LYRICSRADAR: A LYRICS RETRIEVAL SYSTEM BASED ON LATENT TOPICS OF LYRICS

Shoto Sasaki[*1] **Kazuyoshi Yoshii**[**2] **Tomoyasu Nakano**[***3] **Masataka Goto**[***4] **Shigeo Morishima**[*5]
[*]Waseda University [**]Kyoto University
[***]National Institute of Advanced Industrial Science and Technology (AIST)
[1]joudanjanai-ss[at]akane.waseda.jp [2]yoshii[at]kuis.kyoto-u.ac.jp
[3,4](t.nakano, m.goto)[at]aist.go.jp [5]shigeo[at]waseda.jp

ABSTRACT

This paper presents a lyrics retrieval system called *LyricsRadar* that enables users to interactively browse song lyrics by visualizing their topics. Since conventional lyrics retrieval systems are based on simple word search, those systems often fail to reflect user's intention behind a query when a word given as a query can be used in different contexts. For example, the word "tears" can appear not only in sad songs (*e.g.*, feel heartrending), but also in happy songs (*e.g.*, weep for joy). To overcome this limitation, we propose to automatically analyze and visualize topics of lyrics by using a well-known text analysis method called latent Dirichlet allocation (LDA). This enables *LyricsRadar* to offer two types of topic visualization. One is the topic radar chart that visualizes the relative weights of five latent topics of each song on a pentagon-shaped chart. The other is radar-like arrangement of all songs in a two-dimensional space in which song lyrics having similar topics are arranged close to each other. The subjective experiments using 6,902 Japanese popular songs showed that our system can appropriately navigate users to lyrics of interests.

1. INTRODUCTION

Some listeners regard lyrics as essential when listening to popular music. It was, however, not easy for listeners to find songs with their favorite lyrics on existing music information retrieval systems. They usually happen to find songs with their favorite lyrics while listening to music. The goal of this research is to assist listeners who think the lyrics are important to encounter songs with unfamiliar but interesting lyrics.

Although there were previous lyrics-based approaches for music information retrieval, they have not provided an interface that enables users to interactively browse lyrics of many songs while seeing latent topics behind those lyrics. We call these latent topics *lyrics topics*. Several

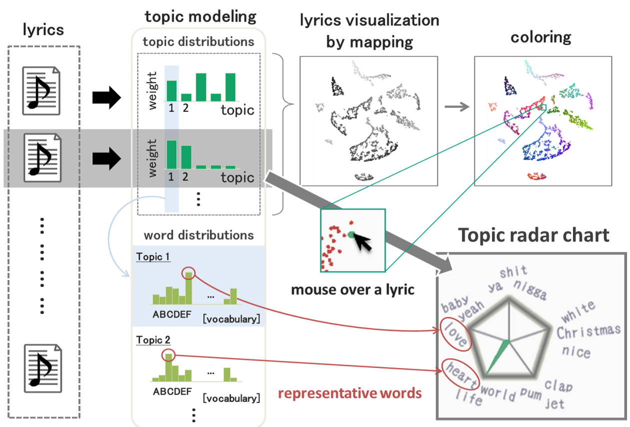

Figure 1. Overview of topic modeling of *LyricsRadar*.

approaches analyzed the text of lyrics by using natural language processing to classify lyrics according to emotions, moods, and genres [2, 3, 11, 19]. Automatic topic detection [6] and semantic analysis [1] of song lyrics have also been proposed. Lyrics can be used to retrieve songs [5] [10], visualize music archives [15], recommend songs [14], and generate slideshows whose images are matched with lyrics [16]. Some existing web services for lyrics retrieval are based on social tags, such as "love" and "graduation". Those services are useful, but it is laborious to put appropriate tags by hands and it is not easy to find a song whose tags are also put to many other songs. Macrae *et al.* showed that online lyrics are inaccurate and proposed a ranking method that considers their accuracy [13]. Lyrics are also helpful for music interfaces: LyricSynchronizer [8] and VocaRefiner [18], for example, show the lyrics of a song so that a user can click a word to change the current playback position and the position for recording, respectively. Latent topics behind lyrics, however, were not exploited to find favorite lyrics.

We therefore propose a lyrics retrieval system, *LyricsRadar*, that analyzes the lyrics topics by using a machine learning technique called latent Dirichlet allocation (LDA) and visualizes those topics to help users find their favorite lyrics interactively (Fig.1). A single word could have different topics. For example, "diet" may at least have two

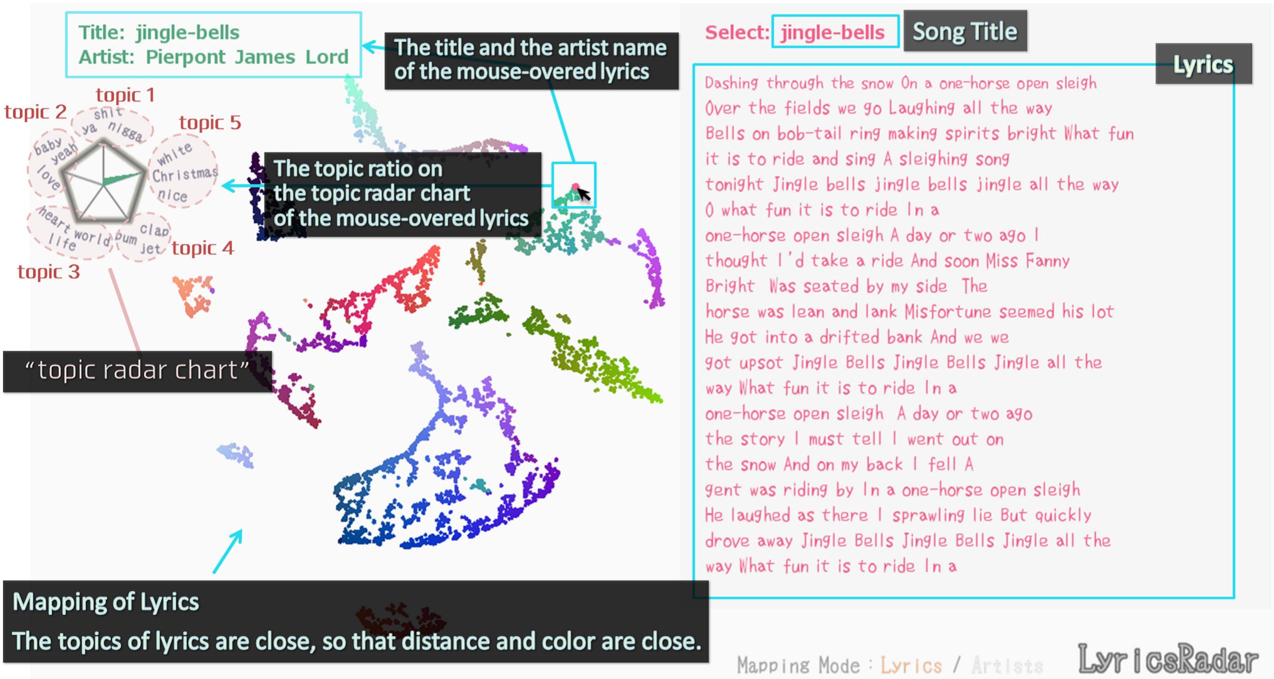

Figure 2. Example display of *LyricsRadar*.

lyrics topics. When it is used with words related to meal, vegetables, and fat, its lyrics topic “food and health” could be estimated by the LDA. On the other hand, when it is used with words like government, law, and elections, “politics” could be estimated. Although the LDA can estimate various lyrics topics, five typical topics common to all lyrics in a given database were chosen. The lyrics of each song are represented by the unique ratios of these five topics, which are displayed as pentagon-shaped chart called as a *topic radar chart*. This chart makes it easy to guess the meaning of lyrics before listening to its song. Furthermore, users can directly change the shape of this chart as a query to retrieve lyrics having a similar shape.

In *LyricsRadar*, all the lyrics are embedded in a two-dimensional space, mapped automatically based on the ratios of the five lyrics topics. The position of lyrics is such that lyrics in close proximity have similar ratios. Users can navigate in this plane by mouse operation and discover some lyrics which are located very close to their favorite lyrics.

2. FUNCTIONALITY OF LYRICSRADAR

LyricsRadar enables to bring a graphical user interface assisting users to navigate in a two dimensional space intuitively and interactively to come across the target song. This space is generated automatically by analysis of the topics which appear in common with the lyrics of many musical pieces in database using LDA. Also a latent meaning of lyrics is visualized by the topic radar chart based on the combination of topics ratios. Lyrics that are similar to a user's preference (*target*) can be intuitively discovered by clicking of the topic radar chart or lyrics representing by dots. So this approach cannot be achieved at all by the conventional method which directly searches for a song by the keywords or phrases appearing in lyrics. Since linguistic expressions of the topic are not necessary, user can find a target song intuitively even when user does not have any knowledge about lyrics.

2.1 Visualization based on the topic of lyrics

LyricsRadar has the following two visualization functions: (1) the topic radar chart; and (2) a mapping to the two-dimensional plane. Figure 2 shows an example display of our interface. The topic radar chart shown in upper-left corner of Figure 2 is a pentagon-shape chart which expresses the ratio of five topics of lyrics. Each colored dot displayed in two dimensional plane shown in Figure 2 means the relative location of lyrics in a database. We call these colored dot representations of lyrics *lyrics dot*. User can see lyrics, its title and artist name, and the topic ratio by clicking the lyrics dot placed on the 2D space, this supports to discover lyrics interactively. While the lyrics mapping assists user to understand the lyrics topic by the relative location in the map, the topic radar chart helps to get the lyrics image intuitively by the shape of chart. We explain each of these in the following subsections.

2.1.1 Topic radar chart

The values of the lyrics topic are computed and visualized as the topic radar chart which is pentagon style. Each vertex of the pentagon corresponds to a distinct topic, and predominant words of each topic (*e.g.*, “heart”, “world”, and “life” for the topic 3) are also displayed at the five corner of pentagon shown in Figure 2. The predominant words help user to guess the meaning of each topic. The center

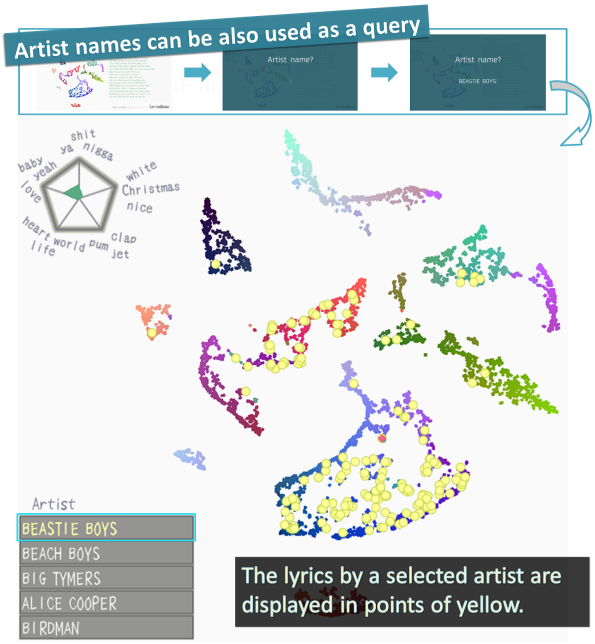

Figure 3. An example display of lyrics by a selected artist.

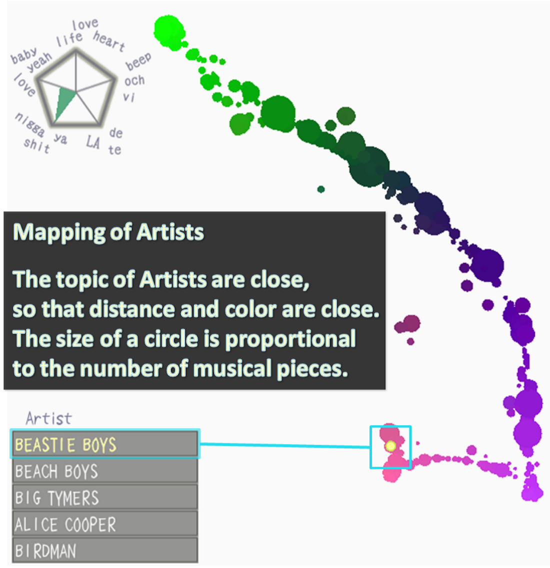

Figure 4. Mapping of 487 English artists.

of the topic radar chart indicates 0 value of a ratio of the lyrics topic in the same manner as the common radar chart. Since the sum of the five components is a constant value, if the ratio of a topic stands out, it will clearly be seen by the user. It is easy to grasp the topic of selected lyrics visually and to make an intuitive comparison between lyrics.

Furthermore, the number of topics in this interface is set to five to strike a balance between the operability of interface and the variety of topics[1] .

2.1.2 *Plane-mapped lyrics*

The lyrics of musical pieces are mapped onto a two-dimensional plane, in which musical pieces with almost the same topic ratio can get closer to each other. Each musical piece is expressed by colored dot whose RGB components are corresponding to 3D compressed axis for five topics' values. This space can be scalable so that the local or global structure of each musical piece can be observed. The distribution of lyrics about a specific topic can be recognized by the color of the lyrics. The dimension compression in mapping and coloring used t-SNE [9]. When a user mouseovers a point in the space, it is colored pink and meta-information about the title, artist, the topic radar chart appears simultaneously.

By repeating mouseover, lyrics and names of its artist and songwriter are updated continuously. Using this approach, other lyrics with the similar topics to the input lyrics can be discovered. The lyrics map can be moved and zoomed by dragging the mouse or using a specific keyboard operation. Furthermore, it is possible to visualize the lyrics map specialized to artist and songwriter, which are associated with lyrics as metadata. When an artist name is chosen, as shown in the right side of Figure 3, the point of the artist's lyrics will be getting yellow; similarly, when a songwriter is chosen, the point of the songwriter's lyrics will be changed to orange. While this is somewhat equivalent to lyrics retrieval using the artist or songwriter as a query, it is our innovative point in the sense that a user can intuitively grasp how artists and songwriters are distributed based on the ratio of the given topic. Although music retrieval by artist is very popular in a conventional system, a retrieval by songwriter is not focused well yet. However, in the meaning of lyrics retrieval, it is easier for search by songwriter to discover songs with one's favorite lyrics because a songwriter has his own lyrics vocabulary.

Moreover, we can make a topic analysis depending on a specific artist in our system. Intuitively similar artists are also located and colored closer in a topic chart depending on topic ratios. The artist is colored based on a topic ratio in the same way as that of the lyrics. In Figure 4, the size of a circle is proportional to the number of musical pieces each artist has. In this way, other artists similar to one's favorite artist can be easily discovered.

2.2 Lyrics retrieval using topic of lyrics

In *LyricsRadar*, in addition to the ability to traverse and explore a map to find lyrics, we also propose a system to directly enter a topic ratio as an intuitive expression of one's latent feeling. More specifically, we consider the topic radar chart as an input interface and provide a means by which a user can give topic ratios for five elements directly to search for lyrics very close to one's latent image. This interface can satisfy the search query in which a user would like to search for lyrics that contain more of the same topics using the representative words of each topic. Figure 5 shows an example in which one of the five topics is increased by mouse drag, then the balance of five topics ratio

[1] If the number of topics was increased, a more subdivided and exacting semantic content could have been represented; however, the operation for a user will be getting more complicated.

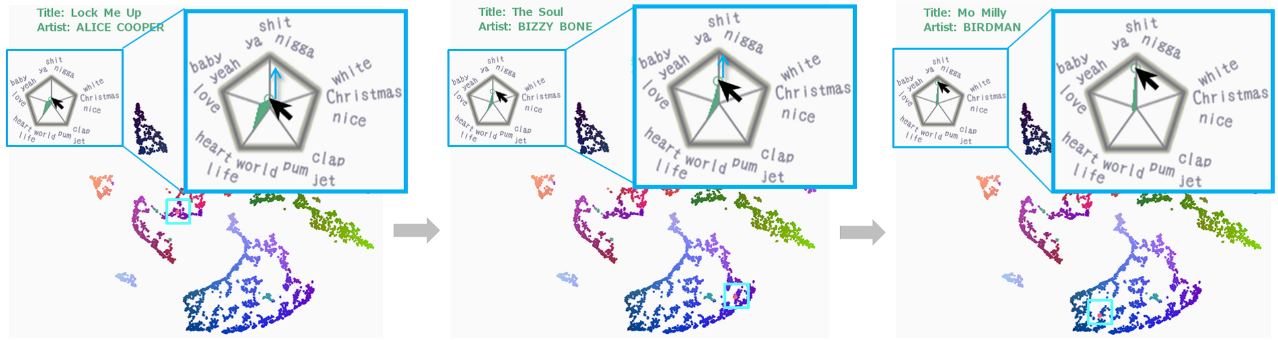

Figure 5. An example of the direct manipulation of the topic ratio on the topic radar chart. Each topic ratio can be increased by dragging the mouse.

has changed because the sum of five components is equal to 1.0. A user can repeat these processes by updating topics ratios or navigating the point in a space interactively until finding interesting lyrics. As with the above subsections, we have substantiated our claims for a more intuitive and exploratory lyrics retrieval system.

3. IMPLEMENTATION OF LYRICSRADAR

LyricsRadar used LDA [4] for the topic analysis of lyrics. LDA is a typical topic modeling method by machine learning. Since LDA assigns each word which constitutes lyrics to a different topic independently, the lyrics include a variety of topics according to the variation of words in the lyrics. In our system, K typical topics which constitute many lyrics in database are estimated and a ratio to each topic is calculated for lyrics with unsupervised learning. As a result, appearance probability of each word in every topic can be calculated. The typical representative word to each topic can be decided at the same time.

3.1 LDA for lyrics

The observed data that we consider for LDA are D independent lyrics $\mathbf{X} = \{\mathbf{X}_1, ..., \mathbf{X}_D\}$. The lyrics $\mathbf{X}_d$ consist of N_d word series $\mathbf{X}_d = \{\mathbf{x}_{d,1}, ..., \mathbf{x}_{d,N_d}\}$. The size of all vocabulary that appear in the lyrics is V, $\mathbf{x}_{d,n}$ is a V-dimensional "1-of-K" vector (a vector with one element containing 1 and all other elements containing 0). The latent variable (i.e., the topics series) of the observed lyrics $\mathbf{X}_d$ is $\mathbf{Z}_d = \{\mathbf{z}_{d,1}, ..., \mathbf{z}_{d,N_d}\}$. The number of topics is K, so $\mathbf{z}_{d,n}$ indicates a K-dimensional 1-of-K vector. Hereafter, all latent variables of lyrics D are indicated $\mathbf{Z} = \{\mathbf{Z}_1, ..., \mathbf{Z}_D\}$. Figure 6 shows a graphical representation of the LDA model used in this paper. The full joint distribution is given by

$$p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\phi}) = p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\phi})p(\mathbf{Z}|\boldsymbol{\pi})p(\boldsymbol{\pi})p(\boldsymbol{\phi}) \quad (1)$$

where $\boldsymbol{\pi}$ indicates the mixing weights of the multiple topics of lyrics (D of the K-dimensional vector) and $\boldsymbol{\phi}$ indicates the unigram probability of each topic (K of the V-dimensional vector). The first two terms are likelihood functions, whereas the other two terms are prior distributions. The likelihood functions themselves are defined as

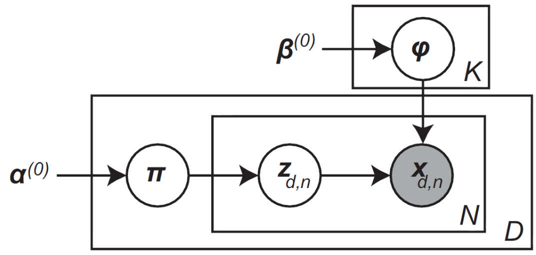

Figure 6. Graphical representation of the latent Dirichlet allocation (LDA).

$$p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\phi}) = \prod_{d=1}^{D}\prod_{n=1}^{N_d}\prod_{v=1}^{V}\left(\prod_{k=1}^{K}\phi_{k,v}^{z_{d,n,k}}\right)^{x_{d,n,v}} \quad (2)$$

$$p(\mathbf{Z}|\boldsymbol{\pi}) = \prod_{d=1}^{D}\prod_{n=1}^{N_d}\prod_{k=1}^{K}\pi_{d,k}^{z_{d,n,k}} \quad (3)$$

We then introduce conjugate priors as

$$p(\boldsymbol{\pi}) = \prod_{d=1}^{D}\mathrm{Dir}(\boldsymbol{\pi}_d|\boldsymbol{\alpha}^{(0)}) = \prod_{d=1}^{D}C(\boldsymbol{\alpha}^{(0)})\prod_{k=1}^{K}\pi_{d,k}^{\boldsymbol{\alpha}^{(0)}-1} \quad (4)$$

$$p(\boldsymbol{\phi}) = \prod_{k=1}^{K}\mathrm{Dir}(\boldsymbol{\phi}_k|\boldsymbol{\beta}^{(0)}) = \prod_{k=1}^{K}C(\boldsymbol{\beta}^{(0)})\prod_{v=1}^{V}\phi_{k,v}^{\boldsymbol{\beta}_v^{(0)}-1} \quad (5)$$

where $p(\boldsymbol{\pi})$ and $p(\boldsymbol{\phi})$ are products of Dirichlet distributions, $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\beta}^{(0)}$ are hyperparameters, and $C(\boldsymbol{\alpha}^{(0)})$ and $C(\boldsymbol{\beta}^{(0)})$ are normalization factors calculated as follows:

$$C(\mathbf{x}) = \frac{\Gamma(\hat{x})}{\Gamma(x_1)\cdots\Gamma(x_I)}, \quad \hat{x} = \sum_{i=1}^{I} x_i \quad (6)$$

Also note that $\boldsymbol{\pi}$ is the topic mixture ratio of lyrics used as the topic radar chart by normalization. The appearance probability $\boldsymbol{\phi}$ of the vocabulary in each topic was used to evaluate the high-representative word that is strongly correlated with each topic of the topic radar chart.

3.2 Training of LDA

The lyrics database contains 6902 Japanese popular songs (J-POP) and 5351 English popular songs. Each of these songs includes more than 100 words. J-POP songs are selected from our own database and English songs are from *Music Lyrics Database v.1.2.7*[2] . J-POP database has 1847 artists and 2285 songwriters and English database has 398 artists. For the topic analysis per artist, 2484 J-POP artists and 487 English artists whose all songs include at least 100 words are selected. 26229 words in J-POP and 35634 words in English which appear more than ten times in all lyrics is used for the value V which is the size of vocabulary in lyrics. In J-POP lyrics, MeCab [17] was used for the morphological analysis of J-POP lyrics. The noun, verb, and adjective components were extracted and then the original and the inflected form were counted as one word. In English lyrics, we use stopwords using *Full-Text Stopwords in MySQL*[3] to remove commonly-used words. However, words which appeared often in many lyrics were inconvenient to analyze topics. To lower the importance of such words in the topic analysis, they were weighted by inverse document frequency (idf).

In the training the LDA, the number of topics (K) is set to 5. All initial values of hyperparameters $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\beta}^{(0)}$ were set to 1.

4. EVALUATION EXPERIMENTS

To verify the validity of the topic analysis results (as related to the topic radar chart and mapping of lyrics) in *LyricsRadar*, we conducted a subjective evaluation experiment. There were 17 subjects (all Japanese speakers) with ages from 21 to 32. We used the results of LDA for the lyrics of the 6902 J-POP songs described in Section 3.2.

4.1 Evaluation of topic analysis

Our evaluation here attempted to verify that the topic ratio determined by the topic analysis of LDA could appropriately represent latent meaning of lyrics. Furthermore, when the lyrics of a song are selected, relative location to other lyrics of the same artist or songwriter in the space is investigated.

4.1.1 Experimental method

In our experiment, the lyrics of a song are selected at random in the space as *basis* lyrics and also *target* lyrics of four songs are selected to be compared according to the following conditions.

(1) The lyrics closest to the *basis* lyrics on lyrics map

(2) The lyrics closest to the *basis* lyrics with same songwriter

(3) The lyrics closest to the *basis* lyrics with same artist

(4) The lyrics selected at random

[2] "Music Lyrics Database v.1.2.7," `http://www.odditysoftware.com/page-datasales1.htm`.
[3] "Full-Text Stopwords in MySQL," `http://dev.mysql.com/doc/refman/5.5/en/fulltext-stopwords.html`.

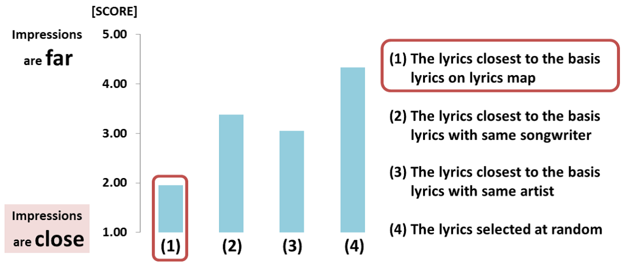

Figure 7. Results of our evaluation experiment to evaluate topic analysis; the score of (1) was the closest to 1.0, showing our approach to be effective.

Each subject evaluated the similarity of the impression received from the two lyrics using a five-step scale (1: closest, 2: somehow close, 3: neutral, 4: somehow far, and 5: most far), comparing the *basis* lyrics and one of the *target* lyrics after seeing the *basis* lyrics. Presentation order to subjects was random. Furthermore, each subject described the reason of evaluation score.

4.1.2 Experimental results

The average score of the five-step evaluation results for the four *target* lyrics by all subjects is shown in the Figure 7. As expected, lyrics closest to the *basis* lyrics on the lyrics map were evaluated as the closest in terms of the impression of the *basis* lyrics, because the score of (1) was closest to 1.0. Results of *target* lyrics (2) and (3) were both close to 3.0. The lyrics closest to the *basis* lyrics of the same songwriter or artist as the selected lyrics were mostly judged as "3: neutral." Finally, the lyrics selected at random (4) were appropriately judged to be far.

As the subjects' comments about the reason of decision, we obtained such responses as a sense of the season, positive-negative, love, relationship, color, light-dark, subjective-objective, and tension. Responses differed greatly from one subject to the next. For example, some felt the impression only by the similarity of a sense of the season of lyrics. Trial usage of LyricsRadar has shown that it is a useful tool for users.

4.2 Evaluation of the number of topics

The perplexity used for the quality assessment of a language model was computed for each number of topics. The more the model is complicated, the higher the perplexity becomes. Therefore, we can estimate that the performance of language model is good when the value of perplexity is low. We calculated perplexity as

$$perplexity(\mathbf{X}) = \exp\left(-\frac{\sum_{d=1}^{D} \log p(X_d)}{\sum_{d=1}^{D} N_d}\right) \quad (7)$$

In case the number of topics (K) is five, the perplexity is 1150 which is even high.

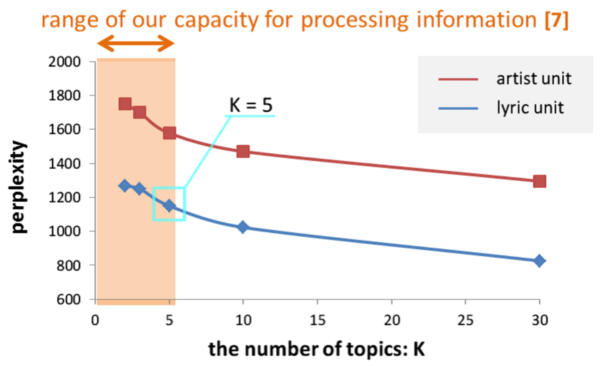

Figure 8. Perplexity for the number of topics.

On the other hand, because Miller showed that the number of objects human can hold in his working memory is 7 ± 2 [7], the number of topics should be 1 to 5 in order to obtain information naturally. So we decided to show five topics in the topic radar chart.

Figure 8 shows calculation results of perplexity for each topic number. Blue points represent perplexity for LDA applied to lyrics and red points represent perplexity for LDA applied to each artist. Orange bar indicates the range of human capacity for processing information. Since there exists a tradeoff between the number of topics and operability, we found that five is appropriate number of topics.

5. CONCLUSIONS

In this paper, we propose *LyricsRadar*, an interface to assist a user to come across favorite lyrics interactively. Conventionally lyrics were retrieved by titles, artist names, or keywords. Our main contribution is to visualize lyrics in the latent meaning level based on a topic model by LDA. By seeing the pentagon-style shape of Topic Radar Chart, a user can intuitively recognize the meaning of given lyrics. The user can also directly manipulate this shape to discover *target* lyrics even when the user does not know any keyword or any query. Also the topic ratio of focused lyrics can be mapped to a point in the two dimensional space which visualizes the relative location to all the lyrics in our lyrics database and enables the user to navigate similar lyrics by controlling the point directly.

For future work, user adaptation is inevitable task because every user has an individual preference, as well as improvements to topic analysis by using hierarchical topic analysis [12]. Furthermore, to realize the retrieval interface corresponding to a minor topic of lyrics, a future challenge is to consider the visualization method that can reflect more numbers of topics by keeping an easy-to-use interactivity.

Acknowledgment: This research was supported in part by OngaCREST, CREST, JST.

6. REFERENCES

[1] B. Logan *et al.*: "Semantic Analysis of Song Lyrics," *Proceedings of IEEE ICME 2004* Vol.2, pp. 827–830, 2004.

[2] C. Laurier *et al.*: "Multimodal Music Mood Classification Using Audio and Lyrics," *Proceedings of ICMLA 2008*, pp. 688–693, 2008.

[3] C. McKay *et al.*: "Evaluating the genre classification performance of lyrical features relative to audio, symbolic and cultural features," *Proceedings of ISMIR 2008*, pp. 213–218, 2008.

[4] D. M. Blei *et al.*: "Latent Dirichlet Allocation," *Journal of Machine Learning Research* Vol.3, pp. 993–1022, 2003.

[5] E. Brochu and N. de Freitas: ""Name That Song!": A Probabilistic Approach to Querying on Music and Text," *Proceedings of NIPS 2003*, pp. 1505–1512, 2003.

[6] F. Kleedorfer *et al.*: "Oh Oh Oh Whoah! Towards Automatic Topic Detection In Song Lyrics," *Proceedings of ISMIR 2008*, pp. 287–292, 2008.

[7] G. A. Miller: "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Journal of the Psychological Review* Vol.63(2), pp. 81–97, 1956.

[8] H. Fujihara *et al.*: "LyricSynchronizer: Automatic Synchronization System between Musical Audio Signals and Lyrics," *Journal of IEEE Selected Topics in Signal Processing*, Vol.5, No.6, pp. 1252–1261, 2011.

[9] L. Maaten and G. E. Hinton: "Visualizing High-Dimensional Data Using t-SNE," *Journal of Machine Learning Research*, Vol.9, pp. 2579–2605, 2008.

[10] M. Müller *et al.*: "Lyrics-based Audio Retrieval and Multimodal Navigation in Music Collections," *Proceedings of ECDL 2007*, pp. 112–123, 2007.

[11] M. V. Zaanen and P. Kanters: "Automatic Mood Classification Using TF*IDF Based on Lyrics," *Proceedings of ISMIR 2010*, pp. 75–80, 2010.

[12] R. Adams *et al.*: "Tree-Structured Stick Breaking Processes for Hierarchical Data," *Proceedings of NIPS 2010*, pp. 19–27, 2010.

[13] R. Macrae and S. Dixon: "Ranking Lyrics for Online Search," *Proceedings of ISMIR 2012*, pp. 361–366, 2012.

[14] R. Takahashi *et al.*: "Building and combining document and music spaces for music query-by-webpage system," *Proceedings of Interspeech 2008*, pp. 2020–2023, 2008.

[15] R. Neumayer and A. Rauber: "Multi-modal Music Information Retrieval: Visualisation and Evaluation of Clusterings by Both Audio and Lyrics," *Proceedings of RAO 2007*, pp. 70–89, 2007.

[16] S. Funasawa *et al.*: "Automated Music Slideshow Generation Using Web Images Based on Lyrics," *Proceedings of ISMIR 2010*, pp. 63–68, 2010.

[17] T. Kudo: "MeCab: Yet Another Part-of-Speech and Morphological Analyzer," `http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html`.

[18] T. Nakano and M. Goto: "VocaRefiner: An Interactive Singing Recording System with Integration of Multiple Singing Recordings," *Proceedings of SMC 2013*, pp. 115–122, 2013.

[19] Y. Hu *et al.*: "Lyric-based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method," *Proceedings of ISMIR 2009*, pp. 122–128, 2009.