A Large Scale Knowledge Base Representing the Base Form of *Kaomoji*

Noriyuki Okumura
Department of Electrical and Computer Engineering, National Institute of Technology, Akashi College,
679-3 Nishioka, Uozumi-cho, Akashi-City, Hyogo-Prefecture, Japan

Keywords: Kaomoji, Emoticon, Original Form, N-gram, Kaomoji's Dictionary, Annotation.

Abstract: In this paper, we construct a large-scale knowledge base representing the base form of *kaomoji* (emoticon) and other elements of *kaomoji*: eye, nose, mouth, and so on, to analyze features of *kaomoji* in detail. Previous methods to analyze *kaomoji* mainly aim to extract *kaomoji* from sentences, paragraphs, or documents, or to classify *kaomoji* into some emotion classes based on the emotion that *kaomoji* shows or potentially includes. We define the base form of *kaomoji* for detailed *kaomoji* analytics. Application systems can estimate another feature of derivative *kaomoji* based on its base form and other elements for sentiment analytics, emotion extraction, or *kaomoji* classification. We annotated about 40,000 kinds of *kaomoji* for constructing a large-scale knowledge base. The total number of extracted base forms is about 3,000. In experimental evaluations based on cosine similarity using N-gram based features and simple Skip-gram based features, we show that the model can estimate the base form of *kaomoji* with an accuracy of about 50%.

1 INTRODUCTION

How can we understand our real intentions, emotions or feelings each other using only written words in computer-mediated media such as Twitter[1], Facebook[2] or something like them? Can we send our gestures using only characters? One of the solutions is *kaomoji*. *Kaomoji* is a typical sequence that can send writer's expression, sign, emotion, and so on to the reader with similar uses as an emoticon, a pictogram, a smiley, and a stamp using only written words. *Kaomoji* is widely used especially in Japanese culture. We observed over 100,000 kinds of *kaomoji* in the web pages written in the Japanese language.

Why *kaomoji* come into widespread use as communicating writer's emotions, gestures and so on in Japan? Because Japanese culture is the most high-context culture all over the world. People living in the high-context culture can communicate with each other without expressing their opinions clearly on a face-to-face basis because speakers and listeners have a cultural context in common(Hall, 1976). On the other hand, people living in the low-context culture (for example, English-speaking people, German-speaking people, and so on) communicate with each other using a clear explanation because they do not have a cultural context in common. It is hard that people do communication with each other at the character-based communication on the internet when using high-context language such as Japanese. People can estimate the intension of speaker's utterance with observing speaker's expression, gestures, and accent at the face-to-face communication. Character-based communication such as internet services has the potential to make readers misunderstand their real intention because of the limited information to express their intention. Because of these situations, *kaomoji* come into widespread use as communicating writer's implicit messages in the character-based communication.

kaomoji do exist over 100,000 kinds of their variations as previously noted. Nowadays, the number of *kaomoji* increase day by day. The preferred situation is to keep renewing a large-scale knowledge base *kaomoji* such as the proposed knowledge base in this paper, but it is distant idea because constructing large-scale knowledge base needs manual annotations. Therefore, we need the method that estimates the meaning that *kaomoji* expresses by extracting its base form[3] such as lemmatization of verbs and the other elements of it. For example, we can extract pos-

[1] https://twitter.com

[2] http://www.facebook.com

[3] In this paper, we use this words (the base form) for *kaomoji* despite the existence of "the original form" or "infinitive"

Okumura N.
A Large Scale Knowledge Base Representing the Base Form of Kaomoji.
DOI: 10.5220/0006517002460252
In *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KEOD 2017)*, pages 246-252
ISBN: 978-989-758-272-1

Table 1: Rules and regulations to extract/make the base form of *kaomoji*.

kaomoji	Original Form	Rules and regulations
(ˆ␣ˆ)	(ˆˆ)	Remove all spaces
\(ˆ_ˆ)/	(ˆ_ˆ)	Remove parts without borders, eyes, mouth and nose
	(ˆ_ˆ)	Transform characters from two-byte characters to one-byte characters
(′‘)	()	If there are no corresponding one-byte characters, then the characters remain in two-byte characters (in this case, the object is ω)
(T_T)/˜˜˜	(T_T)	Do not transform characters from uppercase/lowercase to lowercase/uppercase (for instance, do not transform (T_T) to (t_t))
(>_-)	(>_<), (-_-)	If *kaomoji* is asymmetry, then make it symmetry
(;)	(;;)	If the eyes or some parts are covered with arms or something, then complement them using the other side of objective parts
(ˆ_ˆ;	(ˆ_ˆ)	If *kaomoji* have odd border, then complement it using the other side of border
((;)	(), ()	If *kaomoji* have multiple faces, then extract each original form
)ˆˆ)	(ˆˆ)	If borders are asymmetry, then make border's curve to outside

itive sentiment from *kaomoji* such as (ˆ_ˆ) because we usually use this face as "smiling." If we attach the new element (*) to this *kaomoji*, we transform (ˆ_ˆ) into (*ˆ_ˆ*). (*ˆ_ˆ*) has more positive sentiment than (ˆ_ˆ). In fact, (*) has the role to emphasize positive sentiment. By the same token, we can extract negative sentiment from *kaomoji* such as () because we usually use this face as "anger." If we attach the new element (*) to this *kaomoji*, we transform () into (**). (**) has more negative sentiment than (). In fact, (*) has the role to emphasize positive/negative sentiment. We can guess the sentiment of (*ˆ_ˆ*) or (**) using the sentiment information of (ˆ_ˆ) or () because we know the role of (*) whether we have or do not have any information of (*ˆ_ˆ*) or (**).

In this paper, we annotate about 40,000 kinds of *kaomoji* collecting from websites[4]. We extract about 3,000 kinds of the base form of *kaomoji* using annotated *kaomoji*. We also investigate whether N-gram based model can estimate the base form of *kaomoji*. In experimental evaluations based on cosine similarity using N-gram based features and simple Skip-gram based features, we show that the model can estimate the base form of *kaomoji*' with an accuracy of about 50%.

2 RELATED WORK

The research dealing with *kaomoji* is come across occasionally. The research, however, that focuses on the relation: the base form - the derivative form as our study is in the minority. Ptaszynski et al. proposed CAO: A Fully Automatic Emoticon Analysis System to achieve *kaomoji* extraction with high accuracy(Ptaszynski et al., 2010b; Ptaszynski et al., 2010a). CAO system can manage over 10,000 kinds of *kaomoji*, and the system can extract over a few million of *kaomoji*. On the other hand, they do not focus on *kaomoji*'s original form. CAO system only extracts *kaomoji* from sentences using eye-mouth-eye triplet in this scheme.

Bedrick et al. constructed a robust method to extract *kaomoji* from tweets on Twitter(Bedrick et al., 2012). They proposed the extraction method of *kaomoji* in consideration of symmetric property using Probabilistic context-free grammar (PCFG). In our study, we attach great importance to *kaomoji*'s symmetric property as their research.

Yamada et al. constructed the classification system for about 700 *kaomoji* using N-gram based feature extracted each *kaomoji*(Yamada et al., 2007). This system classifies *kaomoji* to 8 emotions (smile, cry, anger, surprise, confuse, unsatisfied, anxiety, no emotion) with an accuracy of 90%. They, however, focus on only global classification without putting each element of *kaomoji* to the proof.

Tanaka et al. reported the classification system using K-means and Support Vector Machines (SVMs) that classify *kaomoji* to 6 emotions (pleasure, sadness, anger, surprise, action, bitter smile) in the similar way of Yamada's research(Tanaka et al., 2005). These researches mainly focus on emotions that *kaomoji* have, but they do not use each element of *kaomoji*.

Kazama et al. construct a method to extract *kaomoij* from Twitter[5](Kazama et al., 2016). Their method to extract *kaomoji* focuses on Unicode blocks and Unicode character properties. The method can extract over 50 million kinds of *kaomoji* from Twitter (360 million tweets.) The method, however, cannot

[4] For example:http://www.kaomoji.sakura.ne.jp/

[5] This article is submitted to a Japanese conference and it is written in Japanese, however, their extraction method of *kaomoji* marked high accuracy. We dare to introduce this article here.

deal with *kaomoji* that have comments such as (**).[6]

The other research about *kaomoji* are the recommendation method of *kaomoji*(Urabe et al., 2013), the method to extract *kaomoji*'s tense(Onishi and Okumura, 2014) and so on. "Science of Emoticons"(Ptaszynski et al., 2012) has a detailed knowledge of general *kaomoji* extraction method.

3 CONSTRUCTION OF LARGE-SCALE KNOWLEDGE BASE REPRESENTING THE BASE FORM OF *KAOMOJI*

In this study, we annotate 70,106 kinds of *kaomoji* that are collected from websites. Annotators add tags to *kaomoji*'s each element using the rules in the following subsections. The terms to annotate *kaomoji* is from 1st October 2015 to 20th March 2016. Annotators are six persons including five males and a female. As a result, we annotated 43,373 *kaomoji* out of all of the collected *kaomoji* in the experimental period.

3.1 Definition of the Base Form of *Kaomoji*

It is hard that we analyze the meaning of each element of *kaomoji* because *kaomoji* have various elements. Therefore, we define the base form of *kaomoji* such as stemming of verbs or lemmatization for easy handling in natural language processing. The basic elements of the base form of *kaomoji* are the triplet (eye-mouth-eye: sometimes move away from mouth towards the nose) that Ptaszynski et al. proposed. Besides, we focus on the borders of *kaomoji* because almost *kaomoji* have borders. Table 1 shows the rules and regulations to extract/make the base form of *kaomoji*.

We extracted 3,110 base forms of *kaomoji* in this work. Table 2 shows the sample of the base form of *kaomoji* in order of prevalence. The number of the base forms that have only one derivative *kaomoji* is 1,173. Therefore, 1,973 kinds of the base form of *kaomoji* have two or more derivative *kaomoji*. We show the details of each *kaomoji*'s element in the following subsection.

3.2 Annotated Elements of *Kaomoji*

We annotated the collected *kaomoji* to extract the base form of *kaomoji*. We also annotated each element that falls under the category of eye, nose, mouth, cheek, ear, forehead, border, arm, other body elements, caption, onomatopoeia, and the other expression. In this section, we show the examples of these elements of *kaomoji*.

Table 2: Examples of *kaomoji*'s original form.

The base form	Freq.	The base form	Freq.
(_)	1032	(ˆ-ˆ)	495
(-_-)	610	(00)	471
()	589	(w)	466
(__)	540	(o|o)	452
(>_<)	529	()	426

3.2.1 Eyes

The smallest components of *kaomoji* are eyes. Therefore, this research pushes aside emoticons such as orz or back shot that have no eyes because these emoticons do not show their faces. *kaomoji* sometimes have signs of inequality as their eyes such as (>_<). These signs indicate that eyes harmonized with arms. We extract these signs as eyes in this study. Table 3 shows the sample of eyes.

Table 3: Eyes: the elements of the base form of *kaomoji*.

kaomoji	Eyes	*kaomoji*	Eyes
(--)	- -	()	

3.2.2 Mouth, Nose

The next important elements of *kaomoji* are mouth and nose. The expressions dramatically increase with mouth or nose as the elements of *kaomoji*. On the other hand, a nose exists instead of mouth or a mouth harmonized with a nose. Therefore, we extract a mouth and a nose as core elements of *kaomoji*'s original form by consensus of annotators. Table 4 shows the samples of mouth and nose.

Table 4: Mouth and Nose: the second elements of the base form of *kaomoji*.

kaomoji	Mouth	*kaomoji*	Nose
((( )))		(*ˆˆ*)	ˆˆ
(′′)	-	()	@@

[6] means "Hello."

3.2.3 Borders

Borders of *kaomoji* are made mostly of a parenthesis. In other words, a parenthesis is not an essential element of *kaomoji* regarding discriminating *kaomoji* with a concept of IDF(Inverse Document Frequency in information retrieval or natural language processing). Borders, however, change the expression of *kaomoji* or the combination of borders express the movement of *kaomoji*. We extract borders as the element parts of the base form of *kaomoji* because they have possibilities to express a subtle sense. Table 5 shows the samples of borders.

Table 5: Borders: the third elements of the base form of *kaomoji*.

kaomoji	Borders	*kaomoji*	Borders
( )	( )	_	

3.2.4 Cheek, Ears, Forehead

The face has the other parts such as cheek, ears and forehead except for the elements as noted above. We healthy people have these parts without any exceptions. On the other hand, *kaomoji* have these parts for the rare occasion(frequency of these parts is low). We extract the base form of *kaomoji* exclusive of these parts because these parts have the role of decorating *kaomoji*'s original form. Table 6 shows the samples of cheek and ears.

Table 6: The samples of cheek and ears.

kaomoji	Cheek	*kaomoji*	Ears
(*ˆˆ*)	* *	Ł(())Ł	Ł Ł
(;)	;	( )	

3.2.5 Arms

The word "*kaomoji* has two japanese words: "kao" and "moji." The word "kao" means the "face" and the word "moji" means the "characters or symbols" in Japanese language under normal circumstances. Whether arms, legs or some parts exclusive of the parts of face belong to *kaomoji* remains a matter of debate. In fact, there are many something like *kaomoji* that have arms, legs, and so on. We extract arms as the parts that add value to *kaomoji* in this study. Arms consist of symmetry parts or one side parts. Therefore, we extract arms without considering symmetry property. Table 7 shows the samples of arms.

Table 7: The samples of arms.

kaomoji	Arms	*kaomoji*	Arms
|0|		(--)	

3.2.6 Other Elements

Many various elements show something exclusive of face or body as the elements of *kaomoji*. These elements show the particular situation or emphasizing emotions. Whether these parts belong to *kaomoji* remains a matter of debate in a similar way to arms. In this paper, the existential reason of *kaomoji* is to resolve the issue of context in Japanese culture as we once remarked in Section 1. Therefore, this study defines *kaomoji* as combined expression of faces and the other particular situations. Table 8 shows the sample of the elements that shows particular situation. In this case, ˜ shows the situation: "Do you have a cup of Japanese tea?".

Table 8: The other elements exclusive of face and arms.

kaomoji	Elements
(-_-)˜	˜

The manageability of these elements differs depending on the position that *kaomoji* have these elements on the left side or right side. For example, it is harder for CAO system or Bedrick's PCFG to extract sequences forward *kaomoji* than following *kaomoji*. The reason is that *kaomoji* are used as the punctuation of sentences. We can extract the sequences following *kaomoji* using CAO system or Bedrick's PCFG because we usually use *kaomoji* at the end of sentences. On the other hand, it is unobvious where we should return to extract sequences forward *kaomoji*. Our knowledge base of *kaomoji* discriminates the sequences that are at the left side or the right side.

3.2.7 Caption, Onomatopoeia

Kaomoji have every character and symbol as the elements. These elements construct "sentence" or "onomatopoeia" in some instances. Sentences or onomatopoeia have detailed information such as emotions or situations. We extract this information from *kaomoji* in a proactive way. For example, o(ˆ-ˆ)o shows the wonderful situation with onomatopoeia[7].

[7] means wonderful.

3.2.8 Other Expression

There are the characteristic elements of *kaomoji* other than described above. For example, (((())))))) shows shuddering situation with repeated use. Repeated use of a certain element shows this expression. We extract these repeated use element as the other expression of *kaomoji*.

We Japanese people regularly use two-byte characters in daily life. It is necessary to discriminate between two-byte characters and one-byte characters. For instance, (ˆˆ) and are similar looking faces because these *kaomoji* consist of same parts. These are, however, different faces in a strict sense because of the former consist of one-byte characters, the latter consist of two-byte characters. The latter looks like well-rounded than the former. In this paper, we transform characters from two-byte characters to one-byte characters when we extract/make *kaomoji*'s original form because of choking off the manifold.

4 EXPERIMENTS FOR ESTIMATING THE BASE FORM OF *KAOMOJI*

We can refer the base form of *kaomoji* and the other elements of *kaomoji* using the large-scale knowledge base of *kaomoji* constructed in this paper. There is, however, a high possibility of encountering unknown *kaomoji* because of the number of *kaomoji* increase day by day. We except UTF *kaomoji* despite the increase in *kaomoji* because we find out the UTF *kaomoji* that are outside our study. Figure 1 shows the samples of UTF *kaomoji*.

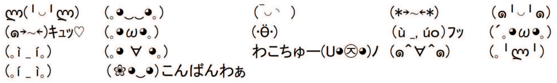

Figure 1: *kaomoji* using UTF-8.

This paper investigates the method to estimate the base form of *kaomoji*. If the method can estimate the base form of *kaomoji*, then the method can also estimate the information of *kaomoji* using the other parts exclusive of the base form. In other words, the method estimate emotions, situations, and gestures using the base form of *kaomoji* and the other elements of it. We describe the method of estimating original form using N-gram based features, simple Skip-gram features, and the combination of these features in the following section.

4.1 Experiment to Estimate the Base Form of *Kaomoji* using N-gram Model

N-gram based features are widely used in natural language processing. We can use this simple model for analyzing *kaomoji*. We can calculate the similarity between *kaomoji* using N-gram based features by separating *kaomoji*. In this paper, we calculated the accuracy of estimating the base form of kaomoji using N-gram based model (N=1 to 5)(Eq. 1). In the case of *Kaomoji* that have multiple base forms, we count it a correct answer if one of the each base form is extracted.

$$Accuracy = \frac{num\ of\ kaomoji\ estimated\ correctly}{num\ of\ all\ annotated\ kaomoji\ (43,373)} \quad (1)$$

Cosine similarity is used to calculate similarity between *kaomoji*(Eq. 2).

$$Cosine(A,B) = \frac{A \cdot B}{|A||B|} \quad (2)$$

Our method calculates similarity using Eq. 2 between input *kaomoji* and all of extracted/made the base forms from large-scale knowledge base of *kaomoji* (Section 3.1). The base form of *kaomoji* that scored the highest similarity is the estimated base form. Table 9 shows the result of estimation using N-gram based features.

Table 9: Accuracy of estimating the base form of *kaomoji* using N-gram based single feature.

N-gram	Accuracy
1gram	0.244
2gram	0.386
3gram	**0.440**
4gram	0.336
5gram	0.019

The highest accuracy is 0.440 (using Trigram). On the other hand, the lowest accuracy is 0.019 (using 5-gram). 5-gram model seems to be useless for estimating original form. We investigate the estimation of *kaomoji*'s original form exclusive of 5-gram in the following section.

4.2 Experiment to Estimate the Base Form of *Kaomoji* using Simple Skip-gram Model

Skip-gram is used in Word2Vec model proposed by Mikolov et al(Mikolov et al., 2013). Skip-gram is the model that use co-occurrence words in configured window size. If the window size equals 6, then for example ”I go to school by bus.” is separated to ”I - go”, ”I - to”, ”go - bus” , and so on. In our study, we use more simple Skip-gram model. We extend N-gram model to simple Skip-gram model that does not use every combination of co-occurrence words(characters) but only use skipping N words (characters). If the window size equals 3(Skip 1 character model), then for example ”I go to school by bus.” is separated to ”I - to”, ”go - school”, ”to - by”, ”school - bus” by using our simple Skip-gram model. We can change the expression of *kaomoji* by inserting a certain element. For instance, we transform (ˆ_ˆ) into (*ˆ_ˆ*) by inserting (*). Simple Skip-gram model is effective about this situation. Table 10 shows the result using naive Skip-gram(Skip N characters N = 1 to 3). The highest accuracy is 0.270(Skip 1 character model). However, the accuracy is lower than N-gram based model.

Table 10: Accuracy of estimating the base form of *kaomoji* using simple Skip-gram based single feature.

naive Skip-gram	Accuracy
Skip 1 character	**0.270**
Skip 2 characters	0.039
Skip 3 characters	0.018

4.3 Experiment to Estimate the Base Form of *Kaomoji* using the Combination of N-gram and Simple Skip-gram based Features

In the above experiment, we use only individual features such as only unigram, only trigram and so on. We examine the multiple features to estimate the base form of *kaomoji*. 5-gram model is removed in this experiment as previously noted.

Table 11 shows the result of combination model(N-gram based features). The highest accuracy is 0.433(bigram + trigram model). However, this accuracy is lower than using only trigram model despite we form a hypothesis that the combination model improves the accuracy.

Table 12 shows the result of combination model (N-gram and naive Skip-gram model). The highest accuracy is 0.489 (bigram + trigram + Skip 1 character model) with contrary to expectations.

Figure 2 shows each the highest accuracy of estimation using N-gram and simple Skip-gram based features.

Table 11: Accuracy of estimating the base form of *kaomoji* using combination of N-gram based features.

Combinations	Accuracy
1gram+2gram	0.381
1gram+3gram	0.406
1gram+4gram	0.339
2gram+3gram	**0.433**
2gram+4gram	0.410
3gram+4gram	0.425
1gram+2gram+3gram	0.422
1gram+2gram+4gram	0.401
1gram+3gram+4gram	0.414
2gram+3gram+4gram	0.427
1gram+2gram+3gram+4gram	0.431

Table 12: Accuracy of estimating the base form of *kaomoji* using combination of N-gram based features and naive Skip-gram based features.

Combinations	Accuracy
2gram+Skip 1 character	0.470
3gram+Skip 1 character	0.404
2gram+3gram+Skip 1 character	**0.489**

5 CONCLUSION

In this paper, we annotated *kaomoji* and constructed a large-scale knowledge base of *kaomoji*. The total number of annotated *kaomoji* is 43,373 out of our collected *kaomoij*. As a result, we extracted 3,110 kinds of the base form of *kaomoji*. In experimental evaluations, we achieved 0.489 accuracy of estimating the base form *kaomoji* using bigram + trigram + Skipping 1 character model. Ih the case of random estimation, the maximum probability is 1,032 / 43,373 (=0.024) that Table 2 shows. We consider that our experiments are efficient for estimating the base form of *kaomoij*.

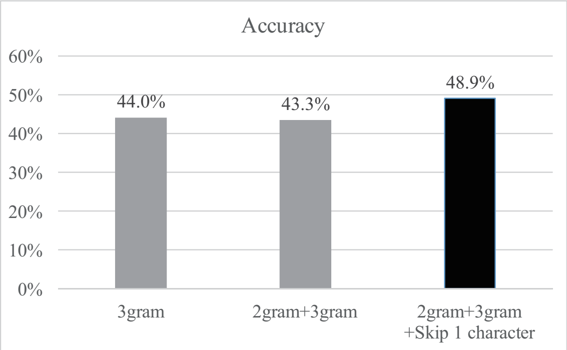

Figure 2: Result of experiments using N-gram and simple Skip-gram based fearures.

As future works, we will finish annotating remaining kaomoji. We will attempt to construct classification method using deep learning tools such as Chainer, TensorFlow, and so on to improve the accuracy of estimation (c.f. emoji2vec(Eisner et al., 2016)). In addition, we have to annotate *kaomoji*'s emotions based on Plutchik model(Plutchik, 1980). Because we do not extract emotions that *kaomoji* shows in this paper.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number 15K21592.

REFERENCES

Bedrick, S., Beckley, R., Roark, B., and Sproat, R. (2012). Robust kaomoji detection in twitter. In *Proceedings of the Second Workshop on Language in Social Media*, LSM '12, pages 56–64, Stroudsburg, PA, USA. Association for Computational Linguistics.

Eisner, B., Rocktäschel, T., Augenstein, I., Bosnjak, M., and Riedel, S. (2016). emoji2vec: Learning emoji representations from their description. *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*, pages 48–54.

Hall, E. T. (1976). *Beyond Culture*. Anchor Books.

Kazama, K., Mizuki, S., and Sakaki, T. (2016). Study of sentiment analysis using emoticons on twitter. *The 30th Annual Conference of the Japanese Society for Artificial Intelligence, 3H3-OS-17a-4*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.

Onishi, C. and Okumura, N. (2014). An inverstigation of the usage of kaomoji for emotions judgment and kaomoji recommendation. In *The 13th IASTED International Conference on Artificial Intelligence and Applications AIA2014. #816-014*.

Plutchik, R. (1980). Chapter 1 - a general psychoevolutionary theory of emotion. In Plutchik, R. and Kellerman, H., editors, *Theories of Emotion*, pages 3 – 33. Academic Press.

Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R., and Araki, K. (2010a). Cao: A fully automatic emoticon analysis system. In Fox, M. and Poole, D., editors, *AAAI*. AAAI Press.

Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R., and Araki, K. (2010b). Cao: A fully automatic emoticon analysis system based on theory of kinesics. *IEEE Transactions on Affective Computing*, 1(1):46–59.

Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R., Araki, K., and Momouchi, Y. (2012). *Science of Emoticons*. IGI Global.

Tanaka, Y., Takamura, H., and Okumura, M. (2005). Extraction and classification of facemarks. In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, IUI '05, pages 28–34, New York, NY, USA. ACM.

Urabe, Y., Rafal, R., and Araki, K. (2013). Emoticon recommendation for japanese computer-mediated communication. In *Semantic Computing (ICSC), 2013 IEEE Seventh International Conference on*, pages 25–31.

Yamada, T., Tsuchiya, S., Kuroiwa, S., and Ren, F. (2007). Classification of facemarks using n-gram. In *2007 International Conference on Natural Language Processing and Knowledge Engineering*, pages 322–327.