The Development and Evaluation of Automatic Grading System of Japanese Summary

Na Niu & Hongyan Wang
School of Foreign Languages, Shenzhen University, Shenzhen, China

ABSTRACT: Summary writing is an effective foreign language teaching method. Though, it is difficult to achieve on the teaching process, since instructors do not have sufficient time for individual detailed evaluation and feedback. Following with the Japanese macro rating standards, the micro rating and application research develops slowly in the limited way of subjective analysis to evaluate the aspects of the summary. Little is focused on high operational, comprehensive, accurate, and objective scoring method. In this paper, based on the macro scoring theory of Japanese summary, and a Japanese automatic scoring method from the micro perspective, we develop an objective quantitative automatic scoring system for Japanese summary. The system is evaluated and validated by comparing with the scores of Japanese instructors. The results show that the system has certain reliability and accuracy.

Keywords: Japanese summary, automatic grading, natural language processing, singular value decomposition

1 INTRODUCTION

Summary writing is a common form in foreign language writing. It is necessary for readers to generalize and condense the main content of articles with concise and refined language without changing the central idea, genre and structure of the original text. Compared with the other foreign language writing, it can train and test the author on the higher level of comprehension, judgment, generalization and expression ability efficiently. Therefore, in Japanese teaching, summary writing is an effective means of teaching and testing, which is an important way to improve the professional knowledge and skills. Summary writing is very difficult to achieve in the teaching process, because of the large number of students, and teachers do not have enough time for each student to provide detailed evaluation and feedback, which is to improve the quality of summary writing.

David J. Steinhart, an American scholar, developed an English summary automatic scoring system, Summary Street, to improve learners’ writing ability in 2001. For the input summary system can score and provide the feedback about content, coherence, and number of words. At present, there are macro evaluation criteria for Japanese summary score, but micro research and application research are developing slowly, only through subjective analysis to judge all aspects of the summary, simply good or bad. If there is a highly operational, reliable, and accurate automatic scoring system, which can provide reference and support for Japanese teachers, the feedback bottleneck problem will be solved. Therefore, this paper based on latent semantic analysis, from Natural Language Processing’s perspective, establishes a highly reliable, comprehensive and accurate mathematical method, developed an automatic measurement system of Japanese summary, and evaluated the accuracy and reliability of this system.

2 LATENT SEMANTIC ANALYSIS

Latent Semantic Analysis (LSA) is a theory and method for extracting and representing the test (Landauer & Dumais, 1997). LSA is fully automatic mathematical technique for extracting and inferring relations of expected contextual usage of words in passages of discourse. It is not a traditional natural language processing or artificial intelligence program; it uses no humanly constructed dictionaries, knowledge bases, semantic networks, grammars, syntactic parsers, or morphologies, or the like and takes as its input only raw text parsed into words defined as
unique character strings and separated into meaningful passages or samples, such as sentences or paragraphs. The adequacy of LSA’s reflection of human knowledge has been established in a variety of ways. Human beings can acquire information knowledge as much as possible by learning the context of vocabulary in the text. Also, through the analysis of large number of representative texts using latent semantic analysis model, get on all the vocabulary knowledge, can be the equivalent of a high school level of knowledge.

The first step of LSA is to represent the text as a matrix in which each row stands for a unique word. Each column stands for a text passage or context. Next, LSA applies Singular Value Decomposition (SVD) to the matrix. In SVD, a rectangular matrix is decomposed into the product of three other matrices. By choosing the appropriate dimension reduction level, three new matrices are obtained. After the product of three new matrices, a new matrix is established. Finally, by calculating the information of each vector of the new matrix, the important sentences and words of the article are obtained.

LSA can measure the potential similarity between texts, predict the coherence of text and understand the content. However, LSA cannot judge the grammar and word order. The article scoring limitation based on LSA is that the word order and grammar cannot be evaluated.

3 DEVELOPING AUTOMATIC SCORING SYSTEM BASED ON LSA

3.1 Theoretical basis and characteristics of the system

Japanese scholar Sakuma (1997) offered scoring method and standard of Japanese summary. That is the content, coherence, structure, grammar, sentences, words. Content, as the most important evaluate standard, determines the summary’s quality and score. The limitation of the current system evaluates the completeness of the content only, for the most of the summary. The measurement method of coherence is being validated by data, which is implemented in the system after optimization. By the latent semantic analysis method, this paper explores the content evaluation standard mathematically and its effect of its reliability and accuracy.

3.2 Development language and environment of system

The system is developed using C# and running on Windows98. Use ChaSen\(^1\) for the Japanese morphological analysis, and use MATLAB (R2009a) for numerical analysis, singular value decomposition, and information computing. At present, the system can score two articles about content and words number. Article 1 is Explore the Dinosaur, and article 2 is Will Human Being Perish?

3.3 The development process of system

As shown in Figure 1, develop the system SAES: Summary Auto-evaluate System. Firstly, construct a text corpus about the summary topic, and Japanese morphological analysis by ChaSen. Secondly, Extract the corresponding word-by-document matrix, and then analysis the matrix using LSA. Finally, the key sentences and keywords of each article are calculated, and the score of the summary is given. In this paper, article 1: Explore the Dinosaur as an example to explain every step.

![Diagram](http://example.com/figure1.png)

Figure 1. The Japanese summary grading process.

3.3.1 Establishing the topic matrix

Collect corpus from textbooks, books, encyclopedic dictionaries, newspapers and so on. Article 1 collected articles about dinosaurs from the network Encyclopedic Dictionary. The corpus contained 13022 words, and 208 sentences. Article 2 collected articles from three books: Human or Die? Environment and Hu-

---

\(^1\)ChaSen is Japanese morphological analysis software developed by Natural Language Processing Institute of Nara Institute of advanced science and technology. Currently, for free [http://chasen-legacy.sourceforge.jp/](http://chasen-legacy.sourceforge.jp/).

\(^2\)Morpheme is the smallest linguistic unit in Japanese, and the word in Japanese is composed of one or more morphological elements. Morpho-logical parsing is one of the basic techniques of Natural Language Processing. It uses computer to segment Japanese sentences into word sequences and distinguishes the word.
man, and Human and Environment. The corpus contained 10,173 words, and 223 sentences.

3.3.2 Extracting words expressing content
Before building a matrix, it is necessary to extract words that represent the content of the article. In this paper, use ChaSen to Japanese morphological analysis. In the morphological analysis of words, extract the general nouns, conjunctive nouns, proper nouns, adverb, conjunction and unknown nouns. For example, proper nouns (ex: 瀬戸内海) represent the content, adverb (ex: 今後) represents the time sequence, conjunction (ex: そして) represents article structure, unknown nouns (ex: イグアンドン) represents the article’s content. Table 1 is the result of morphological analysis of the first sentence (私たちは図鑑や博物館で恐竜を見ることができます.) in article 1.

### Table 1. The result of morphological analysis.

| Word    | Property     |
|---------|--------------|
| 私      | 名詞-一般     |
| たち    | 名詞-接尾-一般 |
| は      | 助動詞-係助詞 |
| 図鑑    | 名詞-一般     |
| や      | 助動詞-係助詞 |
| 博物館  | 名詞-一般     |
| で      | 助動詞-助詞   |
| 恐竜    | 名詞-一般     |
| を      | 助動詞-格助詞-一般 |
| 見る    | 動詞-自立 一段 |
| こと    | 名詞-一般     |
| が      | 助動詞-一般   |
| でき    | 動詞・運用形  |
| ます    | 助動詞・ます基本形 |
| …      | 記号・句点    |

3.3.3 Building matrix of word-sentence
The original matrix X established about article 1 is shown in Table 2. The understanding of the matrix is as follows: each line represents a word, and each column represents a sentence vector. Each unit represents the frequency of the lexical occurrence in the sentence. The word “dinosaur” appeared once in the first sentence of original article and appeared once in the first sentence of original article and twice in the third sentence of summary. The matrix is 565 × 239.

### Table 2. Matrix X for corpus, original article and summary.

|          | (corpus) 1…208 | (original article) 01…028 | (summary) S1 S2 S3 |
|----------|----------------|---------------------------|-------------------|
| 恐竜     | 1 0            | 1 1                        | 1 1 2             |
| 中生代    | 1 0            | 0 0                        | 0 0 0             |
| 恐竜     | 1 0            | 0 0                        | 0 0 0             |
| 見る      | 0 0            | 0 0                        | 0 0 0             |
| こと      | 0 0            | 0 0                        | 0 0 1             |

3.3.4 Reconstruct the original matrix
To eliminate the low frequency values in the matrix and highlight the characteristics of each vector, the singular value decomposition of the original matrix X is needed. The singular value decomposition method is shown in Formula (1). In SVD, the matrix X(mxn) is decomposed into the product of three other matrices. One component matrix describes the original row entities as vectors of derived orthogonal factor values, another describes the original column row entities in the same way, and the third is a diagonal matrix containing scaling values such that when the three components are matrix-multiplied, the original matrix is reconstructed.

\[
X = U S V^T
\]

### Table 3. Reconstructed matrix (X').

|          | (corpus) 1…208 | (original article) 01…028 | (summary) S1 S2 S3 |
|----------|----------------|---------------------------|-------------------|
| 恐竜     | 1.02…0.06     | 0.92…1.01                 | 1.00…2.00         |
| 中生代    | 0.17…0.24     | 0.03…0.03                 | 0.03…0.03         |
| 新たな    | 0.02…0.02     | 0.01…0.07                 | 0.02…0.18         |
| たび      | 0.02…0.02     | 0.01…1.07                 | 0.02…0.14         |

3.3.5 The correlation coefficient between summary and important content of the original article
Through the analysis of each vector of the reconstructed matrix, the average information can be obtained, and the average information determines the importance of the sentence in the article. The greater the average information of the sentence, the more important it is in the article, that is, the key sentence of the article. In the reconstructed matrix (Table 4), the vertical is the eigenvector of each sentence, and the method of calculating the information is shown in Formula (2). H represents information; W represents the feature vector of sentences. The 6 key sentences are extracted using Formula (2) from large to small: O27, O28, O14, O4, O16, O9.
The correlation coefficient between the summary and the important content of the original text is shown in Table 4. The average value of the red value is the score of the summary.

Table 4. The correlation coefficient between the summary and the important content of the original text.

3.4 System introduction

Based on the above process, the automatic scoring system SAES is developed. The system can grade two articles, and choose one before scoring the summary. The system running is shown in Figure 2. The system outputs the evaluation results of Article 1, as shown in Figure 3.

4 SYSTEM EVALUATION

4.1 Evaluation method

Through the summary writing survey, the summary was collected, and the reliability and accuracy of the system were verified by comparing with the scores of Japanese instructor.

4.2 Summary survey

Article 1 of summary survey is carried on for 57 Japanese college students and article 2 is carried on for 52 Japanese college students. The summary for analysis both is 52. At the same time, 5 Japanese teachers (Osaka Japanese Language College) scored 52 summaries with five stage scoring methods (5 good, 4 slightly better, 3 general, 2 slightly worse, 1 poor).

4.3 Assessment results

Article 1 of the correlation coefficient between Japanese teacher ratings and the system score is 0.73, as shown in Figure 4. Article 2 is 0.67, as shown in Figure 5.

The variance analysis of Article 1 is shown in Table 5, and the result of F test is: $F = 57.15 > 4.03 = F_{0.05}$; the variance analysis of Article 2 is shown in Table 6, and the result of F test is: $F = 40.59 > 4.03 = F_{0.05}$, F is statistically significant.

Table 5. Variance analysis of article 1.

|                  | Degrees of freedom | Sum of squares | Mean square | F       | P     |
|------------------|--------------------|----------------|-------------|---------|-------|
| Treatment        | 1                  | 6.037          | 6.037       | 57.149  | 8.02E-10 |
| Error            | 50                 | 5.281          | 0.105       |         |       |
| Total            | 52                 | 11.319         |             |         |       |
Table 6. Variance analysis of article 2.

| Degrees of freedom | Sum of squares | Mean square | F     | P       |
|--------------------|----------------|-------------|-------|---------|
| Treatment          | 1              | 3.610       | 3.610 | 40.591  | 5.78E-08 |
| Error              | 50             | 4.447       | 0.088 |         |          |
| Total              | 52             | 8.057       |       |         |          |

5 CONCLUSION

This paper uses latent semantic analysis to discuss the objective scoring method of Japanese summary, attempts to develop an automatic scoring system for Japanese summary writing, and verifies the reliability and accuracy of this system by comparing with the scores of Japanese teachers. The following conclusions can be drawn.

Firstly, the correlation coefficient between the scores of the system and the Japanese teachers was 0.73 of the articles 1 and 0.67 of the article 2 which had a strong correlation. Therefore, the system score is feasible, the scoring results is accuracy, can change part of the subjective scoring method.

Secondly, latent semantic analysis for Japanese Natural Language Processing has particularity, and needs to combine the characteristics of the Japanese language and other Japanese language analysis tools, to improve the accuracy of the results.

Thirdly, break the technical difficulties in the Natural Language Processing process: establish an appropriate corpus, selection of singular values (S) in matrix reconstruction, the importance of words and weight calculation method, the similarity of texts.

Fourthly, at present, the system cannot evaluate all aspects of the summary, and the structure and grammar scores need to other ways to achieve. The scoring system is limited to two types of articles. In the future, the number and type of articles should be increased, and the experimental evaluation should be reevaluated. After the completion of the above topics, an automatic system can be achieved, which can be applied to Japanese teaching, to evaluate the summaries comprehensively and provide feedback.

ACKNOWLEDGEMENT

This paper is the result of the project: Automatic Measurement Method of Japanese Textual Coherence (Project number: GK201606) of Foreign Languages, Shenzhen University.

REFERENCES

[1] Foltz, P. W., Kintsch, W. & Landauer, T. K. 1998. The measurement of textual coherence with Latent Semantic Analysis. Discourse Processes, 25(2&3): 285-307.
[2] Giora, R. 1985. What’s a coherent text. in Sozer, E. (ed.) Text Connexity, Text Coherence. Hamburg: Helmut Buske Verlag.
[3] Halliday, M.A.K. & Hasan, R. 1976. Cohesion in English. London: Longman.
[4] Landauer, T.K., Foltz, P.W., & Lahman, D. 1998. An introduction to Latent Semantic Analysis, Discourse Processes, 25: 259-284.
[5] Tsunenori Ishioka 2004. Latest Trends in Automated Essay-Scoring Systems. Action metrology 31 volume No 2, 67-87.
[6] Muneko Kimura 1982. Japanese language teaching methods Bonjinsha.
[7] Nagao Satoshi 1979. The generality of article theory for Japanese Language Education. Educational Publishing.
[8] Manabu Okumura 2003. Automatic Summarization. Kyoritsu Publishing.
[9] Sakuma Mayumi 1992. Article and Sentence—The Paragraph’s context—Japanese Language.11-4.
[10] Sakuma Mayumi 1997. The expression type and Scoring method of summary—A comparison of foreign students and Japanese College Students—Heisei6~Heisei8 Scientific Research outlay Research Achievements Report Hitsuji Book.
[11] Tsubakimoto, Nakamura, Kishi, Akabori 2004. The applicable possibility of reading guidance using Latent Semantic Analysis. Educational Psychology Society of Japan 4 times collected papers 6, 549.
[12] Yoshimoto Kiyohisha 2005.Summary ability of grinding and expository text’s guidance. MeiJi book.
[13] Maowenwei 2009. A summary of the current situation of the construction of Japanese Corpus. Japanese learning and research6, 42-4.
[14] Shijianjun, Xuyiping 2003. Corpus and Japanese research. Japanese learning and research 4, 7-11.