

TOWARDS SUPPORT FOR UNDERSTANDING CLASSICAL MUSIC: ALIGNMENT OF CONTENT DESCRIPTIONS ON THE WEB

Taku Kuribayashi* Yasuhito Asano Masatoshi Yoshikawa
Graduate School of Informatics, Kyoto University, Japan

choco.ms@gmail.com, {asano, yoshikawa}@i.kyoto-u.ac.jp

ABSTRACT

Supporting the understanding of classical music is an important topic that involves various research fields such as text analysis and acoustics analysis. Content descriptions are explanations of classical music compositions that help a person to understand technical aspects of the music. Recently, Kuribayashi et al. proposed a method for obtaining content descriptions from the web. However, the content descriptions on a single page frequently explain a specific part of a composition only. Therefore, a person who wants to fully understand the composition suffers from a time-consuming task, which seems almost impossible for a novice of classical music. To integrate the content descriptions obtained from multiple pages, we propose a method for aligning each pair of paragraphs of such descriptions. Using dynamic time warping-based method along with our new ideas, (a) a distribution-based distance measure named $w2DD$, and (b) the concept of *passage expressions*, it is possible to align content descriptions of classical music better than when using cutting-edge text analysis methods. Our method can be extended in future studies to create applications systems to integrate descriptions with musical scores and performances.

1. INTRODUCTION

When listening to classical music, we can enhance our understanding of the music by reading descriptions of the contents of the music simultaneously, which is even truer for those who are not experts of the field of music, such as amateur players in a college orchestra. Those people would want to read content descriptions written by experts when they play or listen to a composition.

A content description of classical music is defined as an objective description related to the structure of the composition that explains specific parts of it, often using technical terms and the names of instruments [19]. Reading those passages along with the music can help people to understand what the part we are listening to means technically,



© Taku Kuribayashi, Yasuhito Asano, Masatoshi Yoshikawa. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** Taku Kuribayashi, Yasuhito Asano, Masatoshi Yoshikawa. “Towards Support for Understanding Classical Music: Alignment of Content Descriptions on the Web”, 16th International Society for Music Information Retrieval Conference, 2015.

which is difficult to understand without preliminary knowledge. An example of a content description of Beethoven’s Symphony No. 9,¹ is the following. “The opening theme, played pianissimo over string tremolos, so much resembles the sound of an orchestra tuning, many commentators have suggested that it was Beethoven’s inspiration.” This example of a content description explains what instruments (strings) are doing technically (pianissimo, tremolos) in a specific part (the opening theme).

Books and the web are two major sources of content descriptions of classical music. Any person having such an interest can find important musical knowledge by reading books such as the well-known *A History of Western Music* [14], which includes not only historical knowledge of the development of western music but also abundant references to other important books. Some encyclopedias contain descriptions of orchestral compositions.

Nevertheless, books have several important limitations. One is that they can hold only a few descriptions. Another problem is that once books are published, they cannot be updated easily or consistently. Although classical music compositions are not increasing to any great degree, the performances are increasing constantly. With the rise of the internet and international communication, there are more descriptions of music and performances. Different perspectives and ways of analysis continue to appear. With the form of printed publications, it is difficult to update the increasing amount of information continuously.

The web is an alternative source of information, offering resources such as “DW3 Classical Music Resources” [11] or Wikipedia. However, it is often difficult to find sufficient information to understand some compositions. Conventional search engines are unsuitable for the vertical search for content descriptions because their results often include commercial websites that do not describe the contents of the compositions. Kuribayashi et al. [19] proposed a method that we can use to collect descriptions from the web. Content descriptions gathered from a number of web pages using their method can be classified into two categories: ones that describe the overall contents of the music, and ones that describe specific parts of the composition. We call the latter ones *partial content descriptions*. Both are essential for technical understanding of the music, although it is often difficult to understand where in the composition partial content descriptions explain. Furthermore,

¹[http://en.wikipedia.org/wiki/Symphony_No._9_\(Beethoven\)](http://en.wikipedia.org/wiki/Symphony_No._9_(Beethoven)), viewed on Jan. 4, 2013

*Current Affiliation: Accenture Japan Ltd.

a single page seldom includes partial content descriptions explains every important part of the composition; a page might describe the introduction in detail, while another page might explain the final part mainly. Therefore, it can be helpful to integrate pieces of information in partial content descriptions from different sources, that is, to check how they complement each other.

We propose a method for the alignment of every pair of paragraphs which are partial content descriptions in different web pages. As a dataset, we manually extract paragraphs corresponding to partial content descriptions from the content descriptions collected from multiple web pages by the method of Kuribayashi et al. [19]. Each alignment clarifies which sentence in a paragraph matches a sentence in the other paragraph. We can understand the music more easily and efficiently by seeing the alignments which integrate the pieces of information in them than merely by reading a single web page. Actually, showing the alignment is beneficial in many situations, as for (1) beginners, (2) experts who want to support those beginners, and (3) future applications.

(1) For beginners, the alignment can help them integrate pieces of information from different websites. If beginners have difficulties understanding one website or feel the need for more information related to a specific part of the composition, they can look at the information that corresponds to the specified part of the description.

(2) For those with a specialized knowledge of classical music, there is always a demand that they want to support beginners as they come to understand music. Web services such as YouTube have several videos that are designed to help beginners to understand classical music. However, preparing all the materials necessary for the explanation is a task that is both difficult and time-consuming. Showing the alignment of sentences provides materials that can support greater understanding. Therefore, showing such an alignment is an important aid to experts who try to support beginners.

(3) In the future, we seek to develop a system that integrates our methods and studies of the analysis of music and music scores; the most important feature is to align content descriptions with the music itself. Beginners will especially benefit from this system because the hardest task for beginners is to ascertain which part of the music the partial content descriptions are referring to. The first step of this ultimate application is analysis of the sentences and their mutual alignment.

The main contributions of this paper are as follows.

- We proposed a novel method named w2DD+PE for aligning partial content descriptions based on dynamic time warping using the following two ideas: (a) the distribution divergence of semantic vectors of words, and (b) *passage expressions*.
- We presented a way to show the aggregated results of our methods for collecting and aligning partial content descriptions.

2. RELATED WORK

This paper deals with various fields of study, including analysis of music, temporal information, multi-document summarization, and parallel corpus discovery; the subject of this paper is the analysis of temporal information in music, and the methodology utilizes the ideas of multi-document summarization and parallel corpus discover models. We will take a look at some of the previous works related to each field of study.

2.1 Musical Knowledge

Music has remained an important topic of research from various aspects, including acoustics, music theory, and psychology. We list a few related works that are closely related to understanding the support and analysis of music.

In the area of understanding support and collecting musical knowledge, Fineman [11] reported a project called “DW3 Classical Music Resources.” The project was a collection of web links that gathered various forms of knowledge related to classical music for college students majoring in music. The link quality was scrutinized by experts, making it easier for students to obtain information that cannot be found easily via conventional web searches. Unfortunately, the project was ceased in 2007.

Other works that are related to the future application of this research include the following. Some studies have been made to analyze the structure of the music itself, such as research by Sumi et al. [36], which created a system for inference of the chord from other data such as the base pitch. Maezawa et al. [24] proposed a system that links the performance and the interpretation of the composition. Using these studies with our research, it would be possible in the future to analyze and extract the music structure and link it to the content descriptions of the composition.

2.2 Temporal Information

As Alonso et al. state in [2], temporal information is an important factor in information retrieval in general. Researches that deal with temporal information in natural language are being studied widely, as in [34], [27], [23], [3], [16]. In order to extract temporal expression, Schilder et al. [32] use finite state transducer (FST); Strötgen et al. [33] use regular expressions, and Mani et al. [26] use machine learning. Chambers et al. [7] focus on the relationships between events, whereas Lapata et al. [20] concentrate on the relationships between expressions in a single sentence. Kimura et al. [17] propose a system that shows chronologically organized information obtained by web searching on a single person. Schilder et al. [32] extract temporal information from news articles.

As we see from these examples, researches on temporal information have various aspects, including many viewpoints on the subject and granularity. In our research, we deal with temporal information in one composition, which is generally an hour or two at the longest.

2.3 Multi-document Summarization

To gain knowledge from multiple sources, summarization of information is an important technique. One type of summarization, the extractive method, chooses subsets of the original document to convey the meaning of the whole text. In the task of multiple document summarization, numerous approaches have been taken. Mani et al. [25] take a graph-based approach, and many of recent studies follow similar ideas [39] [8] [5] [13] [10]. Other approaches include Bayesian models [9] [6], topic models [15], rhetoric-based models [4], and cluster-based models [30] [37].

2.4 Parallel Corpus Discovery

Because we are interested in discovering potential alignments from different documents, we will take a look at previous works that utilize techniques for investigating parallel corpora. Caroline et al. [21] apply dynamic time warping to movie subtitles to construct parallel corpora for machine translation. Previous works on finding parallel texts from bilingual, often non-parallel, corpora include [12], [29], [38] and [35].

3. ALIGNMENT OF DESCRIPTIONS

3.1 Collection of Content Descriptions

We adopted a method of Kuribayashi et al. [19] to collect content descriptions from the web. Their method utilizes labeled latent Dirichlet allocation (labeled LDA) [31] which is a supervised learning to classify documents probabilistically. They proposed eight classes of descriptions (one of them corresponds to content descriptions) contained in the pages obtained by inputting names of compositions to a search engine, and trained labeled LDA with manually-classified 1540 pages. Note that information other than text, such as images or HTML tags, is removed from these pages using `nwc-toolkit`². Applying the trained labeled LDA to paragraphs obtained by inputting the name of a composition to a search engine, we can collect paragraphs corresponding to content descriptions.

From our investigation of a number of content descriptions gathered by this method, we found out that partial content descriptions in a single paragraph are ordered chronologically for a composition. Therefore, the sequence alignment of those sentences is more suitable for integrating information of partial content descriptions than other methods such as matching sentences.

3.2 Bootstrapping Method for Acquiring Passage Expressions

It is quite a difficult task to identify what part of a musical piece that a description corresponds, because most partial content descriptions do not contain measure numbers. Here we try to obtain as much information related to the correspondence of one expression to another. For instance, if we have two descriptions “The first theme is played by

solo flute” and “The lyrical first subject appears after the introduction,” we can see the relationship between them; because the words “theme” and “subject” are semantically similar in this context, we can align these two sentences and understand that the theme is lyrical and is played by the flute. If we have another sentence talking about “solo flute,” we can also infer the relationship of that description to the two sentences above. We have to identify what types of nouns point to the parts of music, which we call *passage expressions*, in order to perform this inference. If we are able to obtain those expressions, then we would be able to use them to align sentences that correspond to the same part of the composition. In the future, we might also be able to employ them to mapping of the actual parts of music by finding measure numbers or giving some information manually.

To obtain the passage expressions, we focus on the grammatical structure of content description sentences. In content descriptions, the most basic structure of sentences is subject-verb-object, where the verb describes the relationship between two passage expressions (subject and object). Therefore, we use a bootstrapping method as in [1], using the relation between the subject and the predicate to extract appropriate nouns. Because a simple bootstrapping method tends to produce noises in the results, we also proposed filtering methods to reduce those noises.

The corpus for the bootstrapping method is 2300 paragraphs which are the top 100 paragraphs obtained by applying the method of Kuribayashi et al. [19] to each of 23 compositions.

First, we prepared an initial list of 14 nouns and 29 verbs for the bootstrapping method. Then we expanded that list when two of the triplet of the subject, the verb and the object (or the object of the preposition) were already in the list, by adding the third word. We did not add the third word when the subject was a personal pronoun (“I”, “we”, “you”, “he”, or “she”) because the word was inappropriate in almost all cases.

Instead of adding all the words that appear in the triplet, we eliminate words that do not fulfill certain conditions to reduce noise words that are not relevant to content descriptions. The following filtering methods incorporate the results of the labeled LDA-based method [19] and `word2vec` [28]³ which converts a word to a vector based on the co-occurrence of words in a corpus; the similarity of words can be calculated using the vectors corresponding the words.

L-LDA Words that are stop words or that do not appear in the training data of labeled LDA in the methods of [19] are not added.

word2vec Words that are below the threshold (0.128 and 0.3) of `word2vec` similarity. `Word2vec` using the same corpus as the one used for our bootstrapping explained above. The `word2vec` similarity used here is defined as the maximum of the similarities between the word and the seed nouns of the bootstrapping method.

²<http://nwc-toolkit.googlecode.com/svn/trunk/docs/tools/text-extractor.html>

³<https://code.google.com/p/word2vec/>

L-LDA && word2vec Only words that fulfill both of the above two are added. The threshold of the word2vec score is 0.128.

L-LDA || word2vec Words that fulfill either one of the two above are added. The threshold of the word2vec score is 0.3.

3.3 Alignment Method using Dynamic Time Warping

We propose a method called “word sets to Distribution Distance-based alignment using Passage Expressions” (w2DD +PE) for finding an alignment of pairs of paragraphs of content descriptions. This method is based on dynamic time warping (DTW), a well-known technique for finding an alignment of two sequences. Applying DTW to paragraphs requires a distance measure of two sentences. We propose a new distance measure employing (a) the distribution of word vectors of each sentence, and (b) passage expressions.

3.3.1 w2DD

A simple measure of distance between two sentences is to take the average of semantic vectors of words in the sentence calculated using word2vec and calculate the cosine distance. However, adopting the average loses much information about how the vectors distribute. Therefore, it is required to propose a new method to capture the feature of the distribution corresponding to each sentence.

The fundamental idea of our new method, named w2DD, is to measure the distance between two sentences by the distance between the distributions of their corresponding vectors. We firstly reduce each vector to a small number of dimensions using principal component analysis because the 200 dimensions obtained by word2vec are too numerous to handle. The number of dimensions is determined empirically, and eleven dimensions were sufficient for the cumulative proportion of 70%. Secondly, we convert the 11-dimension vectors of each sentence into a histogram, in order to apply a distance measure for a pair of probabilistic distributions. For the conversion, we divide each dimension into halves (resulting in 2^{11} subspaces) and count the number of vectors in each subspace; the sequence of the numbers forms the obtained histogram. We tried splitting each dimension into 2, 3, and 4, but the result did not change at all, so we chose 2. Then we calculated the distance of the pair of histograms by the Jensen–Shannon divergence using the following formula:

$$JSD(P \parallel Q) = \frac{1}{2} \left(\sum_x \log \frac{P(x)}{R(x)} + \sum_x \log \frac{Q(x)}{R(x)} \right) \quad (1)$$

where P and Q are the histograms corresponding to the two sets of vectors, x is each subspace, $P(x)$ is the number of vectors of P in x divided by the total number of vectors of P , and $R(x) = \frac{P(x)+Q(x)}{2}$.

3.3.2 Passage Expressions

First, to utilize the information of passage expressions in sentences containing no such expressions, we merge such

sentences into the previous sentence having a passage expression. Then, we calculate the distance of two sentences s_1 and s_2 as follows. Let $sim(p_1, p_2)$ be the cosine similarity between the semantic vectors of passage expressions p_1 in s_1 and p_2 in s_2 . The distance $Dist(s_1, s_2)$ is

$$Dist(s_1, s_2) = \begin{cases} \alpha JSD(s_1, s_2) \\ \quad + (1 - \alpha)(1 - \max_{p_1, p_2}(sim(p_1, p_2))) \\ \quad \text{(if } \max_{p_1, p_2}(sim(p_1, p_2)) \neq 0) \\ JSD(s_1, s_2) \quad \text{(otherwise)} \end{cases} \quad (2)$$

where $JSD(s_1, s_2)$ is the value calculated as in Section 3.3.1. If either one of the paragraphs is without a passage expression, then $\max_{p_1, p_2}(sim(p_1, p_2)) = 0$. Therefore only the Jensen–Shannon divergence matters. Also, α is the coefficient factor, which was set to 0.2, 0.4, 0.6, 0.8, and 1.0.

4. EVALUATION

4.1 Procedure

An input of the alignment is a pair of paragraphs which are partial content descriptions explaining a common section in a composition. The labeled LDA-based method of Kuribayashi et al. [19] is able to collect content descriptions, although it is not able to extract partial content descriptions from them. Consequently, our data set consists of 32 paragraphs (135 sentences) manually extracted from the top 100 paragraphs for each of 10 classical music compositions obtained by their method; the number of pairs are 41. The extraction and assignment of each paragraph to a section is based on keywords corresponding to sections, such as “movement” (the most basic divisions of a music composition), “exposition” and “development” (common structures within a movement). The keywords are selected for the sonata form, and a selection specialized for other types of classic music, “theme and variations” for example, is also possible. A method for automatic extraction and assignment is a candidate of future studies.

To see how each of our ideas work, we used the following variants of methods for calculating the distance between two sentences.

Baseline1 the cosine distance of the averages of word2vec vectors.

Baseline1+PE the cosine distance of the following 400 dimension vectors for the two sentences s_1 and s_2 . The first 200 dimensions are the average of the word2vec vector of each sentence. The second 200 dimensions are the word2vec vector of passage expression p_1 for s_1 or p_2 for s_2 , respectively; p_1 and p_2 are the pair of the closest expressions in terms of word2vec cosine similarity.

Baseline2 the cosine distance calculated using sentence2vec.⁴ This is an implementation of *Paragraph Vector* proposed by Le and Mikolov [22],

⁴<https://github.com/klb3713/sentence2vec>

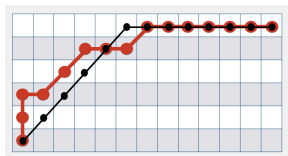


Figure 1. Example of how to calculate F -measure.

which is an advanced method of word2vec that incorporates the order of words in a sentence to represent its semantics. Their experiments showed that *Paragraph Vector* performs better than previous methods for several tasks; word vector averaging, Naive Bayes, SVMs, and recursive neural network for a sentiment analysis task; vector averaging, bag-of-words, and bag-of-bigrams for an information retrieval task.

Baseline2+PE the cosine distance calculated using sentence2vec incorporated with the passage expression vector by the same way Section 3.3.2.

w2DD the method described in Section 3.3.1

w2DD+PE the method described in Section 3.3.2.

For Baseline1+PE, Baseline2+PE, and w2DD+PE, we used the filtered list of passage expressions described in Section 3.1.

The ground truth for the alignment of each pair of paragraphs was created manually by one of the authors who is an enthusiast of classical music, with the help of various books and websites on the compositions. We evaluate each method using precision, recall, and F -measure. We explain below how they are calculated employing Figure 1 which illustrates an example of the result of alignment of two paragraphs. The red dots represent the manual alignment result, and black dots indicate the output a method; in the manual alignment, the first sentence of paragraph 1 (x -axis) corresponds to the first, second, and third sentences of paragraph 2 (y -axis), the second sentence of paragraph 1 corresponds to the third sentence of paragraph2, and so on. The precision is the number of matching red and black dots over the number of black dots, 9/13 in this case. The recall is the number of matching red and black dots over the number of red dots, 9/15 in this case. The F -measure is defined by the following equation.

$$F = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{3}$$

4.2 Results

Tables 1, 2, and 3 present the experimental results. Comparing the “No PE” row with the others in each table, we see that employing PE improves the results in general. Comparing the three tables, we see that w2DD performs much better than baseline methods. Especially, w2DD+PE with L-LDA ($\alpha = 0.2$) and w2DD+PE with L-LDA && word2vec ($\alpha = 0.2$) are the methods that resulted in the best F -measure (shown in bold in the Table 3).

Table 1. Results of Baseline1 (No PE) and Baseline1+PE.

Method	Precision	Recall	F -measure
No PE	0.595	0.534	0.563
No Filtering	0.608	0.633	0.620
L-LDA	0.618	0.647	0.632
word2vec (0.128)	0.582	0.618	0.600
word2vec (0.3)	0.562	0.607	0.584
L-LDA && word2vec	0.592	0.629	0.610
L-LDA word2vec	0.591	0.615	0.602

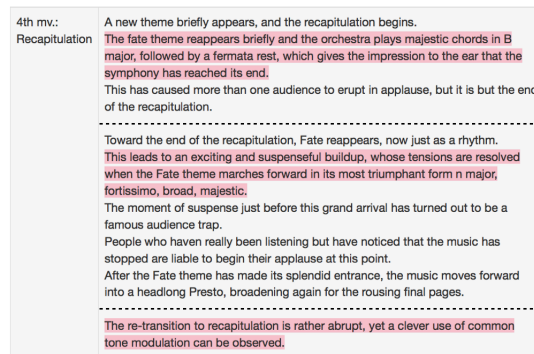


Figure 2. Visualization of Tchaikovsky’s Symphony No.5 (Each block of sentences separated by dotted lines is from a single web page.)

Because the baseline methods “compress” the word vectors in a sentence into a single vector, they are considered to lose much information of the words. On the other hand, w2DD keeps the information on how varied the words are in the sentence.

The numbers of passage expressions employed in L-LDA and L-LDA && word2vec were 30 and 26, respectively. Passage expressions were generally effective as mentioned above, while higher alpha often made the performance worse. These results would indicate that the word distribution employed in w2DD is more important than passage expressions.

4.3 Visualization

To present the results of our methods to users for understanding support, we created a prototype of a system to visualize them in a table form, whose examples can be accessed online.⁵ Each row corresponds to a part of music. Figure 2 shows a single row corresponding to the “4th movement” of the table for Tchaikovsky’s “Symphony No.5.” In this row, there are three blocks of sentences separated by dotted lines, each of which indicates a paragraph retrieved from one web page. As we hover the cursor over one of the sentences, the sentences of other descriptions that are aligned with that sentence by our method is highlighted (shown as pink in the figure).

The recapitulation part builds up the tension and ends

⁵<http://bit.ly/1vHMkgm>

Table 2. Results of Baseline2 (No PE) and Baseline2+PE.

Method	α	Precision	Recall	<i>F</i> -measure
No PE	-	0.610	0.550	0.578
No Filtering	0.2	0.621	0.562	0.590
	0.4	0.630	0.568	0.597
	0.6	0.603	0.544	0.572
	0.8	0.474	0.417	0.444
L-LDA	0.2	0.659	0.598	0.627
	0.4	0.670	0.604	0.635
	0.6	0.627	0.565	0.594
	0.8	0.474	0.417	0.444
word2vec (0.128)	0.2	0.619	0.562	0.589
	0.4	0.630	0.568	0.597
	0.6	0.603	0.544	0.572
	0.8	0.474	0.417	0.444
word2vec (0.3)	0.2	0.625	0.565	0.593
	0.4	0.627	0.565	0.594
	0.6	0.603	0.544	0.572
	0.8	0.474	0.417	0.444
L-LDA && word2vec	0.2	0.659	0.598	0.627
	0.4	0.670	0.604	0.635
	0.6	0.627	0.565	0.594
	0.8	0.474	0.417	0.444
L-LDA word2vec	0.2	0.616	0.559	0.586
	0.4	0.627	0.565	0.594
	0.6	0.603	0.544	0.572
	0.8	0.474	0.417	0.444

up with a brief stop. From the three highlighted descriptions, it is readily apparent that common tone modulation is used cleverly in the recapitulation; the fate theme engenders a suspenseful buildup; and a fermata rest follows the majestic chords in B major. By reading the aligned descriptions that are retrieved from multiple pages, a more detailed and thorough view of the part of the music can be obtained than by reading just one description.

5. CONCLUDING REMARKS

As described in this paper, we proposed methods for supporting the understanding of classical music using mutual alignment of partial content descriptions. Our method w2DD+PE uses word sets to Distribution Distance (w2DD) and the concept of *passage expressions*, which are expressions that serve as the key to identification of which parts of the music the descriptions correspond to. Although the concept of passage expressions is unique to the field of classical music, w2DD can be applied to other domains of text data. It is one of our future tasks to apply w2DD to other datasets.

Future studies will be undertaken to create an application system that can help beginners to appreciate classical music. By integrating our methods with studies of musical analysis such as [36] and [24], or other applications of music-related information retrieval such as [18], it is expected to be possible to support beginners in their efforts to understand and enjoy music.

Table 3. Results of w2DD (No PE) and w2DD+PE.

Method	α	Precision	Recall	<i>F</i> -measure
No PE	-	0.746	0.688	0.716
No Filtering	0.2	0.671	0.733	0.701
	0.4	0.671	0.733	0.701
	0.6	0.690	0.748	0.718
	0.8	0.667	0.718	0.691
L-LDA	0.2	0.680	0.780	0.727
	0.4	0.678	0.774	0.723
	0.6	0.675	0.760	0.715
	0.8	0.669	0.745	0.705
word2vec (0.128)	0.2	0.639	0.721	0.678
	0.4	0.639	0.721	0.678
	0.6	0.651	0.718	0.683
	0.8	0.658	0.718	0.687
word2vec (0.3)	0.2	0.648	0.736	0.689
	0.4	0.648	0.736	0.689
	0.6	0.652	0.739	0.693
	0.8	0.659	0.745	0.699
L-LDA && word2vec	0.2	0.680	0.780	0.727
	0.4	0.678	0.774	0.723
	0.6	0.673	0.757	0.712
	0.8	0.669	0.745	0.705
L-LDA word2vec	0.2	0.637	0.733	0.681
	0.4	0.637	0.733	0.681
	0.6	0.641	0.727	0.681
	0.8	0.651	0.736	0.691

6. ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number 15K00423 and the Kayamori Foundation of Informational Science Advancement.

7. REFERENCES

- [1] E. Agichtein and L. Gravano. Snowball: Extracting relations from large plain-text collections. In *Proc. of the 5th ACM conference on Digital Libraries*, pages 85–94, 2000.
- [2] O. Alonso, M. Gertz, and R. Baeza-Yates. On the value of temporal information in information retrieval. In *ACM SIGIR Forum*, volume 41-2, pages 35–41, 2007.
- [3] O. Alonso, M. Gertz, and R. Baeza-Yates. Clustering and exploring search results using timeline constructions. In *Proc. of the 18th CIKM*, pages 97–106, 2009.
- [4] J. Atkinson and R. Munoz. Rhetorics-based multi-document summarization. *Expert Systems with Applications*, 40(11):4346–4352, 2013.
- [5] E. Canhasi and I. Kononenko. Weighted archetypal analysis of the multi-element graph for query-focused multi-document summarization. *Expert Systems with Applications*, 41(2):535–543, 2014.
- [6] A. Celikyilmaz and D. Hakkani-Tür. Discovery of topically coherent sentences for extractive summarization. In *Proc. of HLT '11*, pages 491–499, 2011.
- [7] N. Chambers, S. Wang, and D. Jurafsky. Classifying temporal relations between events. In *Proc. of the 45th*

- Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 173–176, 2007.
- [8] J. Christensen, Mausam, S. Soderland, and O. Etzioni. Towards Coherent Multi-Document Summarization. In *Proc. of HLT-NAACL '13*, pages 1163–1173, 2013.
- [9] H. Daumé III and D. Marcu. Bayesian query-focused summarization. In *Proc. of the 44th Annual Meeting of the ACL*, pages 305–312, 2006.
- [10] R. Ferreira, L. de Souza Cabral, F. Freitas, R. D. Lins, G. de França Silva, S. J. Simske, and L. Favaro. A multi-document summarization system based on statistics and linguistic treatment. *Expert Systems with Applications*, 41(13):5780–5787, 2014.
- [11] Y. Fineman. DW3 Classical Music Resources: Managing Mozart on the Web. *Libraries and the Academy*, 1(4):383–389, 2001.
- [12] P. Fung and P. Cheung. Mining verynon-parallel corpora: Parallel sentence and lexicon extraction via bootstrapping and em. In *Proc. of EMNLP*, pages 57–63, 2004.
- [13] G. Glavaš and J. Šnajder. Event graphs for information retrieval and multi-document summarization. *Expert Systems with Applications*, 41(15):6904–6916, 2014.
- [14] D. J. Grout, C. V. Palisca, et al. *A History of Western Music*. Number Ed. 5. WW Norton & Company, Inc., 1996.
- [15] A. Haghighi and L. Vanderwende. Exploring content models for multi-document summarization. In *Proc. of the NAACL '09*, pages 362–370, 2009.
- [16] J. Hobbs and J. Pustejovsky. Annotating and reasoning about time and events. In *Proc. of AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, volume 3, pages 74–82, 2003.
- [17] R. Kimura, S. Oyama, H. Toda, and K. Tanaka. Creating personal histories from the Web using namesake disambiguation and event extraction. In *Web Engineering*, pages 400–414. 2007.
- [18] P. Knees, T. Pohle, M. Schedl, and G. Widmer. A music search engine built upon audio-based and web-based similarity measures. In *Proc. of the 30th SIGIR*, pages 447–454, 2007.
- [19] T. Kuribayashi, Y. Asano, and M. Yoshikawa. Ranking method specialized for content descriptions of classical music. In *Poster Proc. of the 22nd WWW*, pages 141–142, 2013.
- [20] M. Lapata and A. Lascarides. Learning Sentence-internal Temporal Relations. *Journal of Artificial Intelligence Research (JAIR)*, 27:85–117, 2006.
- [21] C. Lavecchia, K. Smaïli, and D. Langlois. Building parallel corpora from movies. In *Proc. of the 4th NLPCS*, 2007.
- [22] Q. V. Le and T. Mikolov. Distributed representations of sentences and documents. *Proc. of the 31st ICML*, pages 1188–1196, 2014.
- [23] X. Ling and D. S. Weld. Temporal Information Extraction. In *The AAAI Conference on Artificial Intelligence*, pages 1385–1390, 2010.
- [24] A. Maezawa, M. Goto, and H. G. Okuno. Query-By-Conducting: An interface to retrieve classical-music interpretations by real-time tempo input. In *Proc. of the 11th ISMIR*, pages 477–482, 2010.
- [25] I. Mani and E. Bloedorn. Multi-document summarization by graph search and matching. In *Proc. of AAAI'97/IAAI'97*, pages 622–628, 1997.
- [26] I. Mani, M. Verhagen, B. Wellner, C. Lee, and J. Pustejovsky. Machine learning of temporal relations. In *Proc. of the 44th Annual Meeting of the ACL*, pages 753–760, 2006.
- [27] P. Mazur and R. Dale. Wikiwars: A new corpus for research on temporal expressions. In *Proc. of the EMNLP 2010*, pages 913–922, 2010.
- [28] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *Proc. of Workshop at ICLR*, 2013.
- [29] D. S. Munteanu and D. Marcu. Extracting parallel sub-sentential fragments from non-parallel corpora. In *Proc. of the 44th Annual Meeting of the ACL*, pages 81–88, 2006.
- [30] D. R. Radev, H. Jing, M. Styś, and D. Tam. Centroid-based summarization of multiple documents. *Information Processing & Management*, 40(6):919–938, 2004.
- [31] D. Ramage, D. Hall, R. Nallapati, and C.D. Manning. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Proc. of EMNLP, Volume 1*, pages 248–256, 2009.
- [32] F. Schilder and C. Habel. From temporal expressions to temporal information: Semantic tagging of news messages. In *Proc. of the Workshop on Temporal and Spatial Information Processing-Volume 13*, page 9, 2001.
- [33] J. Strötgen and M. Gertz. HeideTime: High quality rule-based extraction and normalization of temporal expressions. In *Proc. of the 5th International Workshop on Semantic Evaluation*, pages 321–324, 2010.
- [34] J. Strötgen and M. Gertz. Multilingual and cross-domain temporal tagging. *Language Resources and Evaluation*, 47(2):269–298, 2013.
- [35] F. Su and B. Babych. Measuring comparability of documents in non-parallel corpora for efficient extraction of (semi-)parallel translation equivalents. In *Proc. of the Joint Workshop on ESIRMT and HyTra*, EACL 2012, pages 10–19, 2012.
- [36] K. Sumi, K. Itoyama, K. Yoshii, K. Komatani, T. Ogata, and H. G. Okuno. Automatic chord recognition based on probabilistic integration of chord transition and bass pitch estimation. In *Proc. of the 9th ISMIR*, pages 39–44, 2008.
- [37] X. Wan and J. Yang. Multi-document summarization using cluster-based link analysis. In *Proc. of SIGIR '08*, pages 299–306, 2008.
- [38] D. Wu and P. Fung. Inversion transduction grammar constraints for mining parallel sentences from quasi-comparable corpora. In R. Dale, K. Wong, J. Su, and O. Kwong, editors, *Natural Language Processing — IJCNLP 2005*, LNCS 3651, pages 257–268, 2005.
- [39] L. Zhao, L. Wu, and X. Huang. Using query expansion in graph-based approach for query-focused multi-document summarization. *Information Processing & Management*, 45(1):35–41, 2009.