Order Estimation of Japanese Paragraphs by Supervised Machine Learning

Satoshi Ito, Masaki Murata, Masato Tokuhisa
Department of Information and Electronics
Tottori University
4-101 Koyama-Minami, Tottori 680-8552, Japan
Email: {s092007,murata,tokuhisa}@ike.tottori-u.ac.jp

Qing Ma
Department of Applied Mathematics and Informatics
Ryukoku University
Seta, Otsu, Shiga 520-2194, Japan
Email: qma@math.ryukoku.ac.jp

Abstract*— In this paper, we propose a method to estimate the order of paragraphs by supervised machine learning. We use a support vector machine (SVM) for supervised machine learning. The estimation of paragraph order is useful for sentence generation and sentence correction. The proposed method obtained a high accuracy (0.86) in the order estimation experiments of the first two paragraphs of an article and achieved the same accuracy as manual estimation. In addition, it obtained a higher accuracy than the baseline methods in the experiments using two paragraphs of an article. We performed feature analysis and we found that adnominals, conjunctions, and dates were effective for the order estimation of the first two paragraphs, and the ratio of new words and the similarity between the preceding paragraphs and an estimated paragraph were effective for the order estimation of all pairs of paragraphs. Moreover, we compared the order estimation of sentences and paragraphs and clarified differences. For the order estimation of the first two paragraphs, paragraph order estimation would be easier than sentence order estimation because paragraphs have more information than sentences. For the order estimation of all pairs of paragraphs, paragraph order estimation would be more difficult than sentence order estimation because a story may conclude in a paragraph.

I. Introduction

The estimation of sentence order (sometimes referred to as sentence ordering) is a problem that stems from sentence generation and sentence correction [1]–[3]. When generating text that consists of multiple sentences/paragraphs, arranging them in an appropriate order is necessary to understand the text easily. In this study, we employ supervised machine learning to estimate the appropriate order. In addition, we utilize a high-performance support vector machine (SVM) for supervised learning.

Previous studies of the sentence/paragraph order estimation with supervised learning include research by Uchimoto et al. [4] and Hayashi et al. [5], considering word order and sentence order estimations, respectively. Thus, we consider the order estimation of paragraphs.

In this study, we generate two types of problems: original order and reverse order for pairs of paragraphs extracted from a corpus (newspapers). We determine the correct order by machine learning. Furthermore, we analyze features that facilitate paragraph order estimation. This study is conducted in Japanese.

The characteristics of this study are described as follows.

- This study employs supervised learning for paragraph order estimation.
- In our supervised method, training data can be automatically constructed from a corpus (without tags). Our method does not require a manual construction of training data.
- In our proposed method using supervised learning, we can find important information in paragraph order estimation by examining the features. In our experiments, we found that adnominals, conjunctions, dates, the ratio of new words and the similarity between the preceding paragraphs and an estimated paragraph were effective for paragraph order estimation.
- When estimating the order of the first two paragraphs, we obtained a high accuracy rate (0.86) using the proposed method. The accuracy rate was equal to that of manual estimation.
- When estimating the order of two adjacent paragraphs and the order of two paragraphs (pairs of all paragraphs), the accuracy rates of the proposed method were 0.63 and 0.67, respectively. These are higher than those of baseline methods assuming that a paragraph having more nouns in common with the preceding paragraphs is likely to be the first of the pair.
- We compare the order estimation of sentences and paragraphs to clarify differences. In our experiments, we found that paragraph order estimation would be easier than sentence order estimation for the order estimation of the first two paragraphs and paragraph order estimation would be more difficult than sentence order estimation for the order estimation of all pairs of paragraphs.

II. Related work

Uchimoto et al. performed a study of sentence generation to estimate the order of words on the basis of the phrase dependency information using the maximum entropy method [4]. They assumed that the word order in a corpus is correct and therefore built the training data for the word order from the corpus. Their method does not require a manual construction of training data.

For sentence order estimation in newspaper articles, Hayashi et al. performed a study employing supervised machine learning with a large number of features [5]. They selected two sentences from newspaper articles as a pair and generated one sentence pair in the original order (positive example) and another in the reverse order (negative example). They estimated sentence order by judging whether a sentence pair was positive or negative using supervised machine learning. In addition, they referred to a study by Uchimoto et al. and automatically constructed the data for machine learning from a corpus. In their experiments, they utilized three cases for order estimation: the first two sentences in a paragraph, two adjacent sentences in a paragraph, and all pairs of sentences in a paragraph. Furthermore, they compared their results with those from Lapata's study using a probability technique [6] and reported that they obtained higher performance than Lapata's method.

The aforementioned studies considered word or sentence order estimation. In contrast, our study considers paragraph order.

Lapata regarded existing sentences as training data and calculated the probabilities of features appearing in two adjacent sentences in the training data [6]. By utilizing the total product of probabilities, she calculated the probability that the second sentence was placed after the first sentence and determined the sentence order based on the probability. She utilized verb order, common nouns, and sentence structures from two sentences as features.

In her study, she did not employ machine learning for order estimation. In contrast, our study employs machine learning.

For constructing summaries from multiple documents [7]–[10], Okazaki et al. performed a study to estimate the order of extracted sentences [9]. By considering the order of sentences in an original text prior to constructing a summary, they estimated the order of extracted sentences by utilizing the original order. Danushka et al. also studied sentence order estimation for constructing summaries from multiple documents [10]. Their estimation employed supervised machine learning with various features, such as time information, the semantic closeness of content, and the order of sentences in the original documents before constructing summaries.

In these studies, the information from the original documents was utilized before constructing summaries. In contrast, our study does not utilize such information. If the sentence/paragraph order can be estimated without such information, we can use the method also for tasks other than summarization, which include the correction of sentences/paragraphs that are not in an appropriate order.

III. Task and proposed method

A. The task

The task in this study is as follows. An article is the input and the order of only the first several paragraphs is determined. The order of the remaining paragraphs is not determined. The task is to estimate the order of two paragraphs among the remaining undetermined paragraphs. The information that can be utilized for estimation are the two paragraphs to be estimated and the paragraphs that precede the two target paragraphs (see Figure 1).

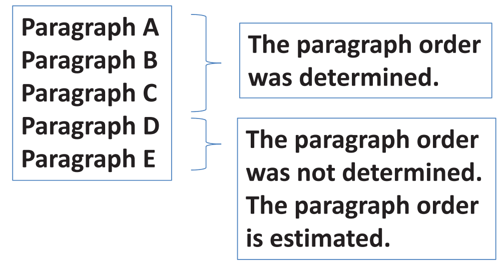

Fig. 1: The model of the task

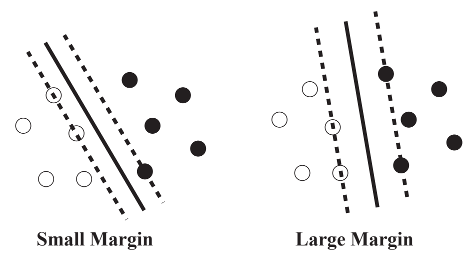

Fig. 2: Maximizing the margin

B. Proposed method

We need to estimate the order of two paragraphs: A and B. These paragraphs are the input to the system, and our method judges whether the order "A→B" is correct by employing SVM.

The training and test data are composed of two paragraphs extracted from a text. From these paragraphs, we construct two sequences: original order and reverse order. The paragraphs in the original order are a positive example, and the paragraphs in reverse order are a negative example. We refer to the studies performed by Uchimoto et al. [4] and Hayashi et al. [5] and automatically construct the training and test data for machine learning from a corpus by assuming that the paragraph order in the corpus is correct.

C. Support vector machine

In this section, we explain the SVM that we use for machine learning.

In SVM, data consisting of two categories are classified by dividing space with a hyperplane. When the margin between examples that belong to one category and the other category in the training data is larger (see Figure 2[1]), the probability of incorrectly selecting categories in open data is believed to be smaller. The hyperplane maximizing the margin is determined,

[1]In the figure, the white and black circles indicate examples that belong to one category and the other category, respectively. The solid line indicates the hyperplane dividing space, and the broken lines indicate planes at the boundaries of the margin regions.

and classification is done by using this hyperplane. Although the basics of the method are as described above, for extended versions of the method in general, the inner region of the margin in the training data can include a small number of examples, and the linearity of the hyperplane is converted to nonlinearity by using kernel functions. Classification in the extended methods is equivalent to classification using the following discernment function, and the two categories can be classified on the basis of whether the output value of the function is positive or negative [11], [12]:

$$\begin{aligned} f(\mathbf{x}) &= sgn\left(\sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \qquad (1) \\ b &= -\frac{max_{i,y_i=-1} b_i + min_{i,y_i=1} b_i}{2} \\ b_i &= \sum_{j=1}^{l} \alpha_j y_j K(\mathbf{x}_j, \mathbf{x}_i), \end{aligned}$$

where $\mathbf{x}$ is the context (a set of features) of an input example; $\mathbf{x}_i$ and $y_i (i = 1, ..., l, y_i \in \{1, -1\})$ indicate the context of the training data and its category, respectively; the function sgn is defined as follows:

$$sgn(x) = \begin{array}{ll} 1 & (x \geq 0), \\ -1 & (otherwise). \end{array} \qquad (2)$$

Each $\alpha_i (i = 1, 2...)$ is fixed when the value of $L(\alpha)$ in Equation (3) is maximum under the conditions of Equations (4) and (5).

$$L(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(\mathbf{x_i}, \mathbf{x_j}) \qquad (3)$$

$$0 \leq \alpha_i \leq C \; (i = 1, ..., l) \qquad (4)$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \qquad (5)$$

K is called a kernel function. Various types of kernel functions can be used; however, in this paper, we use a polynomial function as follows:

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^d, \qquad (6)$$

where C and d are constants set by experimentation. In this paper, C and d are fixed as 1 and 2 for all experiments, respectively.[2] A set of $\mathbf{x}_i$ that satisfies $\alpha_i > 0$ is called a support vector, and the portion used to perform the sum in Equation (1) is calculated by only using examples that are support vectors. We used the software TinySVM [12], developed by Kudoh, as the SVM.

[2]We confirmed that d = 2 produced good performance in preliminary experiments.

TABLE I: Features

ID	Explanation
a1	Words and their parts of speech (POS) in paragraph A (or B).
a2	Words and their POS in the first-half (or second-half) parts of sentences that are divided by a Japanese postpositional particle *wa* in paragraph A (or B).
a3	Whether an adnominal or a conjunction appears at the beginning of paragraph A (or B).
a4	Whether a date (day) appears in paragraph A (or B).
a5	The number of nouns appearing in paragraphs A and B.
a6	The number of nouns appearing in paragraph B (or A) and not appearing in paragraph A (or B).
a7	The difference between the values of a6 when A and B are exchanged.
a8	The number of nouns appearing in paragraph A (or B) and in the first-half parts of sentences that are divided by a Japanese postpositional particle *wa* in paragraph B (or A).
a9	The number of nouns appearing in the first-half parts of sentences that are divided by a Japanese postpositional particle *wa* in paragraph B (or A) and not appearing in paragraph A (or B).
a10	The difference between the values of a8 when A and B are exchanged.
a11	The difference between the values of a9 when A and B are exchanged.
a12	The number of nouns appearing in paragraph A (or B) and in the paragraphs before paragraphs A and B.
a13	The number of nouns appearing in the paragraphs before paragraphs A and B and not appearing in paragraph A (or B).
a14	The difference between the values of a12 when A and B are exchanged.
a15	The difference between the values of a13 when A and B are exchanged.
a16	The number of nouns appearing in the first-half parts of a2 of paragraph A (or B) and in the paragraphs before paragraphs A and B.
a17	The number of nouns appearing in the paragraphs before paragraphs A and B and not appearing in the first-half parts of a2 of paragraph A (or B).
a18	The difference between the values of a16 when A and B are exchanged.
a19	The difference between the values of a17 when A and B are exchanged.
a20	The difference between the number of words (new words) not appearing in the paragraphs before two paragraphs A and B and appearing in paragraph A and the number of words not appearing in the paragraphs before paragraphs A and B and appearing in paragraph B.
a21	The difference between the ratio of new words appearing in paragraph A and that appearing in paragraph B.
a22	The difference between the number of words appearing in the last sentence of paragraph A and in the first sentence of paragraph B and the number of words appearing in the last sentence of paragraph B and in the first sentence of paragraph A.

D. Features used in our proposed method

Here, we explain features (information utilized for classification) that are required for machine learning. The features utilized in this study are shown in Table I. Each feature has additional information indicating whether it appears in the first or second paragraph of the two target paragraphs, denoted as A and B, respectively. To extract words and parts of speech, we utilize the *ChaSen* morphological analyzer [13]. All features are binary-valued.

Some features are explained in more detail as follows.

1) a1: Words and their parts of speech (POS) in paragraph A (or B): The parts of speech used in a1 are a noun, an adjective, an adjectival noun, a verb, an adverb, an adnominal, and a conjunction. Only the words whose parts of speech are the same as those above are used as a feature of a1.

2) a2: Words and their POS in the first half (or second half) of sentences that are divided by a Japanese postpositional particle wa in paragraph A (or B): Because a paragraph comprises plural sentences, a Japanese postpositional particle *wa* often occurs in a paragraph. We divide a paragraph into component sentences. In a sentence including a particle *wa*, we divide the sentence into two parts, the first half and the second half, using the particle *wa*. A sentence without a particle *wa* is entirely handled as a second-half part. We use words and their POS in the first-half parts of sentences as features and those in the second-half parts of sentences as different features. In Japanese, old information is described in the part before *wa*, whereas new information is described in the part after *wa*. The old and new information is related to paragraph order; thus, we use this feature a2 in our method.

3) a3: Whether an adnominal or a conjunction appears at the beginning of paragraph A (or B): When a demonstrative (including *kono* (This), *sono* (Its), and so on) appears at the beginning of a paragraph, it must refer to a word appearing beforehand. In addition, when a conjunction (including *matawa* (Otherwise), *shikashi* (However), and so on) is used, it must be used for the relationship with a previous context. Therefore, when an adnominal or a conjunction appears in the beginning of a sentence, it is believed that there is a preceding paragraph.

4) a4: Whether a date (day) appears in paragraph A (or B): When a paragraph includes important events in newspaper articles, a date (day) is likely to be written in the paragraph. Important events are likely to be written at the beginning of the article. Therefore, a paragraph where a date (day) is written is likely to appear at the beginning of the article. To exploit this tendency, we use feature a4.

5) a5: The number of nouns appearing in paragraphs A and B: Many nouns appear in a paragraph. From this fact, we make a feature to observe the number of nouns (called the common noun number) appearing in both paragraphs A and B. Based on the common noun number, we make the following 15 cases: a range of more than 0, more than 1, ..., more than 9, a range of 0-1, 2-3, ..., 6-7, and a range of more than 7. These 15 cases are used as features.

6) a6: The number of nouns appearing in paragraph B (or A) and not appearing in paragraph A (or B): We calculate the number of nouns appearing in paragraph B (or A) and not appearing in paragraph A (or B). We make some cases based on the number and use them as features, as in a5.

7) a12: The number of nouns appearing in paragraph A (or B) and in the paragraphs before paragraphs A and B: When the content of adjacent paragraphs is similar, the paragraph order is better estimated. From this, we establish the feature a12 such that, between A and B, the paragraph in which there are more common nouns with all preceding paragraphs is judged to appear earlier.

8) a20: The difference between the number of words (new words) not appearing in the paragraphs preceding A and B and appearing in paragraph A, and the number of words not appearing in the paragraphs preceding A and B and appearing in paragraph B: We first calculate the number of words (called new words) not appearing in the paragraphs preceding A and B and appearing in paragraph A. i.e., we calculate the number of new words first appearing in paragraph A. We call this number N_A. We also calculate the same kind of number against paragraph B. We call this number N_B and calculate $N_A - N_B$. Based on the calculated results, we make the following cases; a range of less than 0 and a range of more than 0. The two cases are used as features. In this feature a20, we use only words whose parts of speeches are used in a1.

9) a21: The difference between the ratio of new words appearing in paragraph A and in paragraph B: We calculate the ratio of new words appearing in paragraph A (or B). We call the ratio R_A (or R_B). Here, the ratio of new words is the resultant value dividing the number of new words by the number of all words appearing in the paragraph. We calculate $R_A - R_B$. Based on this, we make features such that a paragraph whose ratio of new words is larger is judged to appear later.

10) a22: The difference between the number of words appearing in the last sentence of paragraph A and in the first sentence of paragraph B, and the number of words appearing in the last sentence of paragraph B and in the first sentence of paragraph A: When paragraph B follows paragraph A, the last sentence of paragraph A and the first sentence of paragraph B will be similar and will have many nouns in common. Feature a22 can verify this. We calculate the subtraction of the number of words appearing in the last sentence of paragraph A and in the first sentence of paragraph B, and the number of words appearing in the last sentence of paragraph B and in the first sentence of paragraph A, and classify the value into three cases: plus, even (0), and minus. The three cases are used as features.

IV. BASELINE METHODS

Information in two adjacent paragraphs will possibly be very similar. Therefore, we utilize the baseline method as follows. The two paragraphs for estimation are denoted as A and B. We count the number of words that appear in the paragraphs immediately preceding paragraphs A and B that also appear in paragraph A (or B). When the number of repeated words in paragraph A is higher than that in paragraph B, "A→B" is the correct order. This is baseline method 1.

In addition, we utilize the baseline method as follows. We calculate the ratio of words (new word ratio) that appear in paragraph A (or B) and do not appear in the immediately preceding paragraphs. When the ratio for paragraph A is higher than that for paragraph B, "B→A" is correct. This is baseline method 2.

In this study, we compare the performance of baseline methods 1 and 2 with the performance of our proposed method.

TABLE II: Accuracy rates of the proposed method and baseline methods 1 and 2

	Proposed method	Baseline methods	
		Baseline 1	Baseline 2
Case 1	0.8560		
Case 2	0.6312	0.5277	0.5257
Case 3	0.6723	0.6181	0.5282

TABLE III: Accuracy rates of the proposed method, baseline methods, and manual estimation

	Our method	Baselines		Subject (Manual)		
		1	2	A	B	Average
Case 1	0.88			0.92	0.84	0.88
Case 2	0.60	0.56	0.54	0.68	0.64	0.66
Case 3	0.65	0.56	0.54	0.84	0.70	0.77

V. EXPERIMENT

A. Experimental conditions

We utilized Mainichi newspaper articles (July 1992) as training data.

We utilized the following three cases for pairs of paragraphs. Case 1: The first two paragraphs in an article. Case 2: Pairs of all adjacent paragraphs in an article. Case 3: Pairs of all paragraphs. Baseline methods 1 and 2 cannot be utilized for Case 1, because the paragraphs preceding the estimated paragraphs are required.

For Case 1, features a12-a21 were not utilized because they require the preceding paragraphs. For Cases 2 and 3, estimating the order by utilizing conjunctions or adnominals is difficult. Thus, in Cases 2 and 3, we did not utilize a3.

For training data, 1,550 paragraph pairs were utilized for Case 1, 29,434 for Case 2, and 80,248 for Case 3.

We used accuracy rates for evaluation. An accuracy rate is the result dividing the number of correctly estimated pairs by the number of input pairs.

B. Comparison between the proposed method and baseline methods

We utilized Mainichi newspaper articles (August 1, 1992) for test data. We utilized 418 paragraph pairs for Case 1, 3,146 for Case 2, and 7,374 for Case 3. Table II shows the accuracy rates of the proposed method and baseline methods 1 and 2.

In Case 1, our proposed method obtained high accuracy (0.8517). In Cases 2 and 3, the accuracies of our proposed method (0.6312 and 0.6723) were not as high as that of Case 1; however, they were higher than those of baseline methods 1 (0.5277 and 0.6181) and 2 (0.5257 and 0.5282).

The baseline methods use the similarity between paragraphs. Our method use many kinds of information on the basis of features used in machine learning. Because the accuracies of our method were higher than those of the baseline methods, We found that the use of many kinds of information was better than using the similarity between paragraphs only.

C. Comparison with manual estimation

We compared the proposed method, the baseline methods, and manual estimation. Manual estimation was separately performed by two individuals (subjects), A and B.

We randomly selected 50 paragraph pairs from Mainichi newspaper articles as test data for Cases 1 to 3. The pairs for Case 1 were from June 1993, Case 2 were from July 1993, and Case 3 were from August 1993.

We show the accuracy of our proposed method, the baseline methods, and manual estimation in Table III. "Average" shows the average accuracy of manual estimations.

In Table III, the performance of the proposed method (0.88) was the same as that of the manual estimation (0.88) in Case 1. In Case 2, the performance of the proposed method (0.60) was higher than those of baseline methods 1 (0.56) and 2 (0.54); however, it was lower than that of the manual estimation (0.66). Because the performance of the manual estimation was also relatively low (0.66) in Case 2, it is believed that estimating order in this case was particularly difficult. As in Case 2, in Case 3, the performance of the proposed method (0,65) was higher than those of baseline methods 1 (0.56) and 2 (0.54); however, it was lower than that of the manual estimation (0.77).

D. Feature analysis

We utilized the following method for feature analysis. We constructed a data item with only one feature and classified it by SVM. The feature with a larger distant against a separating hyperplane is likely to be more important.

We found that in Case 1, adnominals, conjunctions, and dates (features a3 and a4) were effective for order estimation of the first two paragraphs. In Case 2, we found that new word ratios (feature a21) and the similarity between the preceding paragraphs and an estimated paragraph (features a12 and a14) were effective for the order estimation of all pairs of adjacent paragraphs. In Case 3, we found that the number of new words and new word ratios (features a20 and a21) were very effective for the order estimation of all pairs of paragraphs.

E. Comparison of paragraph and sentence order estimation

We examined the difference between the sentence and paragraph order estimation by comparing our results with those for sentence order estimation performed by Hayashi et al. [5]. Cases 1 to 3 correspond in both studies. For Case 1, Hayashi et al. estimated the order of the first two sentences in a paragraph, and the information utilized to perform the estimation was restricted to those sentences. For Case 2, the order of two adjacent sentences in a paragraph was estimated, and the information utilized to perform the estimation was the two sentences and the preceding sentences in the paragraph. For Case 3, the order of two sentences in a paragraph was estimated, and the information utilized to perform the estimation was the two sentences and the preceding sentences in the paragraph. Articles and paragraphs in our study correspond respectively to paragraphs and sentences in Hayashi et al. The

TABLE IV: Accuracy in Hayashi et al.

	Hayashi method	Manual estimation
Case 1	0.79	0.82
Case 2	0.67	0.87
Case 3	0.71	0.72

results from the study by Hayashi et al. are shown in Table IV. As in our study, Hayashi et al. employed supervised machine learning. Their features used in machine learning were similar to ours. Hayashi et al. also performed manual estimation, and these results are included in Table IV.

We compared our results with those of Hayashi et al. for Case 1. Our accuracy rates are higher for Case 1. In this case, the estimation is performed by utilizing only the information in the first two sentences/paragraphs and does not consider information from preceding sentences/paragraphs. Thus, for cases in which the first two sentences/paragraphs have more information, the estimation is more likely to be easy. Paragraph order estimation would be easier than sentence order estimation because paragraphs have more information than sentences.

In the manual estimation of Case 1, the results for a paragraph were higher than those for a sentence. This tendency is the same as in machine learning. We found that for Case 1, estimating paragraph order was easier than estimating sentence order.

Next, we examined the results for Cases 2 and 3. The results from the study by Hayashi et al. were higher than those obtained by the proposed method. In Cases 2 and 3, the information from the sentences/paragraphs preceding the estimated sentences/paragraphs can be utilized so that the relationships with the preceding sentences/paragraphs become important. A story can be concluded in a paragraph; however, it is less likely for a story to be concluded in a sentence. Therefore, a paragraph has fewer hints in the preceding sections than a sentence. It is reasonable to assume that this explains why the results from the study by Hayashi et al. were higher for Cases 2 and 3.

For manual estimation utilizing Cases 2 and 3 parameters, the results for sentences were higher than those for paragraphs. This tendency is the same as in machine learning. We found that for Cases 2 and 3, estimating paragraph order was more difficult than estimating sentence order.

VI. Conclusion

In this study, we proposed a method to estimate the order of paragraphs by employing supervised machine learning. In the experiments on the paragraph order estimation of the first two paragraphs of an article, our proposed method obtained a high accuracy rate of 0.86, obtaining the same accuracy level as manual estimation. From feature analysis, we found that the information on adnominals, conjunctions, and dates was effective for the order estimation of the first two paragraphs. In addition, in the order estimation of all pairs of adjacent paragraphs and all pairs of paragraphs in articles, the proposed method obtained the accuracy rates of 0.63 and 0.67. These accuracy rates were higher than those of baseline methods 1 and 2. We assumed that a paragraph having more nouns in common with the preceding paragraphs is more likely to be the first paragraph of the pair. From feature analysis, we found that adnominals, conjunctions, and dates were effective for the order estimation of the first two paragraphs, and the ratio of new words and the similarity between the preceding paragraphs and an estimated paragraph were effective for the order estimation of all pairs of paragraphs. We compared the order estimation of sentences and paragraphs and clarified differences. For the order estimation of the first two paragraphs, paragraph order estimation would be easier than sentence order estimation because paragraphs have more information than sentences. For the order estimation of all pairs of paragraphs, paragraph order estimation would be more difficult than sentence order estimation because a story may conclude in a paragraph.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 23500178.

References

[1] W. C. Mann and S. A. Thompson, "Rhetorical structure theory: Toward a functional theory of text organization," *Text*, vol. 8, no. 3, pp. 243–281, 1988.

[2] N. Karamanis and H. M. Manurung, "Stochastic text structuring using the principle of continuity," in *Proceedings of the second International Natural Language Generation Conference (INLG' 02)*, 2002, pp. 81–88.

[3] M. Murata and H. Isahara, "Automatic detection of mis-spelled Japanese expressions using a new method for automatic extraction of negative examples based on positive examples," *IEICE Transactions on Information and Systems*, vol. E85–D, no. 9, pp. 1416–1424, 2002.

[4] K. Uchimoto, M. Murata, Q. Ma, S. Sekine, and H. Isahara, "Word order acquisition from corpora," in *COLING '2000*, 2000, pp. 871–877.

[5] Y. Hayashi, M. Murata, L. Fan, and M. Tokuhisa, "Japanese sentence order estimation using supervised machine learning with rich linguistic clues," in *Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013)*, 2013, pp. 1–12.

[6] M. Lapata, "Probablistic text structuring: Experiments with sentence ordering," *Proceedings of the 41st Annual Meeting of the Association of Computational Linguistics*, pp. 542–552, 2003.

[7] J. Carbonell and J. Goldstein, "The use of MMR, diversity-based reranking for reordering documents and producing summaries," in *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1998, pp. 335–336.

[8] K. R. McKeown, J. L. Klavans, V. Hatzivassiloglou, R. Barzilay, and E. Eskin, "Towards multidocument summarization by reformulation: Progress and prospects," in *Proceedings of AAAI/IAAI*, 1999, pp. 453–460.

[9] N. Okazaki, Y. Matsuo, and M. Ishizuka, "Improving chronological sentence ordering by precedence relation," in *Proceedings of the 20th International Conference on Computational Linguistics (COLING 04)*, 2004, pp. 750–756.

[10] D. Bollegala, N. Okazaki, and M. Ishizuka, "A bottom-up approach to sentence ordering for multi-document summarization," *Proceedings of the 44th Annual Meeting of the Association of Computational Linguistics*, pp. 385–392, 2006.

[11] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[12] T. Kudoh, "TinySVM: Support Vector Machines," http://cl.aist-nara.ac.jp/ taku-ku// software/TinySVM/ index.html, 2000.

[13] Y. Matsumoto, A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda, and M. Asahara, "Japanese morphological analysis system ChaSen version 2.0 manual 2nd edition," 1999.