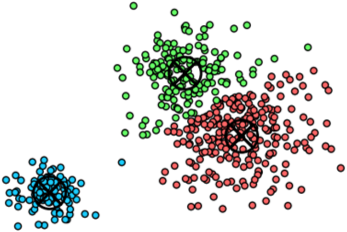
# Comparison of Validity Indexes for Fuzzy Clusters of fMRI data

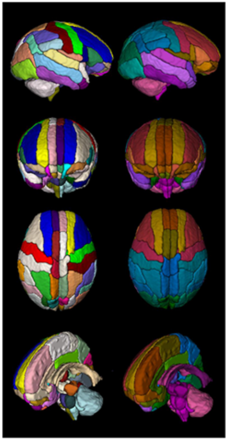
S Martinelli†, AA Vergani‡, and E Binaghi†

†University of Insubria (Varese, Italy)
Center of Research in Image Analysis and Medical Informatics
Department of Theoretical and Applied Science
***
‡Middlesex University (London, UK)

# Contents

# *functional* MRI
# Image Visualization and Processing





# *functional* MRI
## Resting-state and data-driven methods

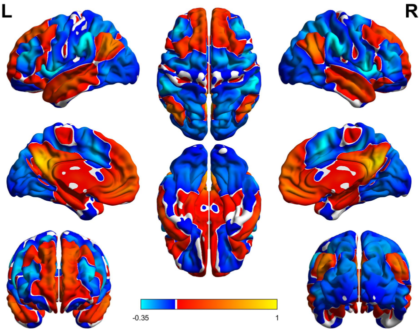



Figure: Example. This image displays **resting-state functional connectivity** as linear correlation for the seed region (Precuenus) in a sample of 1,000 subjects

# Clustering Approaches

### Clustering

Let $X = \{x_1, x_2, \ldots, x_n\}$ a given **dataset** (with $n$ elements), and let $C = \{c_1, c_2, ..., c_K\}$ the set of **cluster**, where $K$ is the desired number of clusters.

# Clustering Approaches

### Clustering

Let $X = \{x_1, x_2, \ldots, x_n\}$ a given **dataset** (with $n$ elements), and let $C = \{c_1, c_2, ..., c_K\}$ the set of **cluster**, where $K$ is the desired number of clusters.

Regardless of the criterion chosen for the partition, the purpose of clustering is to develop a **partition matrix** of size $K \times n$ denoted as $U = [\mu_{ij}]$, with $i = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, n$, where $\mu_{ij}$ is the grade of membership of point $x_j$ to cluster $c_i$.

# Clustering Approaches

### Crisp case (classic logic, $\mathbb{Z}$)

In crisp clustering, each point in the specified dataset belongs to a **single cluster class**. Then $\mu_{ij} = 1$ if $x_j \in c_i$, otherwise $\mu_{ij} = 0$.

# Clustering Approaches

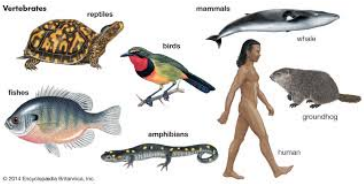
### Crisp case (classic logic, $\mathbb{Z}$)

In crisp clustering, each point in the specified dataset belongs to a **single cluster class**. Then $\mu_{ij} = 1$ if $x_j \in c_i$, otherwise $\mu_{ij} = 0$.

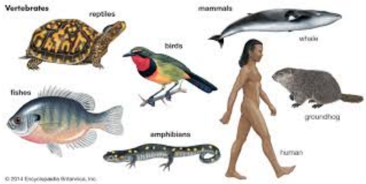
### Fuzzy case (many-valued logic, $\mathbb{R}$)

In fuzzy clustering, a point can be associated with **multiple clusters** with a certain **degree of membership**, and the partition matrix in this case is represented as $U = [\mu_{ij}]$, where $\mu_{ij} \in [0, 1]$ indicates the degree of membership of the $j$-th element to the $i$-th cluster.

# Example of Crisp Partition Matrix
# Which category is an animal?



# Example of Crisp Partition Matrix
# Which category is an animal?




| Animals | Mammals | Amphibians | Birds |
|---|---|---|---|
| Dog | 1 | 0 | 0 |
| Frog | 0 | 1 | 0 |
| Coyote | 1 | 0 | 0 |
| Opossum | 1 | 0 | 0 |
| Eagle | 0 | 0 | 1 |

# Example of Fuzzy Partition Matrix
# Is a bunch of granes an heap?

# Example of Fuzzy Partition Matrix
# Is a bunch of granes an heap?

| granes | heap | no heap |
|---|---|---|
| 1 | 0 | 1 |
| 100 | 0.3 | 0.7 |
| 1000 | 0.5 | 0.5 |
| 10000 | 0.75 | 0.25 |
| 100000 | 0.9 | 0.1 |

# Clustering Approaches

Fuzzy C-Means (Bezdek 1984)

# Clustering Approaches

## Fuzzy C-Means (Bezdek 1984)

**Given a predefined number of classes**, the purpose of the FCM algorithm is to create vectors called **centroids** that minimize the value of the function $J_m$ that is given by the sum of the intra-cluster quadratic error. $J_m$ it is defined as:

$$J_m = \sum_{j=1}^{n} \sum_{i=1}^{K} \mu_{ij}^{m} ||x_j - z_i||^2 \tag{1}$$

$\mathbf{m} > 1$ is the **exponent** to adjust the degree of fuzzy overlap.

# Clustering Approaches

## Fuzzy C-Means (Bezdek 1984)

**Given a predefined number of classes**, the purpose of the FCM algorithm is to create vectors called **centroids** that minimize the value of the function $J_m$ that is given by the sum of the intra-cluster quadratic error. $J_m$ it is defined as:

$$J_m = \sum_{j=1}^{n} \sum_{i=1}^{K} \mu_{ij}^{m} ||x_j - z_i||^2 \tag{1}$$

$\mathbf{m} > 1$ is the **exponent** to adjust the degree of fuzzy overlap.

## Validation step

$\longrightarrow$ do the selected classes (very-well) map on the input data?

# Clustering Approaches

## Fuzzy C-Means (Bezdek 1984)

**Given a predefined number of classes**, the purpose of the FCM algorithm is to create vectors called **centroids** that minimize the value of the function $J_m$ that is given by the sum of the intra-cluster quadratic error. $J_m$ it is defined as:

$$J_m = \sum_{j=1}^{n} \sum_{i=1}^{K} \mu_{ij}^{m} ||x_j - z_i||^2 \tag{1}$$

$\mathbf{m} > 1$ is the **exponent** to adjust the degree of fuzzy overlap.

## Validation step

$\longrightarrow$ do the selected classes (very-well) map on the input data?
$\longrightarrow$ which index(es) should I use?

# Clustering Approaches

## Fuzzy C-Means (Bezdek 1984)

**Given a predefined number of classes**, the purpose of the FCM algorithm is to create vectors called **centroids** that minimize the value of the function $J_m$ that is given by the sum of the intra-cluster quadratic error. $J_m$ it is defined as:

$$J_m = \sum_{j=1}^{n} \sum_{i=1}^{K} \mu_{ij}^{m} ||x_j - z_i||^2 \qquad (1)$$

$\mathbf{m} > 1$ is the **exponent** to adjust the degree of fuzzy overlap.

## Validation step

$\longrightarrow$ do the selected classes (very-well) map on the input data?
$\longrightarrow$ which index(es) should I use?
$\longrightarrow$ how to compare several indexes?

# Clustering Validity Index (CVIs)

## General definition of a validity index

Given a **type** of clustering algorithm, CVIs compute a **relation** between Compactness C and Separation S, and use a **criterion** to determine the optimum.

# Clustering Validity Index (CVIs)

## General definition of a validity index

Given a **type** of clustering algorithm, CVIs compute a **relation** between Compactness C and Separation S, and use a **criterion** to determine the optimum.

## Particular cases

- type: crisp or fuzzy
- relation: $R(C, S) : R = \vee(\times, \backslash, +, -)$
- criterion: optimality towards *min* or *max*

# Clustering Validity Index (CVIs)

## General definition of a validity index

Given a **type** of clustering algorithm, CVIs compute a **relation** between Compactness C and Separation S, and use a **criterion** to determine the optimum.

## Particular cases

- type: crisp or fuzzy
- relation: $R(C, S) : R = \vee(\times, \backslash, +, -)$
- criterion: optimality towards *min* or *max*

1. DBI: C $\backslash$ S | *min* | crisp
2. SDBI: C $\backslash$ S | *min* | fuzzy
3. XBI: C $\backslash$ S | *min* | fuzzy
4. PBMI: C $\times$ S | *max* | crisp
5. FPBMI: C $\times$ S | *max* | fuzzy
6. RLRI: C $+$ S | *min* | fuzzy
7. WSJI: C $+$ S | *min* | fuzzy
8. FSI: C $-$ S | *min* | fuzzy

# Datasets and Methodology

# Datasets and Methodology

Data: resting state fMRI (https://www.nitrc.org/ )

The **Beijing** dataset with 187 healthy subjects (73M /114F; ages 18-25).

# Datasets and Methodology

Data: resting state fMRI (https://www.nitrc.org/ )

The **Beijing** dataset with 187 healthy subjects (73M /114F; ages 18-25).
⇒ Image Preprocessing and Data Reduction

# Datasets and Methodology

### Data: resting state fMRI (https://www.nitrc.org/ )

The **Beijing** dataset with 187 healthy subjects (73M /114F; ages 18-25).
$\Rightarrow$ Image Preprocessing and Data Reduction

### Configurations

The **FCM algorithm** was setted with number of clusters $\mathbf{K} = 2, 3, \ldots, 10$ and weighting exponent $\mathbf{m} = 1.1, 1.2, \ldots, 2.5$.

# Datasets and Methodology

## Data: resting state fMRI (https://www.nitrc.org/ )

The **Beijing** dataset with 187 healthy subjects (73M /114F; ages 18-25).
$\Rightarrow$ Image Preprocessing and Data Reduction

## Configurations

The **FCM algorithm** was setted with number of clusters $\mathbf{K} = 2, 3, \ldots, 10$ and weighting exponent $\mathbf{m} = 1.1, 1.2, \ldots, 2.5$.

## Procedures

- Apply **indexes** (8) on each **configurations** (K $\times$ m) for each **subject** (187)
- Select the **optimal clustering** (K)
- **Compare** the optimal clustering with an experimental ground truth

# Datasets and Methodology

## Data: resting state fMRI (https://www.nitrc.org/ )

The **Beijing** dataset with 187 healthy subjects (73M /114F; ages 18-25).
⇒ Image Preprocessing and Data Reduction

## Configurations

The **FCM algorithm** was setted with number of clusters $\mathbf{K} = 2, 3, \ldots, 10$ and weighting exponent $\mathbf{m} = 1.1, 1.2, \ldots, 2.5$.

## Procedures

- Apply **indexes** (8) on each **configurations** (K $\times$ m) for each **subject** (187)
- Select the **optimal clustering** (K)
- **Compare** the optimal clustering with an experimental ground truth

$\longrightarrow$ How much should we **rely** on the optimality measure of an index?
$\longrightarrow$ Are there **limits** on the optimality criterion?

# Computational Experiments

## Errors of indexes

# Computational Experiments

## Errors of indexes

$$E = |ni - nr| \tag{2}$$

**ni** is the optimal number of clusters **by the index**
**nr** is the candidated number of clusters **by the references**

# Computational Experiments

## Errors of indexes

$$E = |ni - nr| \tag{2}$$

**ni** is the optimal number of clusters **by the index**
**nr** is the candidated number of clusters **by the references**

## References are evidence-based ground-truths of resting-state networks

# Computational Experiments

## Errors of indexes

$$E = |ni - nr| \tag{2}$$

**ni** is the optimal number of clusters **by the index**
**nr** is the candidated number of clusters **by the references**

## References are evidence-based ground-truths of resting-state networks

- **2 fMRI classes**:
  - yes/no of ROIs about the Default Mode Network (**DMN**)
- **4 fMRI classes** associated to
  - Default Mode Network (**DMN**)
  - Visual Network (**VN**)
  - Sensory/Motor Network **(SMN)**
  - Other Resting Networks (**ORN**)

# Type of Experiments

# Type of Experiments

### Experiment 1

Computing the error $E = |ni - nr|$ by using nr **with two classes**

# Type of Experiments

**Experiment 1**

Computing the error $E = |ni - nr|$ by using nr **with two classes**

**Experiment 2**

Computing the error $E = |ni - nr|$ by using nr **with four classes**

# Type of Experiments

### Experiment 1

Computing the error $E = |ni - nr|$ by using nr **with two classes**

### Experiment 2

Computing the error $E = |ni - nr|$ by using nr **with four classes**

### Note: how to make indexes comparable?

Indexes have different **scales** and **directions** of optimization

# Type of Experiments

### Experiment 1

Computing the error $E = |ni - nr|$ by using nr **with two classes**

### Experiment 2

Computing the error $E = |ni - nr|$ by using nr **with four classes**

### Note: how to make indexes comparable?

Indexes have different **scales** and **directions** of optimization

*Rescale* and *costrain* the direction: the lower, the better

# Type of Experiments

### Experiment 1

Computing the error $E = |ni - nr|$ by using nr **with two classes**

### Experiment 2

Computing the error $E = |ni - nr|$ by using nr **with four classes**

### Note: how to make indexes comparable?

Indexes have different **scales** and **directions** of optimization

*Rescale* and *costrain* the direction: the lower, the better

$$Z(* \longrightarrow min) = +\frac{x - \mu}{\sigma} \quad \text{and} \quad Z(* \longrightarrow max) = -\frac{x - \mu}{\sigma} \tag{3}$$

# Comparison Results

Table: Error of the **8 indexes** evaluating clustering with 2 and 4 references

| Index | Experiment 1 | | Index | Experiment2 | |
|---|---|---|---|---|---|
| | E mean | Var | | E mean | Var |

# Comparison Results

Table: Error of the **8 indexes** evaluating clustering with 2 and 4 references

| Index | Experiment 1 | | Index | Experiment2 | |
|---|---|---|---|---|---|
| | E mean | Var | | E mean | Var |
| FSI (-) | 0.58 | 0.13 | WSJI (+) | 0.78 | 0.08 |
| RLRI (+) | 0.65 | 0.18 | RLRI (+) | 1.35 | 0.17 |
| WSJI (+) | 1.28 | 0.10 | FSI (-) | 1.47 | 0.08 |

# Comparison Results

Table: Error of the **8 indexes** evaluating clustering with 2 and 4 references

| Index | Experiment 1 | | Index | Experiment2 | |
|---|---|---|---|---|---|
| | E mean | Var | | E mean | Var |
| FSI (-) | 0.58 | 0.13 | WSJI (+) | 0.78 | 0.08 |
| RLRI (+) | 0.65 | 0.18 | RLRI (+) | 1.35 | 0.17 |
| WSJI (+) | 1.28 | 0.10 | FSI (-) | 1.47 | 0.08 |
| SDBI (/) | 1.54 | 1.50 | SDBI (/) | 1.55 | 0.49 |

# Comparison Results

Table: Error of the **8 indexes** evaluating clustering with 2 and 4 references

| Index | Experiment 1 | | Index | Experiment2 | |
|---|---|---|---|---|---|
| | E mean | Var | | E mean | Var |
| FSI (-) | 0.58 | 0.13 | WSJI (+) | 0.78 | 0.08 |
| RLRI (+) | 0.65 | 0.18 | RLRI (+) | 1.35 | 0.17 |
| WSJI (+) | 1.28 | 0.10 | FSI (-) | 1.47 | 0.08 |
| SDBI (/) | 1.54 | 1.50 | SDBI (/) | 1.55 | 0.49 |
| DBI (/) | 4.76 | 5.02 | DBI (/) | 3.31 | 2.35 |
| XBI (/) | 5.50 | 0.40 | XBI (/) | 3.69 | 0.39 |
| PBMI (×) | 6.22 | 0.42 | PBMI (×) | 4.22 | 0.42 |
| FPBMI (×) | 6.36 | 0.25 | FPBMI (×) | 4.36 | 0.25 |

# Comparison Results

# Conclusions

## Final propostions

# Conclusions

## Final propostions

- **RLRI, WSJI and FSI gained the top three positions**, followed by the SDBI, and then the bottom three DBI, PBMI, FPBMI

# Conclusions

## Final propostions

- **RLRI, WSJI and FSI gained the top three positions**, followed by the SDBI, and then the bottom three DBI, PBMI, FPBMI
- The differences between the CVIs lie in the formalization of **separation component** and in its **relation** the compactness.

# Conclusions

## Final propostions

- **RLRI, WSJI and FSI gained the top three positions**, followed by the SDBI, and then the bottom three DBI, PBMI, FPBMI
- The differences between the CVIs lie in the formalization of **separation component** and in its **relation** the compactness.

## Future works

# Conclusions

## Final propostions

- **RLRI, WSJI and FSI gained the top three positions**, followed by the SDBI, and then the bottom three DBI, PBMI, FPBMI
- The differences between the CVIs lie in the formalization of **separation component** and in its **relation** the compactness.

## Future works

- Test conclusions using also **benchmark** and **synthetic** datasets

# Conclusions

## Final propostions

- **RLRI, WSJI and FSI gained the top three positions**, followed by the SDBI, and then the bottom three DBI, PBMI, FPBMI
- The differences between the CVIs lie in the formalization of **separation component** and in its **relation** the compactness.

## Future works

- Test conclusions using also **benchmark** and **synthetic** datasets
- Extend the evaluation considering the **suboptimes**

# Conclusions

## Final propostions

- **RLRI, WSJI and FSI gained the top three positions**, followed by the SDBI, and then the bottom three DBI, PBMI, FPBMI
- The differences between the CVIs lie in the formalization of **separation component** and in its **relation** the compactness.

## Future works

- Test conclusions using also **benchmark** and **synthetic** datasets
- Extend the evaluation considering the **suboptimes**
- Add **other measures** in addition to $E = |ni - nr|$

# Conclusions

## Final propostions

- **RLRI, WSJI and FSI gained the top three positions**, followed by the SDBI, and then the bottom three DBI, PBMI, FPBMI
- The differences between the CVIs lie in the formalization of **separation component** and in its **relation** the compactness.

## Future works

- Test conclusions using also **benchmark** and **synthetic** datasets
- Extend the evaluation considering the **suboptimes**
- Add **other measures** in addition to $E = |ni - nr|$
- Use a **fuzzy inference system** to merge the measures

# Cheers

Thank you for the attention!

## Reference for other details

Martinelli et al 2019
Comparison of Validity Indexes for Fuzzy Clusters of fMRI Data
*In: Tavares J., Natal Jorge R. (eds) VipIMAGE 2019. VipIMAGE 2019. Lecture Notes in Computational Vision and Biomechanics, vol 34. Springer, Cham*