DOI: 10.2478/dcse-2021-0019

sciendo

Discourse and Communication for Sustainable Education,
vol. 12, no. 2, pp. 96–107, 2021

Actual State of COVID-19 Strategy Meetings

Takao Arai, Kengo Saito and Yuji Hirai
Tokyo Medical University, Hachioji, Japan

Abstract

The authors of this paper applied a new approach combining text mining and principal component analysis (PCA) to objectively determine the actual state of regional COVID-19 strategy meetings and verified its utility. The authors used text mining to analyze meeting minutes and extracted words with high phase ubiquity by co-occurrence analysis. Then, they selected words symbolizing the meeting contents ("report," "prevention," "rules," and "decision") and performed PCA using the occurrence rates of these words as variables. Two principal components (PC1, PC2) were set. For PC1, we observed maximum factor loading for "decision" (0.81) and minimum for "report" (-0.72), so we considered this axis to show the "depth of meeting discussions." For PC2, we observed maximum factor loading for "prevention" (0.81) and minimum for "rule" (-0.76). We considered this axis to show "regional infection status." When we created a plot of all 44 meetings, Phase 1 occurred in quadrants 3 to 4 (knowledge sharing), phase 2 began in quadrant 1 (preparation for spread), and phase 3 shifted to quadrant 2 (response to spread) with significant differences between these phases. Our findings suggest that the actual state of regional COVID-19 strategy meetings could be objectively determined by using a combination of text mining and PCA.

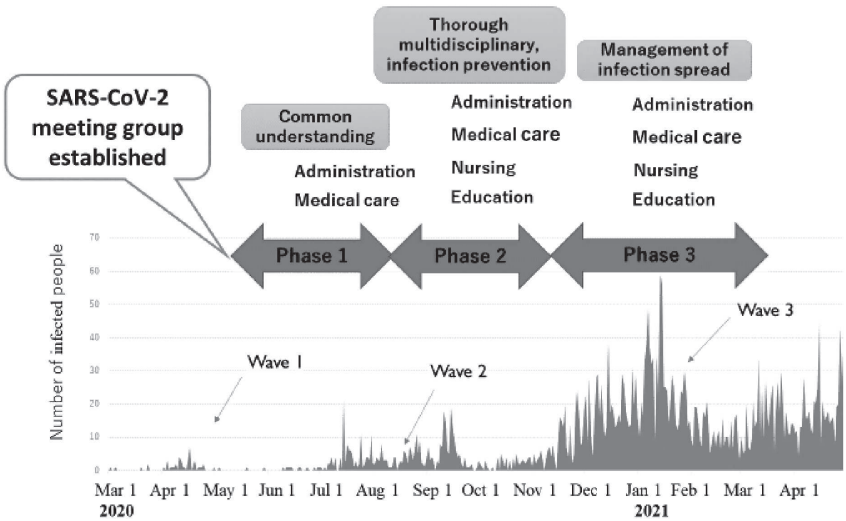
Key words: principal component analysis, text mining, COVID-19, sustainability

Introduction

Starting in May 2020, online SARS-CoV-2 meetings were initiated in Hachioji, Tokyo to discuss and share measures for combatting the COVID-19 pandemic in the region (Figure 1). These meetings were held 44 times over a 1-year period, with regular participation by about 30 stakeholders in the region. At the start of the pandemic (phase 1), knowledge about COVID-19 was shared mainly among health care workers and administrative officers. In the middle stage (phase 2), nursing and education professionals began participating as well, and infection prevention measures were disseminated and implemented.

Figure 1

Establishment of the SARS-CoV-2 Meeting Group. The Meeting Agenda and Participant Categories During Phases 1 to 3 are Shown

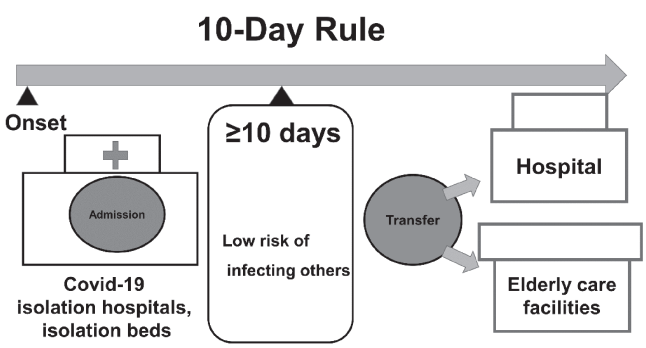


During the subsequent spread of the infection, an arrangement was made to adhere to a "10-day rule". Special isolation beds were prepared for patients diagnosed as having COVID-19, who then used these beds for up to 10 days after disease onset, after which they could be transferred to other hospitals (without isolation capabilities) or elderly care facilities. This arrangement was made to promote improved patient flow (Figure 2). The infectiousness of the original SARS-CoV-2 strain decreases significantly during the first 10 days after disease onset, although it does not reach a value of zero (Jeroen et al., 2021).

However, in order to admit as many COVID-19 patients as possible to the isolation wards, a consensus was reached to share the risk throughout the entire region.

Figure 2

The 10-Day Rule, Which Stipulated That COVID-19 Patients in Isolation Beds be Promptly Transferred to Other Hospitals or Elderly Care Facilities on Day 10 After Disease Onset



However, for local leaders, dealing with an unknown virus necessitated making a series of "blind decisions" that could not be solved by knowledge alone. As the director of the Emergency and Critical Care Center at a regional hospital, the author chaired all the meetings. The participants were faced with a grave dilemma: choosing between social activity and infection control. To better understand the choice facing them, the meeting group focused on ensuring a meaningful discussion and obtaining a common understanding (universal sense of values), rather than rushing to find answers. The meetings proceeded under the idea that a broader perspective would promote individual decision-making and would feed into direct regional infection control.

The meetings allowed the stakeholders involved to gain an actual sense of various outcomes, such as the 10-day rule. We hoped that communicating these types of best practices would provide support for other regional endeavors.

However, the discussions held during these meetings included diverse participants who spoke from a variety of different perspectives. Despite being a participant myself, it was very difficult for me to adequately characterize the details. Therefore, to present the meeting methodology and outcomes in a comprehensible manner that would allow them to be utilized in a variety of settings, we needed to first characterize the contents of the meetings in an objective manner.

However, this did not always work out well in terms of interviews and questionnaires (Janz, 2006; Koch et al., 2016; Scheier & Cattell, 1958). In contrast, quantitative text analysis of the meeting minutes, based on text mining techniques, has been widely adopted for objectively characterizing the actual state of meetings (Blaschke et al., 2003; Dancy-Scott et al., 2018; Przybyła et al., 2016; Renganathan, 2017). Furthermore, in this paper, we sought to determine whether it is possible to objectively grasp the actual state of discussions by making use of a new analysis method that combines text mining and PCA. According to our own exhaustive literature search, this approach has not previously been performed in this area.

The COVID-19 pandemic is not just a health or medical issue. Measures to prevent the spread of the infectious disease will lead to restrictions on people's activities and movements. Also, restrictions on activities and movements will result in a contraction in economic activities. Further, increased demand for masks will lead to increased waste and develop into environmental problems such as increased emissions of carbon dioxide.

Breaking this negative cycle is synonymous with making the economy, society, and the environment sustainable. In other words, measures against the COVID-19 pandemic will not only improve the sustainability of human life and health, but also lead to the sustainability of the economy and the planet. These are in line with the Sustainable Development Goals, or SDGs, which were agreed upon by all member states at the United Nations General Assembly in 2015 to be realized by 2030 (Heasly et al., 2020; Lewis et al., 2019).

The objective of this study was to conduct a PCA of the minutes from local government meetings, which included stakeholders involved in medical care, administration, nursing, and education, as well as to examine whether it is possible to create an objective overview of this discussion.

This local government conference is an initiative related to "community health care collaboration" in "Goal 3: Ensure healthy lives and promote well-being for all at all ages", one of the 17 SDGs (Jodoin & Singer, 2020). The results of this study can contri-

bute to the implementation of Education for Sustainable Development, or ESD, which stands for group work and education by diverse people who will be the creators of the SDGs, by presenting a new method of evaluating the effectiveness of the conference (Mochizuki & Fadeeva, 2010, p. 391; Salīte, 2016).

Research Methods

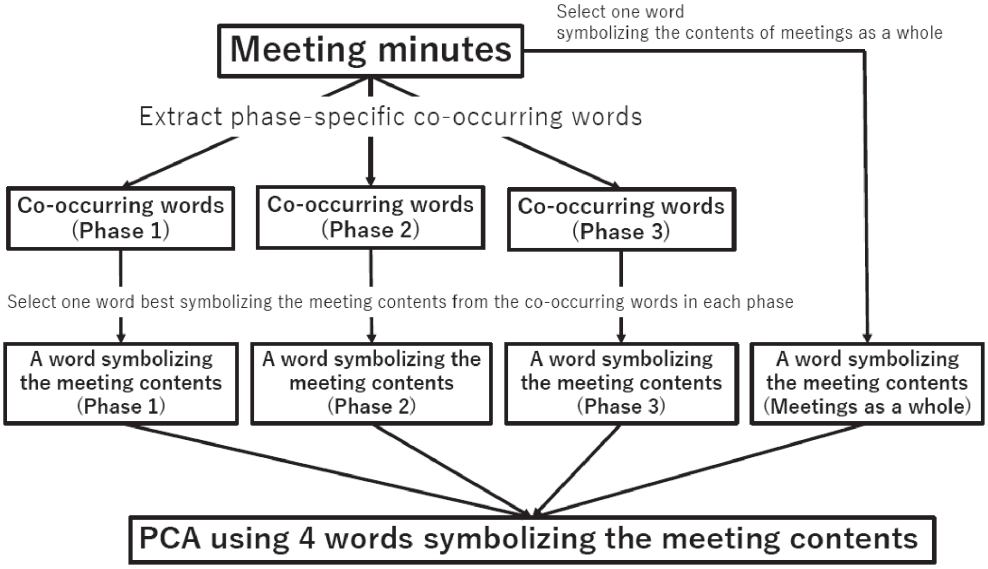
The study was conducted retrospectively. Text mining was applied to the minutes (transcription type) of the COVID-19 countermeasures meetings, and the extracted words were analyzed using PCA.

Analysis of Research Data

The meeting minutes were analyzed using text mining (Blaschke et al., 2003; Dancy-Scott et al., 2018; Przybyła et al., 2016; Renganathan, 2017). As shown in Figure 3, words with high phase distribution were extracted by word co-occurrence analysis (de Quadros et al., 2016; Junge & Jensen, 2020; Maskery et al., 2006; Pavlopoulos et al., 2014), among which words with high phase symbolism (words related to the phase summary) were selected and analyzed using PCA, with the rate of occurrence of those words used as variables (David & Jacobs, 2014; Jolliffe & Cadima, 2016; Martins et al., 2019; Sell et al., 2020). Two principal components were set during the study. Text mining was performed using KH Corder Ver. 3, and statistical analysis was performed using SPSS ver. 27. Ratios were compared using the chi-square test. A p-value of <0.05 was considered significant.

Figure 3

A New Approach Combining Text Mining and PCA



Online Meetings for Regional Cooperation

The SARS-CoV-2 meeting group was established in Hachioji, Tokyo (population approx. 560,000) to discuss the COVID-19 pandemic (Figure 1). Many of the city's medical, nursing, education, and government stakeholders participated in these meetings. Representatives from each of these four sectors communicated the contents of the meetings to their respective hospitals or institutions.

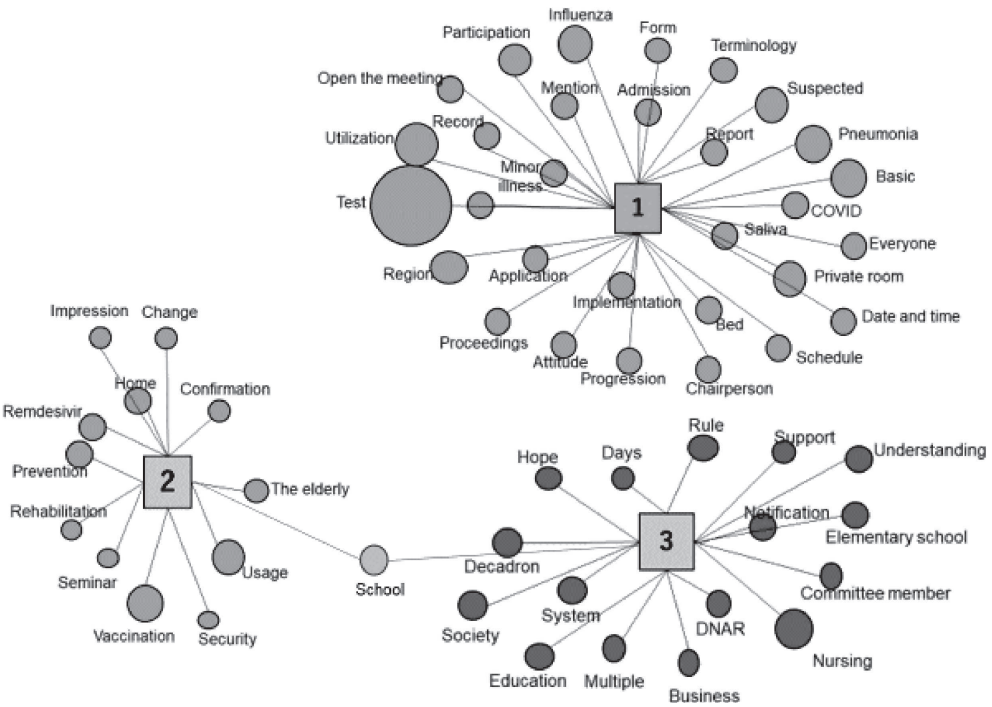
The meetings can be divided into three phases (Figure 1). Phase I (meetings 1–16): Knowledge was shared primarily by health care workers and government officials. Phase II (meetings 17–30): Infection prevention measures were disseminated and implemented by nursing and education professionals in preparation for a winter pandemic. Phase III (meetings 31–44): Countermeasures were discussed in the midst of an infectious disease pandemic.

Research Findings

Figure 4 shows the results of the co-occurrence network analysis (extraction of words with high phase distribution) for each phase.

Figure 4

Co-occurrence Network Diagram for Phases 1 to 3. The Closer the Terms are to the Center, the Higher the Co-occurrence



Among the 60 co-occurring words shown in Figure 4, "report" (phase 1), "prevention" (phase 2), and "rules" (phase 3) were selected by the author as the words that best symbolized the content of the discussion in each phase. Similarly, "decision" was selected as the word that best symbolized the philosophy of all the meetings as a whole.

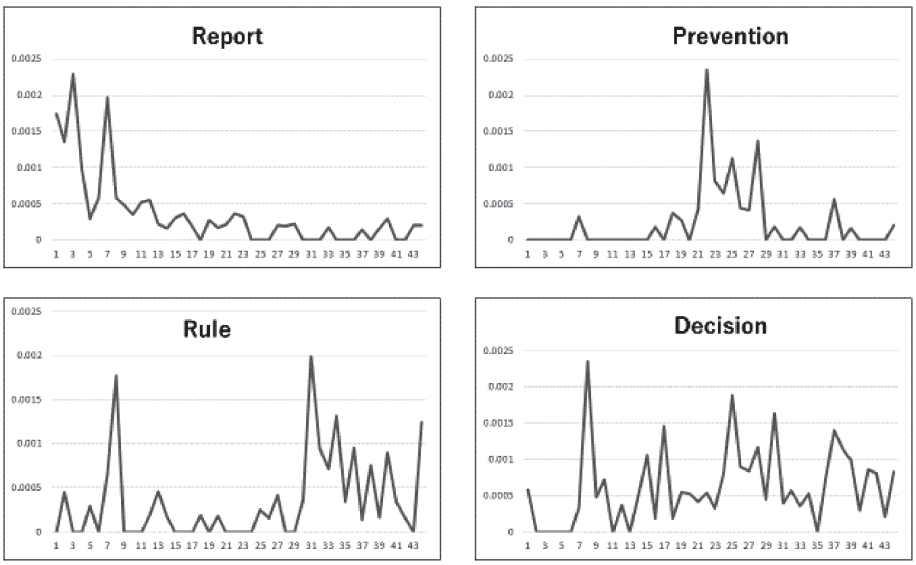
The context in which the above four words were used was examined in the minutes. The word "report" (e.g., "scattered reports of reinfection" and "PCR was reported to be positive for about 50 days at the longest") was used mainly for knowledge sharing. The word "prevention" (or "prevent") was used mainly in the context of infection-prevention measures (e.g., "we ensure that elderly residents wear masks to prevent infections in the facility" and "children are also aware of the prevention measures"). The word "rules" was used mostly in phase 3, in the context of the "10-day rule," an emergency measure to combat the spread of the infection.

The word "decision" was used for behaviors (e.g., "I am worried about the decision to cancel the school trip" and "How do we decide whether to loosen the visitation restrictions at elderly care facilities?") and was often used during discussions aimed at reaching a consensus among participants about decisions.

Figure 5 shows the occurrence rate distribution of these four words. For example, in the case of "report," the occurrence rate indicates the number of times the word "report" was used in the first meeting relative to the total number of words used.

Figure 5

Occurrence Rate Distribution of the Four Selected Words Across All Meetings (1–44)

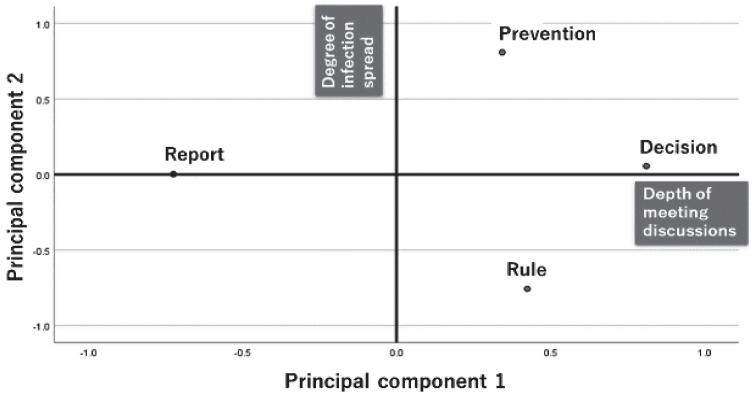


PCA was performed, and the occurrence rates of the four words were used as variables. The contribution of the first principal component was 37.0 % and that of the second principal component was 30.7 %. This means that 67.7 % of the 4 dimensions were aggregated into 2 dimensions.

Figure 6 shows the factor loadings of the four selected words in the first and second principal components.

Figure 6

Plot of Factor Loadings of the Four Selected Words. Due to This Weighting, the First and Second Principal Components had a New and More Comprehensive Meaning



For each meeting (1–44), the values of the first and second principal components can be calculated by the occurrence rate of each word.

In the calculation of the first principal component, "decision" was a positive factor, whereas "report" was a negative factor. In other words, the first principal component increased when the occurrence rate of "decision" was high and when that of "report" was low. The effect of "prevention" on the first principal component was less than half that of "decision." However, in the second principal component, the effect of "prevention" was the largest, followed by "rule," which had a large negative effect. "Report" and "decision" had almost no effect.

Figure 7 is a plot of the positions of meetings 1–44 divided into phases on the plane of Figure 6. The vertical axis is the first principal component, and the horizontal axis is the second principal component.

Figure 7

Plot of the Positions of Meetings 1–44 by Phase

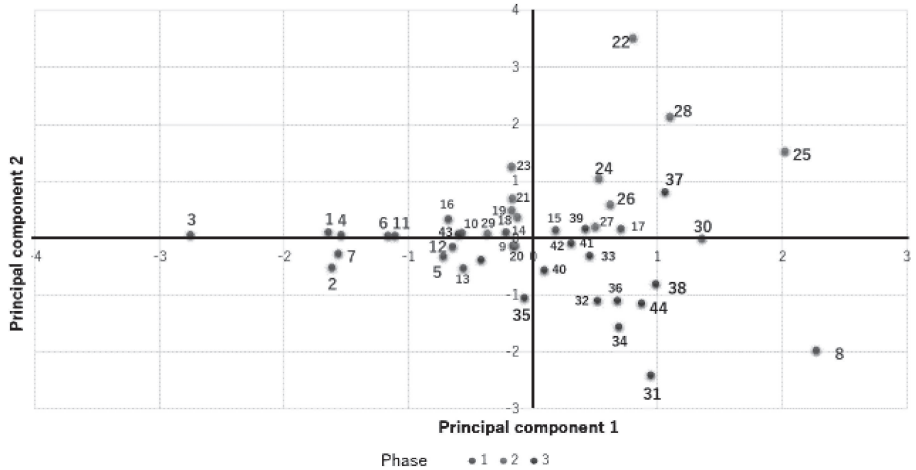


Table 1 shows a cross-tabulation showing which meeting phases in Figure 7 were located in which quadrant.

Table 1

Cross-Tabulation Table Showing the Distribution of Phases and Quadrants

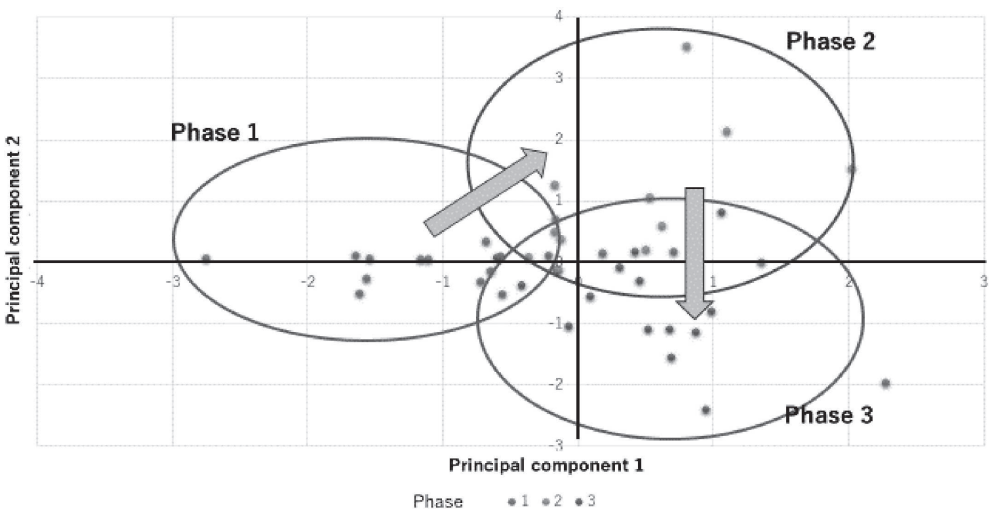
Quadrant	Principal component 1	Principal component 2	Phase 1	Phase 2	Phase 3	Total
1	+	+	2	7	2	11
2	+	–	1	0	9	10
3	–	–	14	6	3	23
4	–	+				
			17	13	14	44

Phase 1 is distributed higher in quadrants 3 and 4, Phase 2 is distributed higher in quadrant 1, and Phase 3 is distributed higher in quadrant 2. A significant difference between these values was verified ($p < 0.05$, chi-squared test).

Finally, Figure 8 shows the transition from phases 1 to 2 to 3 on the analysis plane.

Figure 8

Transition From Phases 1 to 2 to 3 on the Analysis Plane



Discussion

In this study, the authors conducted a PCA, using the occurrence rate of the four selected words that best symbolized the meetings as variables. PCA is a multivariate analysis, and thus aggregates multidimensional information into fewer dimensions (David & Jacobs, 2014; Jolliffe & Cadima, 2016; Martins et al., 2019; Sell et al., 2020). In this study, 67.7 % of the total variance in 4 dimensions was reduced to 2 dimensions.

In the first principal component, there was heavy factor loading for "decision" at 0.81, and low factor loading for "report" at 0.72 (Figure 6). This means that when the occurrence rate of "decision" was high and that of "report" was low, the first principal component increased. In other words, the content of the discussion was not simply a "report" (= knowledge sharing), but rather sought to reach a consensus as well as provide support for individual "decisions." In such "in-depth" discussions, the first principal component increased. The first principal component was considered to be an axis indicating the depth of the discussion during meetings.

In the second principal component, the factor loading for "prevention" was greatest at 0.81, whereas "rule" had a large negative effect of -0.76 (Figure 6). In other words, the second principal component was affected by the implementation of infection "prevention" involving the fields of nursing and education (= preparation to prevent the spread of infection) and the significant operation of the 10-day "rule" (= realistic response to the spread of the infection). The second principal component was considered to be the axis that represents the regional infection status.

Based on these results, meetings that took place in quadrants 3 or 4 are presumed to have still been primarily knowledge-sharing meetings (basic stage). In contrast, meetings that took place in the first quadrant are presumed to have involved more in-depth discussions, when the infection was not yet spreading very quickly. When meetings took place in the second quadrant, it is likely that the discussions were more in-depth, and the infection was spreading rampantly.

The plot results (Figure 7) show an approximate depiction (expectation) of these cases.

In practice, phase 1 begins in quadrants 3 and 4 (knowledge sharing), and phase 2 begins in quadrant 1 (more in-depth discussion, infection not spreading much); there were significant differences between these phases. Phase 3 shifts to the second quadrant (the period that indicates a response to the spread of infection) (Figure 8).

In this way, changes in the actual state of the meetings due to the transition from one phase to the next could be depicted as a dynamic change on a completely new plane. PCA allowed for an objective supplemental characterization of the meetings the authors participated in.

The formation of sustainable education requires the building of a common understanding (universal values) through adequate dialogue and communication between people (Alameda-Hernández, 2008; Foucault, 1981; Hall, 1997). In this regard, it was suggested that the method used in this study could be implemented to evaluate Education for Sustainable Development (ESD) in a more objective manner.

Also, it was suggested that, through this method, the accuracy of the evaluation of education for sustainable development (ESD) could be further improved.

When performing PCA, correlations between variables are necessary (Abegaz et al., 2019; Clark & Ma'ayan, 2011; Zhao et al., 2021). In this paper, to extract the 4 above mentioned words, we focused on the uneven distribution in each phase, and first extracted 60 unevenly distributed words using co-occurrence network analysis (Figure 4). These 60 words were then narrowed down to 3 with an emphasis on the symbolism of the words, based on an awareness of the field. In addition, by adding the most important word describing the contents of all the meetings (i.e., "decision"), we were able to get as close as possible to the real state of the meetings.

Thus, to perform a linguistic analysis of the meeting contents by using PCA, it is necessary to extract words that satisfy both the mathematical correlation (uneven distribution) and the symbolism from a linguistic approach. However, the study of symbolism inevitably involves the subjectivity of the researcher. This was a major dilemma of the study. This means that the reproducibility of the method in this study is unclear, and can therefore be considered a limitation of the present research. It may also be effective to set the variables using only the ubiquity, and discuss semantics only in the final stage of axis interpretation. There is still room to perform comparative studies in future using these methods.

Conclusions

The objective of this study was to examine whether it is possible to create an objective overview of local government covid-19 meetings by making use of a new analysis method that combines text mining and PCA.

In practice, by using text mining to extract the important words from meeting minutes and then aggregating the information into 2 dimensions using PCA, meeting contents were objectivized and explained in a format that was understandable to everyone. Also, meeting participants were able to accurately understand the changes in meeting contents over the course of the pandemic and to explain them to others by using figures for the analytical results. In fact, these meetings enabled the 10-day rule to function effectively, and this analysis method will be useful when expanding such best practices to various other regions.

Education for Sustainable Development (ESD) is essential for achieving sustainable development. By implementing the method used in this study, it will be possible to objectively evaluate basic aspects of ESD, such as dialogue and communication between people, with a higher degree of accuracy. Ultimately, it is expected that this will further facilitate ESD.

Although the reproducibility of the method in this study is still unclear, in any case, this study allowed us to confirm that PCA can be used to effectively supplement conventional text mining methods (e.g., word frequency and co-occurrence analyses). The results of this study will provide a theoretical basis for further studies to establish a new evaluation method for ESD.

References

Abegaz, F., Chaichoompu, K., Génin, E., Fardo, D. W., König, I. R., John, J. M. M., & Van Steen, K. (2019). Principals about principal components in statistical genetics. *Brief Bioinform*, *20(6)*, 2200–2216. https://doi.org/10.1093/bib/bby081

Alameda-Hernández, Á. (2008). SFL and CDA: Contributions of the analysis of the transitivity system in the study of the discursive construction of national identity (case study: Gibraltar). *The Linguistics Journal*, *3*(3), 160–175.

Blaschke, C., Yeh, A., Hirschman, L., & Valencia, A. (2003). ISMB 2003 Text mining SIG meeting report. *Computer Funct Genomics*, *4*, 667–673. https://doi.org/10.1002/cfg.338

Clark, N. R., & Ma'ayan, A. (2011). Introduction to statistical methods to analyze large data sets: Principal components analysis. *Science Signaling*, *4*(190), 3. https://doi.org/10.1126/scisignal.2001967

Dancy-Scott, N., Dutcher, G. A., Keselman, A., Hochstein, C., Copty, C., Ben-Senia, D., Rajan, S., Asencio, M. G., & Choi, J. J. (2018). Trends in HIV terminology: Text mining and data visualization assessment of International AIDS Conference abstracts over 25 years. *JMIR Public Health Surveill*, *4*, e50. https://doi.org/10.2196/publichealth.8552

David, C. C., & Jacobs, D. J. (2014). Principal component analysis: A method for determining the essential dynamics of proteins. *Methods in Molecular Biology*, *1084*, 193–226. https://doi.org/10.1007/978-1-62703-658-0_11

de Quadros, L. G., Kaiser Jr., R. L. K., Neto, M. D. P. G., Campos, J. M., de Santana, M. F., & Ferraz, A. A. B. (2016). Long-term postoperative endoscopic findings after gastric bypass procedure: A co-occurrence analysis. https://doi.org/10.1590/s0004-28032016000400012

Foucault, M. (1981). The order of discourse. In R. Young (Ed.), *Untying the text: A post-structural anthology* (pp. 48–78). Boston: Routledge & Kegan Paul.

Heasly, B., Lindner, J., Iliško, D., & Salīte, I. (2020). From initiatives, to insights, to implementation of the sustainability and securitability agenda for 2030. *Discourse and Communication for Sustainable Education*, *11*(1), 1–4.

Hall, S. (1997). The work of representation. In S. Hall (Ed.), *Representation: Cultural representations and signifying practices* (pp. 13–74). London: Sage and The Open University.

Janz, K. F. (2006). Physical activity in epidemiology: Moving from questionnaire to objective measurement. *British Journal of Sports Medicine, 40*, 191–192. https://doi.org/10.1136/bjsm.2005.023036

Jeroen, J. A., David, A. M., Pieter, L. A., Bart, L. H., Mart, M. L., Nisreen, O., Johannes, P. C., Henrik, E., Diederik, A. M., Jan, J. C., Rogier, A. S., Menno, M. V., Dennis, A. H., Herold, J. M., Annelies, V., Jurriaan, E. M., Gerorgina, I. A., Eric, C. M., Sander, V. B., … Annemiek, A. V. (2021). Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19). *Nature Communications*, *12*, 267. https://doi.org/10.1038/s41467-020-20568-4

Jodoin, J., & Singer, J. (2020). Mainstreaming education for sustainable development in English as a foreign language: An analysis of the image-text interplay found in EFL textbooks in Japanese higher education. In W. L. Filho, A. L. Salvia, R. W. Pretorius, L. L. Brandli, E. Manolas, F. Alves, U. Azeiteiro, J. Rogers, C. Shiel, & A. Do Paco (Eds.), *Universities as living labs for sustainable development* (pp. 545–565). Cham: Springer.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A*, *374*, 20150202. https://doi.org/10.1098/rsta.2015.0202

Junge, A., & Jensen, L. J. (2020). CoCoScore: Context-aware co-occurrence scoring for text mining applications using distant supervision. *Bioinformatics*, *36(1)*, 264–271. https://doi.org/10.1093/bioinformatics/btz490

Koch, M., Lunde, L.-K., Gjulem, T., Knardahl, S., & Veiersted, K. B. (2016). Validity of questionnaire and representativeness of objective methods for measurements of mechanical exposures in construction and health care work. *PLoS One*, *11*, e0162881. https://doi.org/10.1371/journal.pone.0162881

Martins, T. D., Annichino-Bizzacchi, J. M., Romano, A. V. C., & Filho, R. M. (2019). Principal component analysis on recurrent venous thromboembolism. *Clinical and Applied Thrombosis/Hemostasis*, *25*, 1076029619895323. https://doi.org/10.1177/1076029619895323

Maskery, S. M., Zhang, Y., Jordan, R. M., Hu, H., Hooke, J. A., Shriver, C. D., & Liebman, M. N. (2006). Co-occurrence analysis for discovery of novel breast cancer pathology patterns. *IEEE IEEE Transactions on Information Technology in Biomedicine*, *10*(3), 497–503. https://doi.org/10.1109/titb.2005.863863

Mochizuki, Y., & Fadeeva, Z. (2010). Competences for sustainable development and sustainability: Significance and challenges for ESD. *International Journal of Sustainability in Higher Education*, *11*(4), 391–403.

Pavlopoulos, G. A., Promponas, V. J., Ouzounis, C. A., & Iliopoulos, I. (2014). Biological information extraction and co-occurrence analysis. *Methods in Molecular Biology*, *1159*, 77–92. https://doi.org/10.1007/978-1-4939-0709-0_5

Przybyła, P., Shardlow, M., Aubin, S., Bossy, R., de Castilho, R. E., Piperidis, S., McNaught, J., & Ananiadou, S. (2016). Text mining resources for the life sciences. *Database (Oxford)*, *2016*, baw145. https://doi.org/10.1093/database/baw145

Renganathan, V. (2017). Text mining in biomedical domain with emphasis on document clustering. *Healthcare Informatics Research*, *23*, 141–146. https://doi.org/10.4258/hir.2017.23.3.141

Salīte, I. (2008). Educational action research for sustainability: Constructing a vision for the future in teacher education. *Journal of Teacher Education for Sustainability*, *10*, 5–16.

Salīte, I. (2016). Searching for sustainability in teacher education and educational research: Experiences from the Baltic and Black Sea Circle Consortium for educational research. *Discourse and Communication for Sustainable Education*, *6*(1), 21–29.

Scheier, I. H., & Cattell, R. B. (1958). Confirmation of objective test factors and assessment of their relation to questionnaire factors: A factor analysis of 113 rating, questionnaire and objective test measurements of personality. *The Journal of Mental Science*, *104*, 608–624. https://doi.org/10.1192/bjp.104.436.608

Sell, S. L., Widen, S. G., Prough, D. S., & Hellmich, H. L. (2020). Principal component analysis of blood microRNA datasets facilitates diagnosis of diverse diseases. *PLoS One*, *15*, e0234185. https://doi.org/10.1371/journal.pone.0234185

Zhao, J., Li, Z., Gao, Q., Zhao, H., Chen, S., Huang, L., Wang, W., & Wang, T. (2021). A review of statistical methods for dietary pattern analysis. *Nutrition Journal*, *20*, 37. https://doi.org/10.1186/s12937-021-00692-7

Correspondence concerning this paper should be addressed to Takao Arai, Kengo Saito, Yuji Hirai, Hachioji Medical Center, Tokyo Medical University, 1163 Tatemachi, Hachiojishi, Tokyo 193-0998, Japan. Email: takao21@tokyo-med.ac.jp, ksaito@tokyo-med.ac.jp, y-hirai@tokyo-med.ac.jp; Tel.: +81-42-665-5611