impossible due to the large difference in grammar between Japanese and English. In addition to creating logic representation s, the accuracy of each algorithm in this study can be expected to improve by using data in English.

In Algorithm C, an algorithm was created that comprehensively replaces each word with a synonym using a synonym database included in the publicly available Japanese Wordnet. However, there are some synonyms that are not currently listed on Japanese Wordnet. In the future, by expanding the synonym database using a distributed representation cos similarity calculation system, etc., it is thought that the corresponding range of this method will be expanded and the judgment accuracy will be further improved.

In Algorithm D, a logic representation was created from the dependency analysis information. The application to a few sentences prepared manually was successful, but sufficient data could not be verified. As a constraint on creating a logic representation, it is expected that the accuracy will be further improved by utilizing not only the dependency analysis information but also the context data by the natural language machine learning model and the semantic data by the synonym database.

In Algorithm E, a method has been proposed in which logic representation s using matrices are compared with each other by a matrix comparison algorithm to determine compliance or violation of rules. The accuracy will be further improved by changing the numerical value of the matrix according to the corresponding case particle.

5. CONCLUSION

In this study, a natural language processing algorithm for determining image information was proposed for determining malicious intent and violations in nuclear security. In addition, in the verification of the occupational safety rule statement of this algorithm, compliance / prohibition / non-important sentences were classified with high accuracy by two-step binary classification.

In addition, it was confirmed that dependency analysis and synonym correction extract accurate word information from grammatical information and dictionary data. Furthermore, it was shown that the proposed logic representation creation algorithm and the rule compliance / violation judgment algorithm by logic representation comparison can be applied to Japanese sentences.

REFERENCES

[1] Incident and Trafficking Database (ITDB), https://www.iaea.org/resources/databases/itdb

[2] J. Zhang and N. M. El-Gohary, ¥Integrating semantic nlp and logic reasoning into a united system for fully-automated code checking," Automation in construction, vol. 73, pp. 45-57, 2017.

[3] Joseph Redmon,Santosh Divvala,Ross Girshick,Ali Farhadi,” You Only Look Once: Unified, Real-Time Object Detection”, CVPR, pp.779-788, 2016.

[4] Florian Schroff,Dmitry Kalenichenko,James Philbin,”FaceNet: A unified Embedding for Face Recognition and Clustering”, CVPR, pp.815-823, 2015.

[5] Cao, Zhe, et al. "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields." IEEE transactions on pattern analysis and machine intelligence 43.1 (2019): 172-186.

[6] Christian Zimmermann,Thomas Brox,” Learning to Estimate 3D Hand Pose From Single RGB Images”, ICCV, pp.4903-4911,2017.

[7] Shi Chen and Kazuyuki Demachi. "Towards on-site hazards identification of improper use of personal protective equipment using deep learning-based geometric relationships and hierarchical scene graph." Automation in Construction 125 (2021): 103619.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. “BERT: Pre-training of deep bidirectional transformers for language understanding”, arXiv: 1810.04805, 2018.

[9] Livedoor news corpus, https://www.rondhuit.com/download.html

[10] Masayuki Asahara, Hiroshi Kanayma, Takaaki Tanaka, et al., "Universal Dependencies Version 2 for Japanese”, Proc. of 9th Int. Renewable Energy Conf (Irec-Conf.), pp. 1824-1831, 2018.

[11] Masayuki Asahara, and Yuji Matsumoto, “Japanese named entity extraction with redundant morphological analysis”, Proc. of the 2003 human language technology conference of the North American chapter of the association for computational linguistics, pp. 8-15, 2003.

[12] Miller, Geroge A,”WordNet:a lexical database for English.”, Communications of the ACM, 38.11,pp.39-41, 1995.

[13] M. Lewis and M. Steedman. A* CCG Parsing with a Supertag-factored Model. In Proceedings of EMNLP2014, 2014.

[14] Taku Kudoh, Yuji Matsumoto, “Japanese Dependency Analysis Based on Support Vector Machines, EMNLP/VLC 2000.