Discovering Latent Country Words: A Step Towards Cross-Cultural Emotional Communication

Heeryon Cho[1](✉) and Toru Ishida[2]

[1] College of Computer Science, Kookmin University, Seoul 02707, South Korea
heeryon@kookmin.ac.kr
[2] School of Creative Science and Engineering, Waseda University,
Tokyo 1698555, Japan
toru.ishida@aoni.waseda.jp

Abstract. Knowing what concepts are substantial to each country can be helpful in enhancing emotional communication between two countries. As a concrete example of identifying substantial country concepts, we focus on a task of finding latent country words from cross-cultural texts of two countries. We do this by combining word embedding and tensor decomposition: common words that appear in both countries' texts are selected; their country specific word embeddings are learned; a three-way tensor consisting of word factor, word embedding factor, and country factor are constructed; and CANDECOMP/PARAFAC decomposition is performed on the three-way tensor while fixing the country factor values of the decomposed result. We tested our method on a motivating example of finding latent country words from J-pop lyrics from Japan and K-pop lyrics from South Korea. We found that J-pop lyrics words feature nature related motifs such as 'petal', 'cloud', 'universe', 'star', and 'sky', whereas K-pop lyrics words highlight human body related motifs such as 'style', 'shirt', 'head', 'foot', and 'skin'.

Keywords: Cross-cultural text analysis · Tensor decomposition · Word embedding

1 Introduction

Let us suppose that two countries' representatives are trying to resolve a common issue through bilateral discussion, and the discussion is being recorded in the form of meeting minutes. Later, based on the meeting minutes, we want to identify a list of words that are used more substantially by each country's representatives. How can we do this?

This research was supported by the National Research Foundation of South Korea (NRF) grant funded by the South Korean government (NRF-2017R1A2B4011015). This research was partially supported by a Grant-in-Aid for Scientific Research (A) (17H00759, 2017–2020) from Japan Society for the Promotion of Science (JSPS).

H. Nakanishi et al. (Eds.): CRIWG+CollabTech 2019, LNCS 11677, pp. 232–241, 2019.
https://doi.org/10.1007/978-3-030-28011-6_17

One way to find such latent country words is to build a three-way tensor consisting of a list of common, frequent words used by the two countries' representatives in the first dimension (i.e., word factor), the word embedding values of each country's words in the second dimension (i.e., word embedding factor), and the degree of country aspect in the third dimension (i.e., country factor). Since the word embedding representation learns the hidden relationships between words, country-wise speech texts can learn country specific relationship between words. We then perform tensor decomposition on the country-wise word embedding data, while adjusting the country factor values, to find substantial words used by each country's representatives.

Advances in three-way component analysis techniques have realized the summarization of all information in a large three-way data set [1]. Three-way data refer to data that can be arranged in a three-dimensional array, i.e., three-way tensor. Recently, tensor factorization and decomposition has become an important tool for data mining since they can explicitly take into account the multiway structure of the data that would otherwise be lost when analyzing the data by matrix factorization approaches [2]. CANDECOMP (canonical decomposition)/PARAFAC (parallel factors) [3,4] decomposition (hereafter referred to as CP decomposition) is a popular tensor decomposition method that decomposes a tensor as a sum of rank-one tensors [5]. In this study, we apply CP decomposition on the aforementioned three-way tensor while deliberately fixing the country factor values of the decomposed result in order to identify latent country words.

2 Discovering Latent Country Words

To describe our method concretely, we present a motivating example of finding latent lyrics words from J-pop and K-pop lyrics texts. We aim to discover latent country motifs from J-pop/K-pop lyrics by learning country specific word embeddings and applying CP decomposition on the two countries' lyrics word embedding tensor. Figure 1 displays the overview of discovering latent J-pop and K-pop lyrics words. The overall process includes lyrics data collection and tokenization, J-pop/K-pop lyrics word mapping (i.e., language unification) and filtering, country specific word embedding learning, and tensor decomposition.

2.1 Data Gathering and Tokenization

Ten years (2008–2017) worth of yearly top 100 ranking J-pop and K-pop hit song lyrics were gathered as target data. For J-pop lyrics, a total of 1,142 lyrics were crawled from a major Japanese music ranking portal, *Oricon Chart*, and two lyrics search engines, *Uta-Net* and *J-Lyric.net*.[1] For K-pop lyrics, a total of 1,000 lyrics were crawled from a major South Korean music ranking portal, *Melon Chart*.[2] There were more J-pop lyrics than K-pop lyrics since the *Oricon Chart* allowed multiple tied rankings. The two countries' lyrics were then tokenized

[1] https://www.oricon.co.jp/, http://www.uta-net.com, http://j-lyric.net/.
[2] https://www.melon.com/.

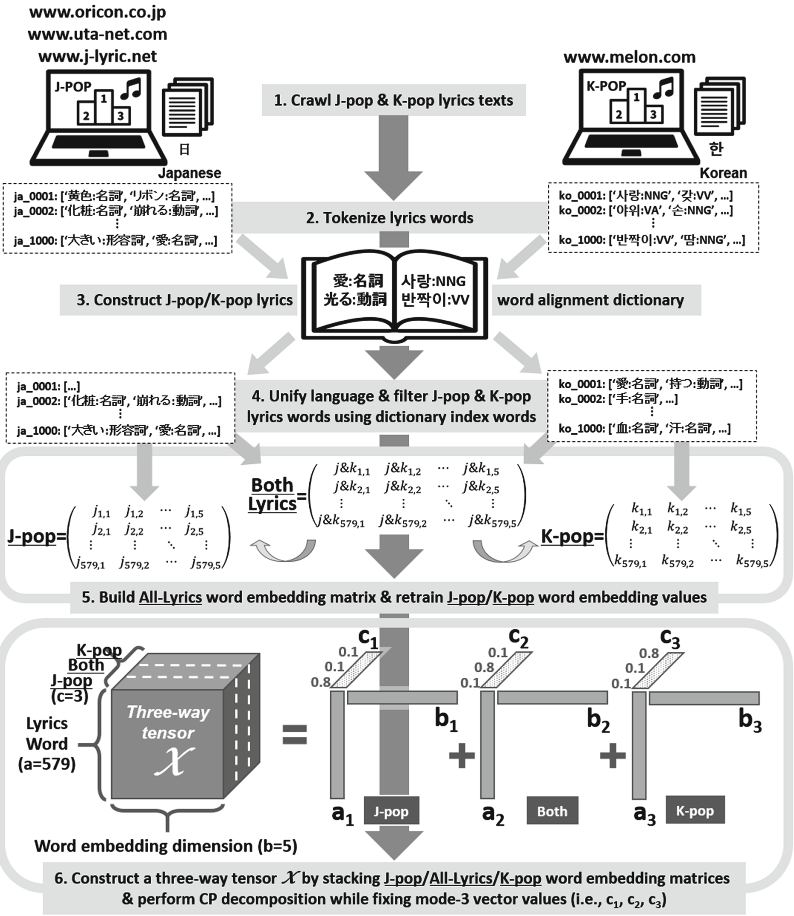
www.oricon.co.jp
www.uta-net.com
www.j-lyric.net
J-POP
日
Japanese
www.melon.com
K-POP
한
Korean
1. Crawl J-pop & K-pop lyrics texts
ja_0001: [‘黄色:名詞’, ‘リボン:名詞’, ...]
ja_0002: [‘化粧:名詞’, ‘崩れる:動詞’, ...]
ja_1000: [‘大きい:形容詞’, ‘愛:名詞’, ...]
2. Tokenize lyrics words
ko_0001: [‘사랑:NNG’, ‘갖:VV’, ...]
ko_0002: [‘아위:VA’, ‘손:NNG’, ...]
ko_1000: [‘반짝이:VV’, ‘땀:NNG’, ...]
3. Construct J-pop/K-pop lyrics
愛:名詞
光る:動詞
사랑:NNG
반짝이:VV
word alignment dictionary
ja_0001: [...]
ja_0002: [‘化粧:名詞’, ‘崩れる:動詞’, ...]
ja_1000: [‘大きい:形容詞’, ‘愛:名詞’, ...]
4. Unify language & filter J-pop & K-pop lyrics words using dictionary index words
ko_0001: [‘愛:名詞’, ‘持つ:動詞’, ...]
ko_0002: [‘手:名詞’, ...]
ko_1000: [‘血:名詞’, ‘汗:名詞’, ...]
J-pop= $\begin{pmatrix} j_{1,1} & j_{1,2} & \cdots & j_{1,5} \\ j_{2,1} & j_{2,2} & \cdots & j_{2,5} \\ \vdots & \vdots & \ddots & \vdots \\ j_{579,1} & j_{579,2} & \cdots & j_{579,5} \end{pmatrix}$
Both Lyrics= $\begin{pmatrix} j\&k_{1,1} & j\&k_{1,2} & \cdots & j\&k_{1,5} \\ j\&k_{2,1} & j\&k_{2,2} & \cdots & j\&k_{2,5} \\ \vdots & \vdots & \ddots & \vdots \\ j\&k_{579,1} & j\&k_{579,2} & \cdots & j\&k_{579,5} \end{pmatrix}$
K-pop= $\begin{pmatrix} k_{1,1} & k_{1,2} & \cdots & k_{1,5} \\ k_{2,1} & k_{2,2} & \cdots & k_{2,5} \\ \vdots & \vdots & \ddots & \vdots \\ k_{579,1} & k_{579,2} & \cdots & k_{579,5} \end{pmatrix}$
5. Build All-Lyrics word embedding matrix & retrain J-pop/K-pop word embedding values
K-pop
Both
J-pop
(c=3)
Lyrics Word (a=579)
Three-way tensor $\mathcal{X}$
Word embedding dimension (b=5)
=
c_1 0.1 0.1 0.8
b_1
a_1 J-pop
+
c_2 0.1 0.8 0.1
b_2
a_2 Both
+
c_3 0.8 0.1 0.1
b_3
a_3 K-pop
6. Construct a three-way tensor $\mathcal{X}$ by stacking J-pop/All-Lyrics/K-pop word embedding matrices & perform CP decomposition while fixing mode-3 vector values (i.e., c_1, c_2, c_3)

Fig. 1. Overall process of discovering latent J-pop/K-pop lyrics words through word embedding learning and CP decomposition.

using Japanese/Korean morphological analyzers. Nouns, verbs, and adjectives were selected as the unit of analyses. Table 1 summarizes the statistics of the gathered data. We see that the J-pop lyrics' vocabulary size is more than twice that of K-pop's as evidenced by the total number of unique words.

Table 1. J-pop/K-pop lyrics data statistics.

Gathered years: 2008–2017	J-pop	K-pop
Total number of lyrics	1,142	1,000
Avg. number of words per lyric	110	77
Total words (total unique)	125,205 (13,086)	77,092 (5,797)

2.2 J-Pop/K-Pop Lyrics Word Mapping and Filtering

Once the nouns, verbs, and adjectives were obtained from the J-pop/K-pop lyrics data, the similar meaning J-pop lyrics words were manually mapped to K-pop lyrics words to build a J-pop/K-pop lyrics word alignment dictionary. This dictionary was used later for unifying the J-pop/K-pop lyrics data into a single language. Table 2 displays three sample entries of the J-pop/K-pop lyrics word alignment dictionary. Approximate English translation is given as reference in the first column. Note that the parts of speech were not distinguished in this dictionary; rather, words with similar concepts were grouped using a '|' delimiter to constitute a single dictionary word (e.g., "get_angry:verb|anger:noun").

Table 2. Sample entry of J-pop/K-pop lyrics word alignment dictionary.

<table>
<tr><th>English translation</th><th>J-pop lyrics word</th><th>K-pop lyrics word</th></tr>
<tr><td>heart:noun</td><td>心:名詞|こころ:名詞|ココロ:名詞</td><td>마음:NNG|맘:NNG</td></tr>
<tr><td>get_angry:verb|anger:noun</td><td>怒る:動詞|怒り:名詞|···</td><td>화내:VV|분노:NNG|···</td></tr>
<tr><td>voice:noun</td><td>声:名詞</td><td>목소리:NNG</td></tr>
</table>

The J-pop/K-pop lyrics word alignment dictionary was constructed as follows. Presented with a list of unique J-pop and K-pop lyrics words and frequency counts, a bilingual Korean with more than ten years' experience of living in Japan first grouped the similar meaning words for each country, and then mapped the grouped words across the two countries. Then, a professional Korean-Japanese translator with twenty years of translation expertise reviewed and corrected the alignment dictionary and further expanded the dictionary. The dictionary building process continued until a sizable dictionary was constructed. From this dictionary, the dictionary words with both countries' frequency counts that were less than six were discarded. As a result, a J-pop/K-pop lyrics word alignment dictionary with 579 J-pop/K-pop lyrics word mappings was constructed. Note that when the delimiter was removed from the grouped dictionary words, the number of atomic J-pop and K-pop lyrics words were 1,065 and 870 respectively.

Using this alignment dictionary, J-pop/K-pop lyrics texts were filtered to contain only the grouped dictionary words. Then, the filtered K-pop lyrics words were converted to J-pop lyrics words by referring to the dictionary mapping. As a result, a list of filtered Japanese grouped dictionary words were obtained for both the J-pop and K-pop lyrics data. For some lyrics, there were no matching dictionary words, and hence, those lyrics were discarded. Consequently, a total of 1,134 and 986 J-pop and K-pop lyrics data were used for constructing the country specific word embeddings.

Table 3 summarizes the statistics of the filtered J-pop/K-pop lyrics word data and the J-pop/K-pop lyrics word alignment dictionary. The overall dictionary coverage was calculated by dividing the number of total filtered words by the total number of words given in Table 1. Even with the limited number of dictionary atomic words, i.e., 1,065 and 870, which constitutes approximately 8% and 15% of unique J-pop and K-pop words respectively (these percentages were calculated by dividing the number of dictionary atomic words by the total unique words in Table 1), the filtered words covered a close to half (48.8%) and three-quarters (74%) of the J-pop and K-pop lyrics data respectively. This indicates that the alignment dictionary covers many of the frequently used words in each country's lyrics. Table 4 lists the top-20 most frequently appearing dictionary noun words in the filtered J-pop/K-pop data. English translation of each word is also given. Although we have grouped the words into similar concepts regardless of the parts of speech, hereafter we select the most representative parts of speech word when presenting our results.

Table 3. Filtered data & dictionary statistics.

	J-pop	K-pop
Total filtered lyrics	1,134	986
Total filtered words	61,094	57,049
Total grouped dictionary words	579	579
Total atomic dictionary words	1,065	870
Overall dictionary coverage	48.8%	74.0%

Table 4. Top-20 total frequent J-pop & K-pop lyrics nouns.

Rank	J-pop lyrics nouns			K-pop lyrics nouns		
	English	Japanese	Korean	English	Japanese	Korean
1	*love*	愛	사랑	*love*	愛	사랑
2	now	今	지금	words	言葉	말
3	everyone	みんな	모두	*day*	日	날
4	*heart*	心	마음	*heart*	心	마음
5	dream	夢	꿈	*moment*	時	때
6	*moment*	時	때	*person*	人	사람
7	hand	手	손	*inside*	中	속
8	*inside*	中	속	night	夜	밤
9	*front*	前	앞	*eyes*	目	눈
10	*eyes*	目	눈	*chest*	胸	가슴
11	*world*	世界	세상	final	最後	마지막
12	*day*	日	날	*tears*	涙	눈물
13	*person*	人	사람	thought	想い	생각
14	future	未来	미래	time	時間	시간
15	sky	空	하늘	*front*	前	앞
16	*chest*	胸	가슴	*world*	世界	세상
17	as is	まま	이대로	one day	一日	하루
18	tomorrow	明日	내일	top	上	위
19	*tears*	涙	눈물	happiness	幸せ	행복
20	thing	事	일	[at one's] side	そば	곁

Tables 5 and 6 list the top-10 most frequent verbs and adjectives appearing in the filtered J-pop/K-pop lyrics data. We see that many of the same words are used frequently in both countries' lyrics (marked in *italics*), e.g., '*love*', '*heart*', '*look*', '*know*', '*good*' and '*sad*'.

Table 5. Top-10 most frequent J-pop & K-pop lyrics verbs.

Rank	**J-pop lyrics verbs**			**K-pop lyrics verbs**		
	English	**Japanese**	**Korean**	**English**	**Japanese**	**Korean**
1	*go*	行く	가다	*look*	見る	보다
2	*look*	見る	보다	*know*	知る	알다
3	*know*	知る	알다	*laugh*	笑う	웃다
4	give	くれる	주다	forget	忘れる	잊다
5	meet	会う	만나다	*live*	生きる	살다
6	come	来る	오다	grasp	掴む	잡다
7	*laugh*	笑う	웃다	*go*	行く	가다
8	change	変わる	변하다	leave	離れる	떠나다
9	believe	信じる	믿다	cry	泣く	울다
10	*live*	生きる	살다	take	取る	들다

Table 6. Top-10 most frequent J-pop & K-pop lyrics adjectives.

Rank	**J-pop lyrics adjectives**			**K-pop lyrics adjectives**		
	English	**Japanese**	**Korean**	**English**	**Japanese**	**Korean**
1	*good*	いい	좋다	*absent*	無い	없다
2	*sad*	悲しい	슬프다	*good*	いい	좋다
3	*far*	遠い	멀다	*painful*	痛い	아프다
4	*painful*	痛い	아프다	many	多い	많다
5	*absent*	無い	없다	*far*	遠い	멀다
6	scared	怖い	두렵다	*sad*	悲しい	슬프다
7	joyful	楽しい	즐겁다	tough	つらい	힘들다
8	kind	優しい	착하다	small	小さい	작다
9	lonely	寂しい	외롭다	late	遅い	늦다
10	hot	熱い	뜨겁다	deep	深い	깊다

2.3 Word Embedding Learning

Using the filtered lyrics word data, the J-pop/K-pop lyrics' country specific word embedding values were learned using the Word2Vec algorithm [7]. Word embedding maps each word to a vector of real numbers. Traditionally, to represent a word in a document, each word occupied one dimension of the document vector; consequently, a vocabulary of ten thousand words required ten thousand dimension vector. Word embedding introduced a way to express each word using a continuous vector space with much lower dimension of few hundreds.

It is known to capture the context of a word such as semantic and syntactic similarity, and the relation with other words. However, to learn useful embeddings, abundant data are needed. Due to the limited amount of the lyrics data, we first learned the initial word embedding values using both the J-pop and K-pop data, and then updated the initial embedding values by separately using the K-pop and J-pop data to build country specific J-pop/K-pop lyrics word embeddings following the work in [6].

With regard to the word embedding dimension size, we tested different dimension size of 5, 10, and 100. Due to the limited amount of data, we found that 5 was the most adequate among the three, and hence set the size of the embedding dimension to 5. Once the J-pop and K-pop lyrics word embedding values were learned, three 579×5 sized Word2Vec matrices, one for J-pop, one for both J-pop and K-pop (i.e., the initial word embedding values), and one for K-pop, having different embedding weights were created. The three Word2Vec matrices were then stacked back to back to create a three-way tensor of size $579 \times 5 \times 3$, which contained the J-pop, Both, and K-pop lyrics word embedding values (see Fig. 1 bottom).

2.4 Performing CP Decomposition

CP decomposition was performed on this three-way tensor using the alternating least squares algorithm to approximately decompose the tensor into a sum of three components of rank-one tensors (Fig. 1 bottom). Although the number of components in CP decomposition are determined by gradually increasing the component size after evaluating the fit of each optimal solution, this approach did not apply to our example task since we formulated a CP decomposition with a predetermined component size of three which matched J-pop, Both, and K-pop. This formulation allowed us to obtain the rank-one vectors (i.e., a_1, a_2, and a_3 in Fig. 1) that corresponded to the dictionary lyrics words reflecting the J-pop, Both, and K-pop lyrics words. To obtain the latent J-pop/K-pop lyrics words, the values of the mode-3 vectors (i.e., the vectors c_1, c_2, and c_3 in Fig. 1 bottom), which represent the country aspect, were adjusted to emphasize the J-pop, Both, and K-pop aspect. We set the mode-3 vector values as [[0.8, 0.1, 0.1], [0.1, 0.8, 0.1], [0.1, 0.1, 0.8]], and proceeded with the CP decomposition.

3 Results and Analyses

Table 7 lists the top-20 latent J-pop/K-pop lyrics noun words discovered using the proposed method. These latent country words were obtained by sorting the a_1 and a_3 rank-one vector values (Fig. 1 bottom) along with the corresponding lyrics words, and selecting the top-20 or bottom-20 lyrics noun words with the larger absolute top-most or bottom-most element value (i.e., the first and the 579th elements' absolute values were compared). We see in Table 7 an underlying theme in the latent J-pop/K-pop lyrics noun words.

Table 7. Top-20 latent J-pop & K-pop lyrics nouns.

Rank	J-pop lyrics nouns			K-pop lyrics nouns		
	English	**Japanese**	**Korean**	**English**	**Japanese**	**Korean**
1	petal	花びら	꽃잎	style	スタイル	스타일
2	*blood*	血	피	brother	兄ちゃん	오빠
3	cherry blossom	桜	벚꽃	bell	ベル	벨
4	light	光	빛	charm	魅力	매력
5	darkness	闇	어둠	shirt	シャツ	셔츠
6	quietness	静か	고요	impossibility	無理	무리
7	cloud	雲	구름	head	頭	머리
8	universe	宇宙	우주	toe	つま先	발끝
9	night sky	夜空	밤하늘	danger	危険	위험
10	winter	冬	겨울	taste	味	맛
11	color	色	색깔	clothes	服	옷
12	fire	火	불	woman	女	여자
13	star	星	별	sweat	汗	땀
14	wave	波	파도	money	お金	돈
15	sky	空	하늘	foot	足	발
16	needle	針	바늘	top	上	위
17	spring breeze	春風	봄바람	skin	肌	피부
18	spark	火花	불꽃	body	身	몸
19	rain	雨	비	*blood*	血	피
20	bottom	下	아래	music	音楽	음악

We see in Table 7 that nature related motifs such as ‘petal’, ‘cherry blossom’, ‘light’, ‘darkness’, ‘cloud’, ‘universe’, ‘night sky’, ‘fire’, ‘star’, ‘wave’, ‘sky’, ‘spring breeze’, and ‘rain’ are present in J-pop lyrics words while human related motifs such as ‘style’, ‘brother’, ‘charm’, ‘shirt’, ‘head’, ‘toe’, ‘clothes’, ‘women’, ‘sweat’, ‘foot’, ‘skin’, ‘body’, and ‘*blood*’ are present in K-pop lyrics words. ‘*Blood*’ is a common noun word listed in both countries.

Tables 8 and 9 list the top-10 latent J-pop/K-pop verbs and adjectives. Three verbs, ‘*fly*’, ‘*dance*’, and ‘*melt*’, appear in both countries. When we exclude these verbs, we are left with nature related verbs for J-pop (e.g., ‘[leaves] fall’, ‘[rain] fall’, ‘glitter’, ‘shine’ and ‘blow’) and human related verbs for K-pop (e.g., ‘wave’, ‘steal’, ‘fit’, ‘suit’, ‘wear’, ‘take off’, and ‘run’). In the case of adjectives, J-pop adjectives focus on color/vision related expressions (e.g., ‘white’, ‘stark white’, ‘red’, ‘blue’, ‘beautiful’, and ‘bright’) while K-pop adjectives focus on human perception based expressions (e.g., ‘cumbersome’, ‘delicious’, ‘stylish’, ‘pretty’, ‘hot’, ‘light’, ‘*cold*’, and ‘shy’). ‘*High*’ and ‘*cold*’ are common adjectives.

For reference, the top-20 latent words that encompass both J-pop and K-pop lyrics (i.e., ‘Both’ a_2 vector in Fig. 1) regardless of the parts of speech are ‘top’, ‘bottom’, ‘cherry blossom’, ‘[leaves] fall’, ‘mirror’, ‘spring breeze’, ‘flower’, ‘half’, ‘hi/bye’, ‘leaf’, ‘[tell a] lie’, ‘distance’, ‘style’, ‘shine’, ‘red’, ‘die’, ‘words’, ‘summer’, ‘love’, and ‘sorry’.

Table 8. Top-10 latent J-pop & K-pop lyrics verbs.

Rank	J-pop lyrics verbs			K-pop lyrics verbs		
	English	Japanese	Korean	English	Japanese	Korean
1	[leaves] fall	散る	흩날리다	wave	振る	흔들다
2	*fly*	飛ぶ	날다	steal	盗む	훔치다
3	[rain] fall	降る	내리다	*dance*	踊る	춤추다
4	glitter	光る	반짝이다	fit	合う	맞다
5	shine	照らす	비추다	[it] suit[-s you]	似合う	어울리
6	*dance*	踊る	춤추다	wear	着る	입다
7	echo	響く	울리다	take off [clothes]	脱ぐ	벗다
8	penetrate	突き抜く	뚫다	*melt*	溶ける	녹다
9	blow	吹く	불다	run	走る	뛰다
10	*melt*	溶ける	녹다	*fly*	飛ぶ	날다

Table 9. Top-10 latent J-pop & K-pop lyrics adjectives.

Rank	J-pop lyrics adjectives			K-pop lyrics adjectives		
	English	Japanese	Korean	English	Japanese	Korean
1	white	白い	하얗다	*high*	高い	높다
2	*high*	高い	높다	cumbersome	面倒い	귀찮다
3	stark white	真っ白い	새하얗다	large	大きい	크다
4	red	赤い	붉다	delicious	美味しい	맛있다
5	blue	青い	파랗다	stylish	かっこいい	멋있다
6	deep	深い	깊다	pretty	かわいい	예쁘다
7	beautiful	美しい	아름답다	hot	熱い	뜨겁다
8	spacious	広い	넓다	light	軽い	가볍다
9	bright	眩しい	눈부시다	*cold*	寒い	춥다
10	*cold*	寒い	춥다	shy	恥ずかしい	부끄럽다

4 Discussion and Conclusion

How can knowing the top-n latent country words help cross-cultural communication? Recall that the proposed method learns the hidden relationships between the words in each country's text in the form of word embeddings, and combining the two countries' word embeddings and performing partially fixed CP decomposition while adjusting the degree of country aspect enables the discovery of the substantial words hidden within each country's text. Such elicitation of the latent country words can aid in understanding the subtle differences underlying the two countries' texts. Revisiting the bilateral meeting minutes example presented at the outset of this paper, by using the proposed method, we could learn the hidden emphasis placed by each country's representatives on the common issue through analyzing the latent country words. We have demonstrated the potential of the proposed approach through the example task of finding latent J-pop/K-pop lyrics words; we discovered that the hidden motifs present in J-pop and K-pop lyrics texts were 'nature' and 'human'.

As the method analyzes the frequent and common words shared by the two countries' texts, the method can be viewed as finding each country's *dominant* concepts from a set of shared concepts; this is very different from finding *distinct* concepts that do not overlap across two countries. Moreover, the proposed method is unsupervised; both the word embedding learning and CP decomposition do not require any labeled training data (we assume that the country information can be automatically collected with the country texts). As long as there exists two countries' texts, the proposed method can be applied. However, the construction of reliable country-specific word embeddings is the key to finding meaningful latent country words, and thus a sufficient amount of two countries' texts is required for successful analysis.

Recent advances in information technology have enabled the easy collection of cross-cultural texts. Nowadays, there are abundant cross-cultural texts generated by two countries' participants that cover a common subject; examples range from two countries' bilateral meeting minutes to movie reviews written by two countries' audiences. We believe the method proposed in this paper can provide a viable data-driven approach to highlighting the hidden emphasis placed on a common subject discussed by the two countries' participants, and in turn aid in understanding the hidden differences underlying the two countries.

Acknowledgments. We thank the anonymous reviewers for many constructive comments. A heartwarming thanks to Yangjean Cho for revising and expanding the J-pop/K-pop lyrics word alignment dictionary.

References

1. Kiers, H.-A.-L., Van Mechelen, I.: Three-way component analysis: principles and illustrative application. Psychol. Methods **6**(1), 84–110 (2001)
2. Mørup, M.: Applications of tensor (multiway array) factorizations and decompositions in data mining. WIREs Data Min. Knowl. Discov. **1**(1), 24–40 (2011)
3. Carroll, J.-D., Chang, J.-J.: Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. Psychometrika **35**, 283–319 (1970)
4. Harshman, R.-A.: Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-modal factor analysis. UCLA Working Papers in Phonetics, vol. 16, pp. 1–84 (1970)
5. Kolda, T.-G., Bader, B.-W.: Tensor decompositions and applications. SIAM Rev. **51**, 455–500 (2009)
6. Cho, H., Yoon, S.-M.: Issues in visualizing intercultural dialogue using Word2Vec and t-SNE. In: Proceedings of 2017 International Conference on Culture & Computing, Kyoto, Japan, pp. 149–150 (2017)
7. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th International Conference on Neural Information Processing Systems, Nevada, USA, pp. 3111–3119 (2013)