Non-Hierarchical Clustering for Distribution-Valued Data

Yoshikazu Terada
Graduate School of Culture and Information Science, Doshisha University.

Hiroshi Yadohisa
Department of Culture and Information Science, Doshisha University.
• Introduction

• Previous dissimilarity measures and clustering for distribution-valued data

• Centroid distribution

• Non-hierarchical clustering

• Applying our method for the weather data

• Conclusion
1.1 Symbolic Data Analysis (SDA)

- In recent years,
  - Development of the Internet
  - Improvement of computer performance
1.1 Symbolic Data Analysis (SDA)

- In recent years,
  - Development of the Internet
  - Improvement of computer performance
  - We deal with "Large data" and "more Complex information".
1.1 Symbolic Data Analysis (SDA)

• In recent years,
  – Development of the Internet
  – Improvement of computer performance
  – We deal with “Large data” and “more Complex information”.

In some cases, Difficult to analyze them by using classical methods.

In some cases, Difficult to describe them by classical data.
1.1 Symbolic Data Analysis (SDA)

• In recent years,
  – Development of the Internet
  – Improvement of computer performance
  – We deal with "Large data" and "more Complex information".

In some cases,
Difficult to analyze them by using classical methods.

In some cases,
Difficult to describe them by classical data.

Some new methods for analyzing them are required.
1.1 Symbolic Data Analysis (SDA)

- In recent years,
  - Development of the Internet
  - Improvement of computer performance
  - We deal with "**Large data**" and "**more Complex information**".

Some new methods for analyzing them are required.

"**Symbolic data analysis**"
1.1 Symbolic Data Analysis (SDA)

- **Symbolic data analysis (SDA)**
  - A extended classical data analysis for more complex data table called “symbolic data table”
1.1 Symbolic Data Analysis (SDA)

- **Symbolic data analysis (SDA)**
  - A extended classical data analysis for more complex data table called “symbolic data table”

  - A more complex data table
  - A cell of that cannot only contain a single quantitative (categorical) value.
1.1 Symbolic Data Analysis (SDA)

- **Symbolic data analysis (SDA)**
  - A extended classical data analysis for more complex data table called “symbolic data table”
  - SDA has been studied as one of useful methods for analyzing large and complex datasets.
1.1 Symbolic Data Analysis (SDA)

- **Symbolic data analysis (SDA)**
  - A extended classical data analysis for more complex data table called “symbolic data table”

- SDA has been studied as one of useful methods for analyzing large and complex datasets.

- **Typical Symbolic data**
  - Interval-valued data
1.1 Symbolic Data Analysis (SDA)

- **Symbolic data analysis (SDA)**
  - A extended classical data analysis for more complex data table called “symbolic data table”

- SDA has been studied as one of useful methods for analyzing large and complex datasets.

- **Typical Symbolic data**
  - Interval-valued data
  - Distribution-valued data
1.1 Symbolic Data Analysis (SDA)

• **Symbolic data analysis (SDA)**
  – A *extended classical data analysis* for more complex data table called “*symbolic data table*”

– SDA has been studied as one of useful methods for analyzing **large** and **complex** datasets.

• **Typical Symbolic data**
  – Interval-valued data
  – **Distribution-valued data**
1.2 Distribution-valued data

- **What is Distribution-valued data?**
  - A cell of such data contains a “distribution”.
    - e.g.) distribution function, density function (histogram) …
1.2 Distribution-valued data

- **What is Distribution-valued data?**
  
  A cell of such data contains a "distribution".

  e.g.) distribution function, density function (histogram) ...

Classical data

\[
X = \begin{bmatrix}
x_{11} & \cdots & x_{1k} & \cdots & x_{1p} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
x_{i1} & \cdots & x_{ik} & \cdots & x_{ip} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
x_{n1} & \cdots & x_{nk} & \cdots & x_{np}
\end{bmatrix}
\]

Variables

Entity
1.2 Distribution-valued data

• What is Distribution-valued data?
  – A cell of such data contains a “distribution”.

  e.g.) distribution function, density function (histogram) ...

Classical data

\[ X = \begin{bmatrix}
  x_{11} & \cdots & x_{1k} & \cdots & x_{1p} \\
  \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_{i1} & \cdots & x_{ik} & \cdots & x_{ip} \\
  \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_{n1} & \cdots & x_{nk} & \cdots & x_{np}
\end{bmatrix} \]

Represented by a single point in \( \mathbb{R}^p \)
1.2 Distribution-valued data

**What is Distribution-valued data?**

- A cell of such data contains a “distribution”.

E.g., distribution function, density function (histogram) ...

---

**Classical data**

\[ X = \begin{bmatrix} x_{i1} & \cdots & x_{ik} & \cdots & x_{ip} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix} \]

Represented by a single point in \( \mathbb{R}^p \)

---

**Distribution-valued data**

\[ X = \begin{bmatrix} x_{i1} & \cdots & x_{1k} & \cdots & x_{1p} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix} \]
1.2 Distribution-valued data

- What is Distribution-valued data?
  - A cell of such data contains a “distribution”.

  e.g.) distribution function, density function (histogram) ...

Classical data

- Represented by a single point in $\mathbb{R}^p$

Distribution-valued data

- e.g.) Histogram, distribution function...

Variables

\[ X = \begin{bmatrix} x_{11} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{i1} & \cdots & x_{ik} & \cdots & x_{ip} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix} \]
1.2 Distribution-valued data

• **When we use distribution-valued data?**
  – Aggregate Large data to more manageable data
  – Describe objects with several values on a one variable
1.2 Distribution-valued data

- **When we use distribution-valued data?**
  - Aggregate large data to more manageable data
  - Describe objects with several values on a one variable

- **Advantages of distribution-valued data**

\[
X = \begin{bmatrix}
  x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\
  x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \\
  x_{i1} & x_{i2} & \cdots & x_{ik} & \cdots & x_{ip} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_{N1} & x_{N2} & \cdots & x_{Nk} & \cdots & x_{Np}
\end{bmatrix}
\]

\[
X^* = \begin{bmatrix}
  \bar{x}_{11} & \bar{x}_{1k} & \cdots & \bar{x}_{1p} \\
  \vdots & \vdots & \ddots & \vdots \\
  \bar{x}_{i1} & \bar{x}_{ik} & \cdots & \bar{x}_{ip} \\
  \vdots & \vdots & \ddots & \vdots \\
  \bar{x}_{n1} & \bar{x}_{nk} & \cdots & \bar{x}_{np}
\end{bmatrix}
\]
1.2 Distribution-valued data

- **When we use distribution-valued data?**
  - Aggregate Large data to more manageable data
  - Describe objects with several values on a one variable

- **Advantages of distribution-valued data**

Initial data

\[ X = \begin{bmatrix}
  x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\
  x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \\
  x_{i1} & x_{i2} & \cdots & x_{ik} & \cdots & x_{ip} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_{N1} & x_{N2} & \cdots & x_{Nk} & \cdots & x_{Np} \\
\end{bmatrix} \]

Classical data

\[ X^* = \begin{bmatrix}
  \bar{x}_{11} & \cdots & \bar{x}_{1k} & \cdots & \bar{x}_{1p} \\
  \vdots & \ddots & \vdots & \ddots & \vdots \\
  \bar{x}_{i1} & \cdots & \bar{x}_{ik} & \cdots & \bar{x}_{ip} \\
  \vdots & \ddots & \vdots & \ddots & \vdots \\
  \bar{x}_{n1} & \cdots & \bar{x}_{nk} & \cdots & \bar{x}_{np} \\
\end{bmatrix} \]

Describe by a single representative value (e.g. mean)
1.2 Distribution-valued data

- **When we use distribution-valued data?**
  - Aggregate Large data to more manageable data
  - Describe objects with several values on a one variable

- **Advantages of distribution-valued data**

![Diagram showing initial data and classical data with aggregate values](image)
1.2 Distribution-valued data

- **When we use distribution-valued data?**
  - Aggregate Large data to more manageable data
  - Describe objects with several values on a one variable

- **Advantages of distribution-valued data**

  ![Diagram](#)

  **Initial data**
  \[
  \mathbf{X} = \begin{bmatrix}
  x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\
  x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_{N1} & x_{N2} & \cdots & x_{Nk} & \cdots & x_{Np}
  \end{bmatrix}
  \]

  **Classical data**
  \[
  \mathbf{X^*} = \begin{bmatrix}
  \bar{x}_{11} & \cdots & \bar{x}_{1k} & \cdots & \bar{x}_{1p} \\
  \bar{x}_{i1} & \cdots & \bar{x}_{ik} & \cdots & \bar{x}_{ip} \\
  \vdots & \ddots & \vdots & \ddots & \vdots \\
  \bar{x}_{n1} & \cdots & \bar{x}_{nk} & \cdots & \bar{x}_{np}
  \end{bmatrix}
  \]

  Describe by a **single** representative value (e.g. mean)

  **Same mean**
1.2 Distribution-valued data

• When we use distribution-valued data?
  – Aggregate Large data to more manageable data
  – Describe objects with several values on a one variable

• Advantages of distribution-valued data

Initial data

\[
X = \begin{bmatrix}
  x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\
  x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_{n1} & x_{n1} & \cdots & x_{n1k} & \cdots & x_{np} \\
  x_{i1} & x_{i2} & \cdots & x_{ik} & \cdots & x_{ip} \\
  x_{N1} & x_{N2} & \cdots & x_{Nk} & \cdots & x_{Np} \\
\end{bmatrix}
\]

Classical data

\[
X^* = \begin{bmatrix}
  \bar{x}_{11} & \bar{x}_{1k} & \cdots & \bar{x}_{1p} \\
  \bar{x}_{i1} & \bar{x}_{ik} & \cdots & \bar{x}_{ip} \\
  \bar{x}_{n1} & \bar{x}_{nk} & \cdots & \bar{x}_{np} \\
\end{bmatrix}
\]

Describe by a **single** representative value (e.g. mean)

- **Same mean**
- **Information loss of Distribution structure**
1.2 Distribution-valued data

• When we use distribution-valued data?
  – Aggregate large data to more manageable data
  – Describe objects with several values on a one variable

• Advantages of distribution-valued data

\[
X = \begin{bmatrix}
  x_{11} & x_{12} & \ldots & x_{1k} & \ldots & x_{1p} \\
  x_{21} & x_{22} & \ldots & x_{2k} & \ldots & x_{2p} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_{n1} & x_{n2} & \ldots & x_{nk} & \ldots & x_{np} \\
  x_{i1} & x_{i2} & \ldots & x_{ik} & \ldots & x_{ip} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_{N1} & x_{N2} & \ldots & x_{Nk} & \ldots & x_{Np}
\end{bmatrix}
\]

\[
X_d = \begin{bmatrix}
  \xi_{11} & \ldots & \xi_{1k} & \ldots & \xi_{1p} \\
  \vdots & \ldots & \vdots & \ddots & \vdots \\
  \xi_{i1} & \ldots & \xi_{ik} & \ldots & \xi_{ip} \\
  \vdots & \ldots & \vdots & \ddots & \vdots \\
  \xi_{n1} & \ldots & \xi_{nk} & \ldots & \xi_{np}
\end{bmatrix}
\]
1.2 Distribution-valued data

- **When we use distribution-valued data?**
  - Aggregate large data to more manageable data
  - Describe objects with several values on a one variable

- **Advantages of distribution-valued data**
1.2 Distribution-valued data

• **When we use distribution-valued data?**
  - Aggregate Large data to more manageable data
  - Describe objects with several values on a one variable

• **Advantages of distribution-valued data**
1.2 Distribution-valued data

- **When we use distribution-valued data?**
  - Aggregate Large data to more manageable data
  - Describe objects with several values on a one variable

- **Advantages of distribution-valued data**

We can consider the relationship between variables.
2.1 Dissimilarity measures for distribution-valued data

- $P$, $Q$ : a probability distribution, respectively
- $p$, $q$ : a density function of $P$, $Q$, respectively
- **Dissimilarity measures for density functions**
  - Kullback-Leibler divergence
    
    - Kullback-Leibler information: $I(P \mid Q) = \int \log \left\{ \frac{p(x)}{q(x)} \right\} q(x) \, dx$
    - Kullback-Leibler divergence: $J(P, Q) = I(Q \mid P) + I(P \mid Q)$
  - Minkowski’s $L^2$ distance (Bock and Diday, 2000)
    
    $d_2(P, Q) = \int (p(x) - q(x))^2 \, dx$
2.1 Dissimilarity measures for distribution-valued data

- $P, Q$: a probability distribution, respectively
- $p, q$: a density function of $P, Q$, respectively

**Dissimilarity measures for density functions**

- **Kullback-Leibler divergence**
  
  - Kullback-Leibler information: $I(P \mid Q) = \int \log \left( \frac{p(x)}{q(x)} \right) q(x) \, dx$
  
  - Kullback-Leibler divergence: $J(P, Q) = I(Q \mid P) + I(P \mid Q)$

- **Minkowski’s $L^2$ distance** (Bock and Diday, 2000)

  \[
  d_2(P, Q) = \int (p(x) - q(x))^2 \, dx
  \]
2.1 Dissimilarity measures for distribution-valued data

- \( P, Q \) : a probability distribution, respectively
- \( p, q \) : a density function of \( P, Q \), respectively

**Dissimilarity measures for density functions**

- **Kullback-Leibler divergence**
  
  ➢ Kullback-Leibler information: 
  
  \[ I(P \mid Q) = \int \log \left( \frac{p(x)}{q(x)} \right) q(x) \, dx \]

  ➢ Kullback-Leibler divergence: 
  
  \[ J(P, Q) = I(Q \mid P) + I(P \mid Q) \]

- **Minkowski’s \( L^2 \) distance** (Bock and Diday, 2000)

  \[ d_2(P, Q) = \int (p(x) - q(x))^2 \, dx \]
2.1 Dissimilarity measures for distribution-valued data

- \( P, Q \): a probability distribution, respectively
- \( F, G \): a distribution function of \( P, Q \), respectively
- **Dissimilarity measures for distribution functions**
  - **Wasserstein metric**
    \[
    d_W(P, Q) = \int |F(x) - G(x)| \, dx = \int_0^1 |F^{-1}(t) - G^{-1}(t)| \, dt
    \]
  - **Mallow’s distance**
    \[
    d_M(P, Q) = \sqrt{\int_0^1 |F^{-1}(x) - G^{-1}(x)|^2 \, dx}
    \]
2.1 Dissimilarity measures for distribution-valued data

- $P, Q$: a probability distribution, respectively
- $F, G$: a distribution function of $P, Q$, respectively
- **Dissimilarity measures for distribution functions**
  - Wasserstein metric
    \[
    d_W(P, Q) = \int |F(x) - G(x)| \, dx = \int_0^1 |F^{-1}(t) - G^{-1}(t)| \, dt
    \]
  - Mallow’s distance
    \[
    d_M(P, Q) = \sqrt{\int_0^1 |F^{-1}(x) - G^{-1}(x)|^2 \, dx}
    \]
2.1 Dissimilarity measures for distribution-valued data

- \( P, Q \): a probability distribution, respectively
- \( F, G \): a distribution function of \( P, Q \), respectively
- **Dissimilarity measures for histogram-valued data**

2.2 Clustering for distribution-valued data

• Irpino and Verde (2006)
  – The hierarchical clustering (Ward’s method) of histogram-valued data using the Wasserstein metric for histogram-valued data

• Irpino et al. (2006)
  – The dynamic clustering of histogram-valued data using the Wasserstein metric for histogram-valued data

• Verde and Irpino (2008)
  – Applying the Mahalanobis–Wasserstein distance for histogram–valued data to the dynamic clustering
2.2 Clustering for distribution-valued data

- **De Souza et al. (2007)**
  - Dynamic clustering methods for **mixed feature-type symbolic data**
  - Pre-processing step for **transforming** mixed feature-type symbolic data **into modal symbolic data**
  - Performing the clustering for
    \[ \xi = \{(\eta_i, w_i) \mid i = 1, \ldots, n\} \]
  - **the transformed data by using the weight vectors**

- **De Carvalho and De Souza (2010)**
  - Unsupervised pattern recognition methods for mixed feature-type symbolic data **using adaptive distances**
2.2 Clustering for distribution-valued data

- **De Souza et al. (2007)**
  - *Dynamic clustering methods* for *mixed feature-type symbolic data*
  - Pre-processing step for *transforming* mixed feature-type symbolic data *into modal symbolic data*
  - Performing the clustering for the transformed data by using the *weight vectors* $\xi = \{ (\eta_i, w_i) \mid i = 1, \ldots, n \}$

- **De Carvalho and De Souza (2010)**
  - Unsupervised pattern recognition methods for mixed feature-type symbolic data *using adaptive distances*
2.2 Clustering for distribution-valued data

Here,

- Define **the centroid of a set of distributions**
- Propose a **non-hierarchical clustering** ($k$-means) method for **more general distribution-valued data** by using the centroid distribution
2.2 Clustering for distribution-valued data

Here,

- Define **the centroid of a set of distributions**
- Propose a **non-hierarchical clustering** ($k$-means) method for **more general distribution-valued data** by using the centroid distribution.

Represented as joint (or marginal) **distribution function** or **density function**.
3.1 Definition of Centroid distribution

• First, we consider the centroid for distribution-valued data.
  
  – $\mathcal{P}$: a set of distributions
  
  – $d$: a dissimilarity measure on $\mathcal{P}$
  
  – $P_i \ (i = 1, 2, \ldots, n)$: elements of $\mathcal{P}$

• Definition of Centroid distribution

We assume $\mathcal{Q} = \{Q \in \mathcal{P} \mid d(P_i, Q) < \infty \ (i = 1, 2, \ldots, n)\} \neq \emptyset$ and define the centroid distribution $P_C$ of distributions $P_i$, satisfying

$$
\sum_{i=1}^{n} d^2(P_i, P_C) = \inf_{Q \in \mathcal{P}} \sum_{i=1}^{n} d^2(P_i, Q).
$$
3.1 Definition of Centroid distribution

- First, we consider the centroid for distribution-valued data.
  - $\mathcal{P}$: a set of distributions
  - $d$: a dissimilarity measure on $\mathcal{P}$
  - $P_i (i = 1, 2, \ldots, n)$: elements of $\mathcal{P}$

- **Definition of Centroid distribution**

  We assume $\mathcal{Q} = \{Q \in \mathcal{P} \mid d(P_i, Q) < \infty (i = 1, 2, \ldots, n)\} \neq \emptyset$ and define the **centroid distribution** $P_C$ of distributions $P_i$, satisfying

  $$\sum_{i=1}^{n} d^2(P_i, P_C) = \inf_{Q \in \mathcal{P}} \sum_{i=1}^{n} d^2(P_i, Q).$$
3.2 Calculation of Centroid distribution

Here, we deal with Minkowski’s $L^2$ distance for distribution functions (or density functions).
- $P, Q$: a probability distribution, respectively
- $p, q$: a density function of $P, Q$, respectively
- $F, G$: a distribution function of $P, Q$, respectively

\[
d_C(P, Q) = \sqrt{\int (F(x) - G(x))^2 \, dx}
\]

(or \[
d_D(P, Q) = \sqrt{\int (p(x) - q(x))^2 \, dx}
\]
3.2 Calculation of Centroid distribution

- Here, we deal with Minkowski’s $L^2$ distance for distribution functions (or density functions).

- When we consider marginal distributions,
  - $P, Q$: a distribution that has marginal distribution $P_j, Q_j$, respectively
  - $p_j, q_j$: a density function of $P, Q$, respectively
  - $F_j, G_j$: a distribution function of $P, Q$, respectively

\[
d_C(P, Q) = \sqrt{\sum_{j=1}^{r} \int (F_j(x) - G_j(x))^2 \, dx} \quad \text{or} \quad d_D(P, Q) = \sqrt{\sum_{j=1}^{r} \int (p_j(x) - q_j(x))^2 \, dx}
\]
3.2 Calculation of Centroid distribution

- **The centroid distribution with** $d_C$ (or $d_D$)
  - $\mathcal{P}_r$ : a set of (continuous) distributions on $\mathbb{R}^r$
  - $P_i (i = 1, 2, \ldots, n)$ : elements of $\mathcal{P}_r$
  - $F_i$ : a distribution function of $P_i (i = 1, 2, \ldots, n)$

\[
\text{If } \mathcal{Q} = \{ Q \in \mathcal{P} \mid d_C(P_i, Q) < \infty (i = 1, 2, \ldots, n) \} \neq \emptyset, \text{ then the centroid distribution } P_C \text{ of } P_i \text{ is given by the distribution that has the distribution function satisfying}
\]

\[
F_C(x) = \frac{1}{n} \sum_{i=1}^{n} F_i(x) \quad (\forall x \in \mathbb{R}^r).
\]
3.2 Calculation of Centroid distribution

• The centroid distribution with $d_C$ (or $d_D$)

If $\mathcal{Q} = \{ Q \in \mathcal{P} \mid d_C(P_i, Q) < \infty \ (i = 1, 2, \ldots, n) \} \neq \emptyset$, then the centroid distribution $P_C$ of $P_i$ is given by the distribution that has the distribution function satisfying

$$F_C(x) = \frac{1}{n} \sum_{i=1}^{n} F_i(x) \quad (\forall x \in \mathbb{R}^r).$$
3.2 Calculation of Centroid distribution

- **The centroid distribution with** $d_C$ (or $d_D$)

If $\mathcal{Q} = \{ Q \in \mathcal{P} \mid d_C(P_i, Q) < \infty (i = 1, 2, \ldots, n) \} \neq \emptyset$, then the centroid distribution $P_C$ of $P_i$ is given by the distribution that has the distribution function satisfying

$$F_C(x) = \frac{1}{n} \sum_{i=1}^{n} F_i(x) \quad (\forall x \in \mathbb{R}^r).$$
4.1 Objective function for clustering

• We propose "a non-hierarchical clustering method" using dissimilarity $d_C$ (or $d_D$) and centroid distribution $P_C$. 

4.1 Objective function for clustering

• We propose

  “a non-hierarchical clustering method” using dissimilarity $d_C (\text{or } d_D)$ and centroid distribution $P_C$.

• Objective function for the clustering

  – $P_j$ ($j = 1, 2, \ldots, n$): distributions
  – $k$: the number of cluster
  – $C_i$ ($i = 1, \ldots, k$): Clusters constructed by $P_j$

  
  \[
  Q_C = \sum_{i=1}^{k} \sum_{j \in C_i} d_C^2(P_j, P_{C_i}) \quad \left(\text{or } Q_D = \sum_{i=1}^{k} \sum_{j \in C_i} d_D^2(P_j, P_{C_i})\right).
  \]
4.1 Objective function for clustering

- We propose "a non-hierarchical clustering method" using dissimilarity $d_C$ (or $d_D$) and centroid distribution $P_C$.

- Objective function for the clustering
  
  - $P_j \, (j = 1, 2, \ldots, n)$: distributions
  - $k$: the number of cluster
  - $C_i \, (i = 1, \ldots, k)$: Clusters constructed by $P_j$

\[
Q_C = \sum_{i=1}^{k} \sum_{j \in C_i} d_C^2(P_j, P_{C_i}) \quad \text{(or } Q_D = \sum_{i=1}^{k} \sum_{j \in C_i} d_D^2(P_j, P_{C_i}) \text{)}.
\]
4.1 Objective function for clustering

• We propose "a non-hierarchical clustering method" using dissimilarity $d_C$ (or $d_D$) and centroid distribution $P_C$.

• Objective function for the clustering
  
  $Q_C = \sum_{i=1}^{k} \sum_{j \in C_i} d_C^2(P_j, P_{C_i})$ (or $Q_D = \sum_{i=1}^{k} \sum_{j \in C_i} d_D^2(P_j, P_{C_i})$).

  - $P_j$ ($j = 1, 2, \ldots, n$): distributions
  - $k$: the number of cluster
  - $C_i$ ($i = 1, \ldots, k$): Clusters constructed by $P_j$
4.2 Non-hierarchical clustering algorithm

• **Clustering algorithm (k-means)**

**Step 1:** Initial seeds \( P_{C_j} \) \((j = 1, 2, \ldots, k)\) are appropriately determined from the objects \( P_i \) \((i = 1, 2, \ldots, n)\) described by distributions (e.g. by using random numbers).
4.2 Non-hierarchical clustering algorithm

- **Clustering algorithm (k-means)**

  **Step 1:** Initial seeds $P_{C_j}$ ($j = 1, 2, \ldots, k$) are appropriately determined from the objects $P_i$ ($i = 1, 2, \ldots, n$) described by distributions (e.g. by using random numbers).

  **Step 2:** Dissimilarity $d_C(P_i, P_{C_j})$ (or $d_D(P_i, P_{C_j})$) from seed $P_{C_j}$ to object $P_i$ is evaluated for $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, k$. 
4.2 Non-hierarchical clustering algorithm

- **Clustering algorithm (k-means)**

**Step 1:** Initial seeds $P_{C_j} (j = 1, 2, ..., k)$ are appropriately determined from the objects $P_i (i = 1, 2, ..., n)$ described by distributions (e.g. by using random numbers).

**Step 2:** Dissimilarity $d_C(P_i, P_{C_j})$ (or $d_D(P_i, P_{C_j})$) from seed $P_{C_j}$ to object $P_i$ is evaluated for $i = 1, 2, ..., n; j = 1, 2, ..., k$.

**Step 3:** The centroid distribution $P_{C_j}$ of each cluster $C_j (j = 1, 2, ..., k)$ is decided as a new seed.
4.2 Non-hierarchical clustering algorithm

- **Clustering algorithm (k-means)**

  **Step 1:** Initial seeds $P_{C_j}$ ($j = 1, 2, \ldots, k$) are appropriately determined from the objects $P_i$ ($i = 1, 2, \ldots, n$) described by distributions (e.g. by using random numbers).

  **Step 2:** Dissimilarity $d_C(P_i, P_{C_j})$ (or $d_D(P_i, P_{C_j})$) from seed $P_{C_j}$ to object $P_i$ is evaluated for $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, k$.

  **Step 3:** The centroid distribution $P_{C_j}$ of each cluster $C_j$ ($j = 1, 2, \ldots, k$) is decided as a new seed.

  **Step 4:** Each object is assigned to the nearest seed.
### 4.2 Non-hierarchical clustering algorithm

- **Clustering algorithm \((k\text{-means})\)**

**Step 1:** Initial seeds \(P_{C_j} (j = 1, 2, \ldots, k)\) are appropriately determined from the objects \(P_i (i = 1, 2, \ldots, n)\) described by distributions (e.g. by using random numbers).

**Step 2:** Dissimilarity \(d_C(P_i, P_{C_j})\) (or \(d_D(P_i, P_{C_j})\)) from seed \(P_{C_j}\) to object \(P_i\) is evaluated for \(i = 1, 2, \ldots, n; j = 1, 2, \ldots, k\).

**Step 3:** The centroid distribution \(P_{C_j}\) of each cluster \(C_j (j = 1, 2, \ldots, k)\) is decided as a new seed.

**Step 4:** Each object is assigned to the nearest seed.

**Step 5:** If it satisfies a stopping rule (e.g. pre-determined maximum iteration number) then stop, else go to Step 2.
5.1 Applying for the weather data at Japan

- **The weather data**
  - Date: 1 to 31 March 2009
  - Observation points: meteorological observatories in the region at Japan
  - Variables: average temperature and humidity (per day)
5.1 Applying for the weather data at Japan

• **The weather data**
  – Date: 1 to 31 March 2009
  – Observation points: meteorological observatories in the region at Japan
  – Variables: average temperature and humidity (per day)

• **Transformed distribution-valued data**

\[
X = \begin{bmatrix}
x_{11} & x_{12} \\
x_{21} & x_{22} \\
\vdots & \vdots \\
x_{n11} & x_{n12} \\
\vdots & \vdots \\
x_{N1} & x_{N2}
\end{bmatrix}
\]
5.1 Applying for the weather data at Japan

- **The weather data**
  - Date: 1 to 31 March 2009
  - Observation points: meteorological observatories in the region at Japan
  - Variables: average temperature and humidity (per day)

- **Transformed distribution-valued data**

![Diagram](image)
5.1 Applying for the weather data at Japan

• **The weather data**
  - Date: 1 to 31 March 2009
  - Observation points: meteorological observatories in the region at Japan
  - Variables: average temperature and humidity (per day)

• **Transformed distribution-valued data**
5.1 Applying for the weather data at Japan

• **Transformed distribution-valued data**
  – If we use histograms, the number of bins or range of bins affect the result.
5.1 Applying for the weather data at Japan

- **Transformed distribution-valued data**
  - If we use histograms, the number of bins or range of bins affect the result.
  
  [Diagram showing histogram with bars and dotted line indicating transformed data]

- Here, we use **the empirical distribution function**.
5.1 Applying for the weather data at Japan

- We also apply the classical ($k$-means) method for the following classical data transformed by using means.

\[
\begin{align*}
\text{Initial data} & : \quad X = \begin{bmatrix}
  x_{11} & x_{12} \\
  x_{21} & x_{22} \\
  \vdots & \vdots \\
  x_{n1} & x_{n2} \\
  \vdots & \vdots \\
  x_{N1} & x_{N2}
\end{bmatrix} \\
\text{classical data} & : \quad X^* = \begin{bmatrix}
  \bar{x}_{11} & \bar{x}_{12} \\
  \vdots & \vdots \\
  \bar{x}_{i1} & \bar{x}_{i2} \\
  \vdots & \vdots \\
  \bar{x}_{m1} & \bar{x}_{m2}
\end{bmatrix}
\end{align*}
\]
5.1 Applying for the weather data at Japan

• We also apply the classical ($k$-means) method for the following classical data transformed by using means.

\[
\begin{bmatrix}
    x_{11} & x_{12} \\
    x_{21} & x_{22} \\
    \vdots & \vdots \\
    x_{n1} & x_{n2} \\
    x_{N1} & x_{N2}
\end{bmatrix}
\]

\[
\bar{X}^* = \begin{bmatrix}
    \bar{x}_{11} & \bar{x}_{12} \\
    \vdots & \vdots \\
    \bar{x}_{i1} & \bar{x}_{i2} \\
    \vdots & \vdots \\
    \bar{x}_{m1} & \bar{x}_{m2}
\end{bmatrix}
\]
• We also apply the classical ($k$-means) method for the following classical data transformed by using means.

\[
X = \begin{bmatrix}
  x_{11} & x_{12} \\
  x_{21} & x_{22} \\
  \vdots & \vdots \\
  x_{n1} & x_{n2} \\
  \vdots & \vdots \\
  x_{N1} & x_{N2}
\end{bmatrix}
\]

\[
X^* = \begin{bmatrix}
  \bar{x}_{11} & \bar{x}_{12} \\
  \vdots & \vdots \\
  \bar{x}_{i1} & \bar{x}_{i2} \\
  \vdots & \vdots \\
  \bar{x}_{m1} & \bar{x}_{m2}
\end{bmatrix}
\]

• We compare the result of the classical method and the proposal method.
5.1 Applying for the weather data at Japan

- **Result of the classical method \((k\text{-means})\)**
  - Classified by **a degree of latitude** and **altitude**
5.1 Applying for the weather data at Japan

- Result of the classical method ($k$-means)
  - Classified by a degree of latitude and altitude
5.1 Applying for the weather data at Japan

- **Result of the classical method** *(k*-means)*
  - Classified by a **degree of latitude** and **altitude**
  - One Observatory located at high altitude is not classified into the blue cluster
5.1 Applying for the weather data at Japan

- **Result of the classical method \((k\text{-means})\)**
  - Classified by **a degree of latitude and altitude**
  - One Observatory located at high altitude is not classified into the blue cluster
5.1 Applying for the weather data at Japan

- **Result of the proposal method**
5.1 Applying for the weather data at Japan

**Result of the proposal method**

- The green cluster --- On the East sea side, mainly
5.1 Applying for the weather data at Japan

- **Result of the proposal method**
  - The green cluster --- On the East sea side, mainly
  - The purple cluster
    - On the Pacific side
5.1 Applying for the weather data at Japan

- **Result of the proposal method**
  - The green cluster --- On the East sea side, mainly
  - The purple cluster
    - On the Pacific side
  - The aqua cluster
    - Inland
5.1 Applying for the weather data at Japan

- **Result of the proposal method**
  - The green cluster --- On the East sea side, mainly
  - The purple cluster
    - On the Pacific side
  - The aqua cluster
    - Inland

These Observatories are located at high altitude
6. Conclusion

• In this presentation,
  – We define the centroid distribution
  – and proposed a non-hierarchical clustering method for more general distribution-valued data by using the centroid distribution

• Possibility that new classification structures are found by using proposal method.

• For the future study,
  – Calculation of the centroid distribution on the other dissimilarity measure.
6. Conclusion

• In this presentation,
  – We define **the centroid distribution**
  – and proposed a **non-hierarchical clustering method for more general distribution-valued data** by using the centroid distribution

• Possibility that new classification structures are found by using proposal method.

• For the future study,
  – Calculation of the centroid distribution on the other dissimilarity measure.

**Thank you very much for your attention!**
Appendix. Altitude of observatories

- Altitude of observatories