PAPER

Modeling Storylines in Lyrics

Kento WATANABE[†a)], **Yuichiroh MATSUBAYASHI**[†], ***Nonmembers***, **Kentaro INUI**[†,††], **Satoru FUKAYAMA**[†††], ***Members***, **Tomoyasu NAKANO**[†††], ***Nonmember***, ***and*** **Masataka GOTO**[†††], ***Member***

SUMMARY This paper addresses the issue of modeling the discourse nature of lyrics and presented the first study aiming at capturing the two common discourse-related notions: storylines and themes. We assume that a storyline is a chain of transitions over topics of segments and a song has at least one entire theme. We then hypothesize that transitions over topics of lyric segments can be captured by a probabilistic topic model which incorporates a distribution over transitions of latent topics and that such a distribution of topic transitions is affected by the theme of lyrics. Aiming to test those hypotheses, this study conducts experiments on the word prediction and segment order prediction tasks exploiting a large-scale corpus of popular music lyrics for both English and Japanese (around 100 thousand songs). The findings we gained from these experiments can be summarized into two respects. First, the models with topic transitions significantly outperformed the model without topic transitions in word prediction. This result indicates that typical storylines included in our lyrics datasets were effectively captured as a probabilistic distribution of transitions over latent topics of segments. Second, the model incorporating a latent theme variable on top of topic transitions outperformed the models without such variables in both word prediction and segment order prediction. From this result, we can conclude that considering the notion of theme does contribute to the modeling of storylines of lyrics.

key words: *bayesian model, generative model, lyrics structure, lyrics understanding, natural language processing*

1. Introduction

Lyrics are an important element in popular music that conveys stories and expresses emotion. Unlike prose text, lyrics are created in consideration of its specific properties such as fitness to rhythm and melody, and rhetoric with rhyme, refrain and repetition [1], [2]. In writing a piece of lyrics for a given piece of music, the writer should select right words such that their syllables fit the rhythm or melody of the music. The writer may also consider using rhymes, refrains or repetitions to color the entire story rhetorically as in the example lyrics shown in Fig. 1, where rhymes can be seen at *night*, *light* and *tight* in Segment 2). Writing lyrics is thus a complex task.

Segment 1 (introduces the story)

From the night we first met
Your smile's been the meaning of my life
With just a glance I knew you were mine

Segment 2 (describes a past event)

I was all alone in the night
And it was you who gave me the light
Let me hold you, oh so tight
Don't let a bit of misunderstanding come in between us

Segment 3 (expresses an emotion)

Don't you never ever say goodbye
Don't tell me that's the thing that's on your mind
We both know we are made for each other
So don't leave me behind

Fig. 1 Lyrics with a storyline (title: *Don't Say Good bye* (RWC-MDB-P-2001 No. 90 from RWC Music Database [17])).

These characteristics of lyrics have been motivating a range of research for computer-based modeling of lyrics and computer-assisted or fully-automated creation of lyrics [3]–[10]. In particular, building a computational model of lyrics is an important research goal. Once a reasonably sophisticated computational model of lyrics is obtained, the model will provide us a better understanding of the nature and structure of lyrics, which will then allow us to consider building computer systems which can enhance the creativity of human lyrics writers. In reality, however, while an increasing number of papers have been published for demonstrating computer systems that automatically generate lyrics or assist human lyricists [3]–[10], research for modeling lyrics and understanding their properties is still limited [11]–[15].

One crucial issue we miss in previous studies is modeling the nature of lyrics as *discourse*. Similar to prose text, a piece of lyrics typically comprises discourse segments; namely, lyrics of popular music typically has *verse*, *bridge* and *chorus* segments [16] and such segments may comprise more fine-grained segments as in Fig. 1. Each segment provides part of the entire story and the segments are organized (or sequentially ordered) so as to constitute a coherent structure as a whole. In spite of its importance, however, no prior study has ever addressed the issue of modeling this discourse-oriented nature of lyrics.

Motivated by this background, in this paper, we report on our novel study for building a computational model of the discourse nature of lyrics. We focus on two notions which characterize lyrical discourse: *storyline* and *theme*. Both notions are described in textbooks on lyrics writing [1], [2].

A segment of lyrics is assumed to have its own pur-

Manuscript received June 10, 2017.
Manuscript revised November 28, 2017.
Manuscript publicized December 22, 2017.

[†] The authors are with Tohoku University, Sendai-shi, 980–8579 Japan.
[††] The author is with RIKEN Center for Advanced Intelligence Project, Tokyo, 103–0027 Japan.
[†††] The authors are with National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba-shi, 305–8568 Japan.
a) E-mail: kento.w@ecei.tohoku.ac.jp

DOI: 10.1587/transinf.2017EDP7188

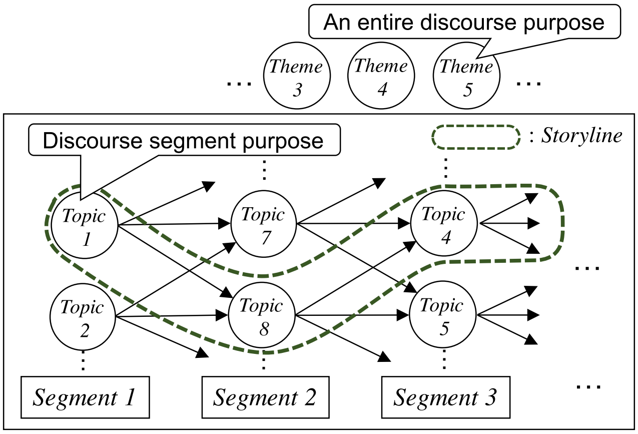


Fig. 2 The notion of storylines and themes in lyrics.

pose, which corresponds to a discourse segment purpose in terms of discourse analysis research [18]. In Fig. 1, for example, Segment 1 introduces the story, Segment 2 retrospects a past event, and Segment 3 expresses an emotion which arises from the retrospection. We model a storyline as such a chain of coherent shifts between discourse segment purposes. Specifically, we capture typical types of discourse segment purposes as *latent topics* by applying topic modeling techniques [19] to a large collection of lyrical texts, and then model typical storylines of lyrics as a probability distribution over the transition of latent topics over successive segments (Fig. 2). On top of storylines, we additionally consider the notion of theme, which we assume to be an entire discourse purpose. We assume that each song has at least one theme and each theme affects the distribution over both topic transitions and word choices. For the lyrics in Fig. 1, for example, our model provides a result with which we can understand its theme as "Sweet Love" and estimates the *theme-sensitive* distributions over topic transitions and word choices.

In order to examine how well our model of lyrics fit real-world data, we experiment with two distinct prediction tasks, word prediction and segment order prediction, and compare four variant models with different settings for considering storylines and themes. In the experiments, the models were trained with unsupervised learning over a large-scale corpus of popular music lyrics for both English and Japanese (around 100 thousand songs). The results demonstrate that the consideration of storylines (topic transitions) and themes contributes to improved prediction performance.

In what follows, we review related work in Sect. 2 and describe our novel method of modeling lyrics in Sect. 3. We then present our experiments in Sect. 4 before concluding this study in Sect. 5.

2. Related Work

2.1 Modeling Structure of Lyrics

Plenty of studies for capturing lyric-specific properties have been reported, where a broad range of music elements including meter, rhythm, rhyme, stressed/unstressed syllables, and accent are studied. Reddy and Knight [14] developed a language-independent rhyme model based on a Markov process that finds rhyme schemes. Greene *et al.* [13] employed a finite-state transducer to assign syllable-stress pattern to all words in each line. Nichols *et al.* [12] identified several patterns in the relationship between the lyrics and melody in popular music by measuring the correlation between textual salience and musical salience. Mayer *et al.* [11] trained a support vector machine to classify music genres using only textual features such as rhyme and part-of-speech patterns. Barbieri *et al.* [5], Abe and Ito [6], and Ramakrishnan and Devi [4] generated English, Japanese and Tamil lyrics that satisfy a given input constraint, such as rhyme, rhythm, and part-of-speech templates. Wu *et al.* [7] applied stochastic transduction grammar induction algorithms to generate fluent rhyming hip hop lyrics. Watanabe *et al.* [15] proposed a computational model that predicts segment boundaries in lyrics by utilizing a self-similarity matrix, which is frequently used in audio-based music structure analysis.

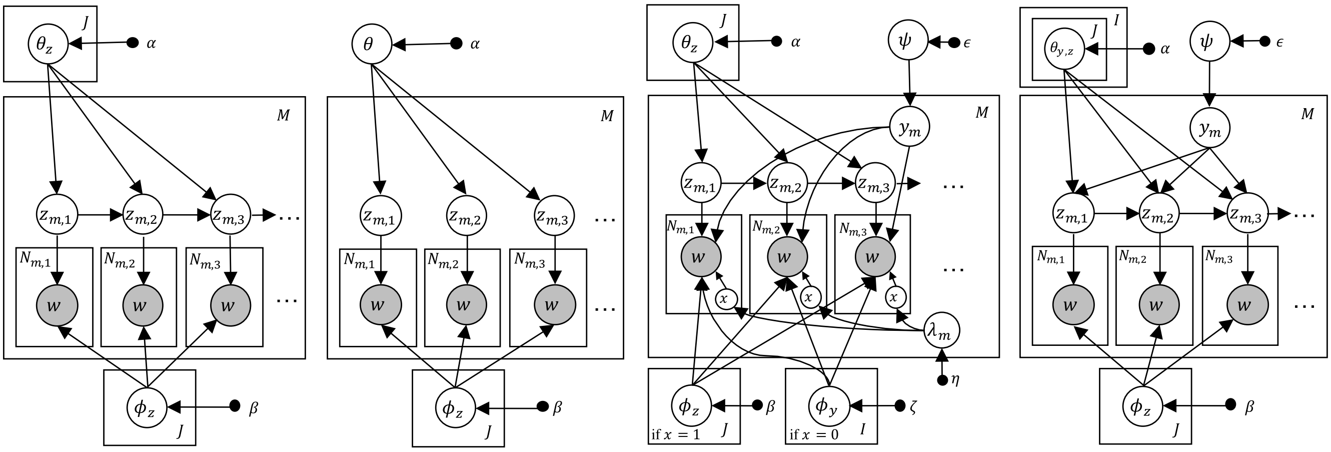
Several studies aim at modeling semantic structure of lyrics. Kleedorfer *et al.* [20] classified lyrics according to topical clusters calculated using nonnegative matrix factorization [21]. Sasaki *et al.* [22] visualized lyric clusters calculated using Latent Dirichlet Allocation (LDA) [19].

These prior studies share the motivation of modeling lyric-specific structure with our study. However, they have not considered capturing the discourse-oriented nature of lyrics whereas our study aims at modeling storylines (i.e. coherent transitions of discourse segment purposes) and themes (i.e. entire discourse purposes) of lyrics.

2.2 Modeling Topic Sequence

Transition of topics has been studied in the context of topic modeling for sequential text data such as newspapers, weblogs, and conversations. Iwata *et al.* [23] proposed a Topic Tracking Model (TTM), an extension of LDA, to models topic sequences. In the TTM, the topic distribution changes at each time. The TTM analyzes changes in user interest (e.g., interest in weblogs and microblogs). Blei and Lafferty [24] proposed the Dynamic Topic Model (DTM), which is similar to the TTM. In the DTM, the prior distribution of topic distribution changes at each time. The DTM analyzes changes in topic over time (e.g., topics in news articles and academic papers). The TTM and DTM have a topic distribution for a specific date (e.g., the DTM can train the topic distribution in a given period). Although the DTM and TTM can represent the topic sequence, extending these model to lyrics is difficult because, in lyrics, a segment's topic is time-independent.

Barzilay *et al.* proposed Content Model (CM), which is typically used for discourse analysis [25], to model topic sequences in documents without date information. CMs are sentence-level hidden Markov models that capture the sequential topic structure of a news event, such as earthquakes.



(a) Content Model (CM)
(b) Mixture of Segment Model (MSM)
(c) Mixture of Unigram and Content Model (MUM-CM)
(d) Mixture of Content Model (MCM)

Fig. 3 Plate notation of base models (a, b) and the proposed combination models (c, d). The shaded nodes are observed variables, dots are hyperparameters, x, y, and z are latent variables, and ψ, θ, ϕ, and λ are probability distributions.

Several studies extended Barzilay's model to dialog acts (e.g., questions and responses) [26], [27]. Ritter *et al.* [26] assumed that an observed sentence is generated from either a dialog act-specific language model (e.g., questions and responses) or a dialog-specific language model (e.g., food and music). Zhai and Williams [27] assumed that an observed word is generated from either a CM or an LDA and modeled the latent structure in task-oriented dialog. In their study, the sequential structure of dialog is modeled as a transition distribution.

We share the core concept as these studies and apply a CM to lyrics to model storylines (See Sect. 3.2). We then extend the CM to capture theme and investigate the effects of considering themes on top of storylines in our experiments.

3. Model Construction

Our final objective is to model the storyline of lyrics. However, precise modeling and representation of storylines remain an open issue. As mentioned previously, lyricists consider the order of topics when creating storylines; if the order changes, the content of the lyrics also changes. Therefore, we assume that a better storyline model can be used to predict the order of segments and the words in lyrics.

Based on the above assumption, we explore different topic sequence models to improve prediction performance. Lyricists often consider the order of topics when they create storylines; therefore, we assume that topic sequences can be represented as a probabilistic distribution of transitions over latent states. Since lyricists often assign a certain role to each segment, we assume that the segment is in one latent variable for a given lyrical content and words are derived from each latent state. Moreover, we assume that lyrics in a song are in one latent variable (i.e., a theme) because lyricists often create storylines according to themes. Based on the above idea, we prepared four data-driven Bayesian models. By comparing the performance of the models, we investigate which encoding method can better model the storyline.

In the following, we first describe the notations used in this study and two baseline methods for modeling the storyline. Finally, we propose two extended combination models to handle theme and storyline simultaneously.

3.1 Preliminaries

We assume that we have a set of M lyrics (songs). The lyric is an index between 1 and M, where M is the number of songs. The m-th lyric contains S_m segments and has a single theme denoted as the latent variable y_m. The theme is an index between 1 and I, where I is the number of themes. The s-th segment contains a bag of words denoted as $\{w_{m,s,1}, w_{m,s,2}, \ldots, w_{m,s,N_{m,s}}\}$, where $w_{m,s,n}$ is an index between 1 and V, where V is the vocabulary size. $N_{m,s}$ is the number of words in the s-th segment of the m-th lyric. In addition, each segment has a single topic denoted as the latent variable $z_{m,s}$. The topic is an index between 1 and J, where J is the number of topics. The storyline is represented as the sequence of a segment's topic denoted as $\mathbf{z}_m = z_{m,1}, z_{m,2}, \ldots, z_{m,S_m}$.

3.2 Base Model 1: Content Model

We use the CM [25] as a baseline model for the storyline of lyrics because this model is the simplest topic transition model that satisfies our assumption that *the topic sequence can be encoded as probabilistic latent state transition*. As shown in Fig. 3 (a), we assume that the storyline can be generated from a topic transition distribution θ_z. For the s-th segment in the m-th lyric, each topic $z_{m,s}$ is generated from the previous topic $z_{m,s-1}$ via the transition probability $P(z_{m,s}|\theta_{z_{m,s-1}})$. This probability is calculated by

J-dimensional multinomial distribution θ_z drawn from Dirichlet distributions with symmetric hyperparameter α. Then the word $w_{m,s,n}$ in each segment is generated from each topic $z_{m,s}$ via topic-specific generative probability $P(w_{m,s,n}|\phi_{z_{m,s}})$. This probability is calculated by V-dimensional multinomial distribution ϕ_z drawn from Dirichlet distributions with symmetric hyperparameter β.

3.3 Base Model 2: Mixture of Segment Model

To investigate the effects of capturing topic transitions, we also build a model that removes topic transitions from the CM (Fig. 3 (b)). We refer to this model as the Mixture of Segment Model (MSM). In the MSM, each segment's topic $z_{m,s}$ is generated via the probability without transition $P(z_{m,s}|\theta)$. This probability is calculated by J-dimensional multinomial distribution θ drawn from Dirichlet distributions with symmetric hyperparameter α. In other words, the MSM has only one probability distribution θ while the CM has J probability distributions θ_z.

3.4 Proposed Model 1: Mixture of Unigram and Content Model

To verify that both theme and topic transition are useful for modeling the storyline, we propose a model that combines the theme and the topic transition simultaneously, and we compare this model to the baseline models. The idea behind this combined modeling is that we can mix a theme-specific model and the topic transition model (i.e., the CM) using linear interpolation assuming that words in lyrics are dependent on both the theme and the topic.

We use the Mixture of Unigram Model (MUM) [28] as the theme-specific model because it is the simplest model that satisfies our assumption; *lyrics in a song are in a single latent variables (i.e., the theme)*. The MUM assumes that theme y_m is drawn from an I-dimensional theme distribution ψ and all words in the lyrics are drawn from V-dimensional multinomial distribution ϕ_{y_m} as shown in Fig. 3 (c).

In the proposed MUM-CM, we define a binary variable $x_{m,s,n}$ that uses either the MUM or the CM when the word $w_{m,s,n}$ is generated. Here, if $x_{m,s,n} = 0$, the word is drawn from the MUM's word distribution ϕ_y, and if $x_{m,s,n} = 1$, the word is drawn from the CM's word distribution ϕ_z. The binary variable x is drawn from a Bernoulli distribution λ_m drawn from a beta distribution with symmetric hyperparameter η. In other words, the words depend on both theme and topic, and the MUM and CM are defined independently in this model.

Figure 3 (c) shows the plate notation of the MUM-CM. The generation process in the MUM-CM is as follows.

1. Draw a theme distribution $\psi \sim Dir(\epsilon)$
2. For each theme $y = 1, 2, \ldots, I$:
   - Draw a distribution of theme words $\phi_y \sim Dir(\zeta)$
3. For each topic $z = 1, 2, \ldots, J$:
   - Draw a topic transition distribution $\theta_z \sim Dir(\alpha)$
   - Draw a distribution of topic words $\phi_z \sim Dir(\beta)$
4. For each lyric $m = 1, 2, \ldots, M$:
   - Draw a theme $y_m \sim Multi(\psi)$
   - Draw a distribution of binary variable $\lambda_m \sim Beta(\eta)$
   - For each segment $s = 1, 2, \ldots, S_m$:
     - Draw a topic $z_{m,s} \sim Multi(\theta_{z_{m,s-1}})$
     - For the n-th word $w_{m,s,n}$ in segment s:
       - ∗ Draw a binary variable $x_{m,s,n} \sim Bernoulli(\lambda_m)$
       - ∗ If $x_{m,s,n} = 0$:
         - · Draw a word $w_{m,s,n} \sim Multi(\phi_{y_m})$
       - ∗ If $x_{m,s,n} = 1$:
         - · Draw a word $w_{m,s,n} \sim Multi(\phi_{z_{m,s}})$

Here, α, β, ϵ, and ζ are the symmetric hyperparameters of the Dirichlet distribution and η is the symmetric hyperparameter of the beta distribution. The generation probability of the m-th lyric is calculated as follows:

$$\begin{aligned} P(m) &= P(x{=}0|\lambda_m) \sum_{y=1}^{I} \Big(P(y|\psi) \prod_{s=1}^{S_m} \prod_{n=1}^{N_{m,s}} P(w_{m,s,n}|\phi_y) \Big) \\ &+ P(x{=}1|\lambda_m) \sum_{\mathbf{z}_{all}} \prod_{s=1}^{S_m} \Big(P(z_s|\theta_{z_{s-1}}) \prod_{n=1}^{N_{m,s}} P(w_{m,s,n}|\phi_{z_s}) \Big) \qquad (1) \end{aligned}$$

where $\mathbf{z}_{all}$ denotes all possible topic sequences. If $s = 1$, θ_{z_0} denotes the initial state probabilities. This equation represents that a word $w_{m,s,n}$ is generated from the MUM according to $P(x = 0|\lambda_m)$ or is generated from the CM according to $P(x = 1|\lambda_m)$.

We use collapsed Gibbs sampling for model inference in the MUM-CM. For a lyric m, we present the conditional probability of theme y_m for sampling:

$$\begin{aligned} P(y_m{=}i|\mathbf{y}_{\neg m}, \mathbf{w}, \epsilon, \zeta) &\propto P(y_m{=}i|\mathbf{y}_{\neg m}, \epsilon) \\ &\cdot P(\mathbf{w}_m|\mathbf{w}_{\neg m}, y_m{=}i, \mathbf{y}_{\neg m}, \zeta) \qquad (2) \end{aligned}$$

where $\mathbf{y}_{\neg m}$ denotes the topic set except the m-th lyric, $\mathbf{w}$ denotes the word set in the training corpus, $\mathbf{w}_m$ denotes the word set in the m-th lyric, and $\mathbf{w}_{\neg m}$ denotes the word set in the training corpus except $\mathbf{w}_m$.

We sample topic $z_{m,s}$ for a segment s of lyric m according to the following transition distribution:

$$\begin{aligned} P(z_{m,s} = j|\mathbf{z}_{\neg(m,s)}, \mathbf{w}, \alpha, \beta) &\propto P(z_{m,s} = j|\mathbf{z}_{\neg(m,s)}, \alpha) \\ &\cdot P(\mathbf{w}_{m,s}|\mathbf{w}_{\neg(m,s)}, z_{m,s} = j, \mathbf{z}_{\neg(m,s)}, \beta) \qquad (3) \end{aligned}$$

where $\mathbf{z}_{\neg(m,s)}$ denotes the topic set except the s-th segment in the m-th lyric, $\mathbf{w}_{m,s}$ denotes the word set in the s-th segment of the m-th lyric, and $\mathbf{w}_{\neg(m,s)}$ denotes the word set in the training corpus except $\mathbf{w}_{m,s}$.

For the n-th word in the segment s in the m-th lyric, we present the conditional probability of its binary variables

Algorithm 1 Model inference for the MUM-CM

```
Initialize parameters in the MUM-CM
for each iteration do
    for each lyrics $m$ in the corpus do
        sample $y_m$ according to (2)
        for each segment $s$ in $m$ do
            sample $z_{m,s}$ according to (3)
            for each word $w$ in $s$ do
                sample $x$ according to (4)
            end for
        end for
    end for
    update hyperparameters by using fixed point iteration
end for
```

Algorithm 2 Model inference for the MCM

```
Initialize parameters in the MCM
for each iteration do
    for each lyrics $m$ in the corpus do
        sample $y_m$ according to (6)
        sample $\mathbf{z}_m$ according to (7) by FFBS
    end for
    update hyperparameters by using fixed point iteration
end for
```

$x_{m,s,n}$ for sampling:

$$\begin{aligned} P(x_{m,s,n} = k|\mathbf{x}_{\neg(m,s,n)}, \mathbf{w}, \eta, \zeta, \beta) \\ \propto P(x_{m,s,n} = k|\mathbf{x}_{\neg(m,s,n)}, \eta) \\ \cdot P(w_{m,s,n}|\mathbf{w}_{\neg(m,s,n)}, x_{m,s,n} = k, \mathbf{x}_{\neg(m,s,n)}, \zeta, \beta) \end{aligned} \tag{4}$$

where $\mathbf{x}_{\neg(m,s,n)}$ denotes the binary variable set except $x_{m,s,n}$. Note that the value of k is always 0 or 1.

We estimate hyperparameters α, β, ϵ, ζ, and η using fixed point iteration [29]. For each sampling iteration, the latent variables x, y, and z are sampled. Then, new hyperparameters are estimated such that the joint probabilities $P(\mathbf{w}, \mathbf{y}|\epsilon, \zeta)$, $P(\mathbf{w}, \mathbf{z}|\alpha, \beta)$, and $P(\mathbf{w}, \mathbf{x}|\eta)$ are maximized, where $\mathbf{y}$, $\mathbf{z}$, and $\mathbf{x}$ denote the latent variable sets in the training corpus.

In summary, the model and parameter inference for the MUM-CM is shown in Algorithm 1, and the update equations for Gibbs sampling are given in Appendix A.

3.5 Proposed Model 2: Mixture of Content Model

In the MUM-CM, we assume that theme and storyline are generated independently. On the other hand, as mentioned in Sect. 1, lyricists often create storylines according to themes. Therefore, here, we propose the Mixture of Content Model (MCM) to verify this intuition. In the MCM, when a theme y is generated, a storyline is generated using the theme-specific topic transition distribution $\theta_{y,z}$.

Figure 3 (d) shows the plate notation of the MCM. The MCM generation process is as follows.

1. Draw a theme distribution $\psi \sim Dir(\epsilon)$
2. For each topic $z = 1, 2, \ldots, J$:
    - Draw a word distribution $\phi_z \sim Dir(\beta)$
    - For each theme $y = 1, 2, \ldots, I$:
        - Draw a topic distribution $\theta_{y,z} \sim Dir(\alpha)$
3. For each lyric $m = 1, 2, \ldots, M$:
    - Draw a theme $y_m \sim Multi(\psi)$
    - For each segment $s = 1, 2, \ldots, S_m$:
        - Draw a topic $z_{m,s} \sim Multi(\theta_{y_m, z_{m,s-1}})$
        - For n-th word $w_{m,s,n}$ in segment s:
            - Draw a word $w_{m,s,n} \sim Multi(\phi_{z_{m,s}})$

The generation probability of lyric m is calculated as follows:

$$P(m) = \sum_{y=1}^{I}\Big(P(y|\psi)\sum_{\mathbf{z}_{all}}\prod_{s=1}^{S_m}\Big(P(z_s|\theta_{y,z_{s-1}})\prod_{n=1}^{N_{m,s,n}} P(w_{m,s,n}|\phi_{z_s})\Big)\Big) \tag{5}$$

where $\mathbf{z}_{all}$ denotes all possible topic sequences. If $z = 1$, θ_{y,z_0} denotes the initial state probabilities. In this model, $P(y|\psi)$ represents the mixture ratio of the CMs.

We use collapsed Gibbs sampling for model inference in the MCM. For the m-th lyric, we present the conditional probability of theme y_m for sampling:

$$\begin{aligned} P(y_m = i|\mathbf{y}_{\neg m}, \mathbf{z}, \alpha, \epsilon) \propto P(y_m = i|\mathbf{y}_{\neg m}, \epsilon) \\ \cdot P(\mathbf{z}_m|\mathbf{z}_{\neg m}, y_m = i, \mathbf{y}_{\neg m}, \alpha) \end{aligned} \tag{6}$$

where $\mathbf{z}$ denotes the topic set in the training corpus, $\mathbf{z}_m$ denotes the topic sequence of lyric m (i.e., $z_{m,1}, z_{m,2}, \ldots, z_{m,S_m}$), and $\mathbf{z}_{\neg m}$ denotes the topic set in the training corpus except $\mathbf{z}_m$.

In the MCM, topic sequence $\mathbf{z}_m$ depends on theme y_m, as shown in Fig. 3 (d). Therefore, when a new theme y is sampled, the MCM must resample all topic sequences in lyric m simultaneously. To sample topic sequence $\mathbf{z}_m$, we present the following conditional probability:

$$\begin{aligned} P(\mathbf{z}_m|\mathbf{z}_{\neg m}, \mathbf{y}, \mathbf{w}, \alpha, \beta) \propto P(\mathbf{z}_m|\mathbf{z}_{\neg m}, \mathbf{y}, \alpha) \\ \cdot P(\mathbf{w}_m|\mathbf{w}_{\neg m}, \mathbf{z}_m, \mathbf{z}_{\neg m}, \beta) \end{aligned} \tag{7}$$

However, enumerating all possible topic sequences is infeasible; thus, we use a Forward Filtering-Backward Sampling (FFBS) method [30] that can sample all latent states in a first-order Markov sequence using dynamic programming. In the FFBS method, the marginal probabilities of a topic sequence are calculated in the forward filtering step. Then, topics are sampled from the obtained probabilities in the backward sampling step. The hyperparameters α, β, and ϵ are estimated using fixed point iteration [29].

In summary, the model and parameter inference for the MCM is shown in Algorithm 2, and the update equations for Gibbs sampling are given in Appendix B.

4. Experiments

Here, we examine the effectiveness of the proposed models. First, we verify that topic transitions are useful for modeling storyline by evaluating the word prediction performance

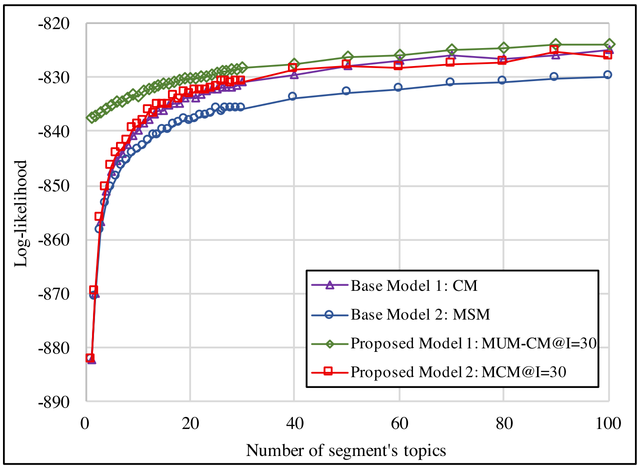


Fig. 4 Log-likelihood on English test data under different segment topic settings (the number of themes I is fixed at 30).

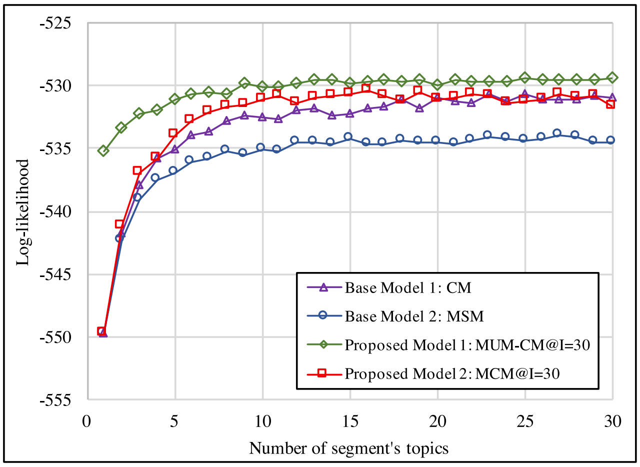


Fig. 5 Log-likelihood on Japanese test data under different segment topic settings (the number of themes I is fixed at 30).

among different models. We then verify that the storyline correlates with the theme performing a segment order prediction task. Finally, we evaluate the proposed models qualitatively by exploring the trained topic transition diagrams.

In our experiments, we originally created two large datasets that contain English and Japanese lyrics of existing songs. One issue that needed to be addressed prior to conducting the experiments was that no existing lyric corpora annotate musical structure (e.g., verse-bridge-chorus tags). In this paper, we assume that segment boundaries are indicated by empty lines inserted by lyricists. In addition, we assume that lyrics with storylines are divided into 6 to 18 segments. The resulting datasets include 80777 lyrics in the English dataset and 16563 lyrics in the Japanese dataset. We randomly split each dataset into 60-20-20% divisions to construct the training, development, and test data.

We trained English-only and Japanese-only models. The collapsed Gibbs sampling ran for 1000 iterations, and the hyperparameters were updated for each Gibbs iteration. For training, we only used content words (nouns, verbs, and adjectives) because we assume that the theme and storyline can be represented using content words.

To extract content words, we use Stanford CoreNLP for English words [31] and the MeCab part-of-speech parser for Japanese words [32].

4.1 Word Prediction Task

To verify that the topic transition and theme are useful properties for storyline modeling, we performed a word prediction task, which measures the test set generation probability. We assume that a better prediction model can capture the storyline of lyrics more effectively. In this experiment, we fixed the number of themes to 30 and computed the test set log-likelihood over the number of segment topics to compare different models.

Figures 4 and 5 show the English and Japanese test set log-likelihood under different segment topic settings. As can be seen, the CM outperforms the MSM, which indicates that typical storylines were effectively captured as the probabilistic distribution of transition over latent topics of segments. Note that the proposed MUM-CM achieves the best performance, which indicates that a better storyline model can be constructed by assuming that the words in lyrics are generated from both theme and topic. The MCM, however, demonstrates only comparable performance to the CM despite that the MCM has a richer parameter space of topic transition distributions.

4.2 Segment Order Prediction Task

In this section, we verify that storyline correlates with theme. Here, we use the order test metric [33], which is used to measure the predictive power of the sequential structure [26], [27]. With the test order metric, the model predicts a reference segment order from all possible segment orders. However, enumerating all possible orders is infeasible; thus, we use the approximation method proposed by Zhai *et al.* [27]:

1. Select N permutations randomly from test data except reference order A.
2. Calculate the $N + 1$ document generative probabilities $P(m)$ whose order is A or N permutations.
3. Choose the hypothesis order A' whose generative probability is the best value in the $N + 1$ orders.
4. Compare the hypothesis order A' with the reference order A to calculate Kendall's tau:

$$\tau = \frac{c^+(A, A') - c^-(A, A')}{T(T-1)/2} \tag{8}$$

where $c^+(A, A')$ denotes the number of correct pairwise orders, $c^-(A, A')$ denotes the number of incorrect pairwise orders, and T denotes the number of segments in a lyric. Here, $N = 50$. This metric ranges from +1 to −1, where +1 indicates that the model selects the reference order and −1 indi-

Table 1 Parameter tuning results with the development set.

Data	Model	I: # of themes	J: # of topics
English lyrics	Base Model 1: CM	none	30
	Proposed Model 1: MUM-CM	120	4
	Proposed Model 2: MCM	70	11
Japanese lyrics	Base Model 1: CM	none	15
	Proposed Model 1: MUM-CM	30	8
	Proposed Model 2: MCM	50	7

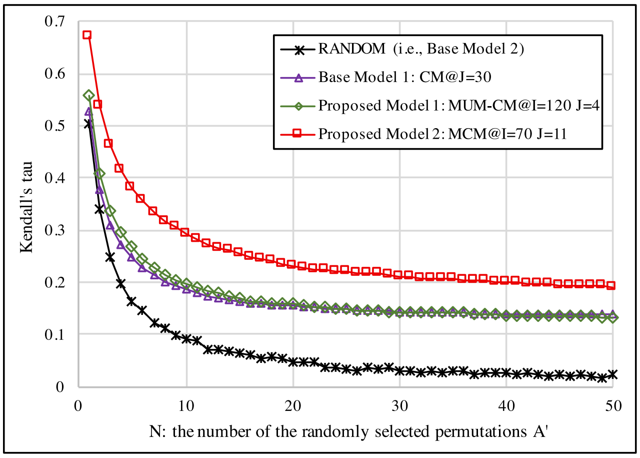


Fig. 6 Average Kendall's τ for English lyrics against the number of random permutations.

cates that model selects the reverse order. In other words, a higher value indicates that the sequential structure has been modeled successfully.

To tune the best parameters (i.e., the number of themes I and number of topics J), we use a grid search on the development set. Table 1 shows the parameters for each model that achieve the best segment order prediction task performance.

As a lower bound baseline, we use a model that randomly selects a hypothesis order A' (i.e., this lower bound is equivalent to the performance of Base Model 2 that does not handle topic transition). To obtain an upper bound for this task, nine Japanese evaluators selected the most plausible order from six orders that include a reference order. Here, $N = 5$ for the human assessments due to cognitive limitations relative to the number of orders. In this manual evaluation, each evaluator randomly selected unknown lyrics. As a result, we obtained 93 orders.

Figures 6 and 7 show Kendall's tau averaged over all English and Japanese test data, respectively. The vertical range shows 95% confidence intervals for the human assessment results. The experimental results indicate that, compared to the lower bound, the proposed models that handle topic transition and theme (i.e., the MUM-CM and MCM) have the predictive power of the sequential structure. This result shows that topic transition and theme are useful properties for storyline modeling. The proposed MCM outperformed all other models on both test sets, while the MUM-CM only demonstrated performance comparable to that of the CM. We also conducted analysis of variance (ANOVA) followed by post-hoc Tukey tests to investigate the differences among these models ($p < 0.05$), drawing the conclusion that the difference between the MCM and the other models is statistically significant. These results show that storyline in lyrics correlates to theme. In contrast to the word prediction task, the MUM-CM has a similar predictive performance as the CM because the MUM-CM has only one topic transition distribution to model the order of segments, which is also the case for the CM.

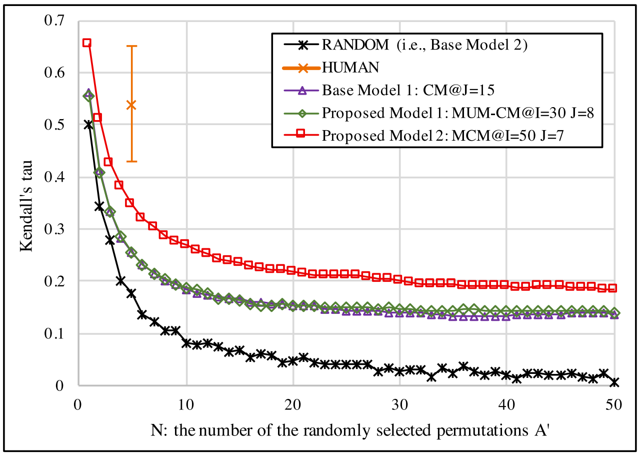


Fig. 7 Average Kendall's τ for Japanese lyrics against the number of random permutations (the vertical range depicts the confidence intervals of the human assessment results).

For Japanese lyrics with $N = 5$, Fig. 7 shows that Kendall's tau for the human evaluation was 0.58 ± 0.11, while the best performance of the model was 0.35. To investigate the cause of this difference, we asked the evaluators to write comments on this task. We found that most evaluators selected a single order by considering the following tendencies.

- Chorus segments tend to be the most representative, uplifting, and thematic segments. For example, the chorus often contains interlude words, such as "hey" and "yo", and frequently includes the lyrical message, such as "I love you". Moreover, the chorus is often the first or last segment; therefore, evaluators tend to first guess which segment is the chorus.
- Verse segments tend to repeat less frequently than choruses.

The human annotators were able to take these factors into account whereas the proposed models cannot consider verse-bridge-chorus structure. This issue could be addressed by combining the storyline of lyrics with the musical structure. We believe this direction will open an intriguing new field for future exploration.

4.3 Analysis of Trained Topic Transitions

Our experimental results indicate that topic transition and theme are useful properties for modeling a storyline. Thus, we are interested in understanding what kinds of themes and topic transitions our model can acquire. Here, to interpret

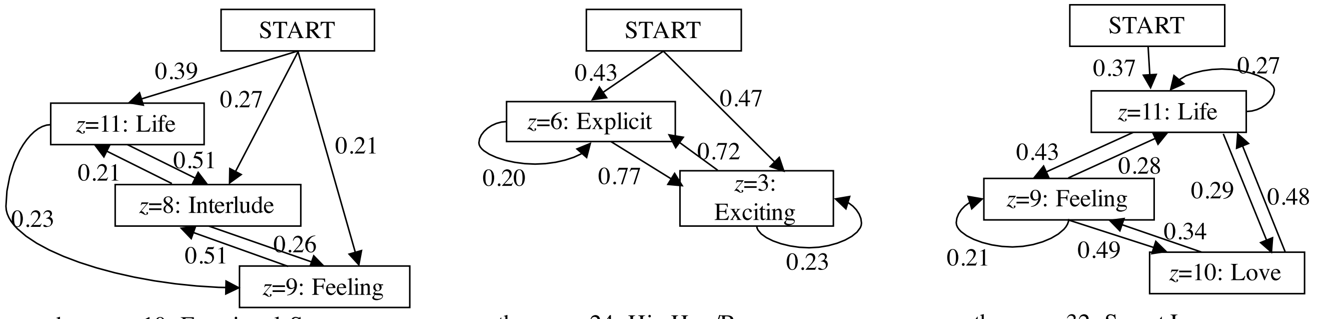
Table 2 Representative words of each topic for English lyrics in MCM@$I = 70$, $J = 11$. The topic label indicates our arbitrary interpretation of the representative words.

z	Label	Representative words in each topic (top 40 words from $P(w\|\phi_s)$)
1	Abbreviation	ah, mi, dem, di, yuh, man, nah, nuh, gal, fus, work, inna, woman, pon, gim, fi, dat, seh, big, mek, weh, u, jump, wah, deh, yah, wid, tek, jah, waan, wine, red, !!!, youth, Babylon, ghetto, neva, hurry, l, nuff
2	Spanish	que, de, tu, el, te, lo, se, yo, un, e, si, por, con, como, amor, una, ti, le, quiero, para, sin, mas, esta, pa, pero, todo, al, solo, las, cuando, hay, voy, corazon, che, soy, je, los, del, vida, tengo
3	Exciting	like, hey, dance, uh, ya, right, body, party, put, shake, move, hand, hot, everybody, boy, beat, floor, c'mon, play, show, 'em, club, bang, drop, huh, lady, bounce, clap, sexy, freak, check, pop, push, low, top, shawty, boom, step, hip, dj
4	Religious	come, day, sing, god, song, lord, hear, Christmas, call, bring, child, new, heaven, beautiful, well, king, name, Jesus, pray, soul, angel, wish, yes, help, year, bear, happy, people, joy, old, son, Mary, bell, peace, father, mother, ring, holy, praise, voice
5	Love	love, feel, need, heart, hold, give, fall, night, dream, world, eye, light, tonight, shine, little, rain, fly, sun, touch, inside, fire, sky, kiss, free, sweet, star, cry, burn, true, close, mine, arm, alive, set, tear, somebody, open, higher, deep, blue
6	Explicit	nigga, shit, fuck, bitch, cause, money, niggaz, ass, hit, real, y', wit, hoe, game, street, em, bout, fuckin, gettin, rap, gun, blow, hood, kid, pay, damn, catch, block, tryin, aint, thug, motherfucker, dick, smoke, straight, house, g, talkin, dog, buy
7	Locomotion	go, get, let, back, ta, take, keep, home, round, turn, run, rock, ride, long, stop, roll, ready, got, road, high, slow, far, music, train, start, town, goin, please, drive, control, radio, fight, fast, car, city, ground, rollin, foot, comin, outta
8	Interlude	oh, la, yeah, ooh, da, whoa, ba, ha, doo, woah, yea, ay, ho, ohh, oooh, mmm, ooo, woo, hoo, oo, dum, ohhh, oh-oh, ahh, ooooh, oooo, wee, la., ohhhh, click, dee, fa, bop, shame, l.a., hmmm, ahhh, drip, trouble, mm
9	Feeling	know, say, time, never, see, make, one, way, think, life, thing, try, find, leave, look, nothing, always, everything, believe, change, lose, live, mind, much, something, wait, better, 'cause, break, wrong, lie, hard, end, word, stay, mean, seem, friend, someone, care
10	Love	na, wan, gon, baby, girl, want, tell, good, bad, alright, talk, crazy, nobody, cuz, im, ai, babe, bye, dont, lovin, fine, feelin, worry, pretty, phone, nothin, fun, thinkin, guy, cos, kind, spend, doin, next, number, sex, treat, cool, honey, cant
11	Life	head, walk, face, stand, watch, die, dead, black, sleep, blood, door, wake, line, wall, kill, water, wind, room, white, sit, hide, grow, bed, fear, lay, rise, hell, sea, meet, scream, pull, death, cut, window, begin, pass, fill, wear, skin, full

Table 3 Representative words of each topic for Japanese lyrics in MCM@$I = 50, J = 7$. The topic label indicates our arbitrary interpretation of the representative words. English words are translated by the authors and original Japanese words are given in parentheses.

z	Label	Representative words in each topic: top 40 words from $P(w\|\phi_s)$
1	English	go, get, let, know, say, night, baby, time, good, way, feel, heart, take, day, dance, make, life, need, party, come, see, tell, dream, everybody, rock, stop, keep, happy, have, give, tonight, please, world, mind, hand, shake, rain, jump, try, your
2	Scene	town (*machi*), night (*yoru*), rain (*ame*), summer (*natsu*), come (*kuru*), window (*mado*), white (*shiroi*), snow (*yuki*), wait (*matsu*), room (*heya*), morning (*asa*), get back (*kaeru*), season (*kisetsu*), fall (*huru*), spring (*haru*), winter (*huyu*), blow (*huku*), wave (*nami*), cold (*tumetai*), hair (*kami*), shoulder (*kata*), memory (*omoide*), back (*senaka*), run (*hashiru*), long (*nagai*), last (*saigo*), shadow (*kage*), sleep (*nemuru*), close (*tojiru*), finger (*yubi*), get wet (*nureru*), remember (*omoidasu*), quiet (*shizuka*), pass (*sugiru*), cheek (*ho*), fall (*otiru*), breath (*iki*), open (*akeru*), car (*kuruma*)
3	Exciting	go (*iku*), front (*mae*), no, sound (*oto*), dance (*odoru*), nothing (*nai*), fly (*tobu*), life (*jinsei*), can run (*hashireru*), begin (*hajimaru*), proceed (*susumu*), stand up (*tatu*), raise (*ageru*), freedom (*jiyu*), era (*jidai*), serious (*maji*), head (*atama*), body (*karada*), ahead (*saki*), power (*chikara*), throw (*suteru*), fire (*hi*), carry (*motu*), high (*hai*), take out (*dasu*), decide (*kimeru*), ride (*noru*), speed up (*tobasu*), Venus, Japan (*nihon*), maximum (*saikou*), rhythm (*rizumu*), non, up, rise (*agaru*), party (*patatii*), wall (*kabe*), companion (*nakama*), girl (*gaaru*), battle (*shobu*)
4	Love	love, love (*ai*), hug (*dakishimeru, daku*), kiss, feel (*kanjiru*), girl, pupil (*hitomi*), ardent (*atsui*), look on (*mitsumeru*), sweet, hold, lonely, sweet (*amai*), kiss (*kisu*), pair (*futari*), smile, stop (*tomeru*), miss, sorrowful (*setsunai*), moon, stop (*tomaru*), heart (*haato*), detach (*hanasu*), overflow (*afureru*), moment (*shunkan*), tempestuous (*hagesii*), moonlight, shine, lovin, touch (*fureru*), little, arm (*ude*), break (*kowareru*), angel (*tenshi*), beating (*kodo*), mystery (*fushgi*), destiny, miracle (*kiseki*), shinin
5	Clean	sky (*sora*), dream (*yume*), wind (*kaze*), light (*hikari*), flower (*hana*), star (*hoshi*), disappear (*kieru*), world (*sekai*), sea (*umi*), future (*mirai*), far (*toi*), voice (*koe*), moon (*tsuki*), shine (*kagayaku*), bloom (*saku*), flow (*nagareru*), sun (*taiyo*), place (*basho*), blue (*aoi*), reach (*todoku*), dark (*yami*), illuminate (*terasu*), cloud (*kumo*), destiny (*eien*), unstable (*yureru*), wing (*tsubasa*), deep (*fukai*), song (*uta*), continue (*tuduku*), sing (*utau*), pass over (*koeru*), shine (*hikaru*), look up (*miageru*), bird (*tori*), finish (*owaru*), color (*iro*), distance (*toku*), high (*takai*), rainbow (*niji*), be born (*umareru*)
6	Lyrical	now (*ima*), mind (*kokoro*), human (*hito*), heart (*mune*), believe (*shinjiru*), word (*kotoba*), oneself (*jibun*), live (*ikiru*), tear (*namida*), forget (*wasureru*), love (*aisuru*), know (*siru*), hand (*te*), cry (*naku*), tomorrow (*ashita*), walk (*aruku*), change (*kawaru*), strong (*tsuyoi*), feeling (*kimochi*), someday (*itsuka*), kind (*yasasii*), everything (*subete*), look (*mieru*), understand (*wakaru*), can be (*nareru*), smile (*egao*), happy (*siawase*), can do (*dekiru*), every day (*hibi*), outside (*soba*), crucial (*taisetsu*), road (*michi*), eye (*me*), look for (*sagasu*), convey (*tutaeru*), time (*jikan*), take leave (*hanareru*), guard (*mamoru*), be able to say (*ieru*)
7	Life	good (*yoi*), say (*iu*), like (*suki*), love (*koi, daisuki*), woman (*onna*), look (*miru*), man (*otoko*), laugh (*warau*), do (*yaru*), today (*kyo*), think (*omou*), spirit (*ki*), face (*kao*), no good (*dame*), listen (*kiku*), phone (*denwa*), tonight (*konya*), friend (*tomodachi*), reach (*tuku*), daughter (*musume*), bad (*warui*), meet (*au*), go (*iku*), appear (*deru*), adult (*otona*), together (*issyo*), good (*umai*), consider (*kangaeru*), die (*sinu*), stop (*yameru*), everyday (*mainichi*), story (*hanashi*), talk (*hanasu*), cheerful (*genki*), drink (*nomu*), human (*ningen*), job (*shigoto*), early (*hayai*)

the proposed MCM, we examine word probabilities $P(w|\phi_z)$ and topic transition probability $P(s|\theta_{y,z'})$ and then visualize topic transition diagrams. To clarify our topic transition analysis, we manually assigned labels to each topic by observing the word list whose generative probability $P(w|\phi_z)$ is a large value. Tables 2 and 3 show the assigned labels and representative words for the topics in the English and Japanese models, respectively. For each topic, we list

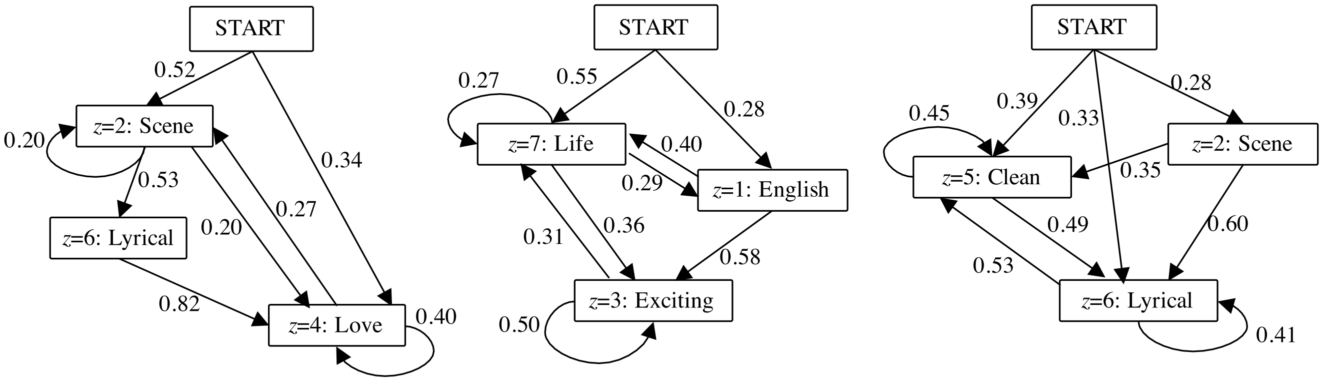


theme y=10: Emotional Story with Interlude

theme y=24: Hip Hop/Rap

theme y=32: Sweet Love

Fig. 8 Examples of English MCM (I = 70, J = 11) transitions between topics for each theme (see Table 2 for word lists). Theme labels are our arbitrary interpretation of their topics and topic transitions.



theme y=6: Sweet Love

theme y=12: Hip Hop/Rap

theme y=14: Hopeful

Fig. 9 Examples of Japanese MCM (I = 50, J = 7) transitions between topics for each theme (see Table 3 for word lists). Theme labels are our arbitrary interpretation of their topics and topic transitions.

the top 40 words in decreasing order of word probability $P(w|\phi_z)$. Figures 8 and 9 show the transition diagrams for some themes in the English and Japanese models, respectively. Here, each arrow indicates higher transition probabilities ($P(s|\theta_{y,z'}) > 0.20$), and each square node indicates the topic z. Note that the initial node ⟨START⟩ indicates the initial state $z = 0$.

We found the following reasonable storylines with the English model (Fig. 8).

- In theme $y = 10$, we see the transition ⟨Life⟩ → ⟨Interlude⟩ → ⟨Feeling⟩. The topic ⟨Interlude⟩ comprises words such as *oh*, *la*, and *yeah* and acts as a bridge between the verse and the chorus.
- In theme $y = 24$, we see that the ⟨ Explicit ⟩ topic tends to shift to ⟨Exciting⟩, which contains words such as *dance*, *sexy*, and *pop*. This topic sequence appears frequently in hip hop/rap songs.
- In theme $y = 32$, we see the transition ⟨Life⟩ → ⟨Feeling⟩ → ⟨Love⟩. We arbitrarily decided the theme label of this topic transition as "Sweet Love". Here, the last topic ⟨Love⟩ tends to shift to the first topic ⟨Life⟩. This indicates that the model captures the repetition structure (e.g., *A*-*B*-*C*-*A*-*B*-*C*, where each letter represents a segment).

We also found the following reasonable storylines with the Japanese model (Fig. 9).

- In theme $y = 6$, we observe the transition ⟨Scene⟩ → ⟨Lyrical⟩ → ⟨Love⟩, which is common in love songs.
- In theme $y = 12$, we see a transition among ⟨Life⟩, ⟨English⟩, and ⟨Exciting⟩, which often appears in Japanese hip hop/rap songs.
- In theme $y = 14$, we see a transition between ⟨Clean⟩ and ⟨Lyrical⟩, which is commonly seen in hopeful songs.

Although we selected these arbitrary diagrams to represent a reasonable storyline, in fact, the self-transition diagrams were trained using other themes. Note that the MCM learns different topic transition distributions according to different themes in an unsupervised manner. This shows that many lyricists consider the topic order and theme as described in textbooks [1], [2].

5. Conclusion and Future Work

This paper has addressed the issue of modeling the discourse nature of lyrics and presented the first study aiming at capturing the two common discourse-related notions: storylines and themes. We assumed that a storyline is a chain of transitions over topics of segments and a song has at least one entire theme. We then hypothesized that transitions over topics of lyric segments can be captured by a probabilistic topic model which incorporates a distribution over transitions of latent topics and that such a distribution of topic transitions is affected by the theme of lyrics.

Aiming to test those hypotheses, this study conducted experiments on the word prediction and segment order prediction tasks exploiting a large-scale corpus of popular music lyrics for both English and Japanese. The findings we gained from these experiments can be summarized into two respects. First, the models with topic transitions significantly outperformed the model without topic transitions in word prediction. This result indicates that typical storylines included in our lyrics datasets were effectively captured as a probabilistic distribution of transitions over latent topics of segments. Second, the model incorporating a latent theme variable on top of topic transitions outperformed the models without such variables in both word prediction and segment order prediction. From this result, we can conclude that considering the notion of theme does contribute to the modeling of storylines of lyrics.

This study has also shaped several future directions. First, we believe that our model can be naturally extended by incorporate more linguistically rich features such as tense/aspect, semantic classes of content words, sentiment polarity, etc. Second, it is also an intriguing direction to adopt recently developed word/phrase embeddings [34], [35] to capture the semantics of lyrical phrases in a further sophisticated manner. Third, verse-bridge-chorus structure of a song is also worth exploring. Our error analysis revealed that the human annotators seemed to be able to identify verse-bridge-chorus structures and use them to predict segment orders. Modeling such lyrics-specific global structure of discourse is an intriguing direction of our future work. Finally, it is also important to direct our attention toward the integration of linguistic discourse structure of lyrics and music structure of audio signals. In this direction, we believe that recent advances in music structure analysis [36], [37] can be an essential enabler.

Acknowledgements

This study utilized the RWC Music Database (Popular Music). This work was partially supported by a Grant-in-Aid for JSPS Research Fellow Grant Number JP16J05945, and CREST and ACCEL, JST. The authors would like to thank Enago and Dr. Paul Reisert for the English language review.

References

[1] D. Austin, J. Peterik, and C.L. Austin, Songwriting for Dummies, Wileys, 2010.

[2] T. Ueda, The writing lyrics textbook which is easy to understand (in Japanese), YAMAHA music media corporation, 2010.

[3] H.R.G. Oliveira, F.A. Cardoso, and F.C. Pereira, “Tra-la-lyrics: an approach to generate text based on rhythm,” Proc. 4th International Joint Workshop on Computational Creativity, pp.47–55, 2007.

[4] A.R. A and S.L. Devi, “An alternate approach towards meaningful lyric generation in tamil,” Proc. NAACL HLT 2010 Second Workshop on Computational Approaches to Linguistic Creativity, pp.31–39, 2010.

[5] G. Barbieri, F. Pachet, P. Roy, and M.D. Esposti, “Markov constraints for generating lyrics with style,” Proc. 20th European Conference on Artificial Intelligence (ECAI 2012), pp.115–120, 2012.

[6] C. Abe and A. Ito, “A Japanese lyrics writing support system for amateur songwriters,” Proc. Asia-Pacific Signal & Information Processing Association Annual Summit and Conference 2012 (APSIPA ASC 2012), PS.2 - SLA.4 Audio & Music Processing (I), 2012.

[7] D. Wu, K. Addanki, M. Saers, and M. Beloucif, “Learning to freestyle: Hip hop challenge-response induction via transduction rule segmentation,” Proc. 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP 2013), pp.102–112, 2013.

[8] P. Potash, A. Romanov, and A. Rumshisky, “Ghostwriter: Using an lstm for automatic rap lyric generation,” Proc. 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), pp.1919–1924, 2015.

[9] M. Ghazvininejad, X. Shi, Y. Choi, and K. Knight, “Generating topical poetry,” Proc. 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), pp.1183–1191, 2016.

[10] K. Watanabe, Y. Matsubayashi, K. Inui, T. Nakano, S. Fukayama, and M. Goto, “Lyrisys: An interactive support system for writing lyrics based on topic transition,” Proc. 22nd International Conference on Intelligent User Interfaces (IUI 2017), pp.559–563, 2017.

[11] R. Mayer, R. Neumayer, and A. Rauber, “Rhyme and style features for musical genre classification by song lyrics,” Proc. 9th International Society for Music Information Retrieval Conference (ISMIR 2008), pp.337–342, 2008.

[12] E. Nichols, D. Morris, S. Basu, and C. Raphael, “Relationships between lyrics and melody in popular music,” Proc. International Society for Music Information Retrieval Conference (ISMIR 2009), pp.471–476, 2009.

[13] E. Greene, T. Bodrumlu, and K. Knight, “Automatic analysis of rhythmic poetry with applications to generation and translation,” Proc. 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP 2010), pp.524–533, 2010.

[14] S. Reddy and K. Knight, “Unsupervised discovery of rhyme schemes,” Proc. 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2011), pp.77–82, 2011.

[15] K. Watanabe, Y. Matsubayashi, N. Orita, N. Okazaki, K. Inui, S. Fukayama, T. Nakano, J. Smith, and M. Goto, “Modeling discourse segments in lyrics using repeated patterns,” Proc. 26th International Conference on Computational Linguistics (COLING 2016), pp.1959–1969, 2016.

[16] J.P.G. Mahedero, Á. MartÍnez, P. Cano, M. Koppenberger, and F. Gouyon, “Natural language processing of lyrics,” Proc. 13th annual ACM international conference on Multimedia, pp.475–478, 2005.

[17] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, “RWC music database: Popular, classical and jazz music databases,” Proc. 3rd of International Society for Music Information Retrieval (ISMIR 2002), pp.287–288, 2002.

[18] B.J. Grosz and C.L. Sidner, “Attention, intentions, and the structure of discourse,” Computational linguistics, vol.12, no.3, pp.175–204, 1986.

[19] D.M. Blei, A.Y. Ng, and M.I. Jordan, “Latent dirichlet allocation,” Journal of Machine Learning Research, vol.3, pp.993–1022, 2003.

[20] F. Kleedorfer, P. Knees, and T. Pohle, “Oh oh oh whoah! towards automatic topic detection in song lyrics,” Proc. 9th International Society for Music Information Retrieval Conference (ISMIR 2008), pp.287–292, 2008.

[21] W. Xu, X. Liu, and Y. Gong, “Document clustering based on non-negative matrix factorization,” Proc. 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, pp.267–273, 2003.

[22] S. Sasaki, K. Yoshii, T. Nakano, M. Goto, and S. Morisihima,

"Lyricsradar: A lyrics retrieval system based on latent topics of lyrics," Proc. 15th International Society for Music Information Retrieval Conference (ISMIR 2014), pp.585–590, 2014.

[23] T. Iwata, S. Watanabe, T. Yamada, and N. Ueda, "Topic tracking model for analyzing consumer purchase behavior," Proc. 18th International Joint Conference on Artificial Intelligence (IJCAI 2009), pp.1427–1432, 2009.

[24] D.M. Blei and J.D. Lafferty, "Dynamic topic models," Proc. 23rd International Conference on Machine learning (ICML 2006), pp.113–120, 2006.

[25] R. Barzilay and L. Lee, "Catching the drift: Probabilistic content models, with applications to generation and summarization," Proc. 2004 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2004), pp.113–120, 2004.

[26] A. Ritter, C. Cherry, and B. Dolan, "Unsupervised modeling of twitter conversations," Proc. 2010 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2010), pp.172–180, 2010.

[27] K. Zhai and J.D. Williams, "Discovering latent structure in task-oriented dialogues," Proc. 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp.36–46, 2014.

[28] K. Nigam, A.K. McCallum, S. Thrun, and T. Mitchell, "Text classification from labeled and unlabeled documents using em," Machine learning, vol.39, no.2-3, pp.103–134, 2000.

[29] T.P. Minka, "Estimating a dirichlet distribution," tech. rep., 2000.

[30] S.L. Scott, "Bayesian methods for hidden markov models," Journal of the American Statistical Association, vol.97, no.457, pp.337–351, 2002.

[31] C.D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S.J. Bethard, and D. McClosky, "The Stanford CoreNLP natural language processing toolkit," Association for Computational Linguistics (ACL 2014) System Demonstrations, pp.55–60, 2014.

[32] T. Kudo, K. Yamamoto, and Y. Matsumoto, "Applying conditional random fields to japanese morphological analysis," Proc. 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), pp.230–237, 2004.

[33] M. Lapata, "Automatic evaluation of information ordering: Kendall's tau," Computational Linguistics, vol.32, no.4, pp.471–484, 2006.

[34] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," Advances in Neural Information Processing Systems 26 (NIPS 2013), pp.3111–3119, 2013.

[35] J. Pennington, R. Socher, and C.D. Manning, "Glove: Global vectors for word representation," Proc. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), pp.1532–1543, 2014.

[36] M. Goto, "A chorus section detection method for musical audio signals and its application to a music listening station," IEEE Trans. Audio, Speech, Language Process., vol.14, no.5, pp.1783–1794, 2006.

[37] J. Paulus, M. Müller, and A. Klapuri, "Audio-based music structure analysis," Proc. International Society for Music Information Retrieval Conference (ISMIR 2010), pp.625–636, 2010.

Appendix A: Equation for MUM-CM Inference

The update equations in Algorithm 1 can be rewritten as (9), (10) and (11). Table A· 1 shows the notations in (9) for collapsed Gibbs sampling of theme y in the MUM-CM inference. Table A· 2 shows the notations in (10) for collapsed Gibbs sampling of topic z in the MUM-CM inference. Table A· 3 shows the notations in (11) for collapsed Gibbs sampling of binary variable x in the MUM-CM inference.

Table A· 1 Notations in (9) for MUM-CM

Notation	**Definition**
$\Gamma(\cdot)$	Gamma function
ϵ, ζ	Hyperparameter
$\mathbf{w}$	Word set in training corpus
$\mathbf{y}_{\neg m}$	Theme set except the m-th lyric
V	Size of the vocabulary
$M_{i,\neg m}$	# of lyrics with theme label i except the m-th lyric
N_m	# of words in the m-th lyric
$N_{i,\neg m}$	# of the word whose theme label is i except the m-th lyric
$N_{m,v}$	# of a word v in the m-th lyric
$N_{i,v,\neg m}$	# of a word v whose theme label is i except the m-th lyric

Table A· 2 Notations in (10) for MUM-CM

Notation	**Definition**
$\mathbb{1}(\cdot)$	Indicator function
α, β	Hyperparameter
$\mathbf{w}$	Word set in training corpus
$\mathbf{z}_{\neg(m,s)}$	Topic set except the s-th segment in the m-th lyric
J	# of topics
$S_{z_{m,s-1}\to j,\neg(m,s)}$	# of segments that trans topic $z_{m,s-1}$ to j except the s-th segment in the m-th lyric
$S_{z_{m,s-1}\to *,\neg(m,s)}$	# of segments with topic $z_{m,s-1}$ except the s-th segment in the m-th lyric
$N_{(m,s)}$	# of words in the s-th segment in the m-th lyric
$N_{j,\neg(m,s)}$	# of words whose topic label is j except the s-th segment in the m-th lyric
$N_{(m,s),v}$	# of a word v in the s-th segment in the m-th lyric
$N_{j,v,\neg(m,s)}$	# of a word v whose topic label is j except the s-th segment in the m-th lyric

Table A· 3 Notations in (11) for MUM-CM

Notation	**Definition**
V	Size of the vocabulary
η, ζ, β	Hyperparameter
$\mathbf{w}$	Word set in training corpus
$\mathbf{x}_{\neg(m,s,n)}$	Binary variable set except the n-th binary variable of the s-th segment in the m-th lyric
$N_{m,\neg(m,s,n)}$	# of words in the m-th lyric except the n-th word of the s-th segment in the m-th lyric
$N_{m,k,\neg(m,s,n)}$	# of words in the m-th lyric with binary label k except the n-th word of the s-th segment in the m-th lyric
$N_{y_m,\neg(m,s,n)}$	# of a word whose theme label is y_m except the n-th binary variable of the s-th segment in the m-th lyric
$N_{y_m,w_{m,s,n},\neg(m,s,n)}$	# of a word $w_{m,s,n}$ with theme label y_m except the n-th binary variable of the s-th segment in the m-th lyric
$N_{z_{m,s},\neg(m,s,n)}$	# of a word whose topic label is $z_{m,s}$ except the n-th binary variable of the s-th segment in the m-th lyric
$N_{z_{m,s},w_{m,s,n},\neg(m,s,n)}$	# of a word $w_{m,s,n}$ with topic label $z_{m,s}$ except the n-th binary variable of the s-th segment in the m-th lyric

Appendix B: Equation for MCM Inference

The update equation in Algorithm 2 can be rewritten as (12). Table A· 4 shows the notations in (12) for collapsed Gibbs

$$P(y_m = i|\mathbf{y}_{\neg m}, \mathbf{w}, \epsilon, \zeta) \propto (M_{i,\neg m} + \epsilon) \cdot \frac{\Gamma(N_{i,\neg m} + \zeta V)}{\Gamma(N_{i,\neg m} + N_m + \zeta V)} \cdot \prod_{v:N_{m,v}>0} \frac{\Gamma(N_{i,v,\neg m} + N_{m,v} + \zeta)}{\Gamma(N_{i,v,\neg m} + \zeta)} \tag{9}$$

$$\begin{aligned} P(z_{m,s} = j|\mathbf{z}_{\neg(m,s)}, \mathbf{w}, \alpha, \beta) \propto\ & \frac{S_{z_{m,s-1}\to j,\neg(m,s)} + \alpha}{S_{z_{m,s-1}\to *,\neg(m,s)} + \alpha J} \cdot \frac{S_{j\to z_{m,s+1},\neg(m,s)} + \mathbb{1}(z_{m,s-1} = j = z_{m,s+1}) + \alpha}{S_{j\to *,\neg(m,s)} + \mathbb{1}(z_{m,s-1} = j) + \alpha J} \\ & \cdot \frac{\Gamma(N_{j,\neg(m,s)} + \beta V)}{\Gamma(N_{j,\neg(m,s)} + N_{(m,s)} + \beta V)} \cdot \prod_{v:N_{(m,s),v}>0} \frac{\Gamma(N_{j,v,\neg(m,s)} + N_{(m,s),v} + \beta)}{\Gamma(N_{j,v,\neg(m,s)} + \beta)} \end{aligned} \tag{10}$$

$$P(x_{m,s,n} = k|\mathbf{x}_{\neg(m,s,n)}, \mathbf{w}, \eta, \zeta, \beta) \propto \frac{N_{m,k,\neg(m,s,n)} + \eta}{N_{m,\neg(m,s,n)} + 2\eta} \cdot \left(\frac{N_{y_m,w_{m,s,n},\neg(m,s,n)} + \zeta}{N_{y_m,\neg(m,s,n)} + \zeta V}\right)^{1-k} \cdot \left(\frac{N_{z_{m,s},w_{m,s,n},\neg(m,s,n)} + \beta}{N_{z_{m,s},\neg(m,s,n)} + \beta V}\right)^{k} \tag{11}$$

$$P(y_m = i|\mathbf{y}_{\neg m}, \mathbf{z}, \alpha, \epsilon) \propto (M_{i,\neg m} + \epsilon) \cdot \prod_{s=1}^{S_m}\left(\frac{\Gamma(S_{i,z_{m,s}\to *,\neg m} + \alpha J)}{\Gamma(S_{i,z_{m,s}\to *,\neg m} + S_{m,z_{m,s}\to *} + \alpha J)} \cdot \prod_{s'=1}^{S_m} \frac{\Gamma(S_{i,z_{m,s}\to z_{m,s'},\neg m} + S_{m,z_{m,s}\to z_{m,s'}} + \alpha)}{\Gamma(S_{i,z_{m,s}\to z_{m,s'},\neg m} + \alpha)}\right) \tag{12}$$

Table A· 4 Notations in (12) for MCM

Notation	Definition
α, ϵ	Hyperparameter
$\mathbf{z}$	Topic set in training corpus
$\mathbf{y}_{\neg m}$	Theme set except the m-th lyric
J	# of topics
$M_{i,\neg m}$	# of lyrics with theme label i except the m-th lyric
$S_{m,z\to *}$	# of segments with topic z in the m-th lyric
$S_{y,z\to *,\neg m}$	# of segments whose topic is z and theme is y except the m-th lyric
$S_{m,z\to z'}$	# of segments whose topic transitions z to z' in the m-th lyric
$S_{y,z\to z',\neg m,}$	# of segments whose theme is y and topic transitions z to z' in the m-th lyric except the m-th lyric

sampling of theme y in the MCM inference.

Kento Watanabe received the B.E. degree in 2013, and M.S. degree in 2015, both from Tohoku University. He is a graduate student in the Department of System Information Sciences, Tohoku University. His research interests include machine learning, natural language processing, and human computer iteration.

Yuichiroh Matsubayashi received his Ph.D degree in Information Science and Technology from the University of Tokyo in 2010. He is currently a research associate at Tohoku University. His research interests include semantic processing of natural language text. He has received several awards including the IPSJ Yamashita SIG Research Award from the Information Processing Society of Japan (IPSJ) and the Best Journal Paper Awards from the Association for Natural Language Processing (ANLP). He is a member of the Association of Computational Linguistics (ACL), ANLP, IPSJ and the Japan Society of Artificial Intelligence (JSAI).

Kentaro Inui received his doctorate degree of engineering from Tokyo Institute of Technology in 1995. He has experience as an assistant professor at Tokyo Institute of Technology and an associate professor at Kyushu Institute of Technology and Nara Institute of Science and Technology, he has been a professor of Graduate School of Information Sciences at Tohoku University since 2010. His research interests include natural language understanding and knowledge processing. He currently serves as the IPSJ director and ANLP director.

Satoru Fukayama received his Bachelor degree in Earth and Planetary Physics in 2008, and Ph.D. degree in Information Science and Technology in 2013, both from the University of Tokyo. He is currently a Senior Researcher at National Institute of Advanced Industrial Science and Technology (AIST), Japan. His main interests are in applications of music information retrieval in automated music generations and choreography synthesis. His approach to research problems is based on machine learning and computational musicology. He has been awarded the Yamashita SIG Research Award from the Information Processing Society of Japan.

Tomoyasu Nakano received the Ph.D. degree in Informatics from University of Tsukuba, Tsukuba, Japan in 2008. He is currently working as a Senior Researcher at the National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan. His research interests include singing information processing, human-computer interaction, and music information retrieval. He has received several awards including the IPSJ Yamashita SIG Research Award from the Information Processing Society of Japan (IPSJ) and the Best Paper Award from the Sound and Music Computing Conference 2013. He is a member of the IPSJ and the Acoustical Society of Japan.

Masataka Goto received the Doctor of Engineering degree from Waseda University in 1998. He is currently a Prime Senior Researcher at the National Institute of Advanced Industrial Science and Technology (AIST), Japan. Over the past 25 years, he has published more than 250 papers in refereed journals and international conferences and has received 46 awards, including several best paper awards, best presentation awards, the Tenth Japan Academy Medal, and the Tenth JSPS PRIZE.