Vol.2019-CE-152 No.18
Vol.2019-CLE-29 No.18
2019/11/17

Class questionnaire analysis system using distributed representation based on Wikipedia Entity Vectors

TAKANORI NEGISHI,[1] YOSHIHIRO ADACHI[1]

Abstract: One of the challenges of conducting collaborative and interactive classes using ICT is developing a highly functional class support system. In this study, we report a system that can classify and search students' opinions (questionnaire responses) in class in real time. The system obtains distributed representations of questionnaire responses on the basis of Wikipedia Entity Vectors and implements a search function, which searches for response sentences containing similar content, and a hierarchical/ non-hierarchical clustering function. This system is educational support software that can be very useful for both teachers and students because it can classify and display all students' opinions and status of understanding in real time from the written responses of student questionnaires.

Keywords: educational support system, questionnaire analysis, distributed representation, Wikipedia Entity Vectors

1. Introduction

The Ministry of Education, Culture, Sports, Science and Technology (MEXT) has promoted the use of ICT in education. MEXT's policy promotes collaborative and interactive teaching innovation using ICT, as discussed in the "Second Basic Education Promotion Plan" [1]. Research on using ICT for education is increasing, supported by this policy. Some examples of such research include a system designed to automatically score descriptive answer sentences and scoring support for report assignments [2][3]. Other examples include research aimed at supporting class progress and student learning during classes [4][5].

Another development in ICT and education is that, in recent years, university classes increasingly use Web services to confirm attendance and accept assignment submissions. These Web services often include a class evaluation questionnaire function and a function for students to answer questions from teachers in class in real time. There are also systems with a function that displays and summarizes students' responses in classes, their opinions, and their impressions in real time [10][11]. These systems have functions such as displaying the answer sentences given by students who answer teachers' questions and counting and displaying the number of students who agree with another student's opinion. However, to the best of our knowledge, there is no system that has a function to classify and retrieve questionnaire response sentences given in a free-description format in real time.

In previous research on class questionnaires, there has been analysis using text mining [12] and analysis using topic models [13]. However, these studies focused on class evaluation questionnaires and did not take questionnaires during class and analyze them in real time.

Research on machine learning and deep learning is also being actively conducted. In the field of natural language processing, many studies on document classification and retrieval based on distributed representations of words using word2vec, proposed by Tomas Mikolov et al. [6] have been reported [7][8]. In the past, such research required the preparation of large amounts of distributed-representation-of-words data. Recently, Wikipedia Entity Vectors [9], which collects distributed representations of words appearing in all of Japanese Wikipedia, has been released, significantly reducing the time required to develop a system using distributed representations of words.

The purpose of this study is to develop a lecture support system that can analyze and classify questionnaire responses in a free-text format collected from students during class in real time. This system analyzes questionnaire response sentences that give opinions, evaluations, and impressions about the class in real time, and visualizes the results. By using this system, teachers can immediately grasp a student's degree of understanding and concentration, opinions, and so forth while proceeding with the lecture. Students can understand how their opinions are positioned relative to the class and can participate in the lectures with enthusiasm by having their opinions appropriately reflected in the class content.

2. Questionnaire analysis system

The outline of the processing of the questionnaire analysis system we developed is described below. Python was used for system implementation.

Step 1: Divide questionnaire responses into word strings with the morphological analysis tool MeCab.
Step 2: Perform data cleansing (removing stop words, converting to basic forms, eliminating duplicates, etc.).
Step 3: Extract the distributed representation of each word from Wikipedia Entity Vectors and calculate said distributed representation's average as the distributed representation of the questionnaire response sentence.
Step 4: Cluster questionnaire response sentences based on distributed expressions. The k-means++ method is used for non-hierarchical clustering; Ward's method and the group average method are used for hierarchical clustering.
Step 5: Obtain a distributed representation of the teacher's question sentence by executing steps 1 to 3. Search questionnaire answer sentences with high cosine similarities to the teacher's question and display them.

This system reads Wikipedia Entity Vector data with

1 Toyo University, Kawagoe, Saitama 350-8585, Japan

KeyedVectors module of the Gensim.models package at startup. Wikipedia Entity Vectors data can be selected at 100 dimensions, 200 dimensions, and 300 dimensions. The feature information of each word increases in proportion to the number of dimensions, but the reading time also increases. The data reading time in our execution environment (CPU: Intel Core i7-6567U, storage: SSD) is about 80 seconds for 100 dimensions, about 160 seconds for 200 dimensions, and about 220 seconds for 300 dimensions. In this study, 100-dimensional data is read only once at startup. Next, the user (teacher) selects an Excel file (xls, xlsx) or a text file of questionnaire response sentences to be analyzed, and steps 1 to 3 are executed. In step 3, the distributed representation of each student's response sentence is calculated by taking the average of the distributed representation vectors acquired from Wikipedia Entity Vectors of the words constituting the sentence. A function for excluding words that are not desired to be included in the distributed representation of the sentence is also implemented at this time. The calculation time for the distributed representations of the questionnaire response sentences differs depending on the data size, but the calculation time for an xls file of about 50 KB (40 to 100 characters in one sentence, 85 sentences in total) was 0.23 seconds.

After that, the user can select Step 4 or Step 5 and execute it repeatedly.

2.1 Clustering questionnaire sentences

This is a function for clustering questionnaire sentences. The k-means++ method implemented with scikit-learn as a non-hierarchical method, and Ward's method and the group average method implemented with scipy as hierarchical methods, are used. As a distance measure, Euclidean distance is used in Ward's method, and cosine similarity is used in the group average method. Words close to the centroid of each cluster are obtained as labels for said cluster on the basis of cosine similarity and displayed in descending order of similarity. At this point in time, an inappropriate word may be included in the labels. For this reason, a function has been implemented in which a user (teacher) registers label candidates in advance and assigns labels to classes from among these candidates.

2.2 Retrieving questionnaire sentences

This is a function that retrieves questionnaire responses that have the same meaning as the entered query sentence. The distributed representation of the input query sentence is acquired, and the cosine similarity between it and the questionnaire response sentences are calculated. The top N questionnaire sentences are displayed in descending order of their similarity. Only sentences that exceed the threshold set by the user are displayed. By default, the threshold is set to 0.7. The upper limit number N of sentences to be displayed can be freely set by the user; it is set to 10 by default.

3. Experiments for each function

3.1 Data sets

Using the following three types of data set, we conducted questionnaire-response clustering and similar-response retrieval experiments.

Type A: This is a set of sentences that are mainly written by the authors and are judged subjectively to be relatively easy to classify. Its content includes sentences related to keywords such as "大学," "服装," and "食事." Each sentence is about 20-50 characters long, and there are 37 sentences in total. Examples of these sentences include "私は東洋大学の学生です" and "彼は埼玉大学の大学院生です," both of which are related to "大学." Statements about "服装" include "花子は赤いワンピースを着ています" and "彼は青いスーツを着用している."

Type B: This is a set of questionnaire response sentences collected during an actual class and judged by the authors to be classifiable to some extent. They are questionnaire response sentences written in a free-description format by the students in response to the prompt "入管法改正の理由," and all sentences give a reason. The length of the answer sentence ranges from one word to about 30 characters, and the total number of sentences is 65.

Type C: This is a set of questionnaire response sentences that was collected during an actual class and judged to be difficult to classify. Most of them describe the students' impressions of the lesson, and each answer sentence consists of about 60 characters. The total number of sentences is 84.

3.2 Clustering experiments

Clustering experiments were conducted for the aforementioned Types A, B, and C data. The labels expected to be assigned to each data set are shown in Table 1. In each clustering, the execution time differs between the first execution and the second and subsequent executions. This seems to be because the processing program takes time to load at the first execution. In the experiments, we confirmed that clustering was executed in 0.01-0.05 seconds after the second time for all data.

Table 1. Types and expected class labels

Types	Expected class labels
A	大学、授業、政治、服装、食事、病気、スポーツ、災害
B	労働力不足、人手不足、東京オリンピック、外国人労働者
C	＊判断困難

3.2.1 Experiment for k-means++ method

When clustering Type A data with 8 clusters by the k-means++ method, sentences that are not suitable for a certain class are included, or classes are biased. After classifying several times with different numbers of clusters, we found that it was possible to classify relatively well with 12 clusters, as shown in Figure 1. The first execution time for clustering by the k-means++ method was 1.56 seconds.

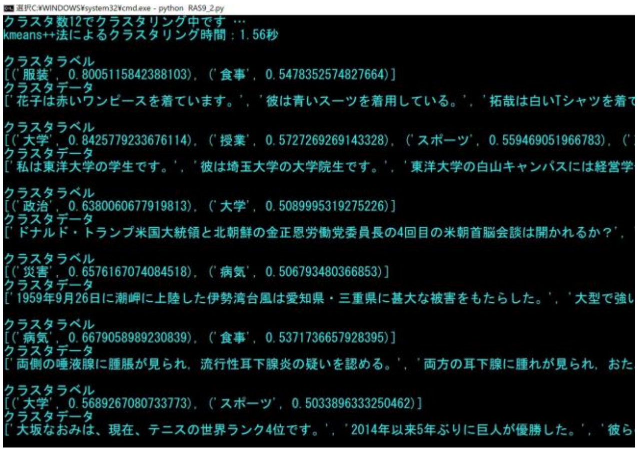


Figure 1. Clustering of Type A data by k-means++ method

Type A data is well classified overall. Two sentences, “両側の唾液腺に腫脹が見られ，流行性耳下腺炎の疑いを認める” and “両方の耳下腺に腫れが見られ，おたふくかぜではないかと疑う,” from medical records described in the literature [8] were found to be classified into the same class.

Next, Figure 2 shows the results for clustering Type B data with 4 clusters.

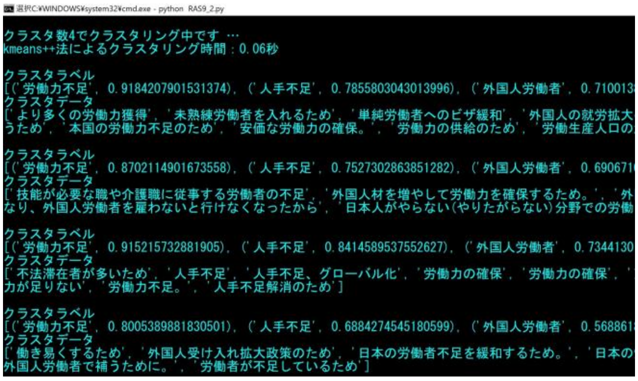


Figure 2. Clustering of Type B data by k-means++ method

When classifying in 4 classes, the “労働力不足” label becomes the most similar label in all clusters. The second most similar label is “人手不足,” and “外国人労働者” is third. In the third class, the “東京オリンピック” label has a similarity of more than 0.5, and the class contains a statement about the Olympics. In addition, response sentences containing the same words “労働力” and “人手不足” were classified into different classes.

Finally, clustering was performed on Type C data with 10 clusters. The results are given in Figure 3.

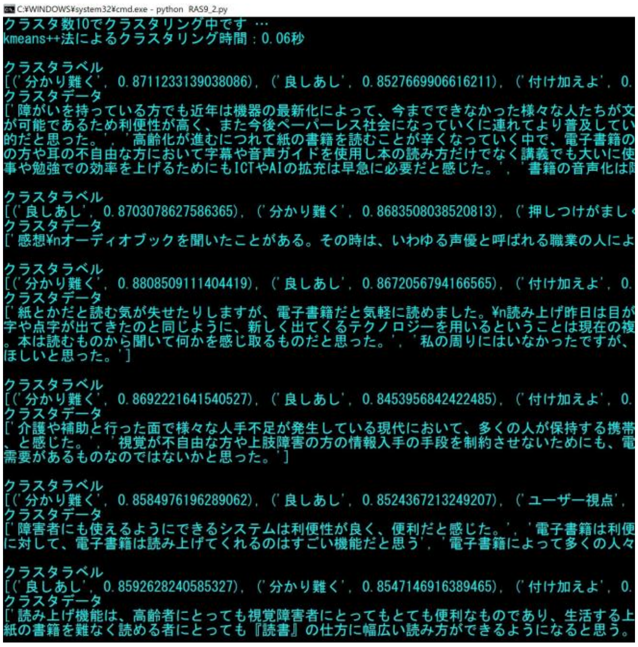


Figure 3. Clustering of Type C data by k-means++ method

In experiments on Type C data, all clusters were highly similar to the “分かり難く” and “良しあし” labels. Looking at each class, it is difficult to determine how the meaning of the answer sentences was classified. Experiments were performed several times with different numbers of clusters, but the results were the same, with the “分かり難く” label being at the top and being classified into classes that were difficult to label.

3.2.2 Experiment for Ward’s method

Clustering was performed on the Type A data using Ward’s method. The generated dendrogram is shown in Figure 4.

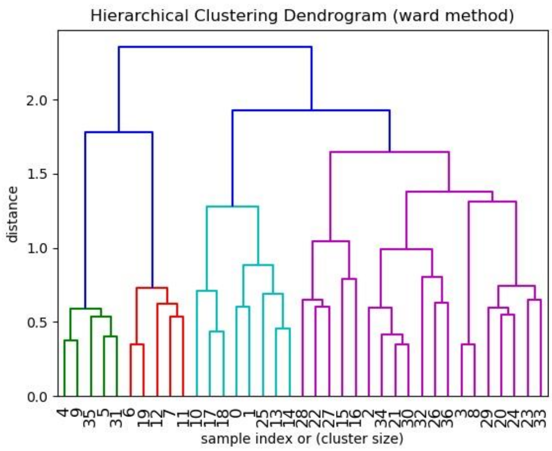


Figure 4. Dendrogram of Type A data by Ward’s method

Figure 4 shows that classification into 8 classes is possible by setting a threshold value of 1.1. Figure 5 shows the results of clustering with the threshold set to 1.1.

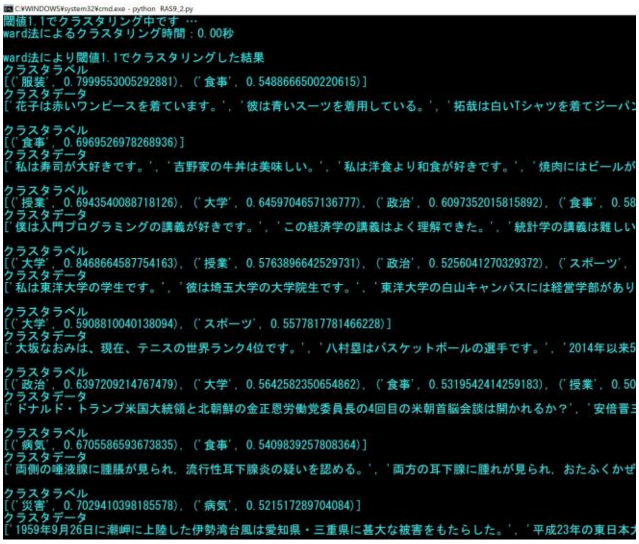


Figure 5. Clustering of Type A data by Ward's method

Type A data is clustered and labeled as expected. Also, as was the case with clustering by the k-means ++ method, "両側の唾液腺に腫脹が見られ, 流行性耳下腺炎の疑いを認める" and "両方の耳下腺に腫れが見られ，おたふくかぜではないかと疑う" are classified into the same class.

Next, clustering by Ward's method was performed on Type B data. Figure 6 shows the generated dendrogram.

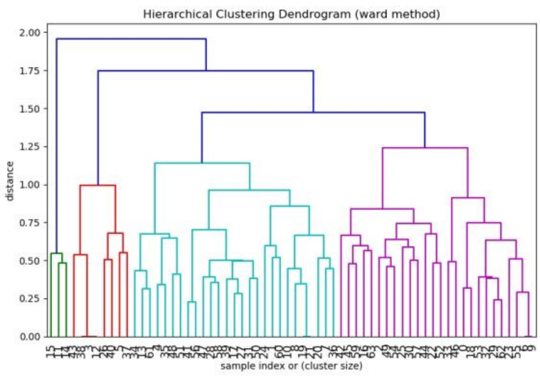


Figure 6. Dendrogram of Type B data by Ward's method

As shown, the data is classified into four classes by setting a threshold value of 1.31. The results of clustering are shown in Figure 7.

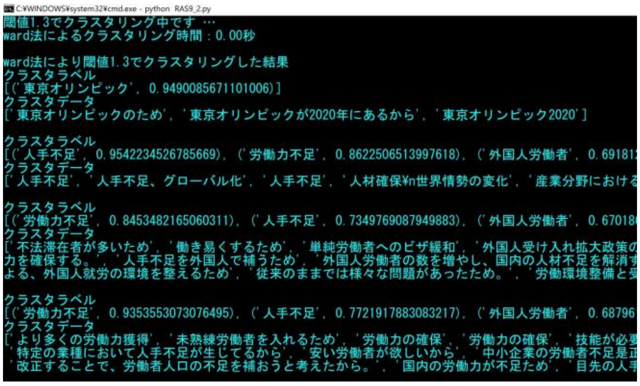


Figure 7. Clustering of Type B data by Ward's method

Sentences related to "東京オリンピック," "人手不足," and "労働力不足" gathered in each class, and the similarity of each label was high, confirming that the sentences were well classified. Sentences related to "外国人労働者" are also concentrated in one class, but the degree of similarity to the label "労働力不足" is high.

Finally, clustering by Ward's method was performed on Type C data. The generated dendrogram is shown in Figure 8.

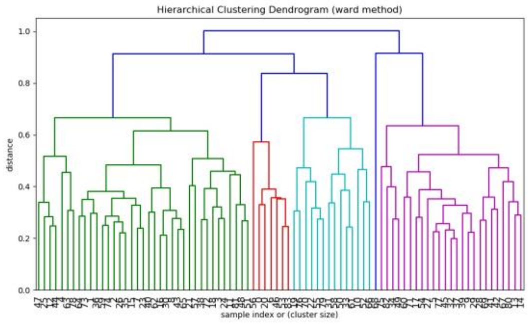


Figure 8. Dendrogram of Type C data by Ward's method

There is one class consisting of only one sentence, but Type C data can be classified into 5 classes with a threshold of 0.7. The classification results are shown in Figure 9.

Figure 9. Clustering of Type C data using Ward's method

In Type C data, most of the answer sentences are difficult for people to classify, and the answer sentences giving impressions and proposals are classified as scattered.

3.2.3 Experiment for group average method

Clustering was performed on Type A data using the group average method. The generated dendrogram is shown in Figure 10.

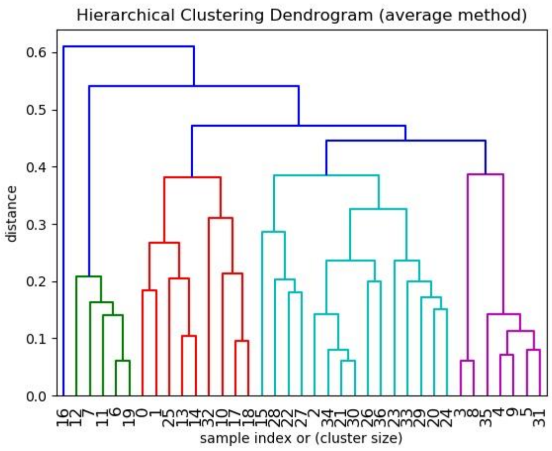


Figure 10. Dendrogram of Type A data by group average method

Figure 10 shows that Type A data can be classified into 8 classes with a threshold value of 0.35. The results of clustering are shown in Figure 11.

閾値0.35でクラスタリング中です …
群平均法によるクラスタリング時間：0.00秒

群平均法により閾値0.35でクラスタリングした結果
クラスタラベル
[('食事', 0.6982205054695243), ('服装', 0.5060136951180889)]
クラスタデータ
['私は寿司が大好きです。', '吉野家の牛丼は美味しい。', '私は洋食より和食が好きです。', '焼肉

クラスタラベル
[('大学', 0.8393068747988054), ('授業', 0.5643430440619102), ('スポーツ', 0.517390420583468
クラスタデータ
['私は東洋大学の学生です。', '彼は埼玉大学の大学院生です。', '東洋大学の白山キャンパスには

クラスタラベル
[('授業', 0.7091127974949453), ('大学', 0.6323769445149368), ('政治', 0.6121744022489869),
クラスタデータ
['僕は入門プログラミングの講義が好きです。', 'この経済学の講義はよく理解できた。', '統計学の

クラスタラベル
[('大学', 0.5689267080733773), ('スポーツ', 0.5033896333250462)]
クラスタデータ
['大坂なおみは、現在、テニスの世界ランク4位です。', '2014年以来5年ぶりに巨人が優勝した。',

クラスタラベル
[('災害', 0.6166701449286681), ('政治', 0.5798020949580929), ('病気', 0.5347005183858065),
クラスタデータ
['ドナルド・トランプ米国大統領と北朝鮮の金正恩労働党委員長の4回目の米朝首脳会談は開かれるか
相次いで発生した熊本地震では震度7が観測された。', '消費税率増加にあわせて、キャッシュレス決
が、御嶽山の噴火が発生した日だ。', '安倍総理大臣は、日本時間の27日夕方、ヨーロッパの首脳らが

クラスタラベル
[('病気', 0.6679058989230839), ('食事', 0.5371736657928395)]
クラスタデータ
['両側の唾液腺に腫脹が見られ，流行性耳下腺炎の疑いを認める。', '両方の耳下腺に腫れが見られ，

クラスタラベル
[('服装', 0.8005115842388103), ('食事', 0.5478352574827664)]
クラスタデータ
['花子は赤いワンピースを着ています。', '彼は青いスーツを着用している。', '拓哉は白いTシャツ

クラスタラベル
[('スポーツ', 0.5965406266959488)]
クラスタデータ
['八村塁はバスケットボールの選手です。']

Figure 11. Clustering of Type A data by group average method

Classes such as “食事,” “大学,” and “服装” are well categorized, but many sentences that are not suitable for the “災害” class are classified as such. As with the two previously discussed clusterings, here also “両側の唾液腺に腫脹が見られ，流行性耳下腺炎の疑いを認める” and “両方の耳下腺に腫れが見られ，おたふくかぜではないかと疑う” are classified into the same class.

Next, clustering by the group average method was performed on Type B data. The generated dendrogram is shown in Figure 12.

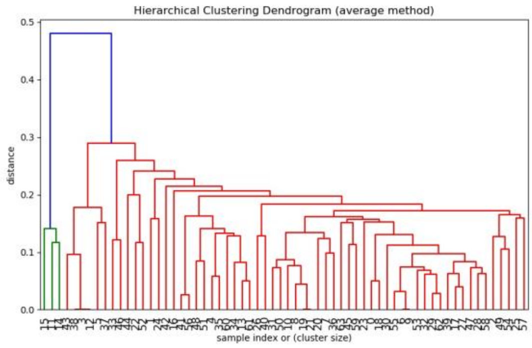
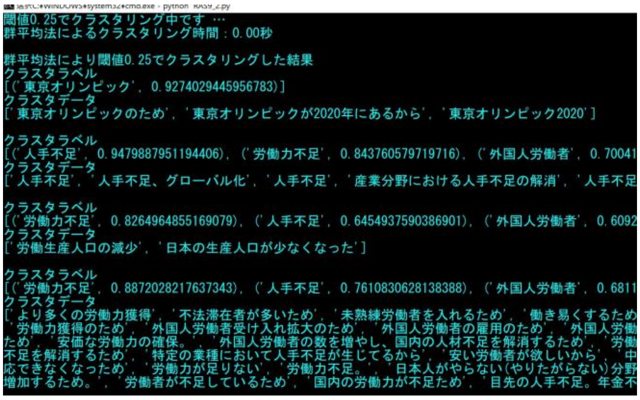


Figure 12. Dendrogram of Type B data by group average method

Figure 12 shows that the clusters are unevenly distributed. The results of clustering with a threshold value of 0.25 for classifying Type B data into four classes are shown in Figure 13.



閾値0.25でクラスタリング中です …
群平均法によるクラスタリング時間：0.00秒

群平均法により閾値0.25でクラスタリングした結果
クラスタラベル
[('東京オリンピック', 0.9274029445956783)]
クラスタデータ
['東京オリンピックのため', '東京オリンピックが2020年にあるから', '東京オリンピック2020']

クラスタラベル
[('人手不足', 0.9479887951194406), ('労働力不足', 0.843760579719716), ('外国人労働者', 0.70041
クラスタデータ
['人手不足', '人手不足、グローバル化', '人手不足', '産業分野における人手不足の解消', '人手不足

クラスタラベル
[('労働力不足', 0.8264964855169079), ('人手不足', 0.6454937590386901), ('外国人労働者', 0.6092
クラスタデータ
['労働生産人口の減少', '日本の生産人口が少なくなった']

クラスタラベル
[('労働力不足', 0.8872028217637343), ('人手不足', 0.7610830628138388), ('外国人労働者', 0.6811
クラスタデータ
['より多くの労働力獲得', '不法滞在者が多いため', '未熟練労働者を入れるため', '働き易くするため
'労働力獲得のため', '外国人労働者受け入れ拡大のため', '外国人労働者の雇用のため', '外国人労働
ため', '安価な労働力の確保。', '外国人労働者の数を増やし、国内の人材不足を解消するため', '労働
不足を解消するため', '特定の業種において人手不足が生じてるから', '安い労働者が欲しいから', '中
応できなくなったため', '労働力が足りない', '労働力不足。', '日本人がやらない(やりたがらない)分野
増加するため。', '労働者が不足しているため', '国内の労働力が不足ため', '目先の人手不足。年金不

Figure 13. Clustering of Type B data by group average method

As shown in Figure12, data was unevenly distributed in one class, and most of the response sentences were related to “外国人労働者” and “労働力不足.” Classes that have a high degree of similarity with “東京オリンピック” or “人手不足” are generally well classified.

Finally, clustering was performed on Type C data by the group average method. The generated dendrogram is shown in Figure 14.

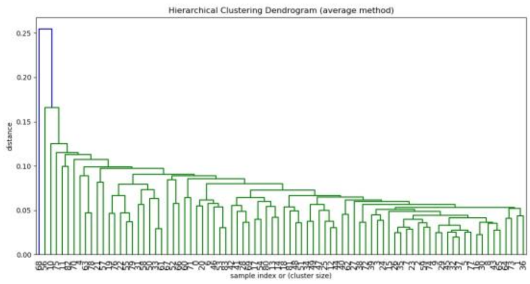


Figure 14. Dendrogram of Type C data by group average method

Similar to Type B data, Type C data also has an uneven cluster distribution. Figure 15 shows the results of clustering with a threshold of 0.114 that classifies Type C data into five classes.

Figure 15. Clustering of Type C data by group average method

Similar to clustering by Ward's method, response sentences that are difficult to label are grouped into one class, and a label with high similarity is "難しい." All the other classes have only one sentence, but the labels other than the words "利便性" and "興味深い" included in response sentences are not suitable for the clusters.

3.3 Response-sentence retrieval experiment

We tested a questionnaire-response retrieval function for query sentences on Type A data. As retrieval targets, we retrieved the response sentences related to "服装" whose similarity to the label was high (at around 0.8) in all three clusterings and the response sentences related to "病気" that we noted previously. Type A data includes five sentences that are expected to be related to "服装," and two sentences that are related to "病気." For "服装," we entered three queries: Q1: "服装に関する話題," Q2: "みんなの着ている服を教えて," and Q3: "服装." For queries regarding "病気，" Q4: "病気の症状を教えて," Q5: "風邪に関する話題," and Q6: "病気" were entered. The similarity threshold for the displayed sentences was set to 0.7. Figure 16 shows the results for the searches related to "服装," and Figure 17 shows the results for the searches related to "病気."

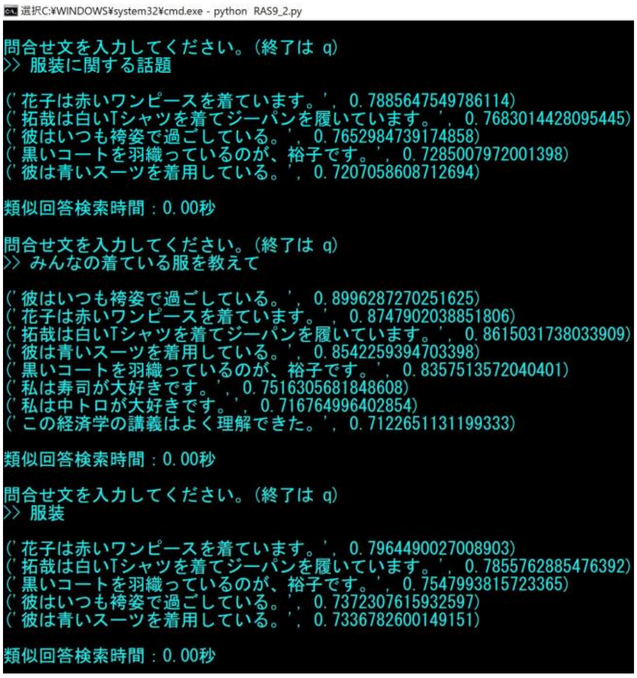


Figure 16. Sentence retrieval results for "服装" with Type A data

The numbers on the right side of the retrieved sentences represent the cosine similarities between the input query and the sentence that has been retrieved and displayed. For Q1: "服装に関する話題," five sentences were retrieved and displayed as expected. The results of Q2: "みんなの着ている服を教えて" include sentences about "食事" and "授業" in addition to sentences about "服装." However, all the retrieved sentences related to "服装" exceeded the similarity of 0.8, and there was a clear boundary with the similarity of the other sentences. For Q3: "服装," as with Q1, only five sentences were displayed, and the results were as expected.

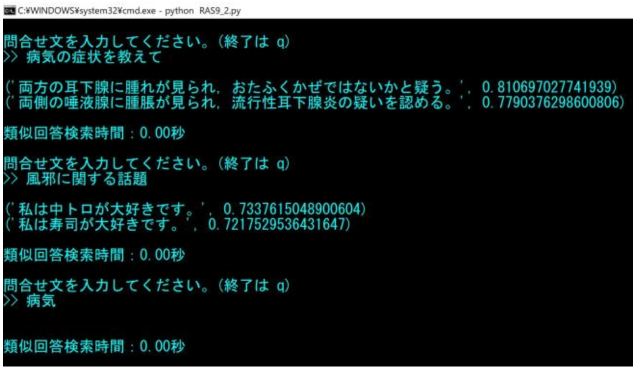


Figure 17. Sentence retrieval results for "病気" with Type A data

For Q4: "病気の症状を教えて," the two sentences we expected were displayed. For query Q5: "風邪に関する話題," the expected sentences were not displayed; instead, only two sentences related to one other topic, "食事," were displayed. In the retrieval for Q6: "病気," similar sentences were not displayed. The execution time for all retrievals was within 0.02 seconds, demonstrating that the search is very fast.

Next, we conducted a questionnaire-response retrieval for Type B data using two queries, Q7: "労働力不足に関する意見," and Q8: "外国人労働者に関する意見." The retrieval results are shown in Figure 18.

Figure 18. Sentence-retrieval results for Type B data

Some of the retrieval results for Q7: “労働力不足に関する意見” are thought to be opinions on “外国人労働者,” and it is difficult to judge whether or not they are appropriate for Q7. However, it is possible to search for answer sentences that have the same meaning as “労働力不足,” such as “労働力の不足” and “労働力の確保.” Regarding Q8: “外国人労働者に関する意見,” it is also possible to search for sentences that do not contain the word “外国人労働者.”

Finally, questionnaire response sentences were retrieved for Type C data. We retrieved sentences containing the words “電子書籍” and “視覚障害者,” which were given in many response sentences. The retrieval results are shown in Figure 19.

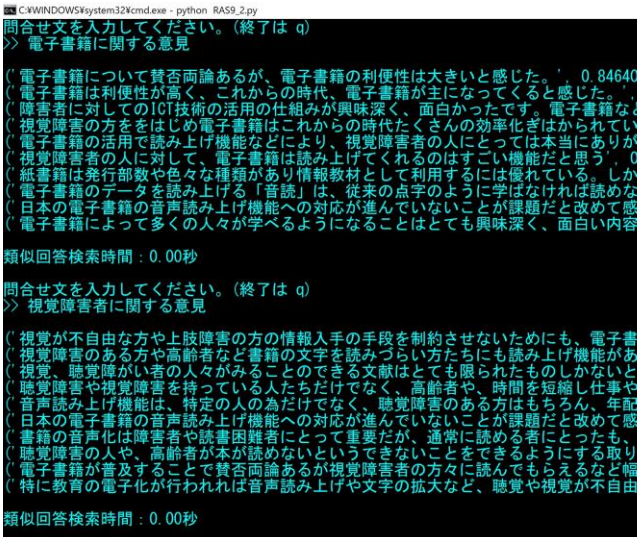


Figure 19. Sentence retrieval results for Type C data

In the retrieval related to “電子書籍,” the word “電子書籍” was included in all the response sentences, but the retrieval function worked sufficiently. In the retrieval for “視覚障害者,” response sentences that included words such as “視覚が不自由な方” and “視覚障がい者” could also be retrieved.

3.4 Discussion of experimental results

In the experiments on Type A data, the results for clustering and sentence retrieval were almost all as expected in all experiments, and there were no problems with the functions and performance of this questionnaire-response-sentence analysis system.

In the experiment on Type B data, the clustering by Ward’s method did not all label the clusters well, but response sentences with similar meanings could be classified well. In the response-sentence retrieval, we retrieved sentences containing words such as “労働力不足” or “外国人労働者,” but at the same time, we could also find sentences that do not contain these words but have the same meaning. Therefore, we can anticipate that flexible retrieval is possible even for actual questionnaire-response-sentence retrieval.

Experiments on Type C data did not give the results that were expected for all clusterings. The cluster labels were ambiguous labels that were difficult for users to understand, such as “難しい” and “分かり難く.” One possible reason for this is that if too many words are included in one sentence, its distributed representation vector will be averaged, and its meaning will approach an ambiguous one.

4. Concluding Remarks

In this study, we developed a lesson support system that analyzes, classifies, and retrieves questionnaire response sentences collected during class on the basis of the distributed representation derived from Wikipedia Entity Vectors. This system takes about one and a half minutes to load Wikipedia Entity Vectors at startup. The execution time of each function is less than 2 seconds at the first execution time, and less than 0.01 seconds from the second execution onward. Also, it is possible to classify and retrieve the questionnaire response sentences in real time while the system is operating (that is, during lectures). Therefore, this class support system is potentially very useful for collaborative and interactive class support.

In the future, we would like to implement a function to enable teachers to classify and search sentences as desired using deep learning in this lesson support system. At this stage, the system runs on the command line, but in the future, we will implement a GUI that can be operated intuitively and visually.

Acknowledgments

We are conducting this research as part of the “ICT Education and Research Project” of the “Development of AI Functions for Education” group at the Institute of Social Sciences, Toyo University. We would like to thank Prof. Satoru Matsubara of Toyo University for allowing us to use the class questionnaire data.

Reference

[1] The Ministry of Education, Culture, Sports, Science and Technology, “The Second Basic Plan for the Promotion of Education,” 2013.

[2] Masayuki Kameda, Tsunenori Ishioka, Dongyue Louis Liu, “*Tantou kijutsushiki mondai kaitoubun no saiten shien shisutemu JS4 no shisaku*” [Prototype JS4 of Scoring Support System for Answers to Short-answer Questions], Proceedings of the Twenty-third Annual Meeting of the Association for Natural Language Processing, pp. 1137–1140, 2017 (in Japanese).

[3] Midori Matsumoto, Toshio Yamaguchi, Kenji Tsujino, Noriyoshi Matsumoto, Junji Kinoshita, “*Gakusei repo-oto no kyouiku hyouka*

shien tsu-uru ni yoru hyou de-eta no kaiseki” [Analysis of Tables by Grading Support System of Reports], IPSJ SIG Technical Report, Vol.2015-CE-129, No.23, 2015 (in Japanese).

[4] Shoya Ota, Hiroyuki Tominaga, “*Puroguramingu enshuu ni okeru hojosha no junkai shidou no tame no taburetto PC jou no shien tu-uru – shou kontesuto keishiki no shokyuu C enshuu deno jissen ni okeru tu-uru no sousa rogu no bunseki*” [A Support Tool of Instruction Assistants with Tablet PC in Classroom Round for Programming Exercises – Analysis of Operation Logs in Introductory C Exercises Based on Small Contest Style –], IPSJ SIG Technical Report, 2016-CE-137, No. 1, 2016 (in Japanese).

[5] Hirokazu Bandoh, Masaki Yamashita, Hidekazu Kaminishi, Takashi Tatenuma, Hiroko Umemura, Chiaki Fujiyama, Kazuya Ohashi, Nobuhiro Sakata, “*jugyou chuu ni kyoushi no kako no PC sousa wo sanshou dekiru gamen seni sanshou tu-uru no teian to shisaku*” [Prototype of Screen Transition Reference Tool for Student to Refer Past Screenshots of Instructor's PC in Class], IPSJ SIG Technical Report, 2016-CE-133, No. 21, 2016 (in Japanese).

[6] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, “Efficient Estimation of Word Representations in Vector Space,” Workshop Proceedings of International Conference on Learning Representations, 2013.

[7] Jey Han Lau, Timothy Baldwin, “An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation,” Proceedings of the 1st Workshop on Representation Learning for NLP, 2016.

[8] Masaaki Tanaka, “*Tango no bunsan hyougen wo motiita bunsyo bunrui*” [Document Classification using Distributed Representation of Words as Features], *Kawasaki Journal of Medical Welfare*, Vol. 28, No.1, pp.167–178, 2018 (in Japanese).

[9] Inui-Suzuki Laboratory, Tohoku University, “Wikipedia Entity Vectors,” http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/ (accessed 2019/10/06).

[10] eventos, https://eventos.tokyo (accessed 2019-10-15).

[11] respon, https://respon.jp (accessed 2019-10-15).

[12] Kouji Etchu, Toshiko Takata, Hidetoshi Kinoshita, Akinobu Ando, Kiyoshi Takahashi, Kenichi Tabata, Masaaki Oka, Kimiharu Ishizawa, “*Tekisuto mainingu ni yoru jugyou hyouka anke-eto no bunseki: kyouki nettowa-aku ni yoru jiyuu kijutsu no kashika no kokoromi*” [Analysis of Questionnaires for Class Evaluation by Text Mining: Attempt to Visualize Free Description by Co-occurrence Network], Bulletin of Information Processing Center, Miyagi University of Education, Vol. 22, pp. 67–74, 2015 (in Japanese).

[13] Hideya Matsukawa, Makiko Oyama, Chiharu Negishi, Yoshiko Arai, Chiaki Iwasaki, Hiroshi Hotta, “Analysis of the Free Descriptions Obtained through Course Evaluation Questionnaires Using Topic Modeling,” *Educational Technology Research*, Vol. 41, Issue 1, pp. 125–137, 2019.