

Modeling Discourse Segments in Lyrics Using Repeated Patterns

Kento Watanabe¹, Yuichiroh Matsubayashi¹, Naho Orita¹, Naoaki Okazaki¹,
Kentaro Inui¹, Satoru Fukayama², Tomoyasu Nakano²,
Jordan B. L. Smith², Masataka Goto²

¹Graduate School of Information Sciences, Tohoku University

²National Institute of Advanced Industrial Science and Technology (AIST)

{kento.w, y-matsu, naho, okazaki, inui}@ecei.tohoku.ac.jp,
{s.fukayama, t.nakano, jordan.smith, m.goto}@aist.go.jp

Abstract

This study proposes a computational model of the discourse segments in lyrics to understand and to model the structure of lyrics. To test our hypothesis that discourse segmentations in lyrics strongly correlate with repeated patterns, we conduct the first large-scale corpus study on discourse segments in lyrics. Next, we propose the task to automatically identify segment boundaries in lyrics and train a logistic regression model for the task with the repeated pattern and textual features. The results of our empirical experiments illustrate the significance of capturing repeated patterns in predicting the boundaries of discourse segments in lyrics.

1 Introduction

Lyrics are an important element of popular music. They provide an effective means to express the message and emotion of music. Similar to prose text, lyrics are a discourse: i.e., they are, typically, a sequence of related lines, rather than an unconnected stream of lines of arbitrary order. Thus, like texts, lyrics also have a discourse structure consisting of discourse segments. Each discourse segment exhibits an individual topic in discourse, and the transition of topics over successive discourse segments constitutes a flow (Austin et al., 2010; Watanabe et al., 2014). Unlike prose text, lyrics have their own peculiar properties, such as frequent repetition of identical or similar phrases and extensive use of rhyme and refrain (Austin et al., 2010), as illustrated in Figure 1. Analogous to prose text, the lyrics in Figure 1 can be viewed as a sequence of discourse segments, wherein each segment is depicted by a colored box. Segments ① and ③ appear repeatedly (e.g., segments ④ and ⑧ are identical to segment ①), which is not typically observed in prose text.

Our goal is to reveal the discourse structure of lyrics in popular music by quantitatively analyzing a large-scale lyrics corpus. The motivations behind the goal are as follows. First, there are hardly any studies that have focused on the data-oriented research of the discourse structure of lyrics; however, multiple studies have focused on the structure of music audio (Paulus et al., 2010, for a review). Second, a better understanding of the nature of lyrics structure combined with the existing theories of music audio could lead to a more comprehensive theory of the overall nature of popular music. Third, understanding the nature of lyrics structure will allow us to devise a variety of useful computer systems, such as systems that automatically generate lyrics, assist human lyrics writers, or evaluate the quality of lyrics.

As a first but crucial step toward achieving this goal, this study explores the nature of the discourse structure of lyrics with a focus on the role of repeated patterns as an indicator of segment boundaries. We choose discourse segments as our primary focus because exploring discourse segments is a necessary step toward a comprehensive understanding of lyrics structure. To address this issue, we consider the task of computationally predicting the boundaries of discourse segments in lyrics under the assumption that a better prediction model would allow us to better understand the nature of the discourse structure of lyrics.

By conducting a large-scale data analysis, we examine our primary hypothesis that *discourse segments in lyrics strongly correlate with repeated patterns*. For example, if a sequence of lines in lyrics has a

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>

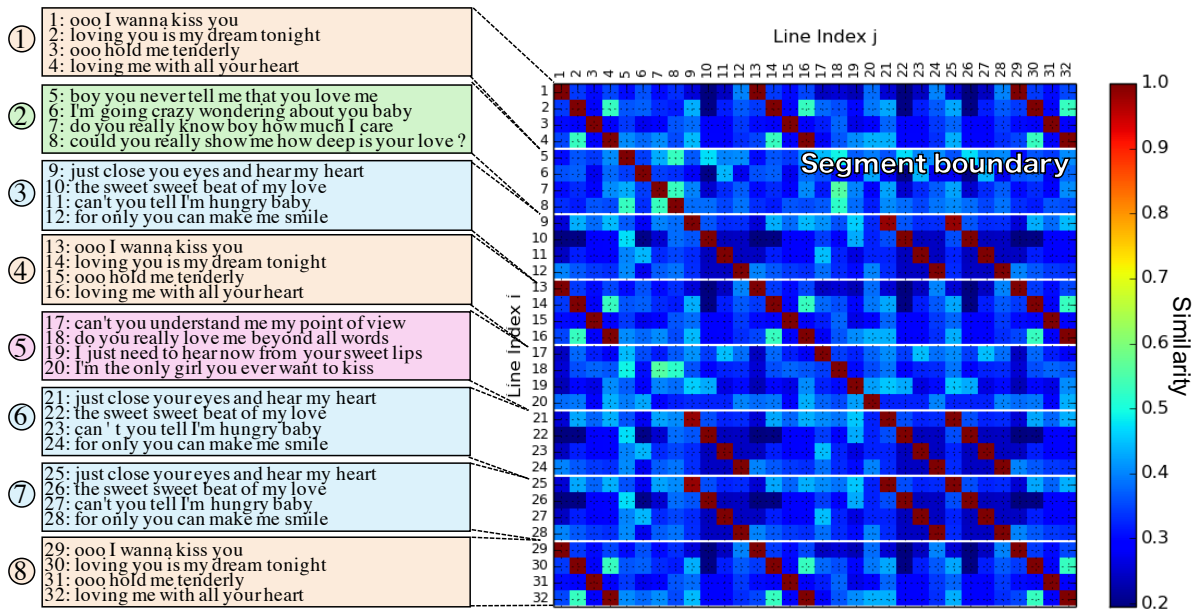


Figure 1: An example of lyrics and corresponding self-similarity matrix (title of lyrics: *How Deep Is Your Love?* (RWC-MDB-P-2001 No. 81 from RWC Music Database (Goto et al., 2002)))

repetition, such as $abcdefabc$ (each letter represents a line, with repeated letters being repeated lines), we expect the boundaries of the discourse segments tend to agree with the boundaries of the repeated parts as in $|abc|def|abc|$, where “|” indicates a boundary.

To examine the extent to which these repeated patterns capture the segment structure of lyrics, we use a large-scale corpus of popular music lyrics that contains more than 140,000 songs. This is, to the best of our knowledge, the first study that takes a data-driven approach to exploring the discourse structure of lyrics in relation to repeated patterns.

One issue to be addressed before conducting the corpus study is that no existing corpus has annotated the discourse structure of lyrics. In this study, we preliminarily assume that discourse segment boundaries are indicated by empty lines inserted by lyrics writers. We admit that empty lines may not be “true” discourse segment boundaries and discourse segments may exhibit a hierarchical structure (e.g., verse–bridge–chorus structure). These issues could be better addressed by combining the analysis of the discourse structure of lyrics with the structural analysis of music. We believe this direction of research will open an intriguing new field for future exploration.

The remainder of this study is organized as follows. Section 2 reviews related work into the discourse structure of lyrics, with particular focus on the segmentation using repeated patterns. Section 3 presents the first quantitative analysis of the distribution of repeated lines and segments in lyrics and suggests cues that could help to identify the segment boundaries. Section 4 describes our computational model, which predicts the boundaries of discourse segments in lyrics using repeated patterns. Section 5 presents experimental results that show the importance of repeated patterns in predicting the boundaries of discourse segments in lyrics. Section 6 concludes and discusses future work. Note that throughout this study, we use the term *segment* to refer to a discourse segment in lyrics.

2 Related work

This section reviews related work into the discourse structure of lyrics, with particular focus on the segmentation of lyrics using repeated patterns.

Text segmentation is a classic text retrieval problem, and there exists a rich body of research into text segmentation in natural language processing. Various linguistic cues have been suggested to identify text boundaries such as expressions that frequently appear at the end of segments (Beeferman et al., 1999), contextual/topical changes (Choi, 2000; Malioutov and Barzilay, 2006; Riedl and Biemann, 2012),

and word/entity repetition (Kan et al., 1998; Reynar, 1999).

Although we share the same motivation as these studies, these text segmentation methods do not consider repeated patterns of phrasal segments because this type of repetition is nearly always absent in prose text. On the other hand, segments in lyrics often have repetitions (Austin et al., 2010) as shown in Section 3.1. We aim to capture the segment structure of lyrics using repeated patterns.

Previous computational work into lyrics segmentation has focused on identifying the segment labels of lyrics that are already segmented. For example, the structure of lyrics can be represented using labels A–B–C–A–B in which each letter refers to a group of lines; e.g., A might represent a chorus that appears twice. Barate et al. (2013) proposed a rule-based method to estimating such structure labels of *segmented* lyrics. Our task differs from this task in that we aim to estimate the segment boundaries of *unsegmented* lyrics using machine learning techniques.

In contrast to the segmentation of lyrics, much previous work has analyzed and estimated the segment structure of music audio signals using repeated musical parts such as verse, bridge, and chorus (Foote, 2000; Lu et al., 2004; Goto, 2006; Paulus and Klapuri, 2006; McFee and Ellis, 2014). To automatically identify these repeated musical parts in music audio signals, a self-similarity matrix (SSM) as shown in Figure 1 is often used. Repeated segments lead to high-valued lines in the off-diagonals of the matrix, and these patterns are used to identify the structure. To capture segments in lyrics using repeated patterns, we apply the SSM to lyrics. Lyrical repetition is known to be an important property of lyrics (Austin et al., 2010), and we expect that repetition patterns would also appear in lyrics as they do in audio signals.

In summary, no previous computational work has exactly focused on the segmentation of lyrics using repeated patterns. Section 3 presents the first quantitative analysis of the distribution of repeated lines and segments in lyrics, and suggests cues that help identify segment boundaries.

3 Statistics of repeated patterns and segment boundaries

As an initial step toward modeling the discourse structure of lyrics, we examine the distribution of segments in lyrics by focusing on repeated patterns. We first show the basic distributions of lyrics and suggest potential cues to indicate segment boundaries in lyrics. To examine the distribution of repeated patterns in lyrics and their relation to segment boundaries, we use a large scale lyrics database that contains 144,891 songs¹.

3.1 The basic distribution of lyrics

Among the 144,891 songs in the lyrics database, there are 5,666,696 lines and 969,176 segments in total, with segment breaks inferred from empty lines. Per song, there are 39.11 lines and 6.69 segments on average. Most songs have at least one repeated line (84.79% using an exact criterion; 90.34% using a lenient matching criterion of normalized edit distance ≥ 0.8 , explained in the next section). A fair number of songs also have at least one repeated segment (exact match: 37.73%, lenient match: 54.57%). Per song, 13.73 lines and 0.52 segments (both exact match) are repeated at least once on average. These distributions show that repetition of lines and segments occurs frequently in lyrics, in line with our expectations. Next, we suggest potential repeated patterns to help in identifying segment boundaries.

3.2 Correlation between repeated patterns and segment boundaries

To examine what kinds of repeated patterns would help identify segments in lyrics, we use the SSM, similar to previous work into the segmentation of music audio signals (Section 2). Figure 1 shows an example SSM. Throughout this study, we represent the i^{th} line in lyrics as l_i ($1 \leq i \leq L$), where L is the number of lines of the lyrics. A degree of similarity between l_i and l_j , i.e., $\text{sim}(l_i, l_j)$, is represented

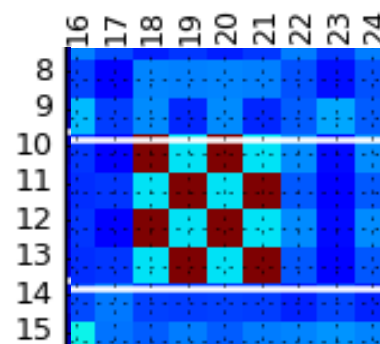


Figure 2: Negative example against pattern (1)

¹Music Lyrics Database. <http://www.odditysoftware.com/page-datasales1.htm>

Repeated pattern	Prior and conditional probabilities	Value
Pattern (1)	$P(\text{Boundary appears})$	0.1455 (824286/5666696)
	$P(\text{Boundary appears} \mid \text{at the starting/ending of a diagonal line})$	0.2218 (339020/1530824)
Pattern (2)	$P(\text{Boundary does not appear})$	0.8545 (4842410/5666696)
	$P(\text{Boundary does not appear} \mid \text{within a diagonal line})$	0.9273 (751195/810098)
Pattern (3)	$P(\text{Adjacent lines appear within a segment})$	0.8507 (4697520/5521806)
	$P(\text{Adjacent lines appear within a segment} \mid \text{adjacent lines are similar})$	0.9439 (218524/231518)
Pattern (4)	$P(\text{Boundary appears after } l_i)$	0.1455 (824286/5666696)
	$P(\text{Boundary appears after } l_i \mid l_i \text{ is similar to the last line of lyrics})$	0.4230 (125659/297069)
	$P(\text{Boundary appears after } l_i \mid l_{i+1} \text{ is similar to the first line of lyrics})$	0.4189 (46531/111079)

Table 1: Correlation between each repeated pattern and segment boundary

as an intensity at a cell where the i^{th} row and j^{th} column overlap. Using the normalized edit distance $\text{NED}(l_i, l_j)$, we compute the degree of similarity (Yujian and Bo, 2007): $\text{sim}(l_i, l_j) = 1 - \text{NED}(l_i, l_j)$. The red diagonal lines in Figure 1 are the result of exact line repetitions². The white horizontal lines in Figure 1 indicate the true segment boundaries.

After manually examining more than 1,000 lyrics and their SSMs, we suggest the following four types of repeated patterns as indicators of segment boundaries.

- (1) **The Start and end points of a diagonal line are segment boundaries.** Some repeated segments correspond to the red diagonal lines. For example, in Figure 1, segment ① is repeated twice (segments ④ and ⑧), and each repetition, starting at l_{13} and l_{29} , can be observed as a diagonal line from l_1 to l_4 . This suggests that some segments could be divided at the start and end points of such a diagonal line.
- (2) **A segment boundary does not appear within a diagonal line.** This is related to (1). A segment boundary does not normally appear within a diagonal line because each diagonal line often corresponds to a segment.
- (3) **Similar adjacent lines appear within a segment.** Line-level repetitions that are adjacent, such as rhymes and refrains, tend to occur within a segment. For example, line l_7 rhymes internally with l_8 where these lines appear within a segment because $\text{sim}(l_7, l_8)$ indicates moderate similarity.
- (4) **A line similar to the first or last line of a song is an indicator of a segment boundary.** Lines similar to the first or last line of lyrics tend to be repeated at segment boundaries. For example, in Figure 1, the first and last lines of the song, i.e., l_1 and l_{32} , are exactly the same as the first and last lines of segments ④ and ⑧. This is because the first and last lines of lyrics tend to be part of a chorus section that is often repeated throughout the lyrics.

To examine the extent to which these four patterns correlate with segment boundaries, we compute the prior and conditional probabilities of each pattern using the full lyrics database. Table 1 shows that all conditional probabilities are greater than their corresponding prior probabilities. These results suggest that the above repeated patterns reasonably capture segment boundaries, supporting the use of repeated patterns for modeling the segment structure of lyrics.

Note that pattern (1) does not hold for many cases. Figure 2 illustrates a typical negative example against pattern (1). This figure includes three diagonal lines, but the shorter lines do not agree with a segment boundary at either end. Similar cases are abundant partly because even a repetition of a single line is identified as a diagonal line in this experiment. This problem implies that a single occurrence of our local repeated pattern is not a sufficient clue for identifying a segment boundary. The conflict between patterns (1) and (2) is also shown in Figure 1, where a segment boundary implied by pattern (1) bisects a diagonal line from (l_{25}, l_9) to (l_{32}, l_{16}) , which goes against pattern (2). This motivates us to build a machine learning-based model to capture combinations of multiple clues. The subsequent section describes how we represent these repeated patterns as features for predicting segment boundaries.

²The diagonal line of $\text{sim}(l_i, l_i)$ is ignored for analysis because it conveys no information.

4 Computational modeling of segment patterns in lyrics

To confirm the validity of our four repeated patterns for segment structures, we address the novel task of detecting segment boundaries in lyrics. Given the lyrics of a song where all segments are concatenated (no empty lines), the task is to identify the segment boundaries of the lyrics reproducing the empty lines. We formalize this task as a binary classification problem to predict the end ($y = 1$) or continuation ($y = 0$) of a segment between lines l_i and l_{i+1} . We model the conditional probability $p(y|i)$ using logistic regression with two different types of features: (1) *repeated patterns in lyrics* and (2) *textual expressions appearing at the line boundaries*.

4.1 Repeated patterns

We propose four subtypes of repeated pattern features (RPF1, RPF2, RPF3, and RPF4) corresponding to the four hypotheses presented in Section 3.2. Here, matrix M denotes the SSM of the lyrics. Each element $m_{i,j}$ represents the similarity between lines l_i and l_j , i.e., $m_{i,j} = \text{sim}(l_i, l_j)$.

RPF1 The first repeated pattern (*the beginning or end point of a diagonal line in an SSM is a clue for a segment boundary*) is formalized as follows. Given two lines i and j , we expect that there exists a boundary after both of these lines if the lines are similar/dissimilar, but $i + 1$ and $j + 1$ are opposite (dissimilar/similar). For a given line i , Equation 1 enumerates a set of lines j ($1 \leq j \leq L$) where there may be boundaries after line i and every line j :

$$g_\lambda(i) = \{j \mid (m_{i,j} - \lambda)(m_{i+1,j+1} - \lambda) < 0\} \quad (1)$$

Here, λ is a threshold for detecting similarity and dissimilarity. The left side of Figure 3 illustrates four likely boundaries for line $i = 24$ with the threshold $\lambda = 0.6$: $g_{0.6}(24) = \{8, 12, 20, 28\}$.

Using the function $g_\lambda(i)$, we define feature functions $f_\lambda^{(\text{RPF1}\#)}(i)$ and $f_\lambda^{(\text{RPF1v})}(i)$ that assess how likely it is that line i is located at the beginning or end points of diagonal lines in the SSM:

$$f_\lambda^{(\text{RPF1}\#)}(i) = |g_\lambda(i)| \quad (2)$$

$$f_\lambda^{(\text{RPF1v})}(i) = \frac{1}{|g_\lambda(i)|} \sum_{j \in g_\lambda(i)} |m_{i,j} - m_{i+1,j+1}| \quad (3)$$

To sum up, $f_\lambda^{(\text{RPF1}\#)}(i)$ counts the number of likely boundaries after line i and other lines j , and $f_\lambda^{(\text{RPF1v})}(i)$ computes the mean of the similarity differences at likely boundaries after line i and other lines j . We define multiple features with different threshold values λ .

RPF2 The second repeated pattern (*a segment boundary does not appear inside of a diagonal line of an SSM*) is formalized analogously to RPF1. Given two lines i and j , we expect that lines i and j are *points of continuity* if lines i and j are similar and $i + 1$ and $j + 1$ are also similar. For a given line i , Equation 4 enumerates a set of lines j ($1 \leq j \leq L$) where i and j are points of continuity:

$$c_\lambda(i) = \{j \mid m_{i,j} \geq \lambda \wedge m_{i+1,j+1} \geq \lambda\} \quad (4)$$

The middle of Figure 3 shows an example of continuous points (here, $c_{0.6}(10) = \{22, 26\}$ in this example).

Similar to RPF1, Equations 5 and 6 count the number of continuous points and the mean of the similarity differences at continuous points, respectively.

$$f_\lambda^{(\text{RPF2}\#)}(i) = |c_\lambda(i)| \quad (5)$$

$$f_\lambda^{(\text{RPF2v})}(i) = \frac{1}{|c_\lambda(i)|} \sum_{j \in c_\lambda(i)} |m_{i,j} - m_{i+1,j+1}| \quad (6)$$

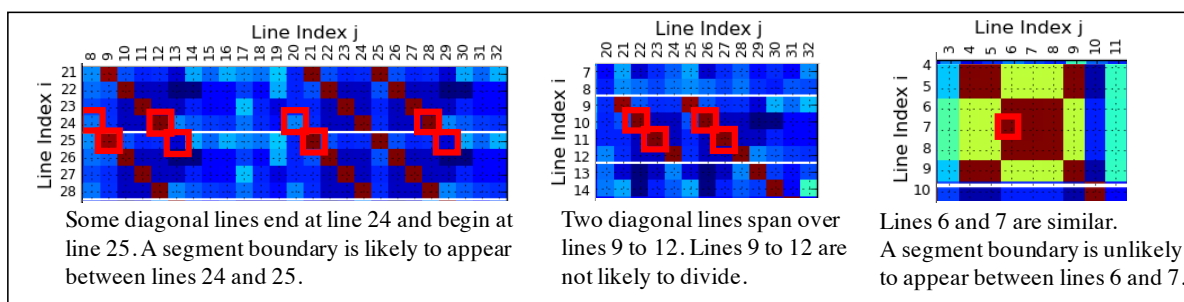


Figure 3: Repeated pattern features: RPF1, RPF2, and RPF3

Position	-3	-2	-1	0 (Line Break)	1	2	3
Word	oh	oh	!!		I	love	you
POS tag	UH	UH	SYM		PRP	VBP	PRP

6 words between i^{th} line and $i+1^{\text{th}}$ line

TF1_Uni-gram(-3) = "oh"	TF1_Bi-gram(-2) = "oh_oh"	TF1_Tri-gram(-2) = "oh_oh_!!"
TF1_Uni-gram(-2) = "oh"	TF1_Bi-gram(-1) = "oh_!!"	TF1_Tri-gram(-1) = "oh_!!_I"
TF1_Uni-gram(-1) = "!!"	TF1_Bi-gram(0) = "!!_I"	TF1_Tri-gram(1) = "!!_I_love"
TF1_Uni-gram(1) = "I"	TF1_Bi-gram(1) = "I_love"	TF1_Tri-gram(2) = "I_love_you"
TF1_Uni-gram(2) = "love"	TF1_Bi-gram(2) = "love_you"	
TF1_Uni-gram(3) = "you"		

Textual Feature 1 (TF1): 15 word N-grams

TF2_Uni-gram(-3) = "UH"	TF2_Bi-gram(-2) = "UH_UH"	TF2_Tri-gram(-2) = "UH_UH_SYM"
TF2_Uni-gram(-2) = "UH"	TF2_Bi-gram(-1) = "UH_SYM"	TF2_Tri-gram(-1) = "UH_SYM_PRP"
TF2_Uni-gram(-1) = "SYM"	TF2_Bi-gram(0) = "SYM_PRP"	TF2_Tri-gram(1) = "SYM_PRP_VBP"
TF2_Uni-gram(1) = "PRP"	TF2_Bi-gram(1) = "PRP_VBP"	TF2_Tri-gram(2) = "PRP_VBP_PRP"
TF2_Uni-gram(2) = "VBP"	TF2_Bi-gram(2) = "VBP_PRP"	
TF2_Uni-gram(3) = "PRP"		

Textual Feature 2 (TF2): 15 Part of speech N-grams

Figure 4: Textual features: TF1 and TF2

RPF3 (similarity with a subsequent line) RPF3 encodes the third repeated pattern, i.e., similar adjacent lines belong to the same segment. For a given line index i , this is quantified by the similarity $\text{sim}(l_i, l_{i+1})$:

$$f^{(\text{RPF3})}(i) = m_{i,i+1} \quad (7)$$

The right of Figure 3 shows an example where RPF3 indicates a continuation between lines 6 and 7.

RPF4 (similarity with the first and last lines) The fourth repeated pattern (i.e., a line similar to the first line of the lyrics is likely to be the first line of a segment, and a line similar to the last line of the lyrics is likely to be the last line of a segment) is encoded by two feature functions $f^{(\text{RPF4b})}(i)$ and $f^{(\text{RPF4e})}(i)$:

$$f^{(\text{RPF4b})}(i) = m_{i,1} \quad (8)$$

$$f^{(\text{RPF4e})}(i) = m_{i,n} \quad (9)$$

4.2 Textual expressions

Some textual expressions appear selectively at the beginning or end of a segment. For example, the phrase "So I" often appears at the beginning of a line but rarely appears at the beginning of a segment. To exploit such indications of the beginnings/ends of lines, we propose two textual features (TF1 and TF2).

TF1 (word n-grams at a line boundary) A phrase like "oh oh !!" tends to appear at the end of a segment. In contrast, a phrase like "I'm sorry" may appear at the beginning of a segment. Previous work on sentence boundary estimation has often used n -grams to detect segment boundaries (Beeferman et al.,

Method	P_k (%)	WD (%)	Precision (%)	Recall (%)	F-measure (%)
Random	49.35	53.67	14.29	12.50	13.33
TF*	40.51	44.65	34.95	31.66	33.22
RPF*	27.00	32.16	56.05	59.42	57.68
Proposed (ALL)	27.22	32.22	56.58	60.65	58.55
Ablation test					
–RPF1	31.38	35.89	51.95	51.32	51.63
–RPF2	30.62	36.73	49.22	57.64	53.10
–RPF3	27.46	32.71	55.59	59.40	57.43
–RPF4	27.64	32.68	55.73	59.94	57.76
–TF1_Uni_gram	27.00	31.95	56.90	60.73	58.75
–TF1_Bi_gram	26.84	31.88	56.96	61.41	59.10
–TF1_Tri_gram	27.32	32.53	55.88	61.42	58.52
–TF2_Uni_gram	28.24	31.84	59.40	51.91	55.40
–TF2_Bi_gram	26.89	31.25	58.86	58.12	58.49
–TF2_Tri_gram	26.67	31.40	57.91	60.23	59.05
–TF1_{Uni,Bi}_gram, TF2_Tri_gram (Best Performance)	26.58	31.55	57.40	61.21	59.24

Table 2: Results of ablation tests

1999). Thus, we define word n -gram features (for $n = 1, 2, 3$) around a line boundary. More specifically, we define 15 n -gram features at different positions, listed and illustrated with an example in Figure 4.

TF2 (part of speech n -grams around a line boundary) Parts of speech (POS), such as *particles* or *determiners* do not tend to appear at the end of a sentence, and *conjunctions* do not appear at the beginning of a sentence. We exploit these tendencies by defining features for POS. Similar to TF1, we define POS n -gram features (for $n = 1, 2, 3$) around a line boundary. Specifically, we define 15 POS n -gram features at different positions, as shown in Figure 4.

5 Experiment

We sampled 105,833 English songs from the Music Lyrics Database v.1.2.7 so that each song contains at least five segments. The resulting dataset includes 2,788,079 candidate boundaries and 517,234 actual boundaries. We then split these songs into training (60%), development (20%), and test (20%) sets. For feature extraction, we used the Stanford POS Tagger (Toutanova et al., 2003). To train the segment boundary classifiers, we used the Classias implementation (Okazaki, 2009) of L2-regularized logistic regression. By employing multiple threshold values of λ from 0.1 to 0.9 with a step size of 0.1, we used them all together.

5.1 Performance evaluation metrics

We used two sets of metrics to evaluate the performance of each model for the task. One was standardly used in audio music segmentation, i.e., the precision, recall, and F-measure of identifying segment boundaries. Precision is the ratio of correctly predicted boundaries over all predicted boundaries, recall is the ratio of correctly predicted boundaries over all true boundaries, and F-measure is the harmonic mean of precision and recall. The other set was standardly used in text segmentation literature: P_k (Beeferman et al., 1999) and WindowDiff (WD) (Pevzner and Hearst, 2002). P_k is the probability of segmentation error that evaluates whether two lines l_i and l_j in lyrics fewer than k lines apart are incorrectly concatenated or divided by a segmentation model. P_k is considered a more suitable measure than F-measure in text segmentation because it assigns partial credit to nearly correct estimations. WD is a variant of P_k that resolves a problem of P_k by penalizing false positives. We set the window size k of P_k and WD to

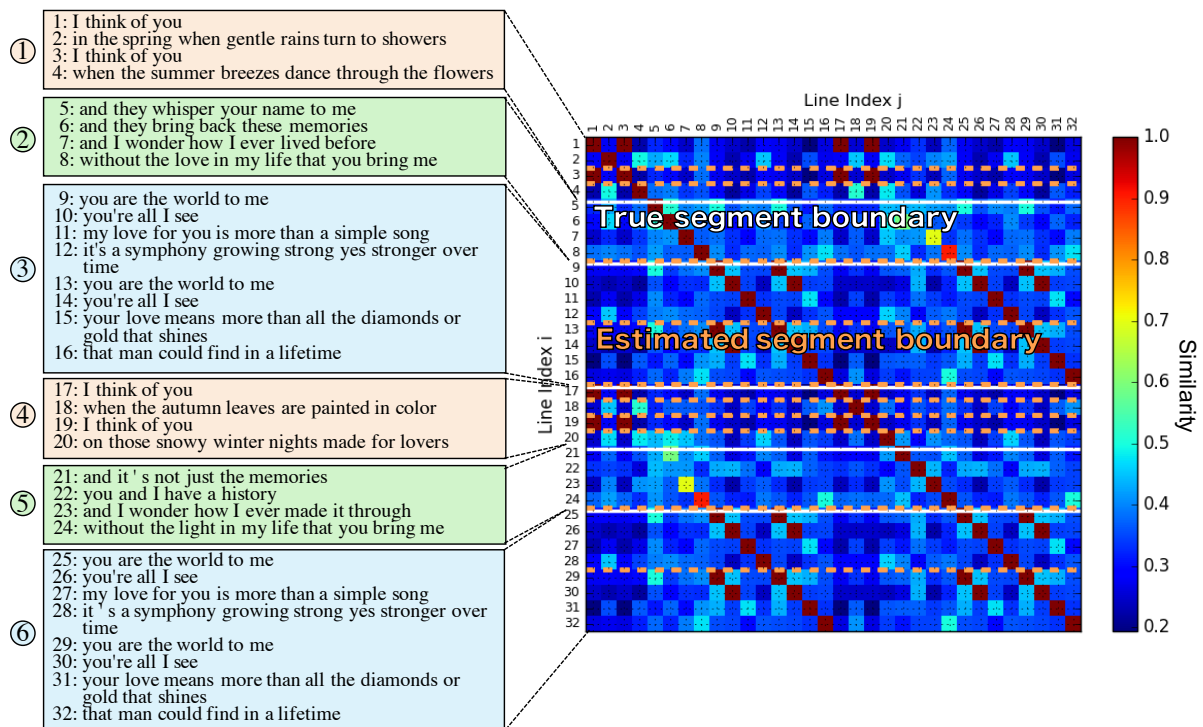


Figure 5: Examples of false positives (title of lyrics: *I think of you* (RWC-MDB-P-2001 No.87 from RWC Music Database (Goto et al., 2002))). White horizontal lines indicate true segment boundaries. Orange horizontal dashed lines indicate predicted boundaries.

one-half the average line length of the correct segments for each song in the test set.

5.2 Contributions of different features

We investigated the contribution of each feature set by conducting ablation tests over different combinations of feature sets. The results are shown in Table 2. **Random** denotes our baseline, a model selecting boundaries with uniform probability $P = 0.186$, the true frequency of boundaries ($P = 517, 234/2, 788, 079$). **RPF*** and **TF*** denote the models with all repeated pattern features and all textual features, respectively. **Proposed** indicates the performance of the model with all proposed features. At the top of Table 2, the F-measure of the proposed method was 58.44, or 45 points higher than that of the random baseline.

The results of the ablation tests are shown in the bottom of Table 2. For example, “-RPF1” indicates that we ablated the feature RPF1 from the proposed method, which uses all of the features. Our best-performing model achieved an F-measure of 59.24 by excluding the TF1 unigram and bigram and TF2 trigram features.

The table shows that each type of our RPF features contributes to performance. Note that these four types are not redundant, and each of our hypotheses yielded positive results. Note that removing RPF1 and RPF2, which are intended to capture long-range repeated patterns, decreased the F-measure by 6.92 and 5.45 points, respectively. This result supports the hypothesis that sequences of repeated lines (diagonal lines in the SSM) are important clues for modeling lyrics segmentation.

In contrast to results reported in text segmentation literature (Beeferman et al., 1999), TF features turned out to be ineffective for lyrics segment boundary estimation, except for TF2 unigram features. One possible reason is that there is a larger variety of expressions used at the beginning or end of a segment in lyrics compared with prose texts. Still, the inclusion of some textual features did lift the performance of the RPF* model by nearly 2 points. Further investigation of TF features is left for our future work.

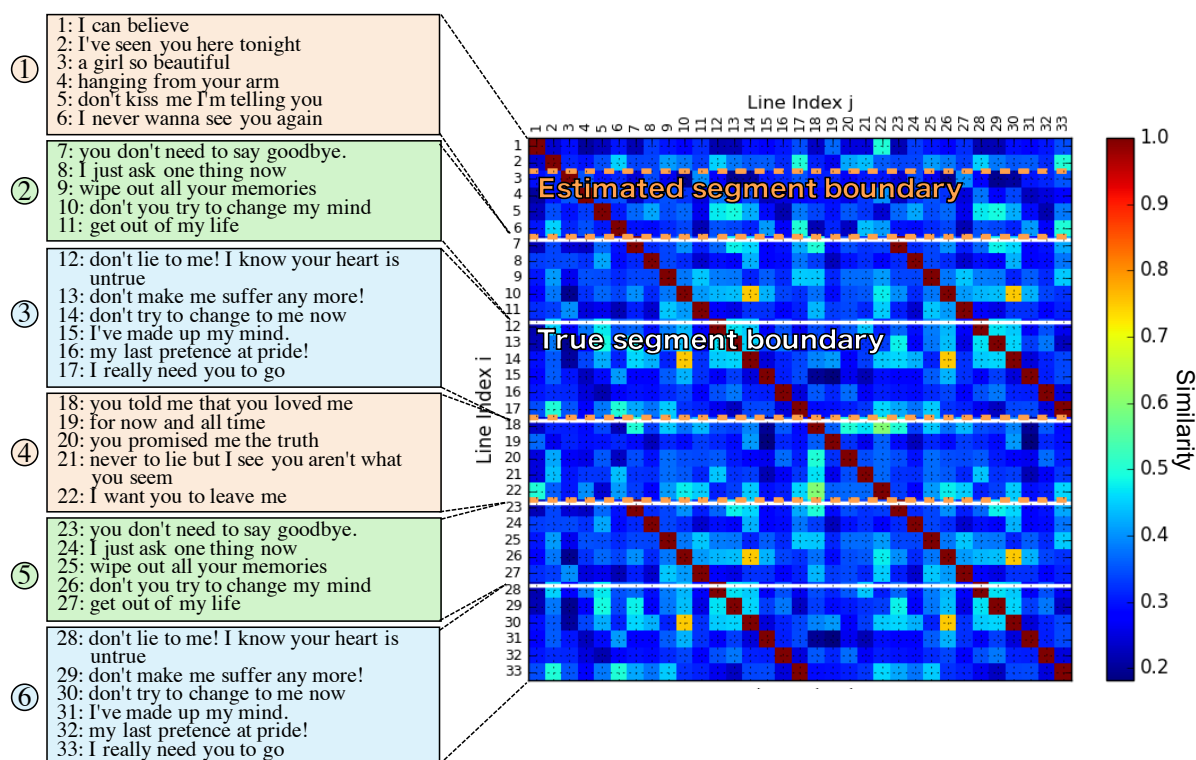


Figure 6: Examples of false negatives (title of lyrics: *Don't lie to me* (RWC-MDB-P-2001 No.97 from RWC Music Database (Goto et al., 2002))). White horizontal lines indicate true segment boundaries. Orange horizontal dashed lines indicate predicted boundaries.

5.3 Error analysis

Figures 5 and 6 give two examples of lyrics and SSMs that illustrate typical errors of our best model. Horizontal dashed lines depict predicted boundaries. As shown in Figure 5, the model sometimes overly divides a true segment into segments as small as single lines, false positives that appear to be due to occurrences of repeated single lines (here, lines 1, 3, 17 and 19). This is not a trivial problem because repetitions of single lines sometimes serve as an important clue. In fact, when restricting diagonal lines to be of the length of two or more lines, we considerably lose recall while gaining precision. More investigation is needed for further improvement.

In contrast to the case of Figure 5, Figure 6 shows a typical example of false negatives. We missed a boundary between, for example, lines 11 and 12. For this boundary, we cannot find any clear repeated pattern indicator. Such cases suggest a limitation of repeated pattern features and the need for further refinement of the model. One direction is to incorporate semantics-oriented state-of-the-art techniques for prose text segmentation such as topic tiling (Riedl and Biemann, 2012).

6 Conclusion and future work

This study has addressed the issue of modeling discourse segments in lyrics in order to understand and model the discourse-related nature of lyrics. We first conducted a large-scale corpus study into the discourse segments of lyrics, in which we examined our primary hypothesis that discourse segmentations strongly correlate with repeated patterns. To the best of our knowledge, this is the first study that takes a data-driven approach to explore the discourse structure of lyrics in relation to repeated patterns. We then proposed a task to automatically identify segment boundaries in lyrics and explored machine learning-based models for the task with repeated pattern features and textual features. The results of our empirical experiments show the importance of capturing repeated patterns in predicting the boundaries of discourse segments in lyrics. In future, we plan to refine the model further by incorporating topic/semantic

information, to extend the modeling of lyric discourse by combining it with audio musical structure, and to embed a resulting model into application systems, such as lyrics generation systems and lyrics composition support systems.

Acknowledgments

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, JSPS KAKENHI Grant Numbers JP15K16045 and JP15H01702, and CREST, JST.

References

- Dave Austin, Jim Peterik, and Cathy Lynn Austin. 2010. *Songwriting for Dummies*. Wileys.
- Adriano Barate, Luca Andrea Ludovico, and Enrica Santucci. 2013. A semantics-driven approach to lyrics segmentation. In *Proceedings of the 8th International Workshop on Semantic and Social Media Adaptation and Personalization*, pages 73–79.
- Doug Beeferman, Adam Berger, and John Lafferty. 1999. Statistical models for text segmentation. *Journal of Machine Learning*, 34(1):177–210.
- Freddy Y.Y. Choi. 2000. Advances in domain independent linear text segmentation. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, pages 26–33.
- Jonathan Foote. 2000. Automatic audio segmentation using a measure of audio novelty. In *Proceedings of International Conference on Multimedia and Expo 2000*, pages 452–455.
- Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka. 2002. RWC music database: Popular, classical and jazz music databases. In *Proceedings of the 3rd of International Society for Music Information Retrieval*, volume 2, pages 287–288.
- Masataka Goto. 2006. A chorus section detection method for musical audio signals and its application to a music listening station. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5):1783–1794.
- Min-Yen Kan, Judith L. Klavans, and Kathleen R. McKeown. 1998. Linear segmentation and segment significance. In *Proceedings of the 6th International Workshop of Very Large Corpora (WVLC-6)*, pages 197–205.
- Lie Lu, Muyuan Wang, and Hong-Jiang Zhang. 2004. Repeating pattern discovery and structure analysis from acoustic music data. In *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*, pages 275–282.
- Igor Malioutov and Regina Barzilay. 2006. Minimum cut model for spoken lecture segmentation. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 25–32.
- Brian McFee and Daniel PW Ellis. 2014. Learning to segment songs with ordinal linear discriminant analysis. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing 2014*, pages 5197–5201.
- Naoaki Okazaki. 2009. Classias: A collection of machine-learning algorithms for classification. <http://www.chokkan.org/software/classias/>.
- Jouni Paulus and Anssi Klapuri. 2006. Music structure analysis by finding repeated parts. In *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*, pages 59–68.
- Jouni Paulus, Meinard Müller, and Anssi Klapuri. 2010. Audio-based music structure analysis. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 625–636.
- Lev Pevzner and Marti A. Hearst. 2002. A critique and improvement of an evaluation metric for text segmentation. *Computational Linguistics*, 28(1):19–36.
- Jeffrey C. Reynar. 1999. Statistical models for topic segmentation. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 357–364. Association for Computational Linguistics.

- Martin Riedl and Chris Biemann. 2012. Topictiling: a text segmentation algorithm based on lda. In *Proceedings of Association for Computational Linguistics 2012 Student Research Workshop*, pages 37–42. Association for Computational Linguistics.
- Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 173–180.
- Kento Watanabe, Yuichiroh Matsubayashi, Kentaro Inui, and Masataka Goto. 2014. Modeling structural topic transitions for automatic lyrics generation. In *Proceedings of The 28th Pacific Asia Conference on Language, Information and Computation*, pages 422–431.
- Li Yujian and Liu Bo. 2007. A normalized Levenshtein distance metric. *Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1091–1095.