Proc. 1st International Conference on
Machine Learning and Data Engineering (iCMLDE2017)
20-22 Nov 2017, Sydney, Australia
ISBN: 978-0-6480147-3-7

Facial Expression Classification Based on Shape Feature of Emoticons

Akira Fujisawa[1], Kazuyuki Matsumoto[1], Minoru Yoshida[1] and Kenji Kita[1]
[1]Tokushima University, Tokushima, Japan

Corresponding author's E-mail: c501547001@tokushima-u.ac.jp

Abstract

Emoticons are used in the situation of textual communication such as web mails and internet forums. Many of the existing studies dealing with classification or extraction of emoticons regard emoticons as a kind of character string and focus on what characters constitute the emoticons or how they are lined up. However, emoticons are used to express human facial expressions, and characters constituting them represent various facial parts such as eyes, nose, mouth, etc. Such characters can be identified as different facial parts depending on their positions, and facial expressions are thought to be represented by the combinations of their shape features. In this study, we classified the facial expressions of emoticons by focusing on the shape features of those emoticons. To deal with shape features of emoticons, we converted emoticons, which are text data, to image data. Emoticons are mainly formed by line segments of characters, and use only black and white colors. Therefore, other factors such as colors and shades were not considered as the feature to classify the facial expressions. In the experiments, we used image features that did not require color information. As the result of comparative experiment with the 1-nearest neighbor method using character features, the facial expression recognition rate is 52% when using the Histograms of Oriented Gradients(HOG) was used as image feature. By this result, proposed method improved recognition rate by 2 % than using baseline .

Keywords: Facial expression classification, Emoticon, Image feature, k-means clustering.

1. INTRODUCTION

Communication with various type of non-verbal expressions has been increasing on Social Networking Services. Among them, emojis and emoticons are type of expressions that can be easily embedded into sentences. Japanese emoticons are unique in their variety. An emoticon is thought to be a kind of deformation that represents a human face by using character sequence.

The existing engineering studies [Suzuki et al.(2006), Urabe et al.(2013)] have proposed several methods that classify facial expressions from emoticons, that extract emoticons from sentences and that recommend emoticons for users to use. These studies focused on the difference between a word sequence consisting of a sentence and emoticon (a symbol sequence) to extract or classify emoticons. However, an emoticon is different from a simple expression using symbols, it is a picture drawn with characters, that is to say, a kind of Ascii Art. Because emoticon convey visually appealing information, a simple matching between the symbol sequences might fail.

In this study, we extracted image features by treating an emoticon as an image and identifying its shape elements. By using this feature, we thought that we could grasp the characteristics of its shape more accurately as more than a simple sequence of symbols, and thus recognize facial expressions from emoticons more precisely.

Many studies on emoticon have tried to extract information from emoticons, recognize facial expressions from emoticons, recommend emoticons to use, etc. As our study aimed to classify facial expressions of emoticons, we would like to introduce some studies related to facial expression recognition from emoticons.

Yamada et al.(2007) attempted recognition of facial expression in emoticons by training the sequence of characters in those emoticons using N-gram. Ptaszynski et al.(2010) developed a “CAO system” for extracting emoticons from text. Okumura(2016) constructed a large-scale emoticon dictionary. Because many emoticon patterns are possible by combining symbols to be added before and after the emoticon, he proposed a method to extract the primitive form of emoticons.

These exiting studies proposed method to recognize facial expressions of emoticons by constructing emoticon dictionaries to match with the target emoticon, recognizing each emoticon as a sequence of characters/symbols or extracting character frequency from each emoticon as a feature. Their approaches are different from our approach as they cannot consider the similarity of the characters that are similar in shape but otherwise different from each other.

2. FEATURE EXTRACTION FROM EMOTICON

In this paper, we used three features: Local Binary Pattern(LBP), Histograms of Oriented Gradients(HOG) and GIST Descriptor to extract shape features of emoticons. These features are often used for image retrieval and object recognition. In the case of emotion, because there is no need to consider the difference of size between the emoticons and the gradients, our approach does not use Scale-Invariant Feature Transform(SIFT) or Random Sample Consensus(RANSAC) that are robust to scale or angle variation.

2.1. Conversion of Emoticon into Image

This subsection describes a process to convert an emoticon into an image. Emoticons are generally expressed in text form. That means if the kind of font used to display the emoticon is different, there will be some differences. This study uses the “sazanami-mincho” font that is often used in the Japanese language environment. To convert an emoticon into an image, we used an image processing application: ImageMagick[1]. Then, we created an image file (jpeg) by setting the background color as white, the font color as black and the sizes of all emoticons as the same. We extracted three kinds of features from this image file: LBP, GIST and HOG.

2.2. Extraction of Image Features

Fig. 1 shows a flow of extracting image features from an emoticon then registering them into a database with annotation of facial expression labels.

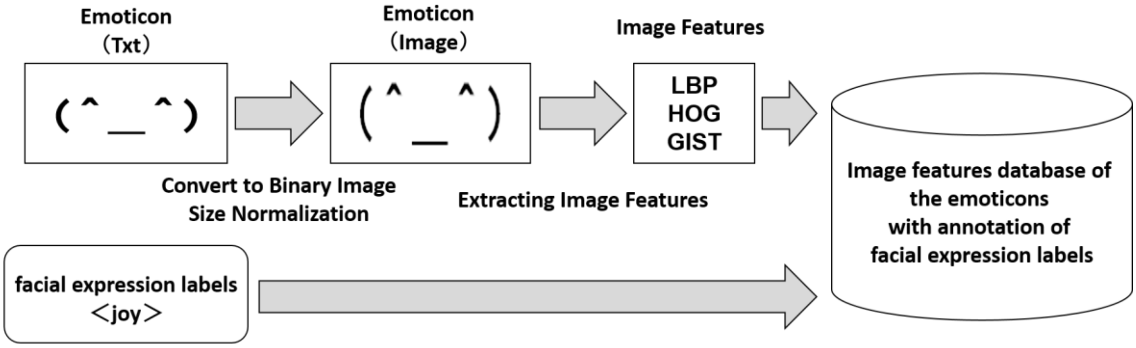

Fig. 1 The flow of construct the image features database of emoticons.

[1] http://imagemagick.org/script/index.php

The image features used in this paper are described below.

2.2.1 Local Binary Pattern(LBP)

Local Binary Pattern (LBP) was proposed by Ojala et al.(1996). The method can extract local features of an image and pattern-based features used for texture analysis. Because the feature is less subject to illumination change and takes low calculation cost, it is sometime used in the field of similar image retrieval.

2.2.2 GIST

Gist, which was proposed by Torralba et al.(2001), is one of the global image features obtained from the whole image. It has been reported that the method using Gist feature can achieve high performance in a task of detecting a copied image. Gist divides the whole image into 4x4 areas and describes scene information by applying the Gabor filter of various directions or frequencies for each area. It is also often used in a task of recognizing a scene.

2.2.3 Histograms of Oriented Gradients (HOG)

The Histograms of Oriented Gradients (HOG) feature, which was proposed by Dalal et al.(2005), is used for object recognition as a gradient-based feature. The method creates a histogram by calculating the gradient direction of brightness over a certain area. Although the HOG feature is not invariant to rotation or scale change, is can detect a person with high accuracy.

2.3. Baseline Feature

We used the character frequency vector and character string itself (character-based feature) as the baseline features. The character frequency vector is a feature that divides an emoticon into characters and then counts the frequency of each character. The similarity among character frequency vectors is calculated by cosine similarity. When the character-based feature is used, an emoticon is treated as a character string and the Levenshtein distance (LD) is calculated.

3. FACIAL EXPRESSION CLASSIFICATION METHOD

Many existing studies related to emoticons classify facial expressions based on six basic expressions proposed by Ekman et al.(1971). Our study aims to classify eight facial expressions including these six basic facial expressions (joy, hate, anger, sorrow, fear (anxiety) and surprise) in addition to 'shame' and 'none' We extracted features from emoticons and created a facial expression classification model by a machine learning method.

The proposed method uses k-Nearest Neighbor Method to check which emoticon is most similar in shape to another emoticon by observing the similarity degree between the extracted features. Then, we propose a method to extract image features from each character that forms an emoticon, not from the total characters of the emoticon itself. The similarities of each part or part order can be considered by this method.

The method calculates the Levenshtein distance between the sequences that are sorted from the image features extracted from each character in an emoticon. Each feature is extracted from the inputted emoticon, then the similarities (cosine similarity, Levenshtein distance) and the features of the emoticons in the training data are compared. Finally, the outputted emoticons are sorted according to the order of similarity and the top *k* emoticons are obtained. The facial expressions ranked on the top *k* emoticons are counted and outputted as the facial expressions of the inputted emoticon.

4. EXPERIMENT

To compare the performance of the proposed method and that of the baseline method, we conducted the evaluation experiment. The following subsection describes the experimental condition. Then, we introduce the experimental results.

4.1. Experimental Condition

4.1.1 Dataset

The dataset used in the experiment consists of the emoticons expressing facial expression that are selected from Kaomoji café(2016), Kaomoji.ru(2012), Kaomoji Copipe(2016), and FACEMARK PARTY(1999).
The dataset are normalized after being removed as parts of the facial contour. The detail of the dataset is shown in Table 1.

Table 1 Experimental dataset

Facial expression	joy	fear	surprise	angry	sorrow	none	hate	shame
Number	431	395	297	189	158	155	138	127

4.1.2 Evaluation Index

We compared the performances of the proposed method and the baseline method by using several evaluation indexes. If the class outputted by the classifier was the same as the correct class, the result was regarded as correct. Then recall rate, precision rate, and F1-score were calculated for each facial expression. We compared, analyzed and discussed the result by using these numerical values as the evaluation indexes. **Error! Reference source not found.** shows the combinations of features and similarities. We used the dataset for the 10-hold cross-validation experiment.

Table 2 Combinations of features and similarities

Method	Feature	Target	Similarity
Baseline	Character frequency	Overall of emoticon	Cosine
	Character string		LD
Proposed method	LBP		Cosine
	HOG		
	GIST		
	LBP	Character basis	LD
	HOG		
	GIST		

4.1.3 Evaluation Experiment

Experimental programs were constructed by Python2.7. To extract HOG, LBP, we used library "scikit-image[2]". To extract GIST, we used lear-gist-python[3].

[2] scikit-image(http://scikit-image.org/docs/dev/api/skimage.html)
[3] lear-gist-python(https://github.com/tuttieee/lear-gist-python/tree/master/sample)

4.2 Experimental Result

In the case that k equal 1, the F1-score of the proposed method and the baseline method are shown in Table 3 and Table 4. The proposed method using the HOG feature obtained high F1-scores, on average, without large differences depending on the facial expressions. Although the LBP feature was calculated with low complexity, it could not classify enough because it was too simple. Because GIST is a global feature, it might have been difficult to distinguish emoticons.

On the other hand, F1-scores by the baseline method differed greatly depending on the facial expressions in both the using character vector and the character string as feature and decreased as *k* increased. The baseline method might have been significantly affected by the distribution of facial expressions of the emoticons.

Table 3 F1-scores of experimental result which target is overall of emoticons(k=1)

	joy	fear	surprise	angry	sorrow	none	hate	shame
LBP	0.165	0.336	0.239	0.193	0.205	0.313	0.183	0.228
HOG	0.368	0.343	0.504	0.381	0.569	0.464	0.274	0.315
GIST	0.310	0.217	0.469	0.361	0.602	0.340	0.225	0.318
Cosine	0.605	0.652	0.488	0.306	0.652	0.400	0.344	0.322
LD	0.614	0.526	0.588	0.515	0.615	0.414	0.295	0.281

Table 4 F1-scores of experimental result which character basis(k=1)

	joy	fear	surprise	angry	sorrow	none	hate	shame
LBP	0.421	0.406	0.426	0.257	0.399	0.249	0.196	0.272
HOG	0.631	0.504	0.621	0.504	0.571	0.444	0.288	0.320
GIST	0.482	0.477	0.532	0.414	0.489	0.287	0.279	0.303

5. DISCUSSIONS

To validate the effectiveness and the problem of the proposed method, the success and failure examples of the facial expression classification are analyzed by showing the obtained similarities between the emoticons. Table 5 shows the top three similar emoticons obtained when we used the HOG feature by calculating the Levenshtein distance on the character basis. The method outputted “anger” as the emoticon most similar to the emoticon of facial expression of “anger”.

Table 5 Success example(HOG + LD)

Input emoticon: (＃▽皿▽)= ３		Facial expression: anger	
Rank	Emoticon	Facial expression	LD
1	(←▼皿▼#)	anger	3.502
2	(;-ω-)=3	hate	3.676
3	(;・∀・)= з	anxiety	4.038

Although the emoticons expressing different facial expressions had been outputted on the second or lower rank, the most similar emoticon had been obtained correctly. It is considered that the features obtained from “▽” and “▼ became similar to each other by using the image feature. The success also might have been caused because a two-byte character “＃” and a single-byte character “\#” had been able to be recognized as similar characters.

6. CONCLUSIONS

This paper proposed a method to classify facial expressions focusing on shape features expressed by the characters forming the emoticons. The proposed method extracts shape image features by converting an emoticon into an image and classifies facial expressions based on similarity between the feature vectors by using the k-Nearest Neighbor Method. The results of the evaluation experiment showed that the proposed method can classify facial expressions more effectively than the baseline method based on the character features.

In the future, we would like to focus on the differences between the symbols that cannot be recognized by image features, and improve the accuracy of the method to classify facial expressions by implementing the algorithm to clarify such differences by providing clustering processing on the characters in advance

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Numbers 15K00425, 15K00309, and 15K16077. And also, this work is partially supported by Tateishi Science and Technology Foundation.

REFERENCES

Suzuki, N, and Tsuda, K(2006). Express emoticons choice method for smooth communication of e-Business, KES 2006, Part II, LNAI 4252, (pp. 296-302).

Urabe, Y, Rzepka, R, and Araki, K(2013). Emoticon recommendation system for effective communication, Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, (pp.1460-1461).

Yamada, T., Tsuchiya, S., Kuroiwa, S. and Ren, F(2007). classification of facemarks using N-gram, 2007 Int. Conf. on Natural Language Processing and Knowledge Engineering, (pp.322-327).

Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R. and Araki, K(2010). CAO: A Fully Automatic Emoticon Analysis System. Proceedings of Twenty-Fourth AAAI Conference on Artificial Intelligence. (pp.1026-1032)

Okumura, N(2016). An Estimation of the Original Form of Kaomoji using Classifier, SIG Technical Reports of Natural Language Processing, Vol. 17, (pp.1-4), (in Japanese)

Ojala. T, Pietikainen. M, and Harwood. D, A(1996). Comparative Study of Texture Measures with Classification Based on Feature Distributions, Pattern Recognition, Vol.19, Issue.1,(pp.51-59).

Olivia, A., and Torralba, A(2001), Modeling the shape of the scene: a holistic representation of the spatial envelope. International Journal of Computer Vision, Vol. 42(3): (pp.145-175).

Dalal. N, and Triggs. B(2005). Histograms of oriented gradients for human detection, IEEE Conference on Computer Vision and pattern Recognition (CVPR), (pp.886-893).

Ekman, P. and Fresen, W, V(1971). Constants across cultures in the face and emotion. Journal of personality and social psychology, Vol.17(2), (pp.124-129).

Kaomoji Café(2016), Accessed via http://kaomoji-cafe.jp/ on 9 Nov 2017.

Kaomoji.ru(2012), Accessed via http://kaomoji.ru/en/ on 9 Nov 2017.

Kaomoji Copipe(2016), Accessed via http://kaomoji-copy.com/ on 9 Nov 2017.

FACEMARK PARTY(1999), Accessed via http://www.facemark.jp/facemark.htm on 9 Nov 2017.