Time-series Visualization of Sentence Analysis and Structural Bias in Meeting Minutes using Large Language Models

Soichiro URITA[1], Tomoya ITO[2]

[1]Graduate School of Engineering, Hachinohe Institute of Technology, [2]Hachinohe Institute of Technology

Introduction

Modern data is multidimensional and highly complex, making visualization essential for understanding structural relationships. Effective visualization can reveal hidden patterns, connections, and insights that would otherwise remain obscured. However, traditional 2D and 3D methods on conventional displays often fail to adequately represent the spatial relationships and connections between topics within such data, limiting their analytical utility.

Recent advancements in virtual reality (VR) technology provide a novel approach to addressing these challenges. VR allows for immersive and interactive visualization environments, enabling users to explore and analyze data in a more intuitive and engaging manner. This study aims to develop a VR-based system that projects topics from parliamentary proceedings into spatial representations, creating a new way to explore and understand complex political data.

The proposed system not only improves the clarity of relationships between topics but also provides a user-friendly and interactive interface for researchers, policymakers, and the general public. By making political data more accessible and engaging, this study contributes to enhancing transparency and encouraging active participation in political discourse.

Dataset Used in This Study

Source: Proceedings of the Japanese National Diet's plenary
Sessions Period: 2018 to 2023
Details: Legislator names, speech content, political party, speech dates, and gender

Method Overview

Preprocessing

- **Data Collection** → Retrieve data using APIs[1)]
- **Cleaning** → Remove symbols and unnecessary information
- **Sentence Splitting** → Use regular expressions to split speeches into individual sentences

Embeddings Conversion

- **Embeddings Conversion** → Convert each speech into embeddings using BERT[2)]

Topic Extraction

- **Dimensionality Reduction** → Perform dimensionality reduction using PCA
- **Clustering** → Classify into 11 clusters using the k-means method

Visualization

- **2D/3D** → Visualize clustering results using Python
- **VR Space** → Display data in VR space using Unity

Figure 1 : Analysis and Visualization Workflow

Results

I implemented the proposed method using Python and Unity. The results are visualized in 2D (top figure), 3D (middle figure), and VR space (bottom figure). By using a VR device, users can explore any location from any perspective.

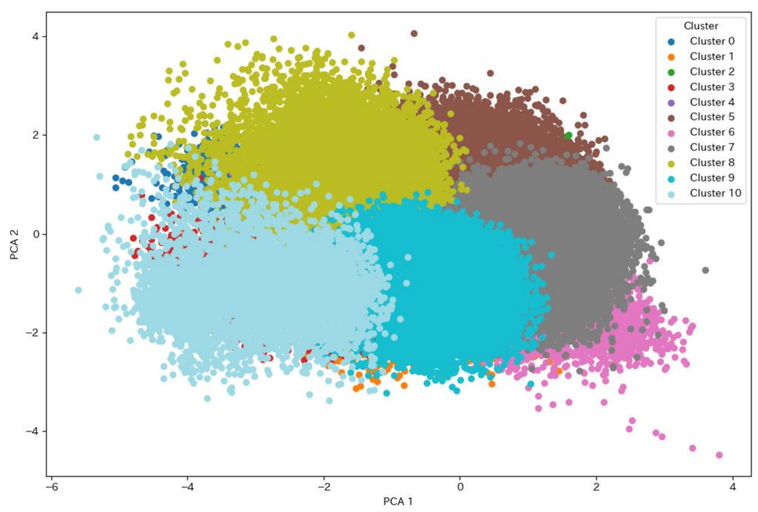

Figure 2 :Example of 2D Visualization

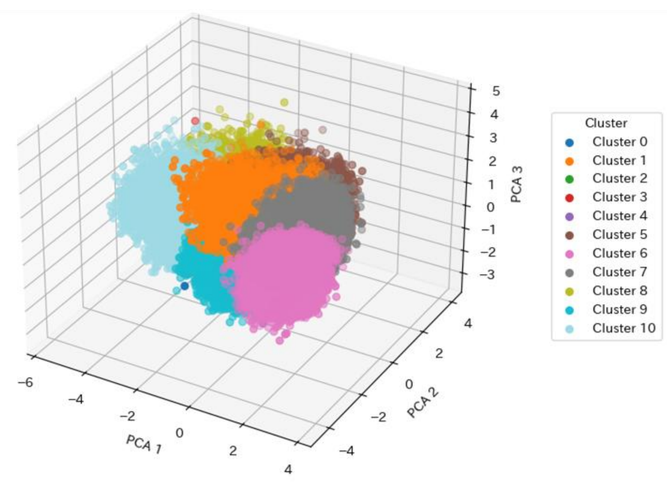

Figure 3 : Example of 3D Visualization

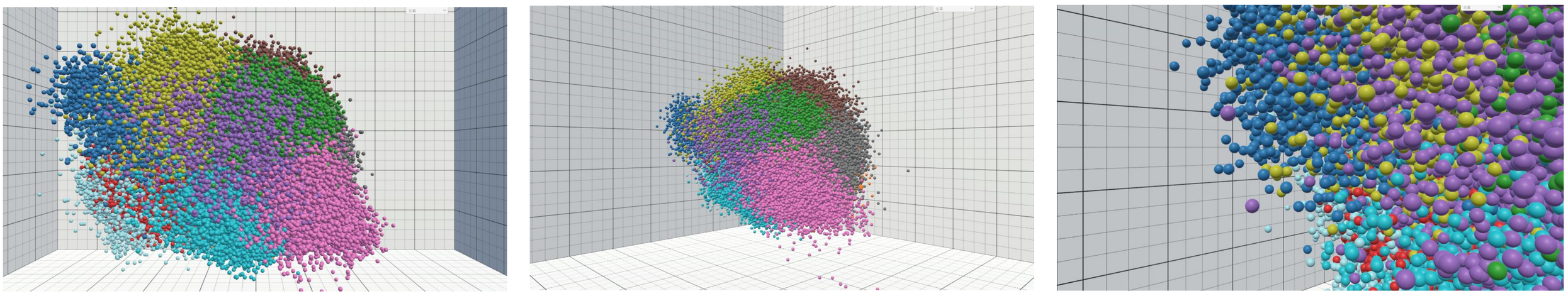

Figure 4 : Example of VR Representation

Future Work

Enhancing Data Analysis and Visualization Accuracy
Explore advanced language models, visualization techniques, and dimensionality reduction methods to enhance accuracy.

Improving VR Visualization through User Feedback
Improve VR data visualization through user feedback and intuitive UI/UX design.

Reference

1) National Diet Proceedings Search System Search APIhttps://kokkai.ndl.go.jp/api.html (accessed 2024-11-27)
2) Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.