JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 38, 605-618 (2022)
DOI: 10.6688/JISE.202205_38(3).0007

# Lyrics Retrieval for Tourist Attractions Based on Shared Word-Embedding Vectors*

YIHONG HAN[1], RYOSUKE YAMANISHI[3] AND YOKO NISHIHARA[2]
[1]*Department of Information Science and Engineering*
[2]*College of Information Science and Engineering*
*Ritsumeikan University*
*Kusatsu, 525-8577 Japan*
*E-mail: is0387ps@ed.ritsumei.ac.jp[1]; nisihara@fc.ritsumei.ac.jp[2]*
[3]*Faculty of Informatics Kansai University*
*Takatsuki, 569-1095 Japan*
*E-mail: ryama@kansai-u.ac.jp*

This paper proposes a lyrics retrieval method based on word-embedding technology to perform music recommendations that matched a tourist attraction. In the proposed method, word importance is calculated by Term Frequency – Inverse Document Frequency (TFIDF) and Smooth Inverse Frequency (SIF). We built a vectorization model from the lyrics corpus using fastText with Continuous Bag of Words (CBOW) and applied this model to both the lyrics corpus and the tourist attraction reviews corpus to create word-embedding vectors for lyrics and tourist attraction reviews. And, the review vectors are integrated for each tourist attraction to generate tourist attraction vectors. Based on the similarity calculation between the tourist attraction vectors and the lyric vectors, the song with the most suitable lyrics for the tourist attraction comes to be the recommended result. Subjective evaluation experiments on the recommendation results of the proposed method was conducted. The results of the experiments showed that the subjects reacted positively to the lyrics of the recommended songs. However, their evaluations of the music of the recommended songs were not as positive as those accorded to the lyrics were. A joint study of acoustic information and text information is an approach to the issue for the future.

***Keywords:*** lyrics retrieval, context aware recommendation, cross domain retrieval, lyrics, reviews, word representation, word-embedding

## 1. INTRODUCTION

Over the past 20 years, new technologies and business models have changed the way we listen to music. Streaming services have taken the place of traditional CD albums and started a new era in the music business. Simultaneously, the tourism industry has also changed dramatically. The concept of "experiential tourism" has gained traction by providing tourists with special experiences. It has become a new challenge for the tourism industry to continually develop. Therefore, the role of music in tourism activities has attracted attention from both music and tourism industries. Of note, research on the effects of music on the deep psychology of human beings is in progress, often in tandem with the marketing and sales industries. For example, the background music in stores influences consumption tendencies and purchase intentions of customers [1] and the image of the stores [2]. Just as music influences consumer behavior, we believe that it is possible to use music to improve tourist experience. In this case, music that matches the tourist destination is required.

Received February 18, 2021; revised April 6, 2021; accepted May 18, 2021.
Communicated by Chia-Hui Chang.
* This work was supported in part by JSPS KAKENHI #16K21482.

This paper proposes a lyrics retrieval method based on word-embedding technology to make music recommendations that matches a tourist attraction. The proposed method performs more accurate lyrics retrieval by using the emotional information of the tourist attraction and the music. In this study, we use word-embedding to select lyrics similar to tourist attraction reviews. By applying a vectorization model built from music lyrics corpus to the tourist attraction reviews, the word-embedding information is shared between the lyrics side and the reviews side. Thus, the tourist attraction review is regarded as a kind of lyric (pseudo-lyrics). The text information from two different fields, lyrics and reviews, are integrated into one common vector space. The lyrics vectors and tourist attraction vectors become comparable in the common vector space. The song with a lyric vector that has a highest similarity to the given tourist attraction vector will be the lyrics retrieval result.

## 2. RELATED WORK

Music Information Retrieval (MIR) has been widely researched. Music retrieval with humming [4] and music genre classification [5, 6] are typical research topics in the MIR field.

Music with singing can be considered as multimedia art and comprises acoustics and linguistics. That is to say, the affection toward music can be caused by the combination of "listening to acoustics" and "understanding lyrics." However, most of the conventional music retrieval systems focus only on acoustic features [7], and few studies focus on the linguistic aspects of music (*i.e.*, lyrics). Let us focus on work related to lyrics, which is the target of this paper. Tsukuda [8] developed *Lyrics Jumper*, which recommends artists whose lyrics have similar topics. Cai [9] has put forward *MusicSense*, which recommends music while reading a document on the Web. In the work by Cai, the affective words extracted from both lyrics and documents on the Web are used to relate the two types of domains with each other. Our proposed method does not focus on some specific words, but the overall similarity between lyrics and reviews by using the word distributed representation model.

Music and location have been related to each other in several existing research works, such as Kaminskas who proposed a location-aware music recommendation system while using a tag-based approach [10] and a knowledge-based approach [11]. In their tag-based approach [10], music and Point of Interest (POI) are related to each other based on the tag given. Their knowledge-based approach [11] constructs a graph that semantically relates music to POI based on the knowledge with DBPedia and ranks the songs for a given POI. Moreover, a hybrid approach of a tag-based and knowledge-based approach has been proposed [12]. Our proposed method is considered as a lyrics-based approach in the research contexts above.

On the other hand, in the field of language processing, with the spread of the Internet, research on languages written freely by ordinary users has been gaining momentum since circa 2010 [13, 14]. This type of research has been attracting increasing attention in recent years [15, 16]. However, there are few studies focus on documents which are considered as a part of art or culture, such as novels and lyrics, exist. This study focuses on the linguistic aspect of music and proposes an interaction model that uses human activity as input in lyrics retrieval.

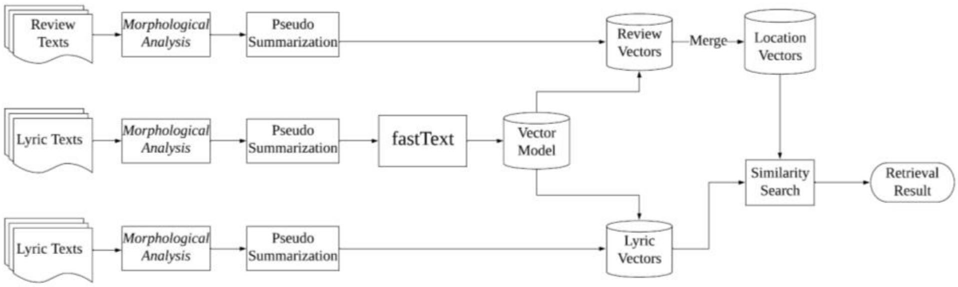


Fig. 1. The flowchart of the proposed method.

## 3. PROPOSED METHOD

Fig. 1 shows the outline of our method. The procedures of the proposed method are as follows. First, two pre-processed corpora of music lyrics and tourist attraction reviews are prepared and a morphological analysis is performed on both corpora due to the special features of the Japanese language. To make the individuality and characteristics in lyrics and reviews more prominent, the method performs a pseudo-summarization to the lyrics and reviews corpus. The summarized texts of the lyrics is used as training material to build a vectorization model. The word-embedding vectors of each lyric and review are generated with the vectorization model generated from lyrics corpus. **By applying the model trained from lyrics corpus to the reviews, the tourist attraction reviews are also regarded as a type of lyric ("pseudo-lyrics"; as it appears below). After converting reviews to pseudo-lyrics, the comparison between lyric vectors and review vectors becomes meaningful. This idea was the key concept of our method and the core part of "cross-domain retrieval."** Then, the method integrated those review vectors which is related to the same tourist attraction and we call the integrated vectors as "location vector." The location vectors are considered to be a quantification of the characteristics of a tourist attraction. Finally, vector similarity is measured between the location vectors and lyric vectors, and the lyric vector with the highest similarity to the given location vector is retrieved. The music information of that lyric is output as the recommendation result.

### 3.1 Text Pre-Processing and Morphological Analysis

The lyrics corpus was fetched from a lyrics site "uta-net"[1] and the tourist attraction corpus was retrieved from a tourism site "jalan."[2] Both corpora are in Japanese. Then, a morphological analysis is performed on both corpora. Unlike the case with English and other European languages, a word always follows the previous word without any space in Japanese texts. Further, the words in Japanese language may have different appearances in dissimilar situations due to tense and voice. The morphological analysis here has two purposes: (1) split a sentence into a sequence of words, and (2) restore each word to its original format so that the word-embedding framework could recognize them as the same word. The morphological analysis is performed by the morphological analysis framework "Sudachi [17]" with the user dictionary "UniDic [18]."

[1] https://www.uta-net.com/
[2] https://www.jalan.net/

### 3.2 Pseudo-Summarization of the Corpora

This section introduces the main idea of pseudo-summarization, and then, it describes the procedures of the pseudo-summarization in detail.

#### 3.2.1 The idea of pseudo-summarization

In this study, both lyrics and reviews of tourist attractions are recognized as textual information that contains human sensitivity. However, the amount of information per word or phrase is not uniform. In the case of lyrics and reviews, some expressions are so unique as to strongly reflect the writer's thoughts or draw out the individuality of lyrics and reviews, while others are not. In this study, we call the former "unique phrases" and the latter "common phrases." The selection of sentences and words by importance in a text document is generally categorized as "text summarization." The traditional methods of summarization change the construction of the documents as it removes the low importance part from the texts. As a result, the context of documents is destroyed or distorted by the summarization process. While word-embedding technology does rely on the context of documents, there is a concern that the traditional summarization method may have a negative impact on word-embedding. To handle this issue, in this study, we propose a "pseudo-summarization" method that uses word importance as the weight of the word rather than as the threshold for filtering any words or sentences. By applying word importance as weight, the proposed method increases the individuality of lyrics and reviews without removing any word, phrase, or sentence.

#### 3.2.2 The procedure of pseudo-summarization

There are two algorithms for calculating the importance of documents: TF-IDF (Term Frequency – Inverse Document Frequency) [19] and SIF (Smooth Inverse Frequency) [20]. TF-IDF is good at picking up the high importance words while SIF performs well at cutting off the low importance words. The main difference between TF-IDF and SIF is the importance values of middle-importance words. TF-IDF always has small values for these words, while SIF keeps their importance high. In addition, the importance value generated by SIF is generally larger than that by TF-IDF. To build a better importance evaluation function that combines the features of TF-IDF and SIF, we design a function in Eq. (1) to take the average of TF-IDF and SIF values of a word:

$$f_\omega = (C \times v_{tfidf} + v_{sif})/2, \tag{1}$$

where $f_\omega$ is the function for calculating the weight $\omega$ of each word in the corpus, $C$ is the coefficient for adjusting the ratio between TF-IDF and SIF values since the former is usually smaller than the latter in most cases. Both TF-IDF and SIF values $v_{tfidf}$ and $v_{sif}$ are normalized to the range [0, 1), and we expect that the calculation results of $f_\omega$ can be controlled to fit the same range. Therefore, the coefficient $C$ is set to 1.7. Fig. 2 shows an example of how TF-IDF and SIF values change when the data is arranged as $f_\omega$ descending order for a sample lyric.

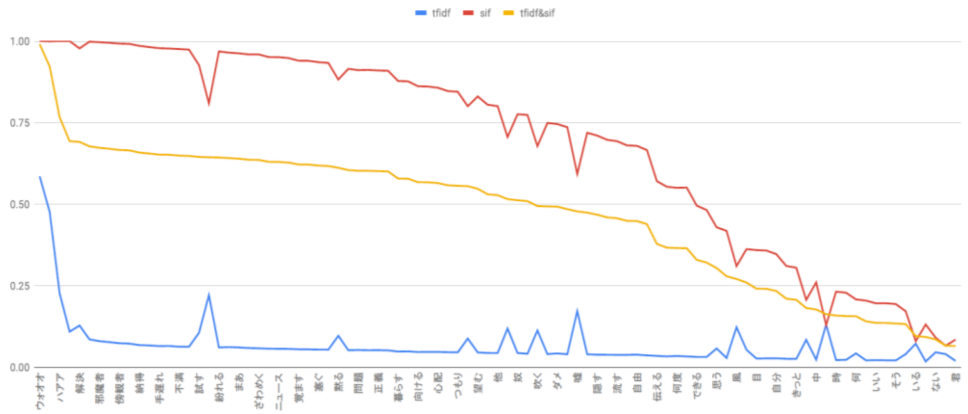


Fig. 2. Value transitions of TF-IDF and SIF importance when data is arranged as $f_\omega$ descending order. (Blue: TF-IDF, Red: SIF, Yellow: $f_\omega$).

### 3.3 The Training of the Vectorization Model

The vectorization model is built with fastText [21-23] and uses the CBOW [24] as the word-embedding model. There are 270,000 lyrics in the lyric corpus. Every word in the lyrics corpus is assigned with a 100-dimension vector.

#### 3.3.1 Comparison between skip-gram and CBOW

Mikolov *et al.* proposed Skip-Gram and CBOW [24] as the training models for word-embedding. Both are designed to learn the underlying word representations for each word using neural networks. In the CBOW model, the distributed representations of context (or surrounding words) are combined to predict the word in the middle. While in the Skip-Gram model, the distributed representation of the input word is used to predict the context. Since the way that Skip-Gram and CBOW deal with the relationship of adjacency words are diametrically opposite, a performance observation of two methods is performed. Two vectorization models with both Skip-Gram and CBOW with the lyrics corpus are generated, several test words are prepared, and the trained model is used to produce a list of words similar to the test words. The words “like(好き)” and “myself(自分)” are used as test words in the performance observation. Table 1 shows the top similar words of the test words “like” and “myself” produced by both models generated using Skip-Gram and CBOW.

It can be confirmed from the results in Table 1 that Skip-Gram prefers association words that are considered to appear usually near the test word, while CBOW prefers synonym words and prioritizes the word that combines the test word in it. In the results of Skip-Gram, the word “嫌い(hate)” and “大ッ嫌い(hate very much)” are treated as words similar to the test word “like(好き)” although “hate” is the antonym of “like” in general.

Exactly, the antonyms words do share similar nuances. However, in this study, the wordembedding method should handle both lyrics and reviews, and the polarity analysis of positive-negative opinion is significant in the field of review analysis. It becomes an issue for Skip-Gram as it confuses the positive and negative emotions & opinions. Thus, our method uses CBOW for training the vectorization model.

**Table 1. The top similar words of test words "like" and "myself", using the wordembedding model built with lyric corpus.**

| Test Word: like (好き) | | Test Word: myself (自分) | |
|---|---|---|---|
| Skip-Gram | CBOW | Skip-Gram | CBOW |
| 大好き (like very much) | 人好き (attractive) | 自分自身 (one's own self) | 自分自身 (one's own self) |
| やっぱり (I knew it) | 猫好き (cat lover) | 認めれる (admit) | 自分事 (one's own issues) |
| 私 (me) | 男好き (attractive to men) | あきらめれる (give up) | 自分らしさ (individuality) |
| 言う (say) | 噂好き (gossipy) | 変われる (can be changed) | 自分本位 (selfishness) |
| 嫌い (hate) | 歌好き (music lover) | 自身 (self) | 自分史 (personal history) |
| 大ッ嫌い (hate very much) | 大好き (like very much) | 自分らしさ (individuality) | ご自身 (one's own self) |
| こっ恥ずかしい (embarrassed) | モノ好き (curious) | はず (must be) | 変われる (can be changed) |
| だいっ嫌い (hate very much) | 女好き (attractive to women) | オクスルコトナク (do not scared) | ご自分 (one's own self) |
| あなた (you) | もの好き (curious) | 決めつける (arbitrarily decide something) | 私自身 (myself) |
| コクハク (confess my love) | やっぱり (I knew it) | 言い聞かせる (persuade oneself) | 認める (admit) |

### 3.4 Vectorization of Lyrics and Tourist Attraction Reviews

This section describes the generation of lyric vectors and reviews vectors. For each word in the lyrics corpus and the reviews corpus, the method tries to fetch the wordembedding vector of that word from the vectorization model. Then an average of all those vectors is taken and the average vector becomes the vector of these lyrics and reviews. The word weight described in Section 3.2 is also applied during the calculation of the vector average. The calculation is performed with the following equation:

$$\bar{V} = \frac{\omega_1 v_1 + \ldots + \omega_i v_i}{N}, \tag{2}$$

where $\omega$ is the word weight mentioned in Section 3.2.2, $v_i$ and $N$ denote the vector of $i$th word and the number of words in a lyric or review, respectively.

### 3.5 The Merger of Review Vectors and the Generation of Location Vectors

The basic idea of vector integrating is similar to the vector average described in Section 3.4. The average of all review vectors that are related to the same tourist attraction is taken. However, the entropy of each review is quite different. It is easy to understand that some reviews have hundreds of words while some reviews only have one sentence or even just one phrase. It is necessary to introduce a method that reflects the difference in entropy when calculating the average. Here, the word number of each review is used to measure the amount of information in each review. The merger of review vectors is performed using Eq. (3):

$$V_l = \frac{\mu_1 v_1 + \ldots + \mu_j v_j}{\mu_1 + \ldots + \mu_j}, \tag{3}$$

where $v_j$ and $\mu_j$ denote the vector of the $j$th review and the number of words in the $j$th review for a tourist attraction, respectively.

**Table 2. Distance mean and group variance before & after the pseudo summarization.**

| | Before | After |
|---|---|---|
| Distance mean | 313.701 | 950.428 |
| Group variance | 1880.538 | 21011.816 |

### 3.6 The Similarity of Lyric and Location Vectors and the Result of Lyrics Retrieval

The lyric vectors and the location vectors are prepared according to the processes laid out in Sections 3.4 and 3.5. The calculation of similarity between location vectors and lyric vectors, and the lyrics retrieval based on the vector similarity are performed according to the following procedure:

1. Specify a tourist attraction and prepare the location vector $V_l$ of that tourist attraction
2. Initialize the variable *max* and *index*
3. Fetch the *i*th lyric vector $L_i$ from the lyric vector group.
4. Calculate the vector similarity *Sim* between $L_i$ and $V_l$.
5. Compare *Sim* with *max*: If *Sim* > *max*, overwrite *max* and *index* with *S* and *i*. If *Sim* ≤ *max*, fetch *i* + 1th lyric vector and go to Step 4.
6. Output the music information (title, artist, genre *etc.*) of *index*th lyric.

The similarity calculation is performed with the following equation:

$$\cos(V_l, L_i) = \frac{V_l \cdot L_i}{|V_l| \times |L_i|}, \tag{3}$$

where $V_l$ is the location vector of the specific tourist attraction and $L_i$ is the lyric vector of music that under observation.

## 4. PRELIMINARY EXPERIMENTS

This section discusses the effects of the pseudo-summaries of lyrics and tourist attraction reviews presented in Section 3.4. The summarization of the lyrics and tourist attraction reviews was predicted to have three effects: (1) emphasizing the individuality and characteristics of lyrics and reviews; (2) making the distinction of lyrics from other lyrics more apparent; (3) promoting the diversity in the final retrieval results.

The effect (1) is supported by a theory noted in Section 3.2 that text summarization removes the “common phrases” and raises the ratio of “unique phrases” in each document. As a result, the important part in each document, which is different from the other documents, was emphasized by the summarization.

We use the distance mean and variance of lyric vector group to measure the distinction of lyrics from other lyrics before and after the pseudo summarization. Table 2 shows the values of distance mean and group variance. After the pseudo summarization, both the distance means and the group variance of lyric vector group increased and the whole vector space was extended. This data supports the effect (2) mentioned above that the pseudo summarization makes the distinction among lyrics more apparent.

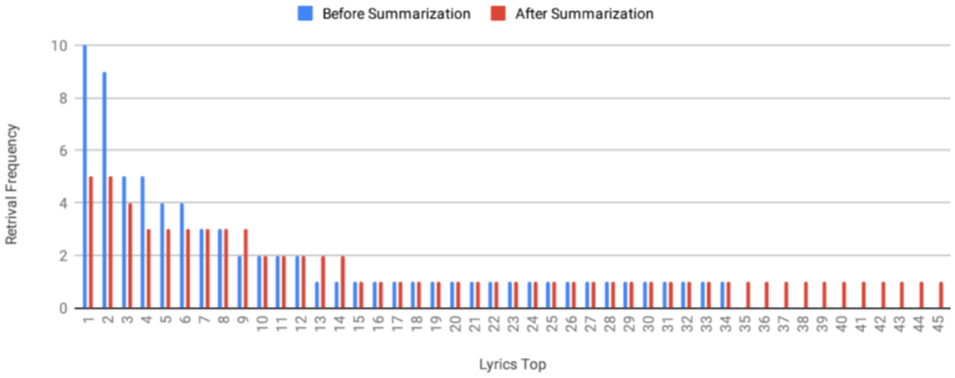


Fig. 3. Distribution before summarization.

To verify the effect (3) of the pseudo summarization, the distribution of the retrieval results, which shows "how many tourist attractions were matched with the same music," was used as an index for the evaluation. The number of tourist attractions used in this evaluation was 73. Fig. 3 shows the distribution of the retrieval results before and after applying the summarization.

According to these results, it was suggested that the summarization emphasized the important part of each lyric, extended the vector space of lyric vector group, and flattened the distribution of retrieval results. It showed that the summarization partially achieved the three purposes mentioned above, and hence the effectiveness of summarization was confirmed.

## 5. EVALUATION EXPERIMENTS

In the evaluation experiment, the satisfactoriness of retrieval results from the proposed method was measured by subjective evaluation with participants.

### 5.1 Procedures of the Experiment

Participants were asked for their favorite music genre and a tourist attraction they knew well in advance. Every participant provided five "tourist attraction & music genre" pairs as the input queries of the retrieval system. The number of participants was 20. The music genres given to the participants to choose are shown in Table 3.

**Table 3. Candidate music genre for the participants.**

| | | | | |
|---|---|---|---|---|
| J-POP | Anime | Rock | Kayo_kyoku | Enka |
| Pop | Sound Track | Hip-hop/Rap | Alternative | R&B |

The experiment was conducted with the following procedures:

1. Generate lyrics retrieval results from the input query, the output will be the experimenttal group.

2. Choose another song as the control group with the same genre randomly as the query.
3. Prepare the URLs that access the lyric and the streaming of music in both experimental and control groups, then send the URLs to the participants.
4. Participants read the lyrics of the music in both experimental and control groups, then evaluate their satisfactoriness on a four-point scale.
5. Participants listen to the music in both experimental and control groups, then evaluate the satisfactoriness from four levels.

The satisfactoriness of the retrieval results is measured as "the similarity between the image of lyric&music and the image of tourist attractions." The four levels in the evaluation are "completely different," "not exactly,", "fit a little," and "perfectly fit." In this paper, we set a positive and negative experiment with a control group of random chosen. The experiment aims to confirm the effectiveness of the recommendation made by proposed method. More advanced comparison with other music or lyrics retrieval methods will be our future focus.

### 5.2 Results of Experiment

Table 4 shows the results of the evaluation experiment. There were 20 participants and each participant submitted five queries of the "tourist attraction & music genre," so there were 100 votes in total in the experiment result. Considering the degree and polarity of the four opinions in the evaluation, we assigned a weight to each opinion to observe the overall evaluation of each experiment item. The point of "completely different," "not exactly,", "fit a little," and "perfectly fit" are set to –1, –0.5, 0.5 and +1. The weighted version of the experiment results is shown in Table 5.

**Table 4. Original version of the experiment results.**

| Experiment Items | Completely Different | Not Exactly | Fit a Little | Perfectly Fit | Total |
|---|---|---|---|---|---|
| Lyrics of experimental group | 9 | 27 | 36 | 28 | 100 |
| Lyrics of control group | 21 | 31 | 34 | 15 | 100 |
| Music of experimental group | 16 | 30 | 38 | 16 | 100 |
| Music of control group | 23 | 31 | 28 | 18 | 100 |

**Table 5. Weighted version of the experiment results.**

| Experiment Items | Completely Different | Not Exactly | Fit a Little | Perfectly Fit | Total |
|---|---|---|---|---|---|
| Lyrics of experimental group | –9 | –13.5 | 18 | 28 | +23.5 |
| Lyrics of control group | –21 | –15.5 | 17 | 15 | –3.5 |
| Music of experimental group | –16 | –15 | 19 | 16 | +4 |
| Music of control group | –23 | –15.5 | 14 | 18 | –6.5 |

## 6. DISCUSSIONS

### 6.1 Discussions about the Results of Experiment

As the overall evaluation shown in Table 5, in the lyrics field, the experiment group achieved a much better satisfactoriness grade than the control group (+23.5 vs –3.5) did.

**Table 6. Top 20 words appeared in the reviews of *Kanemori Red Brick Warehouse*.**

| | | | | |
|---|---|---|---|---|
| 倉庫 (warehouse) | レンガ (brick) | 函館 (Hakodate) | お店 (store) | 行く (go) |
| お土産 (souvenir) | 赤 (red) | 雰囲気 (atmosphere) | とても (very) | 夜 (night) |
| 見る (watch) | 入る (enter) | 思う (think) | 中 (inside) | できる (can) |
| たくさん (many) | 良い (good) | 楽しめる (be able to enjoy) | 観光 (sightseeing) | 場所 (location) |

Meanwhile, in the music field, the experiment group did not perform as well as the lyric field. The gap between the experiment group and the control group was less (−6.5 −4 = −10.5) than it was in the lyric field (−3.5 − 23.5 = −27).

Here is an example of the retrieval results that were evaluated as "perfectly fit" by the participant. The input query is a tourist attraction called *Kanemori Red Brick Warehouse* with the music genre of "J-POP." The lyric of the retrieved music is ベイ・ブリッジ・セレナーデ (Bay Bridge Serenade).[3] Table 6 shows the top 20 words that appear most frequently.

The relevance between the lyrics of retrieved music and the top words from the reviews is clear. For example, the lyric of retrieved music talks about the hero watching the city on a bridge in the night and imagining the girlfriend whom the hero should meet in the future. The main theme in this song combines "night," "watch," and "imagine." Words such as "夜 (night)," "見る (watch) & 観光 (sightseeing)," and "思う (think)" are considered to fit the themes mentioned above.

Thus, in both fields of lyrics and music, the experiment group showed an advantage over the control group and the effectiveness of the proposed method was confirmed. However, the advantage of the experiment group was mainly in the lyrics field.

### 6.2 Discussions about the Performance Difference between Lyrics and Music Fields

The causes for the performance difference between the lyrics and music fields as mentioned in Section 6.1 is considered to be the amount of information that the proposed method accesses compared with the human participants. Music, especially songs, are composed by the human voice, the instrumental accompaniment, the rhythm in the melody, and other elements. Humans build an image of music by accessing not only the lyrics, but also the entirety of the audio information available. While the proposed method builds word-embedding as a representation of music, using only the information of lyrical texts. Compared to the proposed method, people gain much more information from the audio part in a song.

An approach to deal with audio information and to link the audio information to a tourist attraction is needed in order to solve this problem. For example, Ueno *et al.* [25] analyzed the relationship between acoustic features and sensitivity vectors and proposed a method to derive a transformation matrix from acoustic parameters to emotion parameters. Nishimura *et al.* [26] proposed a method for searching among different media using Recurrent Neural Network (RNN). Their proposed method makes it possible to map music (as the time sequence of audio vectors) to lyrics (as document vectors). Once the mapping between acoustic and textual information becomes possible, it is possible to match the acoustic features of the music with the reviews of tourist attractions, using the textual information of the lyrics as an intermediary.

[3] https://www.uta-net.com/song/8048/

**Table 7. Sample data of word embedding similarity and TF-IDF similarity of some recommendation results that evaluated as "Perfectly Fit", with the similarity of TF-IDF based retrieval result.**

| Item No. | Word embedding similarity | TF-IDF similarity | TF-IDF retrieval similarity |
|---|---|---|---|
| 1 | 0.745 | 0.583 | 0.629 |
| 2 | 0.631 | 0.135 | 0.499 |
| 3 | 0.860 | 0.145 | 0.500 |
| 4 | 0.848 | 0.170 | 0.499 |
| 5 | 0.763 | 0.109 | 0.499 |
| 6 | 0.647 | 0.044 | 0.499 |

### 6.3 Discussions about the Similarity Calculation Method: Word Embedding and TF-IDF

We built a document similarity evaluation method based on pure TF-IDF as a comparison to the word embedding method. We unified the lyrics texts and review texts to get a meaningful cross-domain TF-IDF value which based on the same document group. The document vector built by TF-IDF method was following Eq. (5).

$$V_d = (T_{d,1}, T_{d,2}, T_{d,3}, \ldots, T_{d,n}), \tag{5}$$

where $d$, $V_d$, $n$ and $T_{d,n}$ each denotes document number, the vector of $d$th document, the word number in the document group and the TF-IDF value of $n$th word in $d$th document. For those experiment items that (1) the experimental group obtained a "perfectly fit" feedback, (2) the control group obtained a negative ("completely different" and "not exactly") feedback in the evaluation experiment, we calculated the similarity value with the TF-IDF based method mentioned above. For these tourist attractions, we also conducted a lyrics retrieval using the TF-IDF based similarity and recorded the similarity value. Table 7 shows the data of word embedding similarity, TF-IDF similarity, and TF-IDF based retrieval similarity. For items from No.2 to No.6, the TF-IDF based similarity values were very low. For item No.1, the tourist attraction was Ueno Zoo which is famous for giant panda. There were many reviews mentioned about the giant panda in the reviews of Ueno Zoo. Both two lyrics hit by the proposed word embedding method and the test TF-IDF method were songs about giant panda, while there was more word "panda" in the document of TF-IDF retrieved lyric. That showed TF-IDF cared more about the word that appeared in the document so this lyric achieves higher similarity in the TF-IDF retrieval method.

Compared to the entire document group, the word number of each document was very small (hundreds or thousands vs hundreds of thousands). As a result, most of bits in a document vector were filled by zero. Furthermore, in a cross-domain situation, the words appearing in domain A and domain B were always different and the overlap part was small. As a result, the cosine similarity between lyric vectors and location vectors was generally at a very low level (less than 0.1). It was confirmed that, the more zero bits in a lyric vector (means the less word in a lyric), the higher similarity it got, because the location vector also had many zero bits. The TF-IDF retrieved lyric of items from No.2 to No.6 in Table 7 was the same lyric with only one word repeats seven times. We considered that this becomes the reason that this lyric achieves high similarity with many location vectors.

According to the results noted above, we considered that a pure TF-IDF similarity method did not fit the needs of a cross-domain recommendation task.

## 7. CONCLUSION

This paper proposed a lyrics retrieval method based on word-embedding technology to perform music recommendation that is suitable for the information of a tourist attraction. A vectorization model generated from the lyrics corpus to the reviews corpus was applied so that the word-embedding representation was shared on both lyric and review fields. Thus, the tourist attraction reviews were dealt with as pseudo-lyrics and the comparison between the word-embedding vectors of lyrics and reviews became possible and meaningful. The lyrics retrieval was based on the similarity comparison between the lyric vectors and the location vectors formed by the merger of review vectors.

In the evaluation experiment, the experiment group had an advantage over the control group in both lyric and music field. Thus, the effectiveness of the proposed method was confirmed. During the experiment, the retrieval results in the music field did not evoke the same satisfactoriness from the participants that the lyric field did. The reason for this was considered to be the lack of information that the proposed method had available to it. In future work, we will attempt to design an approach to deal with the audio information and link the audio information to tourist attraction through the intermediary of the lyrics field. Also, the performance comparison with other music or lyrics retrieval methods is in our future plan.

## REFERENCES

1. M. Nishii, “Influence of reassurance from environmental music on product selection [in Japanese],” *Journal of Promotional Marketing*, Vol. 10, 2017, pp. 25-39.
2. T. Okayama, “A study on in-store environmental factors as an element of corporate brand in retailing [in Japanese],” *Japan Society of Marketing and Distribution Review*, Vol. 1, 2017, pp. 41-48.
3. Y. Han, R. Yamanishi, and Y. Nishihara, “Music retrieval focusing on lyrics with summary of tourist-spot reviews based on shared word-vectors,” in *Proceedings of International Conference on Technologies and Applications of Artificial Intelligence*, 2020, pp. 73-78.
4. S. Pauws, “Cubyhum: A fully operational query by humming system,” in *Proceedings of the 3rd International Society for Music Information Retrieval Conference*, 2002, pp. 187-196.
5. D. Wang, T. Li, and M. Ogihara, “Are tags better than audio features? the effect of joint use of tags and audio content features for artistic style clustering,” in *Proceedings of the 11th International Society for Music Information Retrieval Conference*, 2010, pp. 57-62,
6. E. Pampalk, A. Flexer, and G. Widmer, “Improvements of audio-based music similarity and genre classification,” in *Proceedings of the 6th International Conference on Music Information Retrieval*, 2005, pp. 628-633.
7. R. Panda, R. Malheiro, and R. P. Paiva, “Musical texture and expressivity features for music emotion recognition,” in *Proceedings of the 19th International Society for Mu-*

*sic Information Retrieval Conference*, 2018, pp. 383-391.
8. K. Tsukuda, K. Ishida, and M. Goto, "A lyrics-based music exploratory web service by modeling lyrics generative process," in *Proceedings of the 18th International Conference on Music Information Retrieval*, 2017, pp. 544-551.
9. C.-M. Chen, M.-F. Tsai, J.-Y. Liu, and Y.-H. Yang, "Using emotional context from article for contextual music recommendation," in *Proceedings of the 21st ACM International Conference on Multimedia*, 2013, pp. 649-652.
10. M. Braunhofer, M. Kaminskas, and F. Ricci, "Location-aware music recommendation," *International Journal of Multimedia Information Retrieval*, Vol. 2, 2013, pp. 31-44.
11. I. Ferna´ndez-Tobias, I. Cantador, M. Kaminskas, and F. Ricci, "Cross-domain recommender systems: A survey of the state of the art," in *Proceedings of the 2nd Spanish Conference on Information Retrieval*, 2012, pp. 187-198.
12. M. Kaminskas, F. Ricci, and M. Schedl, "Location-aware music recommendation using auto-tagging and hybrid matching," in *Proceedings of the 7th ACM Conference on Recommender Systems*, 2013, pp. 17-24.
13. T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: Real-time event detection by social sensors," in *Proceedings of the 19th International Conference on World Wide Web*, 2010, pp. 851-860.
14. E. Aramaki, S. Maskawa, and M. Morita, "Twitter catches the flu: Detecting influenza epidemics using Twitter," in *Proceedings of Conference on Empirical Methods in Natural Language Processing*, 2011, pp. 1568-1576.
15. S. Naz, A. Sharan, and N. Malik, "Sentiment classification on twitter data using support vector machine," in *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence*, 2018, pp. 676-679.
16. M. Yoshida and F. Toriumi, "Information diffusion power of political party twitter accounts during Japan's 2017 election," in *Proceedings of the 10th International Conference on Social Informatics*, 2018, pp. 334-342.
17. K. Takaoka, S. Hisamoto, N. Kawahara, M. Sakamoto, Y. Uchida, and Y. Matsumoto, "Sudachi: a Japanese tokenizer for business," in *Proceedings of the 11th International Conference on Language Resources and Evaluation*, 2018.
18. D. Yasuharu, O. Toshinobu, O. Hideki, Y. Atsushi, M. Nobuaki, U. Kiyotaka, and K. Hanae, "The development of an electronic dictionary for morphological analysis and its application to Japanese corpus linguistics [in Japanese]," *Japanese Linguistics*, Vol. 22, 2007, pp. 101-123.
19. G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill, Inc., USA, 1986.
20. S. Arora, Y. Liang, and T. Ma, "A simple but tough-to-beat baseline for sentence embeddings," in *Proceedings of the 5th International Conference on Learning Representations*, 2017, pp. 1-16.
21. P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, Vol. 5, 2017, pp. 135-146.
22. A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, Vol. 2, 2017, pp. 427-431.
23. A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Je´gou, and T. Mikolov, "Fast-

text.zip: Compressing text classification models," *CoRR*, Vol. abs/1612.03651, 2016.
24. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proceedings of International Conference on Learning Representations*, Vol. 2013, 2013, No. 01.
25. T. Ueno and K. Aikawa, "The relation between emotional representation vectors and acoustic features for sound retrieval [in Japanese]," *Information Processing Society of Japan Special Interest Groups Notes*, Vol. 74, 2008, pp. 211-216.
26. R. Nishimura, M. Higaki, and N. Kitaoka, "Mapping between acoustic vector space and document vector space using RNN-LSTM [in Japanese]," *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Vol. 30, 2018, pp. 628-633.

**Yihong Han** received his B.E and M.E from Ritsumeikan University in Japan in 2019 and 2021, respectively. He is interested in research on multimedia engineering and kansei engineering.

**Ryosuke Yamanishi** received his B.E, M.E and Ph.D. from Nagoya Institute of Technology, Japan, in 2007, 2009 and 2012, respectively. He joined in College of Information Science and Engineering, Ritsumeikan University as a Research Associate in 2012, a Research Assistant Professor in 2013, an Assistant Professor in 2014 and a Lecturer in 2018. During this period, he visited laboratory of computational intelligence of UBC, Canada as a Visiting Assistant Professor. In 2020, he has joined in Faculty of Informatics, Kansai University as an Associate Professor. He is interested in the content-oriented computational culture and arts like music informatics, comic computing, game informatics and computational gastronomy foods and eating activities. He is a member of IEICE, IPSJ, JSAI, JSKE, ASJ, SAS, ACM and ACL.

**Yoko Nishihara** received her B.E, M.E, and PhD degrees from Osaka University in Japan in 2003, 2005, and 2007, respectively. She was a JSPS research fellowship for young scientists (DC and PD). She was an Assistant Professor in the Faculty of Engineering at the University of Tokyo in 2008 and a Lecturer in 2009. She joined the College of Information Science and Engineering at Ritsumeikan University as an Associate Professor, now she is a Professor. She is interested in human-computer interaction and natural language processing. She is a member of IPSJ, JSAI, and JSKE.