CLUSTERING GREETING GESTURES AS CULTURAL BEHAVIOR USING A NOVEL EXPLAINABLE ARTIFICIAL INTELLIGENCE BY CONSIDERING INTERNAL FACTORS TO UNDERSTAND PATTERN CHANGES

ANGGA WAHYU WIBOWO[1], ERI SATO-SHIMOKAWARA[2], KURNIANINGSIH [3],
TAKENORI OBO[4], ICHIRO KOBAYASHI[5], YIHSIN HO[6]
[1, 2, 4] Graduate School of Systems Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino-shi, Tokyo, 191-0065, Japa
[3] Department of Electrical Engineering, Politeknik Negeri Semarang, Jl. Prof. Sudarto, Tembalang, Semarang, 50275, Indone
[5, 6] Faculty of Engineering, Takushoku University, 815-1 Tatemachi, Hachioji-shi, Tokyo, 193-0985, Japan
E-MAIL: angga-wahyu-wibowo@ed.tmu.ac.jp, eri@tmu.ac.jp, kurnianingsih@polines.ac.id,
t.obo@tmu.ac.jp, ikobaya@la.takushoku-u.ac.jp, hyihsin@es.takushoku-u.ac.jp

Abstract:

In recent years, the field of Explainable Artificial Intelligence (XAI) has rapidly advanced. However, no XAI-based clustering technique addresses shifts in interpretability and data distribution considering internal factors, such as politeness assumptions. This study aims to visualize changes in patterns of distinguishing greeting gestures while considering politeness assumptions. We propose Cluster Point Transition (Clepon), a new XAI method used to visualize the cluster patterns of greeting gestures represented as cultural behavior with data transitions. We used Indonesian and Japanese student participants for the greeting gesture data, namely eshaku, keirei, saikeirei, waving a hand, hands in front of the chest with the bow, and hands in front of the chest without the bow, using 3D skeleton angle data. We invite Indonesian and Japanese student participants to analyse politeness assumption as an internal factor. The results show that Clepon successfully visualized the transition of clustered data from raw data to each cluster and then to the center of each cluster using the DBSCAN algorithm. Our findings indicate that an individual's assumptions can influence the clustering pattern results. By integrating internal factors, Clepon provides an effective means of visualizing pattern changes in future.

Keywords:

Clustering, Explainable Artificial Intelligence, Greeting Gestures, Assumptions, and Internal Factors

1. Introduction

Artificial intelligence requires accurate and interpretable models, achieved through machine learning and Explainable Artificial Intelligence (XAI) techniques [?],[?]. XAI mainly advances in supervised learning but needs growth in unsupervised methods like clustering, where interpretability is key to understanding human behavior [?],[?],[?] by visualizing raw data to each cluster centers [?]. Several study proposed to visualize clustering transitions, such as graph-based models [?], neural visualization of evolving clusters [?], inter-cluster dynamics modeling [?], and multi-objective optimization [?]. However, human internal factors, such as individual assumptions or perspectives are often disregarded in clustering techniques [?],[?] especially students [?] to gain a deeper understanding of behavior [?]. Assumptions help distinguish behavior from background knowledge and reliably assess politeness [?],[?]. When internal factors are not considered, explanations risk being misinterpreted. It significantly affect how evaluate explanations. So, interpretability should integrates with human internal factors [?],[?].

Greeting gestures are relate to human behavior and culture, such as eshaku, keirei, and saikeirei in Japan or hands in front of the chest with the bow and hands in front of the chest without the bow in Indonesia. It has meanings such as politeness, which are interpreted through cultural background assumptions [?],[?],[?]. According to [?], this study utilizes the DBSCAN algorithm for its ability to enhance unstructured data. This study proposes Cluster Point Transition (Clepon), a novel XAI-based clustering visualization method that integrates user assumptions of politeness into greeting gestures. This study contributes to enhancing the understanding of greeting gestures as cultural behavior using Clepon.

2. Proposed Method

We collected data 3-dimensional angle using sensor [?],[?] with 15 student from Indonesia and 10 from Japan. Japanese students performed eshaku, keierei, saikeirei, and waving a hand. While Indonesian students performed hands in front of the chest with the bow and hands in front of the chest without the bow. Based on previous studies [?], we uses r_elbow_flex, elbow_flex, unrothum_r1, clav_r2, unrothum_l1, hip_adduction_l, wrist, r_wrist, knee_angle_r, clav_l2, hip_rotation_l, and clav_r3 as gesture data X_{gesture}.

We prepared 10 questionnaire questions for 16 participants from Japanese and Indonesian students, which can be seen in Table 1. We took a sample of greeting gesture data to assess politeness to participants using a likert scale from 1 means impolite, to 5 means polite. Subsequently, the mean, median, and mode were calculated and used as assumed values X_{quest}.

Based on previous studies [?], we use time shift to temporally shift the data Δt (Eq. 1), use the Fast Fourier Transform $\mathcal{F}$ and $\mathcal{F}^{-1}$ to analyze the data in the frequency domain (Eq. 2) and combine with all gesture data X_{gesture}, resulting in the gesture augmented data $X_{\text{gest\_aug}}$ (Eq. 3). We use a scaler to transform each gesture and assumption data, resulting in scaled gesture data $X_{\text{gest\_scaled}}$ (Eq. 4) and scaled assumption data $X_{\text{quest\_scaled}}$ (Eq. 5). Each assumption value corresponding to Japan and Indonesia is combined based on predefined greeting weights and questionnaire weights w_g, w_q, scaled gesture data and scaled assumption data, resulting in combined data X_{combined}, respectively (Eq. 6), Then we processed clustering.

$$X_{\text{shifted}}(t) = X_{\text{gesture}} + \Delta t \tag{1}$$

$$X_{\text{fft}} = \mathcal{F}^{-1}\left(\mathcal{F}\left(X_{\text{gesture}}\right) + \epsilon\right) \tag{2}$$

$$X_{\text{gest\_aug}} = \left[X_{\text{gesture}}\ X_{\text{shifted}}\ X_{\text{fft}}\right] \tag{3}$$

$$X_{\text{gest\_scaled}} = \frac{X_{\text{gest\_aug}} - \mu_{\text{gest}}}{\sigma_{\text{gest}}} \tag{4}$$

$$X_{\text{quest\_scaled}} = \frac{X_{\text{quest}} - \mu_{\text{quest}}}{\sigma_{\text{quest}}} \tag{5}$$

$$X_{\text{combined}} = \left[w_g \cdot X_{\text{gest\_scaled}}\ ;\ \middle|\ ;w_q \cdot X_{\text{quest\_scaled}}\right] \tag{6}$$

TABLE 1. Questionnaire Distribution.

Question	Answer
In your opinion, what is the level of politeness of (greeting gestures) when conducted in Indonesia as a greeting gesture?	An answer scale of 1 assumes impolite to 5 assumes polite
In your opinion, what is the level of politeness of (greeting gestures) when conducted in Japan as a greeting gesture?	An answer scale of 1 assumes impolite to 5 assumes polite
In your opinion, what is the politeness level of (greeting gestures) when conducted with younger people in Japan as a greeting gesture?	An answer scale of 1 assumes impolite to 5 assumes polite
In your opinion, what is the level of politeness of (greeting gestures) when conducted with older people in Japan as a greeting gesture?	An answer scale of 1 assumes impolite to 5 assumes polite
In your opinion, what is the level of politeness of (greeting gestures) when conducted with younger people in Indonesia as a greeting gesture?	An answer scale of 1 assumes impolite to 5 assumes polite
In your opinion, what is the level of politeness of (greeting gestures) when conducted with older people in Indonesia as a greeting gesture?	An answer scale of 1 assumes impolite to 5 assumes polite
In your opinion, can (greeting gestures) be applied in Indonesia for greeting gestures?	An answer scale of 1 assumes disagree to 5 assumes strongly agree
In your opinion, can (greeting gestures) be applied with an intelligent system in Japan for greeting gestures?	An answer scale of 1 assumes disagree to 5 assumes strongly agree
In your opinion, can (greeting gestures) be implemented with a smart system in Indonesia as a greeting gesture?	An answer scale of 1 assumes disagree to 5 assumes strongly agree
How big is the influence of using (greeting gestures) to determine the level of politeness using an intelligent system for greeting gestures?	An answer scale of 1 assumes not big to 5 assumes big

As described in Algorithm 1, The algorithm begins by initialized the value of spread_distance, which is used to

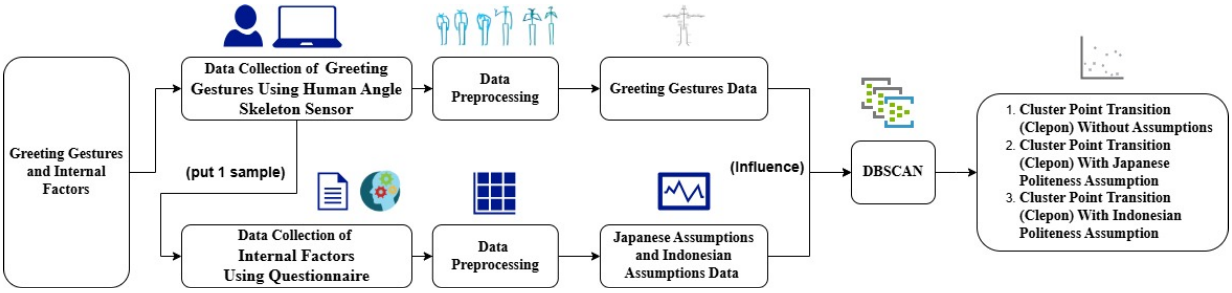


FIGURE 1. Proposed Method

shift the cluster center from the global center. The global data center g is computed as the mean of all points in X, which is used to calculate the global center of all UMAP reduction result points. $X_{\text{shifted\_umap}}$ used to make a copy of the 3D data from UMAP so that the shift is performed without changing the original data. Then, for each cluster ID c in the set of valid cluster labels C, a mask $M = (y = c)$ is created to identify the points belonging to cluster c, where y contains the cluster assignments for all data points. The local cluster center m is computed as the mean of points in X indexed by M. A vector v pointing from the global center g to the cluster center m is then calculated, normalized, and scaled by a predefined spread distance d. This vector v is used to shift all points in cluster c, resulting in an updated spatial configuration stored in $X_{\text{shifted\_umap}}$. If the magnitude of v is zero, no shifting is applied to avoid division by zero. This technique preserves the internal structure of each cluster while improving inter-cluster separation, thereby facilitating more interpretable 3D visualizations.

To produce a transition animation that illustrates the process of mapping data to a customized 3D space, we initialize the initial positions of the points randomly within the minimum and maximum bounds of each axis from the mapping results. The bounds are computed from the shifted 3D data X_{shifted}, which includes the final positions of the points after they have been clustered. The random start positions, denoted as R, are sampled from a uniform distribution within these bounds, specifically in the ranges $[x_{\min}, x_{\max}], [y_{\min}, y_{\max}], [z_{\min}, z_{\max}]$. Then, linear interpolation is performed for 30 steps as n_{step} between the initial random positions and the final positions of each point in the clustered data. This interpolation is calculated using the parameter α, which varies linearly from 0 to 1 as we progress through each step. The positions at each step are stored in P, where $P[f, i]$ represents the

Algorithm 1 Cluster Point Transition (Clepon)

```
set spread_distance
g ← mean(X)
X_shifted_umap ← umap3D
for each cluster ID c ∈ C do
    M ← (y = c)
    m ← mean(X[M])
    v ← m − g
    if ‖v‖ > 0 then
        v ← v/‖v‖ · d
    else
        v ← 0
    end if
    X_shifted_umap[M] ← X[M] + v
end for
n_step ← 30
Compute bounds: x_min, x_max, y_min, y_max, z_min, z_max from X_shifted
R ← UniformRandom([x_min, x_max] × [y_min, y_max] × [z_min, z_max], N)
Initialize P ← zeros(n_step, N, 3)
for i = 1 to N do
    for f = 1 to n_step do
        α ← (f − 1)/(n_step − 1)
        P[f, i] ← (1 − α) · R[i] + α · X_shifted[i]
    end for
end for
```

position of the i-th point at the f-th step. This smooth interpolation forms a gradual transition from the initial random positions to the final clustered distribution, allowing for a clear visualization of the data's transition. This makes the final cluster structure easier to grasp. An illustration of the proposed method can be seen in the Figure 1.

3 Results

Figure 2 shows the results of clustering point transition. The three steps of the visual process are raw data on the left, the clustering process in the middle, and the final clustering results on the right. Figure 2a showed that the Clepon successfully separated the data into six physically distinct clusters, each of which was concentrated and non-overlapping. While each cluster is identified and represented by a distinct hue, the centroid, or cluster center, is denoted by a red star symbol. The clear separation of the groups indicates that the clustering algorithm can successfully distinguish between the relatively diverse properties of each gesture. Furthermore, no outliers were found in the clusters, suggesting it was completed successfully. It is reasonable to believe that several closely spaced clusters indicate gestures similar in appearance or spatial arrangement, such as those between keirei and saikeirei.

Figure 2b shows the final result of the greeting gesture data clustering process considering Japanese assumptions. The results shown show that the clusters formed are very clearly and spatially separated, reflecting that the features used in this process have successfully identified unique patterns in each type of gesture. It can be attributed to the characteristics of Japanese culture that emphasizes consistency and uniformity in culture, including greeting gestures. Dense clusters such as Cluster 1 and Cluster 4 are likely to represent very formal gestures such as keirei or saikeirei, while slightly more dispersed clusters such as Cluster 5 and Cluster 6 could indicate more informal or flexible gestures such as eshaku or waving a hand. The absence of outliers indicates that the data pre-processing process was carried out effectively, ensuring that all data can be cluster appropriately. Overall, these results indicate that the developed model can capture transitions by Japanese assumptions, which is important in the context of cultural understanding.

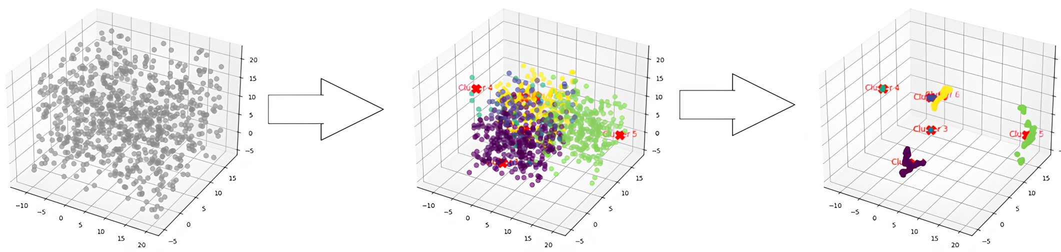
Figure 2c shows the final result of the clustering process considering Indonesian assumptions. The resulting clusters show a very clear and compact spatial separation, with each cluster tightly clustered without any deviating points. This indicates that the features used in the clustering process are very effective in representing the unique characteristics of each gesture. Some clusters, such as Cluster 1 and Cluster 2, appear very dense, reflecting consistent gestures, while other clusters, such as Cluster 4 and Cluster 5, are more scattered, indicating a wider variety of gestures. When compared to previous results that did not consider assumptions, these results show an increase in the separation and clarity of the cluster structure. It means that assumptions can improve the effectiveness of unsupervised learning-based gesture recognition systems.

4. Conclusions

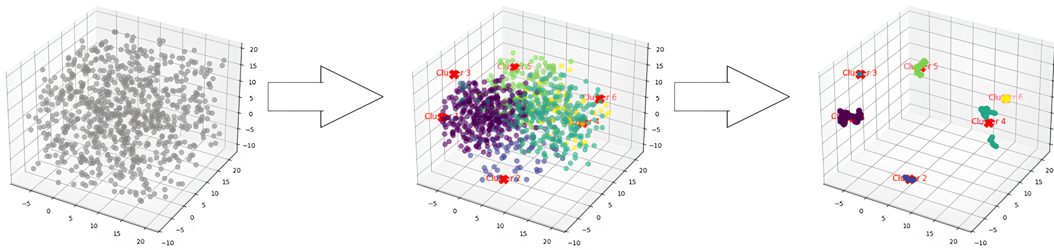
This study presents Clepon, a new XAI technique for transition clustering that works well for examining and displaying cultural greeting gesture variations. Clepon offers an understandable illustration of how discrete clusters emerge during these transitions by fusing transition-based features and illustrating the changes between data points. This study finds that the Clepon can effectively cluster different kinds of gestures according to movement data point and can provide a solid basis for analysis pattern change based on unsupervised learning. By comprehending the variations among clusters, such as Japanese and Indonesian assumptions, we can identify that assumptions can influence the pattern. In the future, Clepon presents new opportunities for the development of XAI in human cultural behavior, enabling the observation and analysis of shifts in behavioral patterns.

Acknowledgements

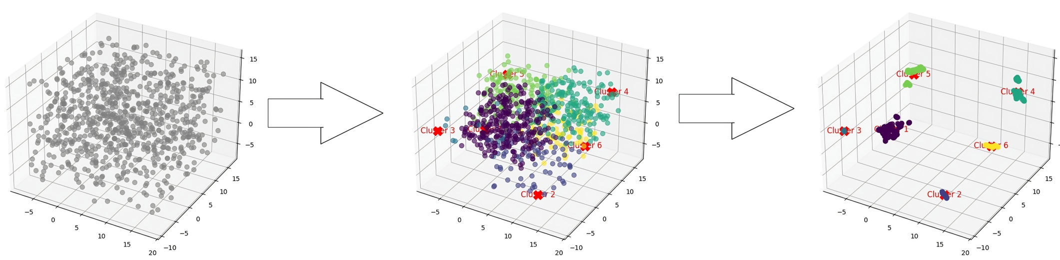
This work was supported by JSPS KAKENHI, Fostering Joint International Research (B), Grant Number JP22KK0162.



(a) Clustering Point Transition From Raw Data to Cluster Center Without Assumptions



(b) Clustering Point Transition From Raw Data to Cluster Center Using Japanese Assumption



(c) Clustering Point Transition From Raw Data to Cluster Center Using Indonesian Assumption

FIGURE 2. Clustering Point Transitions Results