Paper:

A Method for Detecting Harmful Entries on Informal School Websites Using Morphosemantic Patterns

Michal Ptaszynski[*1,†], **Fumito Masui**[*1], **Yoko Nakajima**[*2], **Yasutomo Kimura**[*3], **Rafal Rzepka**[*4], **and Kenji Araki**[*4]

[*1]Department of Computer Science, Kitami Institute of Technology
165 Kouen-cho, Kitami, Hokkaido 090-8507, Japan
[*2]Department of Information Engineering, National Institute of Technology, Kushiro College
2-32-1 Otanoshike-Nishi, Kushiro-shi, Hokkaido 084-0916, Japan
[*3]Department of Information and Management Science, Otaru University of Commerce
3-5-21 Midori, Otaru 047-8501, Japan
[*4]Graduate School of Information Science and Technology, Hokkaido University
Kita 14, Nishi 9, Kita-ku, Sapporo, Hokkaido 060-0814, Japan
[†]Corresponding author, E-mail: ptaszynski@cs.kitami-it.ac.jp
[Received January 18, 2017; accepted August 3, 2017]

This paper presents a novel method of analyzing morphosemantic patterns in language to the detect cyberbullying, or frequently appearing harmful messages and entries that aim to humiliate other users. The morphosemantic patterns represent a novel concept, with the assumption that analyzed elements can be perceived as a combination of morphological information, such as parts of speech, and semantic information, such as semantic roles, categories, etc. The patterns are further automatically extracted from the data containing harmful entries (cyberbullying) and non-harmful entries found on the informal websites of Japanese high schools. These website data were prepared and standardized by the Human Rights Center in Mie Prefecture, Japan. The patterns extracted in this way are further applied to a document classification task using the provided data in 10-fold cross-validation. The results indicate that morphosemantic sentence representation can be considered useful in the task of detecting the deceptive and provocative language used in cyberbullying.

Keywords: cyberbullying detection, morphosemantics, pattern extraction, semantic role labeling, natural language processing

1. Introduction

Communication taking place on the Internet allows users to hide their real names and have online conversations in environments that have a high degree of anonymity. For example, Twitter users usually appear in the form of handles, and Facebook allows users to create accounts with fictitious names. Moreover, on generic electronic bulletin boards (BBS), one can post a message simply as a "guest." This situation has given users the feeling that anything can go unpunished, eventually causing the emergence of the problem of harmful and offending messages appearing on the Internet. Terms such as "spam" (unwanted email messages), "flooding" (sending multiple meaningless messages), or "troll" (person sending messages unrelated to the main thread on an Internet forum) are special terms coined to label such behaviors or the people who engage in them. Antisocial behaviors have existed on the Internet for many years, but the overwhelming majority are ignored by other users because such behaviors are largely innocuous. However, some cases of Internet harassment escalate, become more frequent, and lead to serious consequences, such as depression, self-mutilation, or, in extreme cases, even to suicide of the victim. This problem has recently been officially defined and labeled as cyberbullying (CB). The National Crime Prevention Council states that CB happens "when the Internet, cell phones, or other devices are used to send or post text or images intended to hurt or embarrass another person" [1].

In Japan, the problem has become serious enough to be noticed by the Ministry of Education, which released a guidebook for school personnel explaining how to deal with cases of cyberbullying [2]. Moreover, in 2007, Japanese schools started a movement called Internet Patrol (later: net-patrol), consisting of members of Parent-Teacher Associations (PTA)[1]. The Internet Patrol monitors Internet activities to spot websites containing inappropriate contents. However, since this patrolling is performed manually and on a volunteer basis, the countless quantities of information appearing daily on the Internet make this an extremely difficult task.

To mitigate the problem of cyberbullying, this study aims to ease the burden of the net-patrol members by developing a solution that automatically spots entries classi-

1. An organization composed of parents and school personnel.

fiable as cyberbullying on the Web and reporting them to appropriate organizations. In this paper, we specifically focus on developing a system that automatically detects and classifies entries that constitute cyberbullying[2].

This paper is organized as follows. First, we present previous research on cyberbullying detection. Next, we describe the method and the dataset used in this research. Finally, we explain the evaluation settings, thoroughly analyze the results, and discuss possible improvements.

2. Previous Research

Some of the first robust research on CB was done by Hinduja and Patchin, who carried out numerous surveys on the subject in the USA [3, 4]. They found that the harmful information may include threats, sexual remarks, pejorative labels, or false statements aimed at humiliating others. When posted on a social network, such as Facebook or Twitter, humiliating personal information that defames and ridicules its victims may be disclosed.

Cyberbullying has also been thoroughly studied and analyzed by Dooley, Pyżalski, and Cross [5]. They performed an in-depth comparative analysis of traditional face-to-face bullying and cyberbullying, while Lazuras et al. [6] discussed implications of cyberbullying for teachers in classroom environments.

There have also been a small number of studies on extracting harmful information from the Internet. These studies have focused either on information generally perceived to be harmful or on cyberbullying specifically. For example, Ishisaka and Yamamoto [7] developed a dictionary of abusive expressions based on a large Japanese electronic bulletin board (BBS) *2channel*. In their research, they labeled words and paragraphs in which the speaker explicitly insults other people with words and phrases such as *baka* (stupid), or *masugomi no kuzu* (trash of mass-mudia). Based on which words appeared most often with abusive vocabulary, they extracted abusive expressions from the surrounding context.

Ptaszynski et al. [8] performed affect analysis on a small dataset of cyberbullying entries and found that distinctive features of cyberbullying were vulgar expressions. They applied a lexicon of such words to train a Support Vector Machine (SVM) classifier. With a number of optimizations, the system was able to detect cyberbullying with 88.2% of a balanced F-score (a standard balanced measure of evaluation representing a harmonic mean of Precision and Recall). However, increasing the data caused a decrease in results, which made Ptaszynski et al. conclude that SVMs are not ideal in terms of dealing with frequent language ambiguities typical of cyberbullying. Such ambiguities are of various types. Some of them, such as jargonization (changing a normal word to its colloquial or other form to make it difficult to understand by bystanders) are feasible on a preprocessing level, as long as one has the appropriate lexicon of such jargonized words. However, a problem that is more difficult to deal with is when two or more words are not harmful when they appear separately in one context but gain harmful coloring when used together in different context. This problem was also mentioned by Nitta et al. [9]. To be specific, Nitta et al. mention the example "monkey face." Neither the word "monkey" nor "face" on its own has a harmful meaning. However, when used together as a phrase, they are often used by adolescents in Japan to slander each other. In this context, the SVM classifier is most commonly trained on a bag of words (BoW) language model, which means that the features used for training consist of separate words. Thus, both phraseological connections as well as word order in a sentence are disregarded. This was the reason for the eventual low performance of SVM-based classifiers and the reason for Nitta et al. to attempt different approaches.

A successful approach was proposed by Matsuba et al. [10]. They proposed a method for automatically detecting harmful entries, based on an extended SO-PMI-IR score. This method was originally proposed by Turney [11] to calculate the relevance of a document with positive and negative contents. Matsuba et al. modified it to be able to deal with harmful contents. Using a small number of seed words, Matsuba et al. were able to detect, with an accuracy of 83%, a large number of candidate harmful documents from test data.

Later, Nitta et al. [9] proposed an improvement to Matsuba et al.'s method. They grouped seed words into three categories, namely, abusive, violent, and obscene, to calculate a SO-PMI-IR score, and maximized the relevance of categories. Their method achieved 90% of Precision for 10% Recall. We used both of the above methods as baselines for comparison due to similarities in the applied datasets and experiment settings.

Unfortunately, the method by Nitta et al. [9], based on the *Yahoo!* search engine API, had the problem of a sudden drop in Precision (about 30 percentage points) in the two years after it was originally proposed. This was caused by change in information available on the Internet. In Section 4.5 we discuss the possible reasons for this drop in performance. Recently, Hatakeyama et al. [12] attempted, with considerable success, to improve the method by automatically acquiring and filtering harmful seed words.

Most of the previous research assumed that using vulgar words is the key to detecting cyberbullying entries. However, all of the previous research notices that vulgar words are only one kind of distinctive vocabulary; vulgar words do not appear in all cases. We assume that the harmfulness of an entry does not depend only on such words; harmfulness is expressed through both semantic and grammatical patterns within the structure of a sentence. Therefore, in this research, we do not focus on detecting vulgar words, nor do we restrict the scope of analyzed patterns to words or phrases. Instead, we extend the search to sophisticated patterns with disjoint elements. Moreover, the success of detecting such entries

2. The methods developed in this research are language independent. However, we use data collected in the Japanese language, which was the only data available to us at the time of writing.

relies on how accurately the sentence structure is represented. Thus, in our research, we use a novel representation method incorporating both morphological and semantic information.

3. Morphosemantic Pattern Extraction Method

In this section, we describe our method of extracting morphosemantic patterns from sentences. The method consists of two stages. First, the sentences are represented using a combination of semantic role labeling and morphological information. Second, frequent combinations of such patterns are extracted from training data using an automatic pattern extraction architecture.

3.1. Morphosemantic Patterns

In the first stage of the method, all sentences included in the dataset (see Section 4.1 for details) are represented in morphosemantic structure (MS). From sentences represented this way, morphosemantic patterns (MoPs) are extracted during the second stage.

The idea of morphosemantic structures has been described widely in linguistics and structural linguistics. For example, Levin and Rappaport Hovav [13] distinguish them as one of the two basic types of morphological operations on words, mostly verbs, that modify the Lexical Conceptual Structure (LCS), or the semantic representation of a word. Kroeger [14] applied morphosemantic structure to analyze an Indonesian suffix *–kan*. Later, Fellbaum et al. [15] applied morphosemantic patterns to improve links between the synsets in WordNet. More recently, Raffaelli [16] used morphosemantic patterns to analyze a lexicon in Croatian, a language rich both morphologically and semantically. Also, Nakajima et al. [17] applied morphosemantic structure to extract patterns of sentences referring to the future and applied the patterns to the task of reasoning about the unfolding of future of events in the Japanese language.

In our research, we also used datasets in the Japanese language, and we applied morphosemantic structures for the same reason. Using only one representation (lexical, morphological, or semantic) narrows the spectrum of information encoded in the language.

We generated the morphosemantic model using semantic role labeling with additional morphological information. Below, we describe in detail the process behind the morphosemantic representation of sentences.

First, sentences from the datasets are analyzed using semantic role labeling (SRL). SRL provides labels for words and phrases according to their role in sentence context. For example, in the sentence "John killed Mary," the labels for words are as follows: John=`Actor`, kill[past]=`Action`, Mary=`Patient`. Thus the semantic representation of the sentence is

"`[Actor][Action][Patient]`."

For semantic role labeling in Japanese, we used ASA, a system developed by Takeuchi et al. [18]. ASA provides semantic roles for words and generalizes their semantic representation using an original thesaurus Takeuchi et al. developed for their system. Examples of labels ASA provides for certain words are presented in **Table 1**. Two examples of SRL provided by ASA are presented in **Table 2**.

However, not all words are semantically labeled by ASA. The omitted words include those not present in the thesaurus as well as grammatical particles, which do not have a direct influence on the semantic structure of the sentence but, in practice, contribute greatly to the overall meaning. For such function words, we used a morphological analyzer MeCab [19] in combination with ASA to provide morphological information, such as "proper noun" or "verb," etc. However, in its basic form, MeCab provides morphological information for all words separately. Therefore, compound words are unnecessarily divided. For example, "Japan health policy" is one morphosemantic concept, but in grammatical representation, it takes form of "noun noun noun." Therefore, as a post-processing procedure, we added a set of linguistic rules to specify compound words when only morphological infor-

Table 1. Examples of future-referring words and phrases with their semantic and morphological representation.

Surface	Semantic (Semantic role, Category, etc.) and grammatical representation
mezasu ("aim to")	No change (activity)-action aiming to solve [a problem]-pursuit; Verb;
hōshin ("plan to")	Other; Noun;
mitooshi ("be certain to")	Action; Noun;
kentō ("consider to")	No change (activity)-action aiming to solve [a problem]-act of thinking; Noun;
-*suru* ("to do")	Change-creation or destruction-creation (physical); Verb;
-*iru* ("is/to be")	Verb;

Table 2. One example of sentence analysis by ASA.

Example I: Romanized Japanese (RJ): *Ashita kare wa kanojo ni tegami o okuru darō.* / **Glosses:** Tomorrow he TOP her DIR letter OBJ send will (TOP: topic particle, DIR: directional particle, OBJ: object particle.) / **English translation (E):** He will [most probably] send her a letter tomorrow.

No.	Surface	Label
1	*ashita*	[Time-Point]
2	*kare ha*	[Agent]
3	*kanojo ni*	[Patient]
4	*tegami o*	[Object]
5	*okuru darou*	[State_change]-[Place_change]-[Change_of_place(physical)(between_persons)]-[Movement_of_property_to_others]-[Provide]

Example: What a nice day !				
5-el. pattern:	4-el. patterns:	3-el. patterns:	2-el. patterns:	1-el. patterns:
What a nice day !	What a nice * !	a nice * !	What a	What
	What a nice day *	What a nice	What * !	a
	What a * day !	What a * !	nice * !	nice
	⋮	⋮	⋮	⋮
no. of patterns: (1)	(5)	(10)	’(10)	(5)

Fig. 1. Examples of various-length (=number of elements) patterns extracted by SPEC from one sentence.

mation was provided.

Moreover, as shown in **Table 2**, some labels provided by ASA are too specific (see label No. 5). Therefore, in order to normalize and simplify the patterns, we prioritized label groups as follows.

1. Semantic role (Agent, Patient, Object, etc.)
2. Semantic meaning (State_change, etc.)
3. Category (dog → live animal → animated object)
4. If no label by ASA, perform compound word clustering for parts of speech (e.g., "Japan Health Policy" → `[Noun][Noun][Noun]` → `[Proper_Noun]`)

Furthermore, post-processing in the case of no semantic information is organized as follows.

- If a compound word can be specified, output the part-of-speech cluster (point 4 above).
- If it is not a compound word, output the part of speech for each word.

Below is an example of a sentence generalized using the morphosemantic structure labeling method applied in this research.

- **Sentence** (in Romanized Japanese): *Nihon unagi ga zetsumetsu kigushu ni shitei sare, kanzen yōshoku ni yoru unagi no ryōsan ni kitai ga takamatte iru.*
- **English**: As the Japanese eel has been specified as an endangered species, the expectation that they will be mass produced in full aquaculture is growing.
- **MS**: `[Object][Agent][State_change][Action][Noun][State_change][Object][State_change]`

3.2. Automatic Extraction of Frequent Patterns

Once all sentences are represented in morphosemantic structure, as described in Section 3.1, we use SPEC, a system for the extraction of sophisticated sentence patterns that was developed by Ptaszynski et al. [20]. SPEC, or Sentence Pattern Extraction arChitecture is a system that automatically extracts frequent sentence patterns distinguishable in a corpus, or collection of sentences. First, the system generates ordered non-repeated combinations from all sentence elements. In every n-element sentence, there are k-number of combination groups, such that $1 \leq k \leq n$, where k represents all k-element combinations being a subset of n. The number of combinations generated for one k-element group of combinations is equal to a binomial coefficient, represented in Eq. (1). In this procedure, the system creates all combinations for all values of k from the range of $\{1,\ldots,n\}$. Therefore, the number of all combinations is equal to the sum of combinations from all k-element combination groups, as in Eq. (2).

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \ldots\ldots\ldots\ldots \quad (1)$$

$$\sum_{k=1}^{n}\binom{n}{k} = \frac{n!}{1!(n-1)!} + \frac{n!}{2!(n-2)!} + \cdots + \frac{n!}{n!(n-n)!} = 2^n - 1 \quad \ldots\ldots\ldots\ldots \quad (2)$$

Next, the system specifies whether the elements appear next to each other or are separated by other words by placing the wildcard character asterisk, or "*," between all noncontiguous elements. SPEC uses all patterns generated this way to extract frequent patterns appearing in a given corpus, and it calculates their weight. A few examples of patterns of various lengths extracted by SPEC from one sentence are presented in **Fig. 1**. In the presented examples, patterns are extracted from sentence tokens, such as words, punctuation marks, etc. In our research, we extract patterns from morphosemantic structures.

The weight of patterns can be calculated in several ways. Two features are important in weight calculation. A pattern is more representative of a corpus the longer the pattern is (length k) and the more often it appears in that corpus (occurrence O). Thus, the weight can be calculated by doing the following:

- awarding length (`LA`)[3],
- awarding length and occurrence (`LOA`),
- awarding none (normalized weight, `NW`).

The normalized weight w_j is calculated according to Eq. (3). Normalization is performed so that weights fit in the range from $+1$ to -1, and it is achieved by subtracting 0.5 from the initial score and multiplying the intermediate product by 2. In this way, the weightings are normalized

3. Explanations of all symbols and abbreviations are summarized in an Appendix at the end of the paper.

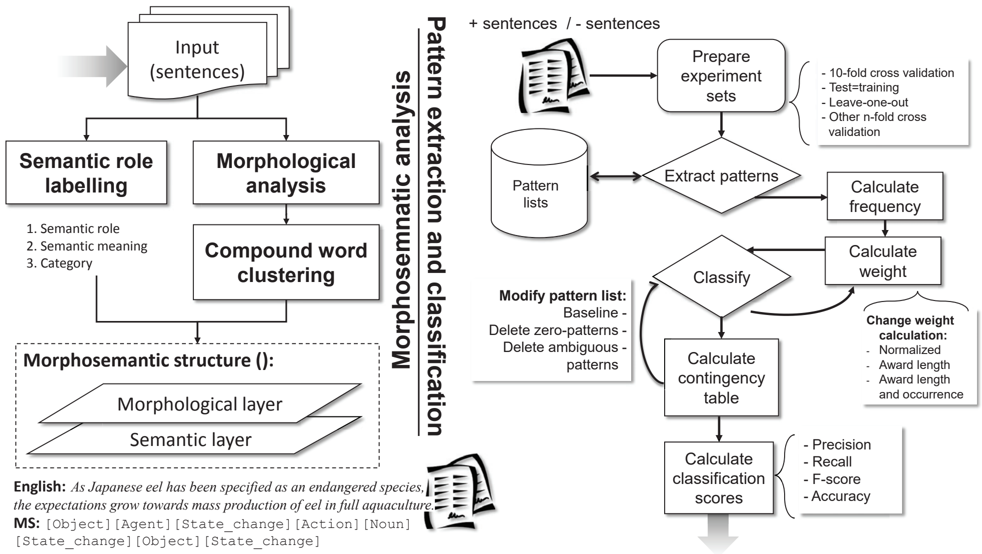

Fig. 2. A graphical summarization of the whole method for automatic extraction of morphosemantic patterns.

in a range from -1.00 (non-harmful) to $+1.00$ (harmful).

$$w_j = \left(\frac{O_{pos}}{O_{pos}+O_{neg}} - 0.5\right) * 2 \quad \text{. . . . .} \quad (3)$$

The list of frequent patterns generated in this way can be further modified. When two collections of sentences of opposite features (such as "harmful vs. non-harmful") are compared, the list will contain patterns that appear uniquely on only one of the sides, e.g., uniquely positive patterns and uniquely negative patterns, or on both, as ambiguous patterns. Thus, a pattern list can be modified by doing the following:

- using all patterns (`ALL`),
- erasing all ambiguous patterns (`AMB`),
- erasing only those ambiguous patterns that appear in the same number on both sides due to their normalized weight being equal to 0. These are later called "zero patterns" (0P).

Moreover, a list of patterns will contain both the sophisticated patterns (with disjoint elements) and more common n-grams. Therefore, the system can be trained on a model using the following:

- all patterns (`PAT`) or
- only n-grams (`NGR`).

All combinations of the above modifications are further tested in the evaluation experiment. A graphical summarization of the entire method is presented in **Fig. 2**.

4. Evaluation Experiment

4.1. Dataset

First, we needed to prepare a dataset. We used the dataset originated by Matsuba et al. [21] and further developed by Matsuba et al. [10]. The dataset was also used by Ptaszynski et al. [8] and recently by Nitta et al. [9]. It contains 1,490 harmful and 1,508 non-harmful entries. The original data, provided by the Human Rights Research Institute Against All Forms of Discrimination and Racism in Mie Prefecture, Japan (later abbreviated to Human Rights Center) [22], contains data from a number of informal school websites in Mie Prefecture, Japan. The harmful and non-harmful sentences were manually labeled by Internet Patrol members according to instructions included in the MEXT manual for dealing with cyberbullying [2]. Some of those instructions are summarized below.

The MEXT definition assumes that cyberbullying happens when a person is personally offended on the Web. This includes disclosing the person's name, personal information, and other things considered as private. Therefore, as the first feature distinguishable as cyberbullying, MEXT defines private names. This includes the following information:

- Private names and surnames, e.g., "John Smith"
  - When a person's name can be clearly distinguished
- Initials and nicknames, e.g., "Mr. P.," "Mi*al Ptasz*ski"
  - When a person's identity can be clearly distinguished
  - When a person's identity cannot be clearly distinguished
- Names of institutions and affiliations, e.g., "That foreign assistant professor from the Kitami Institute of Technology"
  - When a person's identity can be clearly distinguished
  - When a person's identity cannot be clearly distinguished

Table 3. Four examples of cyberbullying entries gathered during Internet Patrol. The upper three represent strong sarcasm despite the use of positive expressions in the sentence. English translations appear below the Japanese content.

>>104 Senzuri koi te shinu nante? sonna hageshii senzuri sugee naa. "Senzuri masutaa" toshite isshou agamete yaru yo. >>104 Dying by 'flicking the bean'? Can't imagine how one could do it so fiercely. I'm gonna worship her as a 'master-bator', that's for sure.
2-nen no tsutsuji no onna meccha busu suki na hito barashimashoka? 1-nen no anoko desuyo ne? kimogatterunde yamete agete kudasai Wanna know who likes that awfuly ugly 2nd-grade Azalea girl? Its that 1st-grader isn't it? He's disgusting, so let's leave him mercifully in peace.
Aitsu wa busakute sega takai dake no onna, busakute se takai dake ya noni yatara otoko-zuki meccha tarashide panko anna onna owatteru She's just tall and apart of that she's so freakin' ugly, and despite of that she's such a cock-loving slut, she's finished already.
Shinde kureeee, daibu kiraware-mono de yuumei, subete ga itaitashii... Please, dieeee, you're so famous for being disliked by everyone, everything in you is so pathetic

As the second feature distinguishable as cyberbullying, MEXT defines any other type of personal information. This includes the following information:

- Addresses, phone numbers, etc., e.g., "165 Koen-cho, Kitami, 090-8507, Japan" or "+81-123-45-6789"
  - When the information refers to a private person
  - When the information is public or refers to a public entity
- Questions about private persons, e.g. "Who is that tall guy wandering around the Computer Science Dept. corridors?"
  - Always considered as undesirable and harmful, including situations in which the object is described in a positive way
- Entries revealing other personal information, e.g., "I hate that guy responsible for the new project against cyberbullying."
  - When a person's identity can be clearly distinguished
  - When a person's identity cannot be clearly distinguished

Also, according to MEXT, vulgar language is distinguishable as cyberbullying due to its ability to convey offenses against particular persons. This has also been confirmed in other literature [4, 8]. Examples of such words are, in English: *sh*t*, *f*ck*, or *b*tch*, in Japanese: *uzai* (freaking annoying), or *kimoi* (freaking ugly).

In the prepared dataset all entries containing any of the above information was classified as harmful. Some examples from the dataset are presented in **Table 3**.

Due to the private nature of the information and the high level of potential harmfulness of some of the data used in this study, the dataset cannot be widely released at present. The dataset was provided to the authors by the Human Rights Center in Mie Prefecture, Japan, with the strict limitation that it be used only by the researchers involved in the project. Furthermore, although the researchers themselves were not obliged to sign any ethical permission to use the dataset, the subjects who collected the data, namely, the Internet Patrol members, usually need to voluntarily accept the potential influence they could gain by reading the harmful Internet content. Moreover, the identities of neither the Internet Patrol members who collected the dataset nor the authors of the original Web entries were revealed or provided to the researchers. Under the agreement with the Human Rights Center, only the contents of the Web entries are known to the researchers. Finally, the researchers were obliged to agree that they would not use any examples revealing any private information and that they would mask or change all the private information in the examples.

4.2. Dataset Preprocessing

As mentioned in Section 3.1, we propose representing the sentences in morphosemantic structure as a novel approach to detecting cyberbullying. However, we needed to verify empirically whether it was useful to use morphosemantics for this kind of data or whether it was sufficient to only choose only one kind of representation. Therefore, we applied the following sentence preprocessing in the experiment.

- **Parts of speech (`POS`):** Words are replaced with their representative morphemes and parts of speech.
- **Semantic roles (`SR`):** Words and phrases are replaced with their semantic representations within the context of sentences (semantic roles).
- **Morphosemantic patterns (`MS`):** The sentences are preprocessed using combined morphological and semantic information.

Four examples of preprocessing are presented in **Table 4**.

4.3. Experiment Setup

The preprocessed original dataset provided three separate training and test sets for the experiment (`POS`, `SR`, `MoPs`). The experiment was performed three times, with 10-fold cross-validation (ten times for each kind of preprocessing) to choose the best option. Using these preprocessed datasets, we performed the classification as follows. Each test sentence was given a score calculated as a sum of the weights of patterns extracted from the training data and matched to the input sentence (Eq. (4)).

$$score = \sum w_j, \quad (1 \geq w_j \geq -1) \quad \ldots\ldots \quad (4)$$

Table 4. Four examples of preprocessing of a sentence in Japanese.

Sentence (in Romanized Japanese): *Nihon-unagi ga zetsumetsu kigu shu ni shitei sa re, kanzen yōshoku niyoru unagi no ryōsan ni kitai ga takamat te iru.* **English translation:** As Japanese eel has been specified as an endangered species, the expectations grow towards mass production of eel in full aquaculture.
Preprocessing examples
1. POS: [Noun][Particle][Noun][Noun][Noun][Noun][Particle][Noun][Verb][Verb][Punct.][Noun][Noun] [Particle][Noun][Particle][Noun][Particle][Noun][Particle][Verb][Particle][Verb][Punct.] **2. POS with compound word clustering:** [Noun][Particle][Noun][Particle][Noun][Verb][Punct.][Noun][Particle][Noun][Particle][Noun][Particle][Noun][Particle][Verb][Particle][Verb][Punct.] **3. Semantic roles:** [Object][Agent][State_change][Action][State_change][Object] [State_change] **4. Morphosemantic structure:** [Object][Agent][State_change][Action][Noun][State_change][Object][State_change]

The results were calculated using standard Precision (P), Recall (R), and balanced F-score (F1), and additionally with standard Accuracy, for the whole threshold span. However, if the initial collection of sentences was biased toward one of the sides, e.g., sentences of one kind were larger in number or longer, there would be more patterns of a certain type. Thus, using a rule of thumb in evaluation, e.g., a fixed threshold above which a new sentence is classified as either harmful or non-harmful, such as "above or below zero," would not provide a sufficiently objective view in the results. Therefore, we additionally performed threshold optimization to find the threshold for which the classifier achieved the highest scores.

For each version of the dataset preprocessing, a 10-fold cross validation was performed. In one experiment, 14 different versions of the classifier were compared. Since the experiment was performed for three different versions of preprocessing, we obtained a total of 420 experiment runs.

There were several evaluation criteria. First, we looked at which version of the algorithm achieved the top scores within the threshold span. We also looked at Break-Even Points (BEP) of Precision and Recall. Finally, we checked the statistical significance of the results. We used a paired t-test because the classification results could represent only one of two classes: harmful or non-harmful. To choose the best version of the algorithm, we separately compared the results achieved by each group of modifications, e.g., "different pattern weight calculations," "pattern list modifications," and "patterns vs. n-grams." We also compared the performance to those of previous methods, which we considered baselines [9, 10, 12].

4.4. Results and Discussion

To summarize the results, we looked at which version of the algorithm achieved the top scores within the threshold span.

First, we looked at the standard balanced F-score to see if a clear winner could be selected by using the simplest measure. The best F-score for all three kinds of preprocessing, namely, Parts-of-speech [POS], Semantic Roles [SR], and Morphosemantic Patterns [MoPs], reached the same maximum of 0.68. Therefore, there was no clear winner, but within this evaluation context, Semantic Roles, achieved the highest balance of F-score and Accuracy for the version of classifier trained on ngrams with length awarded and zero-patterns deleted (NGR-LA-0P). Morphosemantic Patterns were the second highest with classifiers trained on all patterns and unmodified pattern list (PAT-ALL). The results are presented in **Table 5**.

To provide additional support for the results, we also looked into the Break-Even Point of Precision and Recall (BEP). Here similarly, Semantic roles achieved the highest score of 0.67 for classifiers trained on a pattern list with ambiguous patterns discarded (PAT-AMB). MoPs were second-best (0.64) when trained on ngrams with ambiguous patterns discarded (NGR-AMB). This could suggest that, regardless of which preprocessing achieved the highest scores, training the classifier on a pattern list with ambiguous patterns deleted could also result in a high BEP in the future.

In the process of detecting cyberbullying messages, sometimes net-patrol members may want to focus not on finding many suspicious messages but on the most harmful ones, or those which are certainly harmful (to apply for deletion of those in the first place). Therefore, we also looked at the highest Precision within the threshold. The results are presented in **Table 6**.

The highest Precision was achieved by SR (PAT-LA-AMB) and POS (NGR-LA) (both 0.93). As both of these classifier versions incorporated the length of the pattern in the pattern weight calculation (LA), this suggests that, to achieve the highest P, it could be useful to also apply this pattern list modification in the future.

However, for such high P, both SR and POS achieved very low R (0.11 for SR and 0.06 for POS). Interestingly, when it came to the highest P optimized for F, the highest score was achieved by MoPs (P=0.85 for F=0.18, for NGR-ALL *ex aequo* with NGR-0P).

Table 5. Comparison of best F-scores within threshold span and BEP for each version of the classifier. Best classifier version within each preprocessing kind is highlighted in bold type font; the best overall is underlined.

	Highest F-score within threshold												BEP (P=R=F)		
	POS				Semantic Roles				MoPs				POS	Sem Rol	MoPs
	Pr	Re	F1	Acc	Pr	Re	F1	Acc	Pr	Re	F1	Acc			
PAT-ALL	0.53	0.95	0.68	0.55	0.59	0.80	0.68	0.64	**0.61**	**0.76**	**0.68**	**0.64**	0.61	0.67	0.64
PAT-0P	**0.53**	**0.95**	**0.68**	**0.55**	0.59	0.80	0.68	0.64	0.57	0.83	0.68	0.61	0.61	0.66	0.64
PAT-AMB	**0.53**	**0.95**	**0.68**	**0.55**	0.58	0.81	0.68	0.62	0.58	0.82	0.68	0.62	0.61	<u>**0.67**</u>	0.64
PAT-LA	**0.53**	**0.95**	**0.68**	**0.55**	0.60	0.78	0.68	0.64	0.61	0.74	0.67	0.64	0.61	0.67	0.64
PAT-LA-0P	0.52	0.95	0.68	0.54	0.59	0.79	0.68	0.63	0.59	0.80	0.68	0.62	0.59	0.66	0.62
PAT-LA-AMB	**0.53**	**0.95**	**0.68**	**0.55**	0.63	0.73	0.68	0.66	0.58	0.82	0.68	0.62	0.60	0.67	0.63
NGR-ALL	0.52	0.96	0.67	0.53	0.58	0.82	0.68	0.63	0.59	0.82	0.68	0.63	**0.61**	0.67	0.64
NGR-0P	0.52	0.95	0.67	0.54	0.58	0.82	0.68	0.63	0.59	0.81	0.68	0.63	0.57	0.60	0.54
NGR-AMB	0.50	1.00	0.67	0.50	0.54	0.89	0.67	0.57	0.49	1.00	0.66	0.49	0.61	0.67	**0.64**
NGR-LA	0.53	0.94	0.68	0.55	0.63	0.75	0.68	0.66	0.58	0.82	0.68	0.61	0.60	0.67	0.63
NGR-LA-0P	0.52	0.95	0.67	0.54	<u>**0.63**</u>	<u>**0.74**</u>	<u>**0.68**</u>	<u>**0.67**</u>	0.58	0.81	0.68	0.62	0.60	0.62	0.58
NGR-LA-AMB	0.57	0.76	0.65	0.59	0.56	0.82	0.67	0.60	0.56	0.74	0.64	0.58	0.61	0.67	0.63

Table 6. Comparison of best Precision and Accuracy within the threshold span for each version of the classifier. The best classifier version within each preprocessing kind is highlighted in bold type font; the best overall is underlined.

	Highest Precision within threshold												Highest Accuracy within threshold											
	POS				Semantic Roles				MoPs				POS				Semantic Roles				MoPs			
	Pr	Re	F1	Acc	Pr	Re	F1	Acc	Pr	Re	F1	Acc	Pr	Re	F1	Acc	Pr	Re	F1	Acc	Pr	Re	F1	Acc
PAT-ALL	0.78	0.05	0.09	0.52	0.87	0.20	0.33	0.60	0.81	0.15	0.25	0.56	0.58	0.78	0.66	0.60	0.67	0.67	0.67	0.68	0.61	0.76	0.68	0.64
PAT-0P	0.78	0.05	0.09	0.52	0.87	0.28	0.42	0.63	0.81	0.15	0.25	0.56	0.58	0.78	0.66	0.60	0.67	0.67	0.67	0.68	0.61	0.75	0.67	0.64
PAT-AMB	0.80	0.02	0.04	0.51	0.89	0.04	0.07	0.53	0.79	0.17	0.28	0.56	**0.58**	**0.78**	**0.66**	**0.61**	0.64	0.70	0.67	0.66	0.61	0.74	0.67	0.64
PAT-LA	0.78	0.05	0.09	0.52	0.89	0.03	0.06	0.53	0.79	0.17	0.27	0.56	0.58	0.78	0.66	0.60	0.67	0.66	0.67	0.68	0.61	0.74	0.67	0.64
PAT-LA-0P	0.76	0.11	0.20	0.54	0.92	0.03	0.05	0.52	0.72	0.14	0.24	0.56	0.59	0.63	0.61	0.60	0.71	0.56	0.63	0.68	0.61	0.73	0.67	0.64
PAT-LA-AMB	0.76	0.12	0.21	0.54	<u>**0.93**</u>	<u>**0.06**</u>	<u>**0.11**</u>	<u>**0.54**</u>	0.70	0.17	0.27	0.56	0.60	0.56	0.58	0.60	0.63	0.73	0.68	0.66	0.61	0.72	0.66	0.64
NGR-ALL	0.92	0.02	0.04	0.51	0.87	0.20	0.33	0.60	**0.85**	**0.10**	**0.18**	**0.55**	0.63	0.53	0.58	0.61	<u>**0.80**</u>	<u>**0.49**</u>	<u>**0.61**</u>	<u>**0.69**</u>	0.64	0.64	0.64	0.65
NGR-0P	0.92	0.02	0.04	0.51	0.87	0.20	0.33	0.60	**0.85**	**0.10**	**0.18**	**0.55**	0.63	0.53	0.58	0.61	<u>**0.80**</u>	<u>**0.49**</u>	<u>**0.61**</u>	<u>**0.69**</u>	**0.62**	**0.72**	**0.67**	**0.65**
NGR-AMB	0.65	0.21	0.32	0.55	0.69	0.14	0.23	0.55	0.54	0.71	0.61	0.56	0.54	0.83	0.65	0.56	0.54	0.89	0.67	0.57	0.54	0.71	0.61	0.56
NGR-LA	**0.93**	**0.03**	**0.06**	**0.51**	0.88	0.02	0.05	0.52	0.83	0.14	0.25	0.56	0.62	0.54	0.58	0.61	0.78	0.50	0.61	0.69	0.63	0.69	0.66	0.64
NGR-LA-0P	0.91	0.03	0.06	0.51	0.89	0.03	0.05	0.52	0.83	0.15	0.25	0.56	0.59	0.72	0.65	0.61	0.78	0.50	0.61	0.69	0.63	0.68	0.65	0.64
NGR-LA-AMB	0.66	0.31	0.42	0.57	0.75	0.24	0.36	0.59	0.60	0.49	0.54	0.59	0.58	0.71	0.64	0.60	0.56	0.53	0.67	0.60	0.56	0.73	0.63	0.59

Table 7. Results of the paired two-tailed Student's T-test for F-score and Accuracy for the classifier versions that achieved the highest BEP.

	F-score		Accuracy	
	SR	MoPs	SR	MoPs
POS	0.0248* ($p<0.05$)	0.0079** ($p<0.01$)	0.0004*** ($p<0.001$)	0.0001*** ($p<0.001$)
SR		0.3077 ($p>0.05$)		0.2079 ($p>0.05$)

We also looked at standard Accuracy as a supportive mean for evaluation. As in previous results, SR achieved the highest maximum score (0.69), with MoPs being second (0.65).

To confirm that the above results were not a matter of chance, we also calculated statistical significance of the results using the paired two-tailed Student's T-test[4] for F-score and Accuracy results for those classifier versions that achieved the highest BEP. We selected this significance test due to the fact that the classification could result in only one of two labels, either "harmful" or "non-harmful." A comparison of statistical significance in the results is presented in **Table 7**. The differences between POS and SR or MoPs were always statistically significant. This means that when SR or MoPs achieve higher scores than does POS, the improvement can be considered reliable and not bound by chance.

On the other hand, the differences between SR and MoPs were always not statistically significant. This suggests that although SR achieved higher scores than MoPs in some cases, this advantage could be a matter of chance. This means that both SR and MoPs remain viable, and further experiments on larger datasets are required to finally specify which type of dataset preprocessing is more effective.

In the overall summarization, preprocessing by only parts of speech achieved the lowest results. On the other hand, preprocessing by only semantic role labeling (SRL) achieved the highest results. Morphosemantic preprocessing placed in the middle, but closer to the winner due to the lack of statistical significance in differences between results.

This time we used part-of-speech tagging supported with compound word clustering, which lowered the number of morphology-related patterns but increased the occurrence of those that were extracted. As was shown in other research using the SPEC architecture [20, 24], preprocessing methods with lower element generalization, e.g., with example words instead of parts of speech, often also positively influenced the Recall (more patterns were found in general) by increasing the number of patterns. This raises the final overall F-score as a result. Therefore, in the future, we plan to first lower the generalization of POS preprocessing by not using compound word clustering, as this lowering will increase the overall number of patterns. Doing so will help us find out if compound word clustering is useful in cyberbullying detection or whether it hinders the classification. We also plan to use other preprocessing methods, such as tokenization, lemmatization, or combinations of various preprocessing methods used in previous studies, applying the SPEC architecture [24].

However, the results of the current study suggest that semantic information is in general helpful in differentiating between harmful and non-harmful entries appearing on the Internet. This means that harmfulness can be considered a linguistic feature well representable on the semantic layer of language. The fact that semantics plays a great role in conveying harmful meaning is obvious from the point of view of a normal language user. However, this idea had not yet been thoroughly verified scientifically. Some of the previous attempts to find a relationship between semantics and swearing were, for example, done by Kidman in 1993 [25]. Also, preliminary research which to some extent applies the idea of semantic analysis to harmful contents detection has recently been started by Zhao and Mao [26]. However, we were not able to find any previous research that thoroughly studied and explained the relationship between semantics and harmful language. Although our paper does not exhaust the topic, we believe the research presented in this paper, besides proposing an effective method of detecting cyberbullying, also provides perhaps the first quantifiable proof of the relationship between semantics and harmful meaning.

In general, this is good news since semantic information can be helpful in further in-depth analyses of harmful entries to determine the roles of participants in the bullying process, such as the perpetrator (who could correspond to the [Actor] label in SRL) or the victim (who could possibly correspond to the [Patient] label in SRL). We plan to investigate these potential paths in the near future.

4.5. Comparison with Previous Methods

After analyzing various multiple settings for the proposed method, we compared it to previous methods. In the comparison, we used the method by Matsuba et al. [10], Nitta et al. [9], and its most recent improvement by Hatakeyama et al. [12]. However, since the latter extracts cyberbullying relevance values from the Web, apart from comparing to the results reported in the papers, we also repeated their experiment to find out how the performance of the Web-based method had changed during the three years since it was originally proposed. Finally, to make the comparison fairer, we compared our best and worst results. As the evaluation metrics, we used the area under the curve (AUC) on the graph showing Precision and Recall, the same metrics used in the abovementioned research. The results are presented in **Fig. 3**.

The highest overall results for AUC were obtained by the best settings of the proposed method (trained on a pattern list with semantic roles, length awarded in

4. A standard statistical test used in the verification of whether differences between two sets of results are statistically significant or are a matter of chance, proposed originally by W. S. Gossett (1876–1937), also known as "Student" [23].

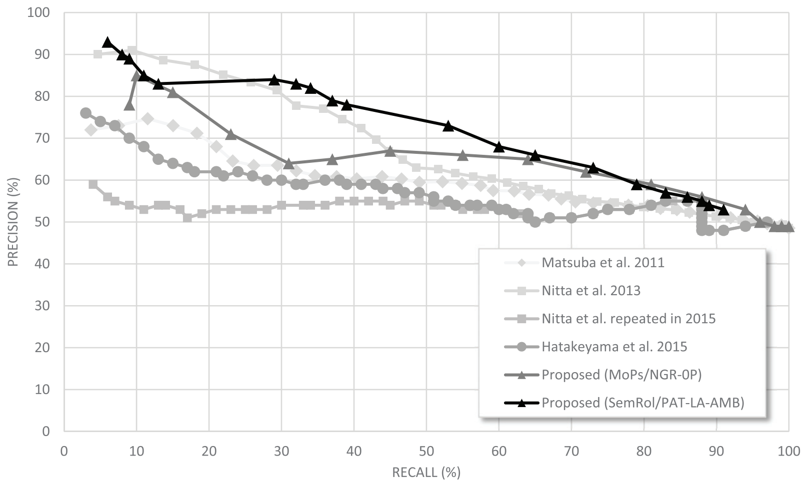

Fig. 3. Comparison of the proposed method (best and worst performance) and previous methods.

weight calculation and ambiguous patterns discarded – SemRol/PAT-LA-AMB), which starts from a high 93% and retains Precision between 90% and 70% for the major part of the threshold. The highest Precision score (93%) was higher than the one by Nitta et al. [9] (91%). Moreover, the Precision-performance of their method decreased more rapidly within the threshold span. However, when we repeated their experiment in 2015, the results of their method dropped significantly. After a thorough analysis of the experimental data, we noticed that most of the information extracted in 2013 was not available in 2015. Hatakeyama et al. [12] in their discussion provide the three most likely reasons for this drop: (1) fluctuation in page rankings (hindering information extraction), (2) the net-patrol movement itself (frequent deletion requests of harmful contents sent to service providers by PTA members), and (3) recent tightening of usage policies by most Web service providers, such as Google [27], Twitter [28] and *Yahoo!* used by Nitta et al. [9]. The two latter reasons are in fact positive ones, and it is difficult to consider an increase in general cyberbullying awareness as a hindrance to the developed system. On the contrary, it would be desirable for the system to be able to adapt to any ongoing changes in the Internet environment. Therefore, the need to modify the information extraction procedure in Nitta et al.'s system can also be considered as an important area for improvement. The initial study in this matter was performed by Hatakeyama et al. [12].

5. Conclusions and Future Work

This paper presented a novel method of detecting harmful entries that fall under the general label of cyberbullying (CB) on the Internet. Cyberbullying refers to the deliberate use of modern Internet technology to slander and harass other people. It is a recently noticed yet severe social problem that affects the mental health of Internet users, sometimes leading its victims to self-mutilation and even suicide.

Previous research on the topic of cyberbullying detection mainly focused on exploiting the frequent appearance of vulgar, violent, or obscene words to detect harmful content. In contrast, the proposed method is based on a novel and unprecedented approach consisting of three general steps: 1) representing the sentence in a morphosemantic structure, 2) automatically extracting sophisticated morphosemantic patterns from sentences, and 3) applying them to the classification of messages on the Internet. It is therefore an attempt to approach the problem of cyberbullying from a completely novel point of view; it is not based on any harmful vocabulary lists but rather on deep sentence structure represented by both morphological and semantic information.

The morphosemantic patterns, containing both semantic and morphological information, were extracted from actual cyberbullying entries, provided by the Human Rights Center. The extraction was done using a combinatorial algorithm to obtain not only traditional word patterns, or n-gram patterns, but also much more sophisticated patterns with disjointed elements. As the results showed that the proposed method outperformed previous approaches, it is sufficiently promising, indicating that morphosemantic sentence representation is useful in the context of detecting the deceptive and provocative language used in cyberbullying. Since the method requires minimal human effort, it can also be considered more efficient.

In the near future, we plan to put the proposed method into practice and propose an improvement to the method originated by Nitta et al. [9] and presently developed by Hatakeyama et al. [12]. Specifically, we will apply the method proposed here to extract specific morphosemantic patterns closely related to cyberbullying content and apply them to the information extraction of the original method.

We also plan to obtain new data to evaluate the method even more thoroughly, applying different classifiers. Finally, we plan to verify the actual amount of CB information on the Internet and reevaluate the method under more realistic conditions.

Acknowledgements

This research was supported by a (JSPS) KAKENHI Grant-in-Aid for the Encouragement of Young Scientists (B) (Project Number: 15K16044).

References:

[1] National Crime Prevention Council, http://www.ncpc.org/cyberbullying [accessed January 18, 2017]

[2] Ministry of Education, Culture, Sports, Science and Technology (MEXT), " "Bullying on the Net" Manual for handling and collection of cases (for schools and teachers)," MEXT, 2008 (in Japanese).

[3] S. Hinduja and J. Patchin, "Bullying beyond the schoolyard: Preventing and responding to cyberbullying," Corwin Press, 2009.

[4] J. W. Patchin and S. Hinduja, "Bullies move beyond the schoolyard: A preliminary look at cyberbullying," Youth Violence and Juvenile Justice, Vol.4, No.2, pp. 148-169, 2006.

[5] J. J. Dooley, J. Pyżalski, and D. Cross, "Cyberbullying Versus Face-to-Face Bullying: A Theoretical and Conceptual Review," Zeitschrift für Psychologie, Vol.217, No.4, pp. 182-188, 2007.

[6] L. Lazuras, J. Pyżalski, V. Barkoukis, and H. Tsorbazoudis, "Empathy and Moral Disengagement in Adolescent Cyberbullying: Implications for Educational Intervention and Pedagogical Practice," Studia Edukacyjne, nr23, pp. 57-69, 2012.

[7] T. Ishisaka and K. Yamamoto, "Extraction of abusive expressions from 2channel," Proc. of the 16th Annual Meeting of the Association for Natural Language Processing (NLP2010), pp.178-181, 2010 (in Japanese).

[8] M. Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka, K. Araki, and Y. Momouchi, "In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis," Int. J. of Computational Linguistics Research, Vol.1, Issue 3, pp. 135-154, 2010.

[9] T. Nitta, F. Masui, M. Ptaszynski, Y. Kimura, R. Rzepka, and K. Araki, "Detecting Cyberbullying Entries on Informal School Websites Based on Category Relevance Maximization," Proc. of the 6th Int. Joint Conf. on Natural Language Processing (IJCNLP 2013), pp. 579-586, 2013.

[10] T. Matsuba, F. Masui, A. Kawai, and N. Isu, "A study on the polarity classification model for the purpose of detecting harmful information on informal school sites," Proc. of the 17th Annual Meeting of the Association for Natural Language Processing (NLP2011), pp. 388-391, 2011 (in Japanese).

[11] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," Proc. of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 417-424, 2002.

[12] S. Hatakeyama, F. Masui, M. Ptaszynski, and K. Yamamoto, "Statistical Analysis of Automatic Seed Word Acquisition to Improve Harmful Expression Extraction in Cyberbullying Detection," Int. J. of Engineering and Technology Innovation, Vol.6, No.2, 2016, pp. 165-172.

[13] B. Levin and M. R. Hovav, "Morphology and Lexical Semantics," The Handbook of Morphology, A. Spencer and A. M. Zwicky (Eds.), pp. 248-271, 1998.

[14] P. Kroeger, "Morphosyntactic vs. morphosemantic functions of Indonesian -kan," Architectures, Rules, and Preferences: Variations on Themes of Joan Bresnan, A. Zaenen, J. Simpson, T. H. King, J. Grimshaw, J. Maling, and C. Manning (Eds.), pp. 229-251, CA: CSLI Publications, 2007.

[15] C. Fellbaum, A. Osherson, and P. E. Clark, "Putting semantics into WordNet's "morphosemantic" links," Human Language Technology. Challenges of the Information Society, Springer Berlin Heidelberg, pp. 350-358, 2009.

[16] I. Raffaelli, "The model of morphosemantic patterns in the description of lexical architecture," Lingue e linguaggio, Vol.12, No.1, pp. 47-72, 2013.

[17] Y. Nakajima, M. Ptaszynski, H. Honma, and F. Masui, "A Method for Extraction of Future Reference Sentences Based on Semantic Role Labeling," IEICE Trans. on Information and Systems, Vol.E99-D, No.2, pp. 514-524, 2016.

[18] K. Takeuchi, S. Tsuchiyama, M. Moriya, and Y. Moriyasu, "Construction of Argument Structure Analyzer Toward Searching Same Situations and Actions," IEICE Technical Report, Vol.109, No.390, pp. 1-6, 2010.

[19] MeCab: Yet Another Morphological Analyzer, http://taku910.github.io/mecab/ [accessed January 18, 2017]

[20] M. Ptaszynski, R. Rzepka, K. Araki, and Y. Momouchi, "Language combinatorics: A sentence pattern extraction architecture based on combinatorial explosion," Int. J. of Computational Linguistics (IJCL), Vol.2, No.1, pp. 24-36, 2011.

[21] T. Matsuba, F. Masui, A. Kawai, and N. Isu, "Detection of harmful information on informal school websites," Proc. of the 16th Annual Meeting of the Association for Natural Language Processing (NLP2010), 2010 (in Japanese).

[22] Human Rights Research Institute Against All Forms for Discrimination and Racism in Mie Prefecture, Japan, http://www.pref.mie.lg.jp/jinkenc/hp/ [accessed January 18, 2017]

[23] Student, "The probable error of a mean," Biometrika, Vol.6, No.1, pp. 1-25, 1908.

[24] M. Ptaszynski, F. Masui, R. Rzepka, and K. Araki, "Emotive or Non-emotive: That is The Question," Proc. of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA 2014), pp. 59-65, 2014.

[25] A. Kidman, "How to do things with four-letter words: A study of the semantics of swearing in Australia," Unpublished Dissertation, University of New England, 1993.

[26] R. Zhao and K. Mao, "Cyberbullying Detection based on Semantic-Enhanced Marginalized Denoising Auto-Encoder," IEEE Trans. on Affective Computing, 2016.

[27] Google Anti-harrassment Policy, https://www.google.com/events/policy/anti-harassmentpolicy.html [accessed January 18, 2017]

[28] Twitter, Building a Safer Twitter, https://blog.twitter.com/2014/building-a-safer-twitter [accessed January 18, 2017]

Appendix A. List of Abbreviations

Descriptions of abbreviations used in this paper in alphabetical order.	
0P	patterns which appear with the same occurrence on both sides of data; called "zero patterns" because their weight is equal to 0
ALL	all patterns or ngrams were used in classification
AMB	ambiguous patterns; refers to patterns which appear on both sides of data with different occurrence
MoPs	morphosemantic patterns; see Section 3.1 for explanation
MS	morphosemantic structure; see Section 3.1 for explanation
NGR	ngrams; refers to only word ngrams extracted from sentences not from all patterns
NGR-0P	zero-ngrams deleted
NGR-ALL	all ngrams used
NGR-AMB	ambiguous ngrams deleted
NGR-LA	ngram length awarded in weight calculation and all ngrams used
NGR-LA-0P	length awarded and zero-ngrams deleted
NGR-LA-AMB	ngram length awarded and ambiguous ones deleted
PAT	patterns; refers to sophisticated patterns with disjoint elements
PAT-0P	zero-patterns deleted
PAT-ALL	all patterns used classification
PAT-AMB	ambiguous patterns deleted
PAT-LA	pattern length awarded in weight calculation and all patterns used
PAT-LA-0P	length awarded and zero-patterns deleted
PAT-LA-AMB	length awarded ambiguous patterns deleted
POS	parts of speech; nouns, verbs, particles, etc.
SR	semantic roles; see Sections 3.1 and 4.2 for explanation

Name:
Michal Ptaszynski

Affiliation:
Assistant Professor, Department of Computer Science, Kitami Institute of Technology

Address:
165 Kouen-cho, Kitami, Hokkaido 090-8507, Japan

Brief Biographical History:
2010- JSPS Post Doctoral Fellowship, Hokkai-Gakuen University
2012- Post Doctoral Researcher, Kitami Institute of Technology
2013- Assistant Professor, Kitami Institute of Technology

Main Works:
- "ML-Ask: Open Source Affect Analysis Software for Textual Input in Japanese," J. of Open Research Software, Vol.5, No.1, pp. 1-16, 2017.
- "Sustainable cyberbullying detection with category-maximized relevance of harmful phrases and double-filtered automatic optimization," Int. J. of Child-Computer Interaction, Vol.8, pp. 15-30, 2016.

Membership in Academic Societies:
- The Association for Computational Linguistics (ACL)
- The Association for the Advancement of Artificial Intelligence (AAAI)
- The Institute of Electrical and Electronics Engineers (IEEE)
- The Association for Natural Language Processing (ANLP)
- The Information Processing Society of Japan (IPSJ)
- The Japanese Society for Artificial Intelligence (JSAI)

Name:
Fumito Masui

Affiliation:
Associate Professor, Department of Computer Science, Kitami Institute of Technology

Address:
165 Kouen-cho, Kitami, Hokkaido 090-8507, Japan

Brief Biographical History:
1990- Joined Oki Electric Industry Corporation
2000- Assistant Professor, Mie University
2009- Associate Professor, Kitami Institute of Technology

Main Works:
- "Informatics to Support Tactics and Strategies in Curling," Int. J. of Automation Technol., Vol.10, No.2, pp. 244-252, 2016.
- "A System for Recommendation of Accommodation Facilities Adaptable to User Interest," Tourism Informatics: Towards Novel Knowledge Based Approaches, Springer-Verlag, 2015.

Membership in Academic Societies:
- The Association for the Advancement of Artificial Intelligence (AAAI)
- The Institute of Electrical and Electronics Engineers (IEEE)
- The Association for Natural Language Processing (ANLP)
- The Information Processing Society of Japan (IPSJ)
- The Japanese Society for Artificial Intelligence (JSAI)
- Institute of Electronics, Information and Communication Engineers
- Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
- Hokkaido Regional Tourism Society
- Japan Curling Association

Name:
Yoko Nakajima

Affiliation:
Assistant Professor, National Institute of Technology, Kushiro College

Address:
2-32-1 Otanoshike-Nishi, Kushiro-shi, Hokkaido 084-0916, Japan

Brief Biographical History:
2016 Ph.D. in Engineering from Kitami Institute of Technology
2016- Assistant Professor, Department of Creative Engineering, National Institute of Technology, Kushiro College

Main Works:
• "A Method for Extraction of Future Reference Sentences Based on Semantic Role Labeling," IEICE Trans. on Information and Systems, Vol.E99-D, No.2, pp. 514-524, 2016.

Membership in Academic Societies:
• Institute of Electronics, Information and Communication Engineers
• The Information Processing Society of Japan (IPSJ)

Name:
Yasutomo Kimura

Affiliation:
Associate Professor, Department of Information and Management Science, Otaru University of Commerce

Address:
3-5-21 Midori, Otaru 047-8501, Japan

Brief Biographical History:
2004 Ph.D. at Information Science and Technology, Hokkaido University
2005- Assistant Professor, Otaru University of Commerce
2007- Associate Professor, Otaru University of Commerce

Main Works:
• "Creating Japanese Political Corpus from Local Assembly Minutes of 47 Prefectures," Coling 2016 Workshop, The 12th Workshop on Asian Language Resources, pp. 78-85, 2016.

Membership in Academic Societies:
• The Association for Natural Language Processing (ANLP)
• The Information Processing Society of Japan (IPSJ)
• Institute of Electronics, Information and Communication Engineers
• The Japanese Society for Artificial Intelligence (JSAI)

Name:
Rafal Rzepka

Affiliation:
Assistant Professor, Graduate School of Information Science and Technology, Hokkaido University

Address:
Kita 14, Nishi 9, Kita-ku, Sapporo, Hokkaido 060-0814, Japan

Brief Biographical History:
1999 M.A. from University of Adam Mickiewicz
2004 Ph.D. from Hokkaido University
2004- Assistant Professor, the Graduate School of Information Science and Technology, Hokkaido University

Main Works:
• "ELIZA Fifty Years Later: An Automatic Therapist Using Bottom-Up and Top-Down Approaches," Machine Medical Ethics, Intelligent Systems, Control and Automation: Science and Engineering, Vol.74, pp. 257-272, 2015.

Membership in Academic Societies:
• The Association for the Advancement of Artificial Intelligence (AAAI)
• The Japanese Society for Artificial Intelligence (JSAI)
• The Association for Natural Language Processing (ANLP)
• Japanese Cognitive Science Society (JCSS)

Name:
Kenji Araki

Affiliation:
Professor, Graduate School of Information Science and Technology, Hokkaido University

Address:
Kita 14, Nishi 9, Kita-ku, Sapporo, Hokkaido 060-0814, Japan

Brief Biographical History:
1988 Ph.D. from Hokkaido University
1988- Professor, Hokkai-Gakuen University
1998- Associate Professor, Division of Electronics and Information Engineering, Hokkaido University
2002- Professor, Graduate School of Information Science and Technology, Hokkaido University

Main Works:
• "Non-Segmented Kana-Kanji Translation Using Inductive Learning," Trans. of the Institute of Electronics, Information and Communication Engineers, D-II, Vol.J79-D-II, No.3, pp. 391-402, 1996 (in Japanese).

Membership in Academic Societies:
• The Association for the Advancement of Artificial Intelligence (AAAI)
• The Institute of Electrical and Electronics Engineers (IEEE)
• The Information Processing Society of Japan (IPSJ)
• The Japanese Society for Artificial Intelligence (JSAI)
• Japanese Cognitive Science Society (JCSS)