Preservation of emotional context in tweet embeddings on social networking sites

Maruyama, Osamu
Faculty of Design, Kyushu University

Yoshinaga, Asato
Graduate School of Design, Kyushu University

Sawai, Ken-ichi
Faculty of Design, Kyushu University

出版情報：Artificial Life and Robotics. 29 (4), pp.486-493, 2024-10-08. Springer
バージョン：

Artificial Life and Robotics
https://doi.org/10.1007/s10015-024-00974-3

ORIGINAL ARTICLE

Preservation of emotional context in tweet embeddings on social networking sites

Osamu Maruyama[1] · Asato Yoshinaga[2] · Ken-ichi Sawai[1]

Received: 1 April 2024 / Accepted: 14 September 2024

Abstract
In communication, emotional information is crucial, yet its preservation in tweet embeddings remains a challenge. This study aims to address this gap by exploring three distinct methods for generating embedding vectors of tweets: word2vec models, pre-trained BERT models, and fine-tuned BERT models. We conducted an analysis to assess the degree to which emotional information is conserved in the resulting embedding vectors. Our findings indicate that the fine-tuned BERT model exhibits a higher level of preservation of emotional information compared to other methods. These results underscore the importance of utilizing advanced natural language processing techniques for preserving emotional context in text data, with potential implications for enhancing sentiment analysis and understanding human communication in social media contexts.

Keywords Emotion · Intensity · Tweet · Embedding vector · BERT · Word2vec

1 Introduction

The representation of emotional types like "joy" and "fear" in our communication is undeniably significant. There are many studies on the conceptualization and analysis of emotion. For instance, Plutchik proposed a model that identifies eight primary emotions (anger, disgust, fear, joy, sadness, surprise, trust, and anticipation) [13], while Ekman initially proposed six basic emotions (anger, disgust, fear, joy, sadness, and surprise) and later expanded the list to include 11 more (amusement, contempt, contentment, embarrassment, excitement, guilt, pride in achievement, relief, satisfaction, sensory pleasure, and shame) [5].

The WRIME (version 1) dataset [6] consists of 43,200 tweets written by 80 individuals, each annotated with 4-point intensity ratings for Plutchik's eight primary emotions by the author and three anonymous annotators. These intensity data were all obtained through a crowdsourcing service. Thus, analyzing the conservation of emotional information in the tweets by creating embedding vectors is an intriguing prospect.

The advent of deep learning technique can provide us with a learned mapping from input to output features, given sufficiently large training datasets. Deep learning has succeeded first in image recognition, and second in the field of natural language processing (NLP) during this third AI boom [9]. In the current field of NLP, words, sentences, paragraphs, and documents are vectorized into distributed representations, also known as embedding vectors [8]. These vectors are used in many NLP tasks like machine translation [1], question answering [16], and text classification [7]. In this study, we considered three different methods to generate embedding vectors of tweets in social networking services (SNS) and analyzed the extent to which emotional information is preserved in the resulting embedding vectors.

These methods were designed based on the models of word2vec [11], and pre-trained and fine-tuned Bidirectional Encoder Representations from Transformers (BERT)

This work was presented in part at the joint symposium of the 29th International Symposium on Artificial Life and Robotics, the 9th International Symposium on BioComplexity, and the 7th International Symposium on Swarm Behavior and Bio-Inspired Robotics (Beppu, Oita and Online, January 24–26, 2024).

✉ Osamu Maruyama
maruyama.osamu.158@m.kyushu-u.ac.jp

Asato Yoshinaga
yoshinaga.asato.294@s.kyushu-u.ac.jp

Ken-ichi Sawai
301ken1@gmail.com

[1] Faculty of Design, Kyushu University, Fukuoka, Japan

[2] Graduate School of Design, Kyushu University, Fukuoka, Japan

Published online: 08 October 2024

[4]. Word2vec is a method for finding embedding vectors of words within given sentences. The mean vector of the word2vec-embedding vectors corresponding to the words in a tweet is formulated as the embedding vector for the tweet. BERT is a ubiquitous baseline model in NLP tasks [14]. We used a pre-trained Japanese BERT model [2], as well as BERT models fine-tuned with eight emotional intensities. The CLS token of the last layer of the BERT models, which is the aggregate representation of the words of the input tweet, is used as the embedding vector of the input tweet.

To get insights into how much emotional information is preserved in these embedding vectors of tweets, we clustered the tweets based on the resulting embedding vectors for each method. We then evaluated the degree of overlap between the resulting clusters and those generated from the eight primary emotional intensities. Our findings indicate that the embedding vectors generated by the fine-tuned BERT model preserve emotional information well. This result is reasonable given that the eight emotional intensities are used in the fine-tuning process. Additionally, we observed that the embedding vectors generated by the word2vec method and a pre-trained BERT model also retain emotional information to some extent, despite not being optimized for the eight emotional intensities.

2 Materials and methods

2.1 Dataset

WRIME (version 1) is a database of 43,200 tweets of 80 writers [6]. Each tweet is characterized by a 4-point intensity scale (0: none, 1: weak, 2: medium, 3: strong) for the eight primary emotions (anger, disgust, fear, joy, sadness, surprise, trust, and anticipation), independently assigned by the author and three anonymous annotators. These eight emotional categories were proposed by Plutchik [13]. All intensity data were obtained through a crowdsourcing service. In this study, we removed the tweets with weak intensities (0 or 1) for all eight emotions. As a result, 18,237 tweets were used for the analysis.

One of the advantages of this dataset is its size. Even with the 18,237 tweets, it surpasses the size of the second-largest dataset, SemEval-2018 [12], which contains 12,634 tweets (see Table 1 in [6]).

2.2 Methods

We considered three methods to generate embedding vectors of tweets, as shown in Fig. 1. Note that all tweets are segmented into morphemes (tokens) using the morphological analysis tool MeCab [17]. The first generator is a word2vec model [11], the second is a Japanese pre-trained BERT model [2], and the third is the BERT model fine-tuned with the eight emotional intensities. Details are described in the following subsections.

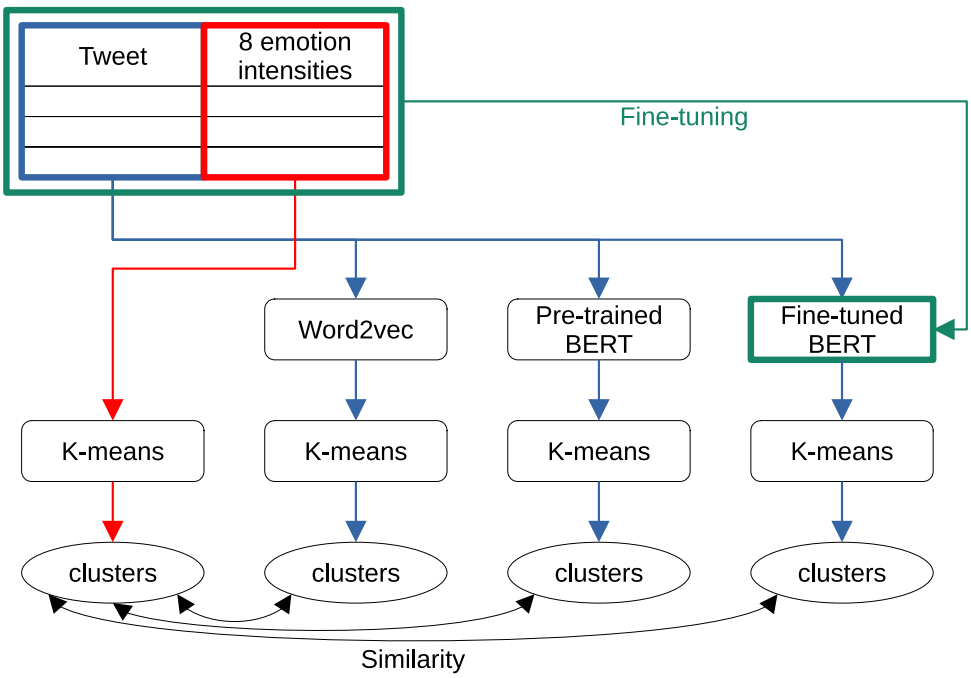


Fig. 1 Overview of our computational experiments. The tweets are vectorized in different ways, **a** word2vec model, **b** pre-trained BERT model, and **c** fine-tuned BERT model. The pre-trained BERT model is the BERT base Japanese (IPA dictionary, whole word masking enabled) [2]. For each method, clusters of tweets based on the resulting embedding vectors are compared with the cluster of tweets determined with the eight primary emotional intensities

2.2.1 Embedding vectors by word2vec

The word2vec method is used to find embedding vectors of words appearing in given sentences [11]. A key feature of the method is that it captures the conceptual relationships among words and represents them as the embedding vectors. We generated the embedding vectors for words in the tweets and defined the mean of these vectors, representing the words appearing in a tweet, as the embedding vector for the tweet. Note that the dimension of the embedding vectors of words is set to be 768.

2.2.2 Embedding vectors by pre-trained BERT

The Japanese pre-trained BERT model we used in this study is the BERT base Japanese (IPA dictionary, whole word masking enabled) [2].

The CLS token of the last hidden layer of the model for an input tweet was extracted and used as the embedding vector of the input tweet, whose dimension is also 768.

2.2.3 Embedding vectors by BERT fine-tuned with eight emotional intensities

We furthermore fine-tuned the pre-trained BERT model by learning the mapping from the tweet sentences to the eight-emotion intensities assigned by three anonymous readers in the WRIME dataset. We conducted the fine-tuning process using 2-fold cross-validation, achieving accuracies of 0.74 and 0.73 on the two test datasets. After fine-tuning, the CLS token from the last hidden layer of the model, when using a tweet as a test instance, was assigned as the tweet's embedding vector.

3 Results

3.1 Clustering tweets by eight emotional intensities

First, using the WRIME dataset, we classified the tweets based on the eight emotional intensities assigned by three anonymous annotators, using K-means clustering. We used the mean of these intensities and did not use the writer's intensities, because the writer's evaluation is more subjective than the annotators' evaluation and depends trivially on the writer's personality and context when the tweet was posted. In fact, the objective (annotators') intensities are reported to be more predictable than the subjective intensities from tweeted sentences using a Japanese BERT model [6].

To obtain the appropriate number of clusters, K, we calculated silhouette coefficient [15] and Davies–Bouldin Index (DBI) [3] in the range from $K = 2$–40. The best K can be larger than eight, because many tweets are characterized by combinations of the eight primary emotions. Unexpectedly, at $K = 6$, the silhouette clustering coefficient had a local peak and Davies–Bouldin Index had the global minimum (data not shown). This implies that the emotional patterns of tweets were not so varied.

The six clusters were characterized by the eight emotions in Fig. 2. This shows the proportion of each of the eight emotions in the six clusters. For example, cluster C_1 consists of tweets with high intensities of "joy," and with relatively low intensities of "anticipation," "surprise," and "trust." These three emotions might be causes of "joy," like by trusting something they then felt "joy." For example, the next tweet has intensity three in "joy" and "anticipation" simultaneously (The original tweet is in Japanese and translated into English by the authors):

> "I thought seriously about changing to an LCC ticket and acted immediately. I couldn't finish it today because of the holder name, but I want to go back tomorrow and will change it! As the LCC is currently running a campaign, I was lucky to get a lot of freebies, thanks to talking with a salesclerk."

The high-intensity emotion and the low-intensity emotions in the remaining clusters are as follows: C_2 sadness (low-intensity emotions: surprise, fear, and disgust); C_3 anticipation (joy); C_4 surprise (joy); C_5 fear (sadness, surprise, and disgust); C_6 disgust (sadness, surprise, anger, and fear). In these clusters, the relationships of high and low-intensity emotions seem reasonable. Figure 3 shows the size of each cluster.

Figure 4 is the UMAP [10] output for the eight emotional intensities where UMAP stands for Uniform Manifold Approximation and Projection and it is a dimensionality reduction technique commonly used in data visualization and analysis. Here, the number of neighbors, a parameter of UMAP, is 10. Each cluster in the figure is represented with

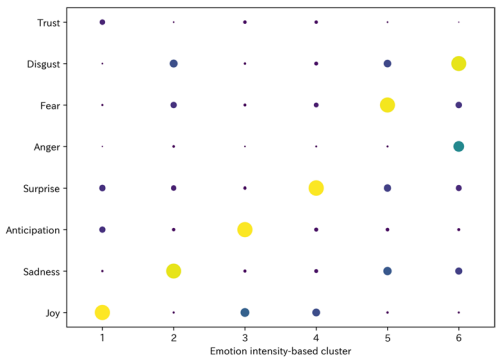


Fig. 2 Graphical representation of the proportions of the eight emotions in the six clusters. Circle size is proportional to the sum of the intensities of each emotion within each cluster

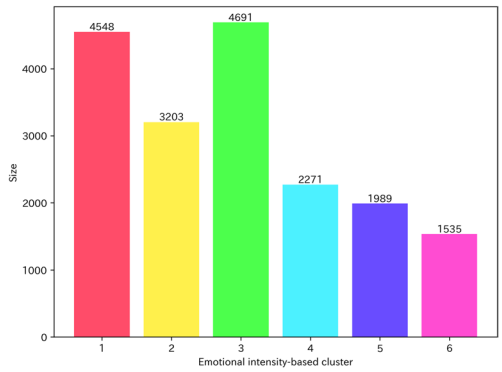


Fig. 3 Size distribution of clusters generated from the eight emotional intensities

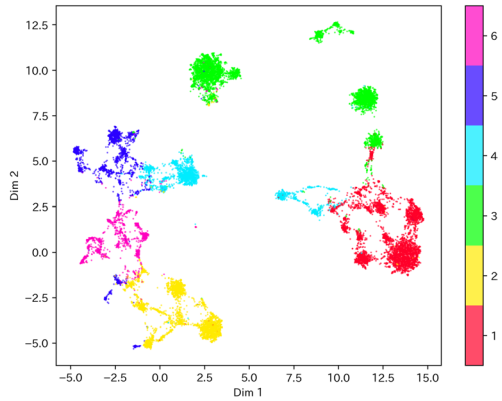


Fig. 4 The UMAP output for the eight emotional intensities, using 10 neighbors. Each cluster is represented by a different color

a distinct color and it can be seen that the clusters are well separated from each other.

3.2 Embedding the embedding vectors of tweets into two-dimensional space by UMAP

Figure 5 shows the two-dimensional UMAP embedding of the embedding vectors of tweets generated using (a) word-2vec model, (b) pre-trained BERT model, and (c) fine-tuned BERT model. The left plots are labeled with the cluster IDs determined by the K-means clustering for $K = 6$ with the generated embedding vectors of the tweets, and the right plots are labeled with the cluster IDs based on the eight emotional intensities. As seen, the UMAPs of the embedding vectors generated by the word2vec and pre-trained BERT models do not conserve the emotional information. On the other hand, the UMAP of the embedding vectors generated by the fine-tuned BERT model conserves the emotional information.

3.3 Generating embedding vectors of tweets

We evaluated how much the emotional information of tweets is conserved in the embedding vectors of tweets. As previously mentioned, we considered three schemes to generate embedding vectors of tweets; each of which are based on (i) the word2vec method, (ii) a pre-trained BERT model, and (iii) a BERT model fine-tuned with pairs of tweets and the corresponding eight primary emotional intensities of the tweets. We clustered the tweets based on these embedding vectors and compared how much the resulting clusters overlap with the clusters derived from the eight primary emotional intensities.

For comparison purposes, we denote the four collections of clusters by $I_1, \ldots, I_6$ (intensity-based clusters), $W_1, \ldots, W_6$ (word2vec-based clusters), $P_1, \ldots, P_6$ (pre-trained model-based clusters), and $F_1, \ldots, F_6$ (fine-tuned model-based clusters).

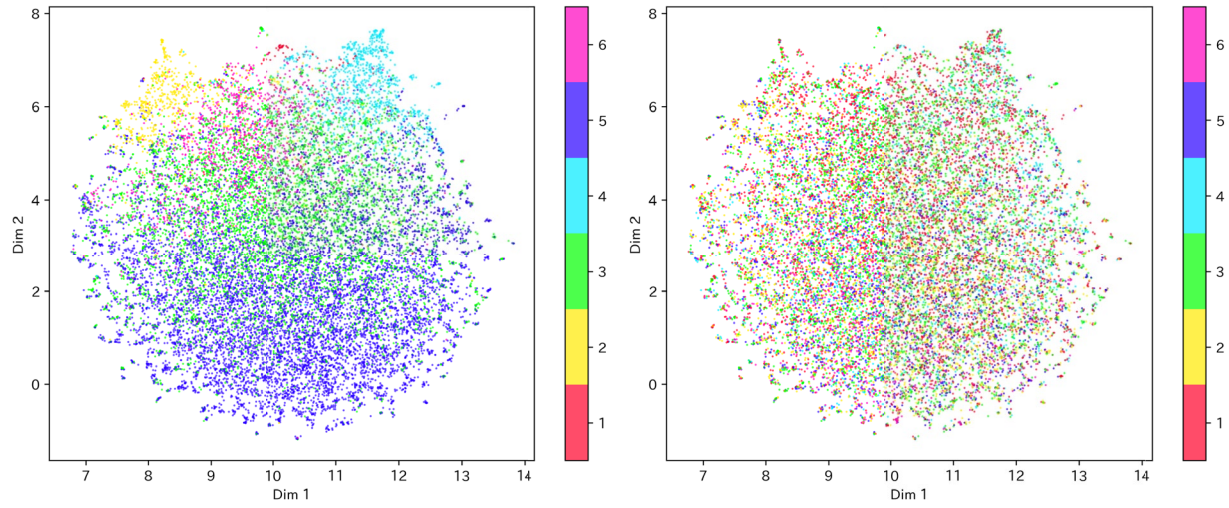
Figure 6a shows the overlap ratio between I_1 to I_6 and W_1 to W_6, normalized by column, i.e., divided by the sum of W_j. Note that these symbols, W_1 to W_6, are heuristically sorted to maximize the sum of the diagonal elements. For I_i ($i = 1$–6), the highest overlap ratio occurs in the diagonal element except for I_6. This means that for most intensity-based clusters, the most similar clusters among the word2vec-based clusters are identifiable. This implies that the emotional intensity information is preserved in the embedding vectors generated by word2vec to some extent. However, comparing the overlap ratios with each other in each W_j (i.e., column), W_1, W_2, W_4, and W_6 have the highest ratios with I_1. In addition, among 18,237 tweets, 48% and 33% are assigned to W_5 and W_3, respectively (see Fig. 6a right graph).

Figure 6b shows the overlap ratio between I_1 to I_6 and P_1 to P_6, normalized by the sum of P_j. This heat map looks fairly close to that of the word2vec approach. However, the cluster size distribution, different from that of the word2vec approach, looks more balanced.

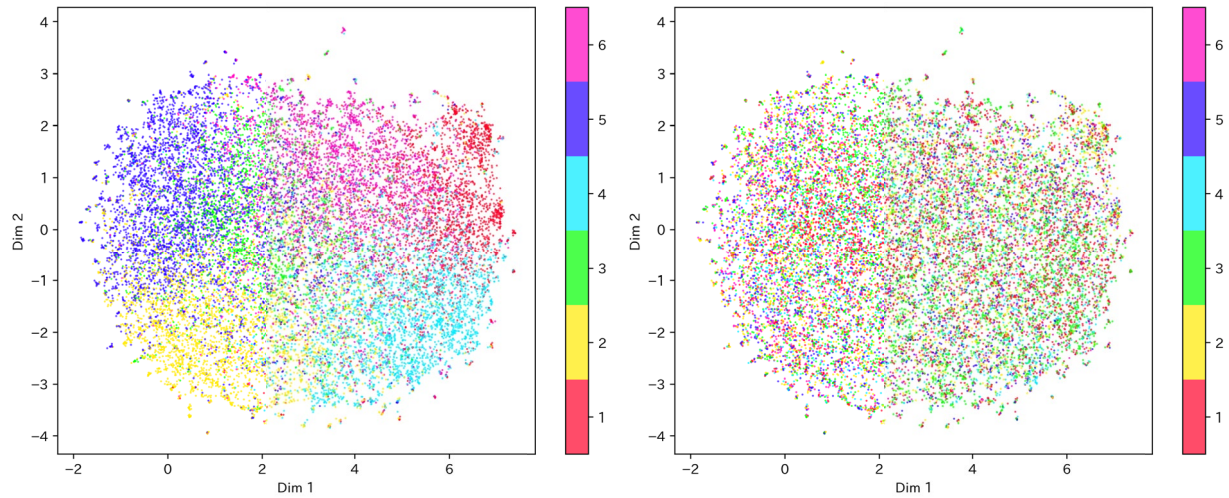
The third heat map between I_1 to I_6 and F_1 to F_6 is shown in Fig. 6c. F_i is largely overlapped with I_i for $i =$ 1–6. This heat map has higher diagonal elements than the two previous heat maps. The difference of the fine-tuned BERT approach from the first two approaches is that the fine-tuned model is optimized with the pairs of tweets and their eight emotional intensities. This means that without such fine-tuning process, the resulting embedding vectors weakly conserve the emotional information.

4 Discussion

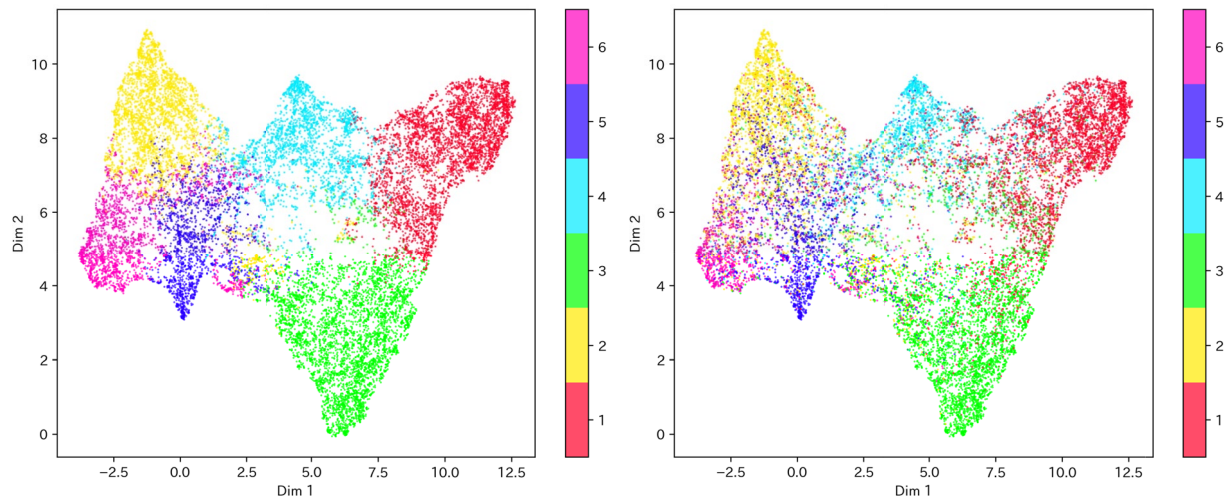
To demonstrate the applicability of the generated embedding vectors of tweets, we present the following example. First, we selected two tweets: one with intensity three in "joy" and



(a) Word2vec

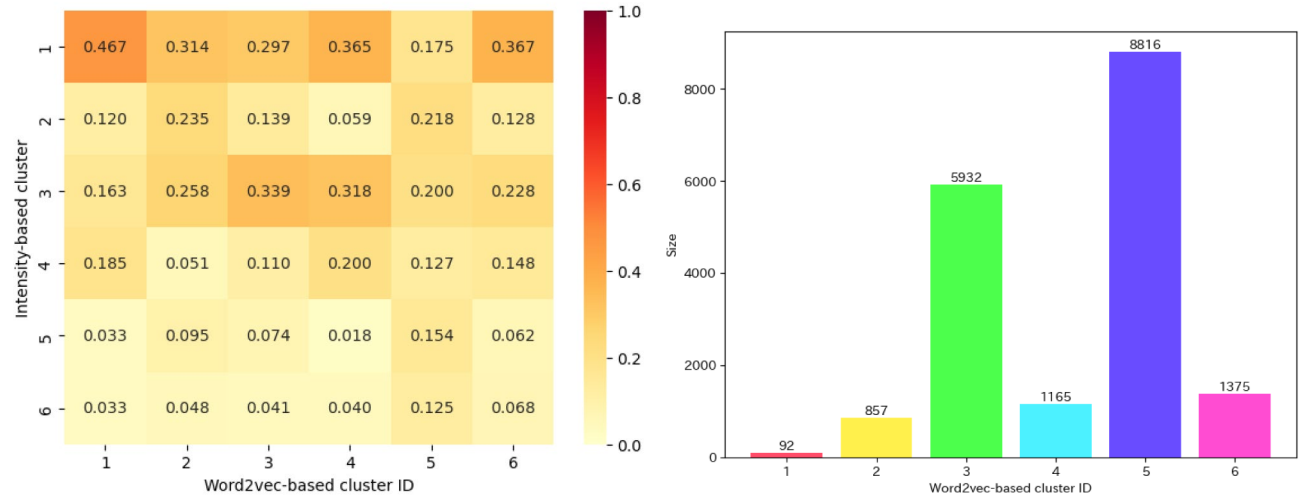


(b) Pre-trained BERT

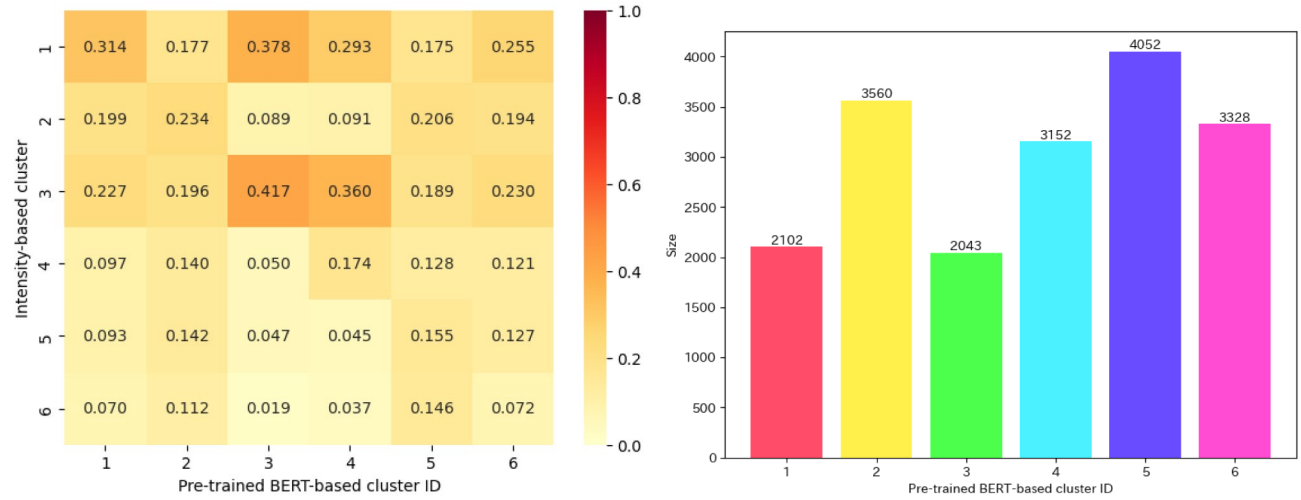


(c) Fine-tuned BERT

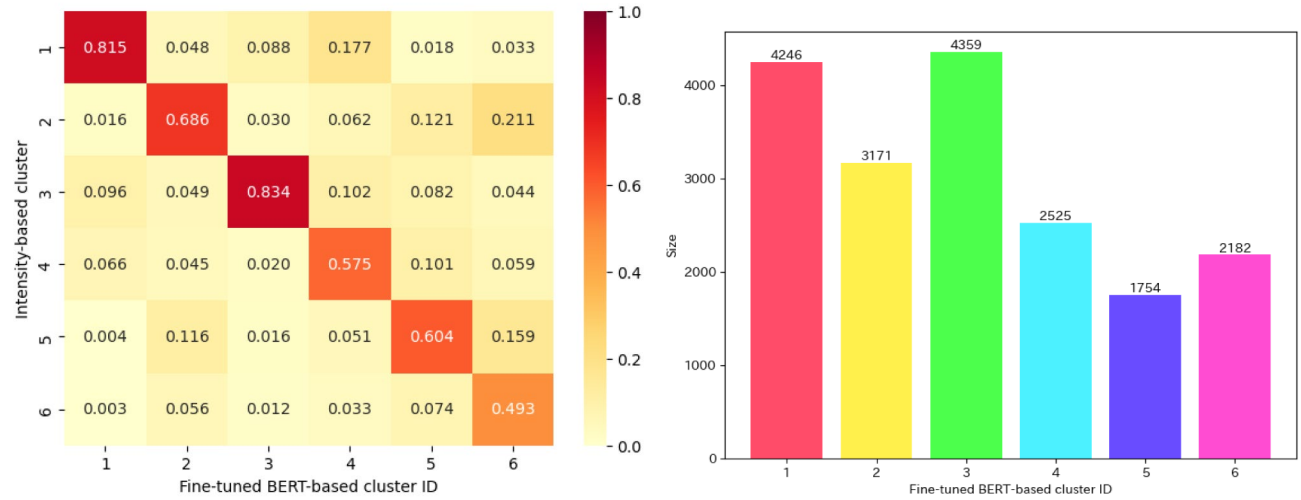
Fig. 5 UMAP for the embedding vectors of tweets generated by **a** word2vec, **b** pre-trained BERT, and **c** fine-tuned BERT. The left column plots are labeled with cluster IDs based on the generated embedding vectors, while the right column plots are labeled with cluster IDs based on the eight emotional intensities



(a) Word2vec



(b) Pre-trained model



(c) Fine-tuned model

Fig. 6 Confusion matrix between clusters based on eight emotional intensities and clusters based on embedding vectors generated by each method (left), and the sizes of clusters based on embedding vectors of tweets generated by **a** word2vec, **b**pre-trained BERT, and **c** fine-tuned BERT (right)

zero in the remaining emotions, and the other with intensity three in "sadness" and zero in the remaining emotions. Subsequently, we identified the tweet whose embedding vector exhibited the closest cosine similarity to the mean of the embedding vectors of the two given tweets. The embedding vectors generated by the fine-tuned BERT model were used in this application. Following are the first and second tweets:

> I'm finally able to start saving money, and I'm so happy... Even though saving money is a hobby of mine, I haven't been able to save anything for nearly the past ten years. Ah, I'm so happy... Hehe...
>
> My wedding ring flowed down the bathtub drain (crying).

Next is the tweet with the highest cosine similarity to the mean of the embedding vectors of the above two tweets:

> Antlers Congratulations, I'm so happy! I can't come home tonight. It's so sad...

Note that "Antlers" is the name of a professional Japanese soccer team. Indeed, it was confirmed that this tweet successfully expresses both emotions of joy and sadness. Specifically, the eight emotional intensities assigned to this identified tweet are 2 in joy, 2 in sadness, and 0 in the others.

5 Conclusion

The 43,200 tweets of 80 writers in the WRIME database are annotated with the intensities of eight primary emotions. In this study, first, we classified the tweets into six clusters based on the eight emotional intensities. This suggests that the major patterns of emotions when we tweet are not highly diverse. Second, we considered three different schemes to generate embedding vectors of tweets, based on word2vec, pre-trained BERT, and fine-tuned BERT. We clustered the tweets based on the resulting embedding vectors and compared them with the clusters based on the eight emotional intensities. We found that the embedding vectors generated by the fine-tuned BERT model effectively preserve the emotional information. This result is reasonable, because the eight emotional intensities are used in the fine-tuning process. Additionally, we observed that the embedding vectors generated by word2vec and pre-trained BERT models also preserve the emotional information to some extent, despite being not optimized the eight emotional intensities.

Availability of scripts The scripts used in this study are available at https://github.com/maruyama-lab-design/Preservation-of-Emotional-Context-in-Tweet-Embeddings. They are implemented in Python 3.11.3 and PyTorch 2.0.1.

References

1. Bahdanau D, Cho K, Bengio Y (2014). Neural machine translation by jointly learning to align and translate. https://doi.org/10.1016/j.jneumeth.2010.04.011
2. cl-tohoku (2020) Bert base Japanese (IPA dictionary, whole word masking enabled). https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking
3. Davies DL, Bouldin DW (1979) A cluster separation measure. IEEE Trans Pattern Anal Mach Intell PAMI–1(2):224–227. https://doi.org/10.1109/TPAMI.1979.4766909
4. Devlin J, Chang MW, Lee K, Toutanova K (2018) BERT: pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805v2
5. Ekman P (1999) Basic emotions. Chap. 3. Wiley, Hoboken, pp 45–60
6. Kajiwara T, Chu C, Takemura N, Nakashima Y, Nagahara H (2021) WRIME: a new dataset for emotional intensity estimation with subjective and objective annotations. In: Proceedings of the 2021 conference of the North American chapter of the Association for computational linguistics: human language technologies. Association for Computational Linguistics, Online. pp 2095–2104. https://doi.org/10.18653/v1/2021.naacl-main.169
7. Kim Y (2014). Convolutional neural networks for sentence classification. https://doi.org/10.3115/v1/D14-1181
8. Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: Proceedings of the 31st international conference on machine learning, 32. pp 1188–1196. PMLR
9. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
10. McInnes L, Healy J, Saul N, Großberger L (2018) UMAP: uniform manifold approximation and projection. J Open Source Softw 3(29):861
11. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. https://arxiv.org/abs/1301.3781
12. Mohammad S, Bravo-Marquez F, Salameh M, Kiritchenko S (2018) Semeval-2018 task 1: affect in tweets. In: Proceedings of the 12th international workshop on semantic evaluation, pp 1-17
13. Plutchik R (1980) Chap. 1. A general psychoevolutionary theory of emotion. In: Plutchik R, Kellerman H (eds) Emotion: theory, research and experience. Theories of emotion, vol I. Academic Press, New York, pp 3–33
14. Rogers A, Kovaleva O, Rumshisky A (2020) A primer in BERTology: what we know about how BERT works. Trans Assoc Comput Linguist 8:842–866
15. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Comput Appl Math 20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7

16. Seo M, Kembhavi A, Farhadi A, Hajishirzi H (2016). Bidirectional attention flow for machine comprehension. https://doi.org/10.18653/v1/N16-1014
17. Taku Kudo: MeCab: yet another part-of-speech and morphological analyzer. https://taku910.github.io/mecab/