Information Engineering Express
International Institute of Applied Informatics
2023, Vol. 9, No. 2, IEE759

Automatic Extractive Summarization for Japanese Academic Papers by LDA

Hideyuki Sawahata *, Tetsuro Nishino *

Abstract

The demand for the automatic summarization of newspaper headlines and articles is increasing, and various studies on automatic summarization are currently being conducted. However, there are only a few studies on the summarization of Japanese documents compared with English documents.

In this study, we verified the effectiveness of existing summarization methods for academic papers written in Japanese. First, we demonstrate the effectiveness of a topic-based extractive summarization method called Latent Semantic Analysis (LSA). We then show that more effective topic-based extractive summarization can be achieved using Latent Dirichlet Allocation (LDA).

Keywords: Automatic Summarization, Extractive summarization, Natural Language Processing, LDA, LSA

1 Introduction

1.1 Background

The current demand for automatic summary generation can vary widely from the automatic generation of headlines and newspaper summaries to that of academic paper abstracts. In addition, document summarization is required for many other scenarios, such as business books and novels.

Automatic summarization methods that use of both supervised and unsupervised learning have been developed in numerous studies on English summarization [1, 2]. For example, supervised summarization methods use neural networks, such as the encoder-decoder model[3]. A pre-training model for automatic summarization called PEGASUS[4] was recently developed.

Unsupervised Learning for English Summarization of graph- and Topic-based methods. An example of a graph-based summarization method is LexRank[5], which uses degree centrality for automatic summarization. Latent Semantic Analysis (LSA)[6] is a topic-based methods. Various LSA-based methods use either the topic data or singular values obtained from LSA to generate summaries.

* University of Electro-Communications, Tokyo, Japan

As mentioned above, many studies have been conducted on automatic summary generation. However, most of these studies have focused on English summaries, and only a few was conducted using documents written in Japanese. Very few studies have been conducted on academic papers written in Japanese. This is because the number of corpora available for research was small. The small number of available corpora makes it difficult to train and construct state-of-the-art models, and it also makes it difficult to validate and research new methods using such models. Additionaly, in the case of Japanese, there are multiple methods of extracting words from documents, that affects the performance and evaluation results of the models used. This may influence the difficulty of conducting on Japanese people and the small number of studies on Japanese people.

1.2 Purpose

Considering this background, this study aimed to verify whether existing automatic summarization methods are effective for Japanese papers. However, only a few Japanese corpora are available to summarizing the training. Therefore, thies study focus on topic-based extractive summarization methods capable of unsupervised auto-summarization. In particular, we focus on extractive automatic summarization methods based on LSA, which have been suggested to work in multiple languages[7]. Furthermore, we focused on topic models other than LSA and the structure of the documents to be summarized to improve extractive summarization methods using topic models in situations where linguistic resources are limited.

In this study, we first verified whether extractive summarization methods based on LSA are effective in Japanese. In addition, we determined whether Latent Dirichlet Allocation (LDA)[8], which shares the same topic model as LSA but use a probabilistic model to improve the flexibility of topic representation, is more effective than LSA. Finally, we examine whether further improvement can be achieved by limiting the scope of automatic summarization to the Introduction and Conclusion.

2 Related Works

2.1 LSA

LSA was the first technique used for statistical latent semantic analysis. In the LSA, singular value decomposition is used to extract the co-occurrence of words, including latent words. LSA was used for word clustering and to calculate the similarity between documents.

In LSA, words are extracted from documents through morphological analysis to generate a word-document matrix M in which the rows and columns correspond to words and documents. After obtaining the word-document matrix M, singular value decomposition is performed as follows:

$$M = U\Sigma V^T \tag{1}$$

to obtain U, σ, and V^T, which are called the left singular, singular, and right singular value matrices, respectively. These three matrices can then be used in various applications for LSA.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

2.2 LDA

LDA is a statistical latent semantic analysis technique similar to LSA. The LDA is a probabilistic model. In LDA, it is assumed that topics generate words and documents are created as collections of words. Specifically, the documents were generated through the following order:

1. Number of words N is decided by Poisson distribution.
2. The parameter θ of the topic distribution is determined from the Dirichlet distribution $Dir(\alpha)$ with α as a parameter.
3. For each N words w_n
   (a) The latent topic z_n for w_n is determined from the multinomial distribution $Multi(\theta)$
   (b) The word w_n is determined from the multinomial conditional probability $p(w_n|z_n, \beta)$ for the latent topic z_n.

The parameter β in the above generation process is a matrix representing the selection probability of each word for each topic. Each matrix element represents the probability of a word appearing in a certain topic.

The probability distribution was defined based on the above generation process and the parameters are learned based on the probability distribution.

2.3 Extractive summarization method using LSA

2.3.1 LSA in extractive summarization

In LSA-based methods for extractive summarization, the topic is extracted from each sentence in the document to be summarized, and the sentence to be selected as the summary sentence is determined based on the topic. Matrix X is generated for the document summary, in which the matrix elements correspond to the number of occurrences of each word in each sentence. Suppose we have the following three sentences.

s_1 He has a dog.

s_2 He walked with the dog.

s_3 He went to the park.

Only nouns and verbs were extracted from the matrix. Matrix X generated for sentences with words in rows and sentences in columns is shown in Table 1.

Singular value decomposition is then applied to matrix X to decompose it into three matrices, as shown in Equation 1.

Extractive summarization using LSA was performed mainly using Σ and V. Because the rows and columns of V correspond to sentences and topics, respectively, V is often used as a summary and criterion for selecting summary sentences in many methods. To select summary sentences using topics, the co-occurrence of words among sentences in the potential selection can be considered such that the entire document is considered in selecting the summary sentences.

A singular value matrix was used as a criterion to determine the topic to focus on in LSA. Extractive summarization methods based on LSA can be classified based on how the right singular value matrix V and singular value matrix Σ are used.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

Table 1: Example of matrix X

	s_1	s_2	s_3
He	1	1	1
has	1	0	0
dog	1	1	0
walked	0	1	0
went	0	0	1
park	0	0	1

2.3.2 Gong & Liu's method

Gong and Liu's method[9] is an extractive summarization method that uses LSA. In this method, sentences are selected up to a default upper limit to generate a summary using the right singular value matrix obtained from the LSA through the following steps.

1. A word-document matrix in which each sentence is a document, is generated for the target document.
2. The right singular value matrix of the word-document matrix was obtained using the LSA.
3. The sentences with the highest values in each column of each row in the right singular matrix are selected from the left of the row until the default number of sentences is selected.
4. The summary sentences are determined as the selected sentences.

This method assumes that the values in the right singular matrix represent the hood to which a sentence is assigned to each topic. Thefore sentences with the highest value for each topic can be selected to summarize the entire document.

2.3.3 Steinberger & Jezek's method

In Steinberger and Jezek's method[10], right singular and singular matrices are used to select sentences used for the summary. The "length" of a sentence in the document is calculated as

$$length_i = \sqrt{\sum_j V_{ij}^T \Sigma_{jj}} \tag{2}$$

Equation 2 implies that a sentence with a higher value in the right singular value matrix corresponding to the value of the singular value matrix is more likely to be selected. That is, a sentence is more likely to be selected if its topic matches that of the document.

2.3.4 Murray et al.'s method

In Murray et al.'s method[11], sentences for summarization are selected from each topic as in Gong and Liu's method, in which the number of sentences that can be selected for a topic is determined by the ratio of the singular value of the topic to the sum of all singular values. In this method, the focus is on the main topic of a document, and sentences with the topic are selected for the summary.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

To generate a summary for a document consisting of three sentences, two sentences are selected from topic0 for the summary by rounding $3 \times 0.76 = 2.28$.

2.3.5 Cross method

The cross method[12] is similar to Steinberger and Jezek's method. For each topic in the right singular value matrix, the average value was calculated, and the value of the element in the column was less than the average value of the column, which was set to 0. The selection tendency is the same as that of Steinberger and Jezek's method but modified so that it is not affected by poorly assigned topics.

2.3.6 Topic method

Similar to the Cross method, in the Topic method[12], each element of the right singular value matrix is set to zero if it is less than the mean value of that row. The "strength" $strength$ between each topic is then calculated and a topic-topic matrix, which indicates the strength between each topic, is generated. The $strength_{ik}$ between topic i and topic k is calculated as

$$strength_{ik} = \begin{cases} \sum_j V_{ij}^T & (i = k) \\ \sum_j \left(V_{kj}^T + V_{ij}^T \right) & (i \neq k) \end{cases} \tag{3}$$

However, if $V_{kj}^T = 0$ or $V_{ij}^T = 0$, then $V_{kj}^T + V_{ij}^T = 0$.

Equation 3 represents the policy for selecting sentences from topics likely to co-occur with other topics in each sentence in the document.

After calculating the strength between each topic, the sum of each row and the sentences is selected in the same way as in Gong and Liu's method, that is, in descending order of the topic with the highest value.

2.4 ROUGE

ROUGE[13] was used to evaluate the match between the words in the reference and the generated summaries. Recall and precision were calculated, and the harmonic mean was used to evaluate the generated summary. The $score_{f1}$ for the generated summary is given by

$$score_{f1} = \frac{2(score_{recall} \times score_{precision})}{score_{recall} + score_{precision}} \tag{4}$$

where $score_{recall}$ is the recall and $score_{precision}$ is the precision.

There are several variants of ROUGE, depending on the method used to calculate the precision and recall. We introduce ROUGE-L. In ROUGE-L, recall and precision are based on the "Longest Common Subsequence" (LCS), defined as the sequence of words that are completely consistent with one another. The LCS is unaffected by the difference in the words as long as the sequence matches.

The recall and precision in ROUGE-L are defined as

$$score_{recall} = \frac{LCS(Ref_{words}, Sys_{words})}{|Ref_{words}|} \tag{5}$$

$$score_{precision} = \frac{LCS(Ref_{words}, Sys_{words})}{|Sys_{words}|} \tag{6}$$

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

where *LCS* is the number of words in the LCS, and Ref_{words} and Sys_{words} indicate the number of words in the reference and generated summaries, respectively.

3 Experiments

3.1 Experimental methods

In this study, we first investigated the effectiveness of extractive summarization methods based on LSA for Japanese sentences, and then investigated the effectiveness of using LDA as a sentence feature. In addition, we investigated the effects of limiting the input for extractive summarization considering the document structure. Specifically, we investigated the effects of limiting the input of sentences to be included in the summary of the "introduction" and "conclusion" sections of academic papers used as extractive summarization targets. Three experiments were conducted.

Experiment 1 Evaluation of the applicability of the extractive summarization method using LSA in Japanese.

Experiment 2 Evaluation of the effectiveness of LDA for extractive summarization for Japanese.

Experiment 3 Evaluation of limiting the input to the "introduction" and "conclusion" sections in the extractive summarization of academic papers.

For each experiment, we used the following five extractive summarization methods:

- Gong & Liu's method
- Steinberger & Jezek's method
- Murray et al's method
- Cross method
- Topic method

We used five experimentals methods because it is difficult to determine whether a parametric method is effective when only one method is used. When the above extractive summarization method is applied using LDA, the topic distribution θ is computed as Σ, and the probability of belonging to a topic for each word β is computed as V.

ROUGE-L was used as the evaluation method for verification experiments. The summaries generated by each extractive summarization method were evaluated by comparing them with the manually generated summaries in each experiment.

Throughout the experiments, we set a limit of 10% of the input sentences on the number of sentences to be generated in the summaries. Additionaly, stop words were used to remove these words.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

Table 2: Score in English and Score in Japanese

	English	Japanese
Gong & Liu's method	0.180	0.181
Steinberger & Jezek's method	0.138	0.193
Murray et al.'s method	0.180	0.184
Cross method	0.182	0.200
Topic method	0.180	0.190

3.2 Experiment 1

3.2.1 Overview

Extractive summarization using the LSA is effective for English and other languages such as Turkish. Therefore, we investigated whether acheving the same performance for individual Japanese documents was possible by comparing the results of English and Japanese summarizations. For the English results, we quote the evaluation results for the Summac dataset in [7]. Summac is a collection of summaries of computer science article pablished at ACL-sponsored conferences. We used the Corpus of the Journal of the Association for Natural Language Processing [14] to evaluate the automatic summarization in Japanese. In this experiment, the main text was set as the target sentence for the summary, and the following items in the text were deleted:

- Equation, figures, and tables in the text
- Ornamental descriptions such as "bf"
- Except for the documents enclosed by "begin" and "end"

The abstract of the paper was used as the reference summary.

3.2.2 Results

Table 2 shows the results of the extractive summarization obtained using the LSA for English and Japanese. Owing to the differences in the target corpora, Japanese and English scores for each method. However, the scores for the two languages were relatively similar, except for Steinberger and Jezek's method, which exhibited large difference.

3.2.3 Observation

In Experiment 1, we checked whether LSA extractive summarization performed as well in Japanese as in English. In this experiment, stop words were used for word removal. This removal, which was based on the frequency of occurrence and parts of speech, may be one of the reasons why the system worked as well for Japanese as it did for English. The results in Table 2 confirm that the scores of several methods are relatively similar. The slight difference in scores was probably due to differences in the target corpora.

The performances for Japanese and English were almost the same because LSA extracts topics based only on the co-occurrence of words without considering word order, for which there are major differences between English and Japanese. Therefore, methods such as LSA, which does not consider word order, can be used regardless of the language. Stop words were removed for both English and Japanese before summaries were generated.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

Table 3: Comparison of LSA and LDA

	LSA	LDA
Gong & Liu's method	0.181	0.197
Steinberger & Jezek's method	0.193	0.208
Murray et al.'s method	0.184	0.202
Cross method	0.200	0.190
Topic method	0.190	0.199

There are some similarities in the stop words between the two languages. For example, particles are frequency used as stop occurring words. Other frequently occurring nouns were also included as stop words. The policies for creating stop words were similar in both languages. This may be the reason for the similar LSA extractive summarization results in both English and Japanese.

3.3 Experiment 2

3.3.1 Overview

After confirming that extractive summarization using LSA applies to the Japanese people, we examined whether LDA is more effective than LSA for extractive summarization. The corpus, extractive summarization methods used for comparison, and the limitations of the generated summaries remained unchanged from those of Experiment 1, except that LSA was replaced with LDA in the extractive summarization methods.

3.3.2 Results

Table 3 presents experimental results. The score was increased by replacing the topic extraction method with LDA in the various extractive summarization methods except for the Cross method.

3.3.3 Observation

Experiment 2 examined whether the LDA was effective for extractive summarization. The LSA score was higher than that of LDA only for the cross method, whereas that of LDA were higher for the other methods. This differences can be attributed to the algorithm used in the Cross method. As described in Section 2.3.5, in each column of the topic-document matrix, elements with values below the average of their corresponding rows were set to 0 before the length calculation. Additionaly, the sum of each column was calculated. However, the sum of the topic distributions is 1 in LDA. Therefore, setting the matrix elements smaller than the mean to zero does not have a significant effect, possibly because topic distribution is not considered in the calculation. Conversely, the higher score achieved by LDA compared to other methods is possibly because the obtained values are probability values. By contrast, negative values may appear in the LSA. The influence of negative values may have caused some important sentences to be considered unimportant.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

Table 4: Results when input was limited to introduction and conclusion only.

	LSA	LDA
Gong & Liu's method	0.233	0.240
Steinberger & Jezek's method	0.226	0.256
Muray et al.'s method	0.228	0.242
Cross method	0.245	0.258
Topic method	0.234	0.233

3.4 Experiment 3

3.4.1 Overview

As the "introduction" and "conclusion" in academic papers are expected to describe the research summary and prospect, we hypothesized that it would be effective to perform extractive summarization by focusing on these two sections. Therefore, in this experiment, we investigated the effectiveness of extractive summarization for academic papers by limiting the input to only the sentences included in the "Introduction" and "Conclusion" sections, based on the structure of the paper. In addition, we investigated whether there were any difference between the results obtained using LSA and those obtained using LDA when the above mentioned input was used. The corpus and evaluation methods used in this experiment are the same as those used in Experiments 1 and 2. During the summarization of each article, only the sentences in the "Introduction" and "Conclusion" were extracted as inputs and combined. The mathematical expressions and figures included in the sections have been removed.

3.4.2 Result

Table 4 presents the results of the study. Compared with Table 3, it can be seen that the scores for both the LSA and LDA increased. In addition, when the focus of the summary target was limited, there was no significant difference between the LSA and LDA for the Topic method; however, the LDA scores were higher for the other methods.

3.4.3 Observation

In Experiment 3, we investigated the effectiveness of limiting the summary to the "introduction" and "conclusion" of the study. The results showed that the scores for both LSA and LDA were higher than those in Experiment 2. In addition, the LDA scores were higher than or equal to those of both LSA and LDA in Experiment 2. The higher score may be attributed to the input of only the "introduction" and "conclusion" sections, which are expected to be easily included in the summary. In an academic paper, the "introduction" usually contains the overall outline of the research, such as a summary of the research, objectives, and proposed methodology, and the "conclusion" contains a summary of the research results, such as a summary of the research and experimental results. The results of Experiment 3 show that the summary in extractive summarization can be effectively generated from parts that are likely to be included in the summary.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

4 Discussion

The following conclusion can be drawn from the above experiments.

- Topic-based extractive summarization using LSA works in both Japanese and English.
- LDA is more effective than LSA for topic-based automatic summarization.
- Using the sentences in the "Introduction" and "Conclusion" sections is effective for extracting abstracts of academic papers written in Japanese.

The performance of extractive summarization by LSA in Japanese was verified and confirmed in Section 3.2 of this study. Furthermore, in Section 3.3, we confirmed that the summaries were improved by using LDA. Therefore, topic-based extractive summarization methods are assumed to be applicable to various languages. In addition, as mentioned in Section 2, LSA and LDA are unsupervised techniques that do not require correct answer data for training. This can be advantageous in summarizing documents using a small language corpus. The higher effectiveness of LDA compared with LSA may be attributed to LDA being a probabilistic topic model. In LSA, the values obtained from singular value decomposition include both positive and negative ones.

In contrast, in the LDA, the probability values obtained were greater than 0 and less than 1. Therefore, in methods that involve calculcating sums and products, values close to zero are less likely to affect the calculations, making it easier to select important sentences. However, because only values between 0 and 1 are obtained, methods in which some values are set to 0 based on the average, such as the Cross method, may not be effective because of these small values. In other words, although LDA is generally effective for extractive summarization, it may not be effective in certain cases. In addition, from the ROUGE scores of the experiments, it can be determined which methods are more effective for LSA and which are more effective for LDA. The ROUGE scores show that the cross method is effective for LSA, and that Steinberger and Jezek's method is effective for LDA. However, when we look at the LDA scores of various methods, Steinberger and Jezek's method is not outstandingly effective. In other words, while the cross method is effective in LSA, all methods appear to work equally well for LDA.

Finally, we discuss the effects of limiting the input by considering the document's structure in Experiment 3. For Experiment 3, we hypothesized that the "introduction" and "conclusion" sections of the main body of an academic paper would be included in its summary and generated a summary from these sections. The higher resultant ROUGE score compared with that of the summary generated from the entire body of the paper confirms the effectiveness of our hypothesis. This higher score may be attributed to the fact that the introduction and conclusion are likely to be included in the summary because they contain important content, such as an overview of the entire study. However, it is highly likely that other important information, such as that on the proposed methodology and specific numerical values obtained from the experiments, will not be included in the generated summary. Therefore, an improvement can be achieved by performing an extractive summarization of this paper's introduction, conclusion, and other important sections.

In Experiment 3, improved extractive summarization results were achieved by considering the document structure. This suggests that the hypothesis for the experiment may also apply to documents other than academic papers, such as essays, which have structure relatively similar to academic papers, in which the document begins with an introduction and

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

ends with a conclusion. However, as this hypothesis is based on the structure of academic papers, it is unlikely to appliy to novels and other documents with dissimilar structures. However, the hypothesis may still apply to documents other than academic papers if a summary section can be inferred.

5 Conclusions

In this study, we focus on unsupervised extractive auto-summarization, which can work with a small corpus to summarize Japanese papers and meet the demand for auto-summarization. Because most automatic summarization methods are applied to English documents, it is unclear whether they perform equally well in Japanese. Therefore, we focused on extractive auto summarization methods based on the LSA, which can work in multiple languages, and verified whether they work in Japanese. In addition, we examined whether LDA, a statistical latent semantic analysis method similar to LSA, is more effective than LSA. Furthermore, we hypothesized that the summary of a paper always includes the contents of the introduction and conclusion sections, and investigated the effectiveness of this hypothesis.

Through several experiments, we found the following.

- Extractive automatic summarization methods based on LSA work well for Japanese.
- LDA is more effective than LSA for topic-based extractive summarization.
- Generating a summary from the introduction and conclusion sections is effective for the extractive summarization of academic papers.

Thus be features obtained using unsupervised learning techniques that do not consider the word order, such as LSA and LDA, are effective in many languages. Furthermore, because the probability values are obtained as features in LDA, unimportant sentences are ignored. However, LDA was not effective in the cross method. It is hypothesized that small differences between the average and probability values may lead to methods that use the averages of features, such as the cross method, being unable to exclude unimportant sentences and, conversely, to exclude important sentences. Therefore, although the LDA is generally effective, it is necessary to consider using LSA based on the specific method employed. In addition, we confirmed that generating summaries from the Introduction and Conclusion sections was effective for the extractive summarization of academic papers. This suggests that the approach may also be effective for documents with structures similar to academic papers. In future studies, verifying whether the automatic summarization of Japanese papers works with other unsupervised methods will be necessary. In addition, because this study was conducted on academic papers written in Japanese, it is necessary to verify the effectiveness of these methods in papers written in other languages.

References

[1] Radityo Eko Prasojo, Mouna Kacimi, and Werner Nutt. Modeling and summarizing news events using semantic triples. In *European Semantic Web Conference*, pages 512–527. Springer, 2018.

[2] Kai Hong, John M Conroy, Benoit Favre, Alex Kulesza, Hui Lin, and Ani Nenkova. A repository of state of the art and competitive baseline summaries for generic news summarization. In *LREC*, pages 1608–1616. Citeseer, 2014.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

[3] Parag Jain, Anirban Laha, Karthik Sankaranarayanan, Preksha Nema, Mitesh M. Khapra, and Shreyas Shetty. A mixed hierarchical attention based encoder-decoder approach for standard table summarization. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 622–627, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[4] Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 11328–11339. PMLR, 13–18 Jul 2020.

[5] Günes Erkan and Dragomir R Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, 22:457–479, 2004.

[6] S. DEERWESTER. Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci.*, 41(6):391–407, 1990.

[7] Ilyas Cicekli Makbule Gulcin Ozsoy, Ferda Nur Alpaslan. Text summarization using latent semantic analysis. *Journal of Information Science*, 2011.

[8] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[9] Yihong Gong and Xin Liu. Generic text summarization using relevance measure and latent semantic analysis. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 19–25, 2001.

[10] Josef Steinberger and Karel Jezek. Using latent semantic analysis in text summarization and summary evaluation. 01 2004.

[11] Gabriel Murray, Steve Renals, and Jean Carletta. Extractive summarization of meeting recordings. In *in Proceedings of the 9th European Conference on Speech Communication and Technology*, pages 593–596, 2005.

[12] Makbule Ozsoy, Ilyas Cicekli, and Ferda Alpaslan. Text summarization of turkish texts using latent semantic analysis. In *Proceedings of the 23rd international conference on computational linguistics (Coling 2010)*, pages 869–876, 2010.

[13] C.-Y. LIN. Rouge : A package for automatic evaluation of summaries. *Proc. Workshop on Text Summarization Branches Out, Post Conference Workshop of ACL 2004*, 2004.

[14] The Asso for Natural Language Processing. Latex corpus of the transcations of the association for natural language processing. `https://www.anlp.jp/resource/journal_latex/index.html`, 2020.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.