Multiobjective Optimization based Clustering Techniques

Dr. Sriparna Saha

Associate Professor

Department of Computer Science and Engineering

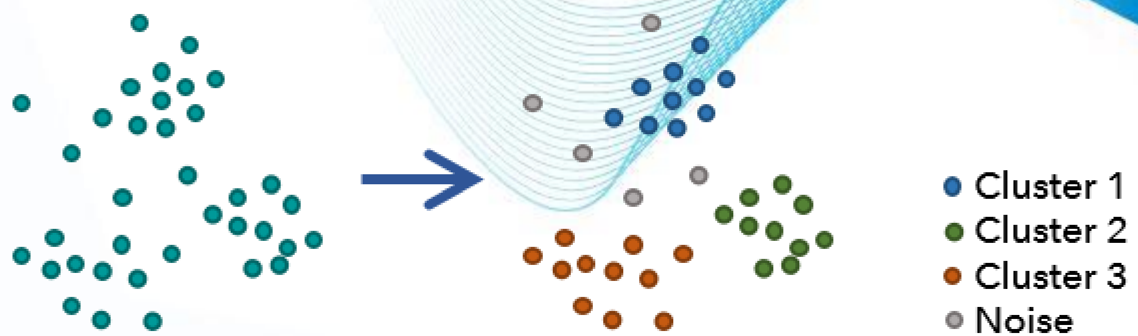
IIT Patna

Patna-800 013, India

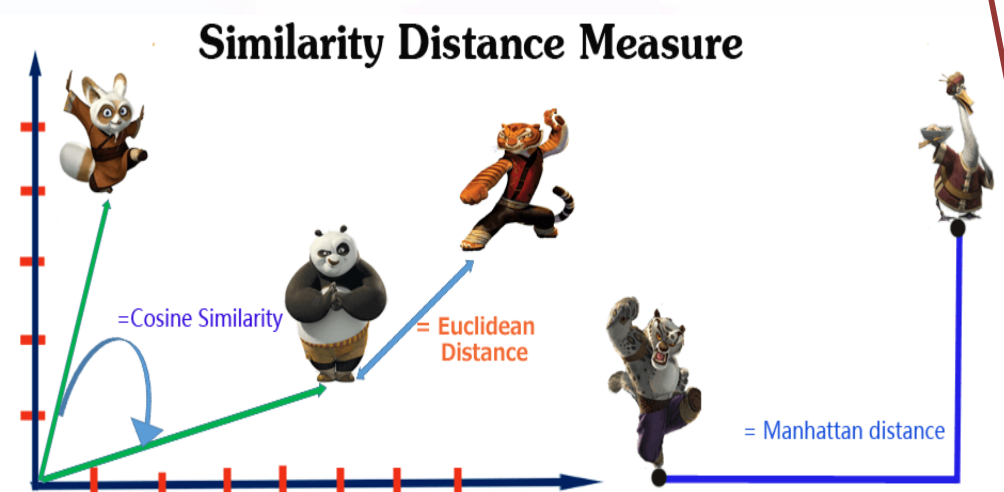
Email: sriparna@iitp.ac.in

Webpage: www.iitp.ac.in/~sriparna

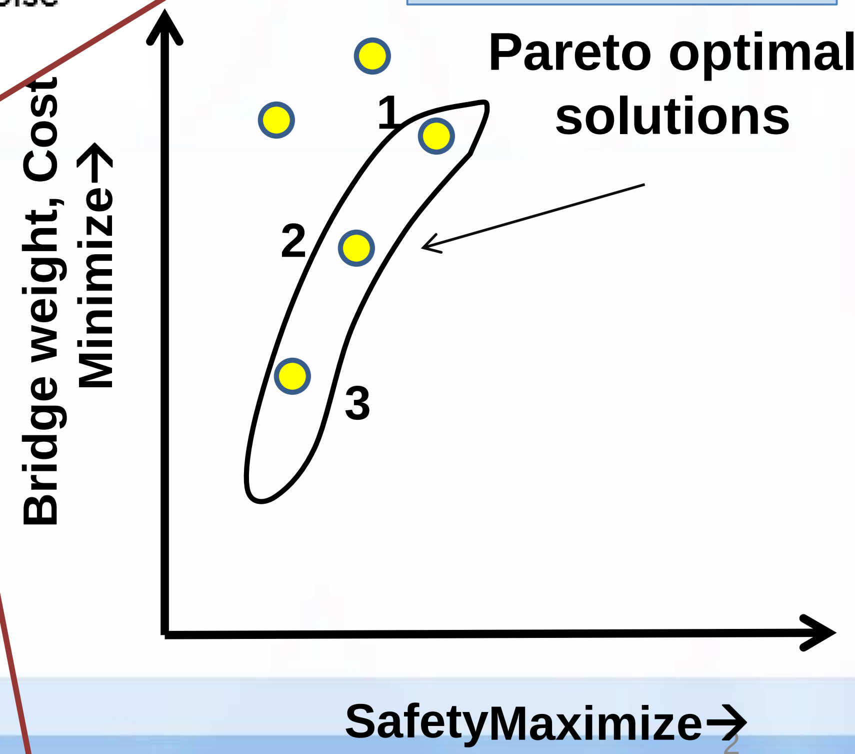
Key concepts



Clustering

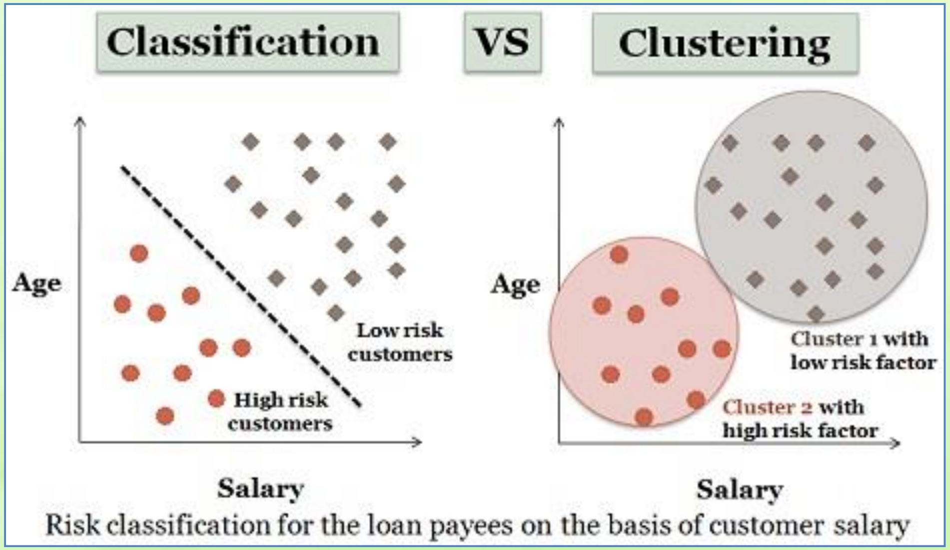


Multi-objective optimization



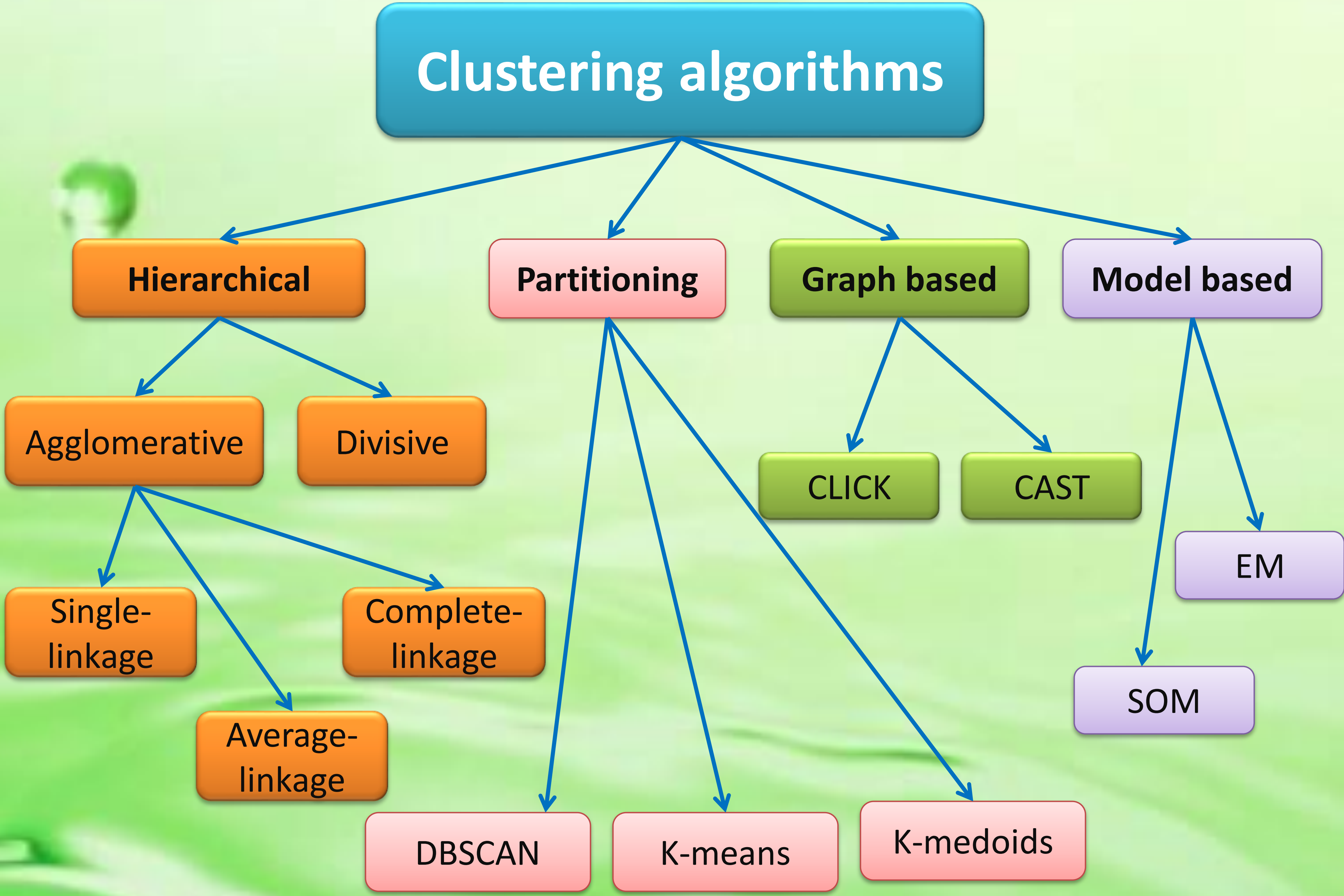
Clustering: The unsupervised classification

- Clustering is one kind of classification→ Unsupervised (no prior information about data labels; *sometimes labeled data is difficult to obtain*)



- Classification → In general it denotes supervised classification (set of labeled data required for training)

Clustering


Clustering algorithms
Hierarchical
Partitioning
Graph based
Model based
Agglomerative
Divisive
CLICK
CAST
EM
Single-linkage
Complete-linkage
Average-linkage
SOM
DBSCAN
K-means
K-medoids

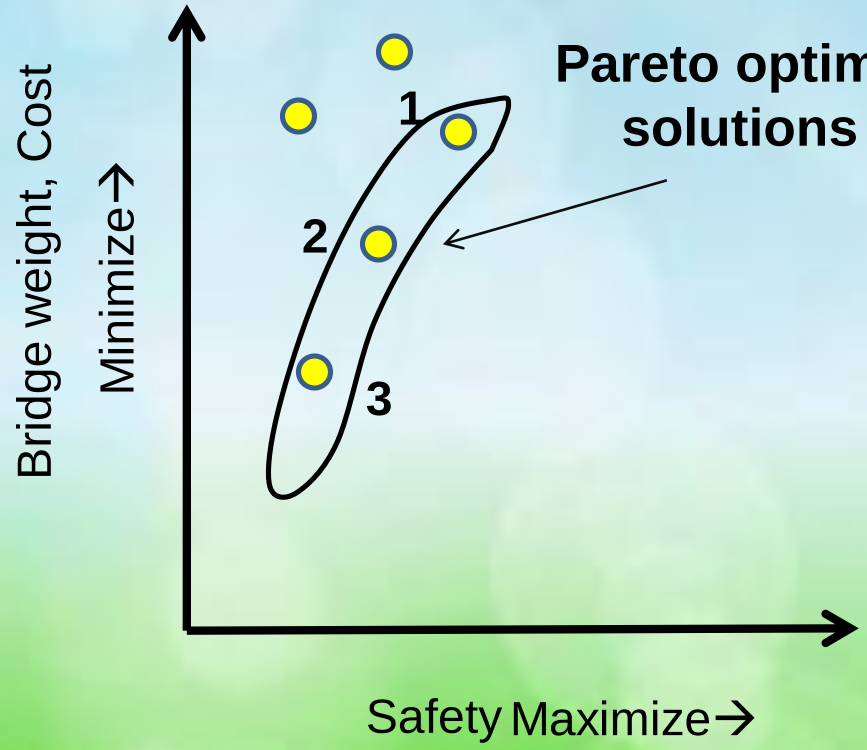
Optimization

Single-objective optimization (SOO)

Vs.

Multi-objective optimization (MOO)

Advantages of MOO



- Produces a set of non-dominating solutions →**Pareto front**
- Flexibility in choosing amongst the solution set based on any user-defined criteria.

Formal Definition of Multiobjective Optimization

- The multiobjective optimization can be formally stated as:
  - Find the vector of decision variables

    $$x=[x_1, x_2 \ldots\ldots x_n]^T$$

    which will satisfy the m inequality constraints:

    $$g_i(x) >= 0,\ i=1,2,\ldots.m,$$

    And the p equality constraints

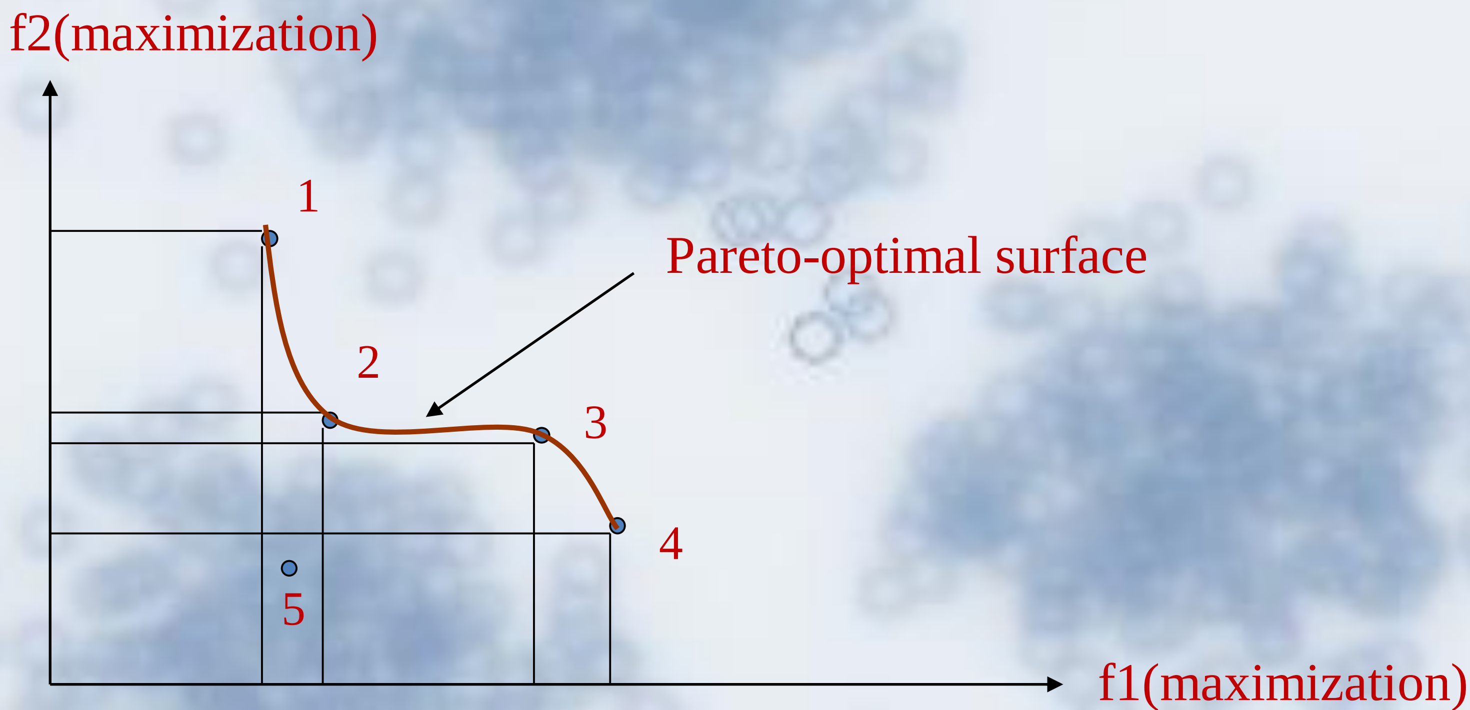
    $$h_i(x)=0\ ,\ i=1,2,\ldots.p.$$

    And simultaneously optimizes M objective functions

    $$f_1(x), f_2(x) \ldots.\ f_M(x).$$

  - No concept of global optimum
  - Produce a set of trade-off solutions
    - Pareto optimal set

Example of Dominance and Pareto-Optimality



- Here solutions 1, 2, 3 and 4 are non-dominating to each other.
- 5 is dominated by 2, 3 and 4, not by 1.

Existing Evolutionary MOO Strategies

- A multi-objective optimization algorithm must achieve:
  1. ***Guide the search towards the global Pareto-Optimal front.***
  2. ***Maintain solution diversity in the Pareto-Optimal front.***

- EAs
  - **Search and optimization tools**
  - **Provide near-optimal solutions for complex, hard, multimodal problems**.

- Multiobjective EAs are more popular primarily because of their *population based nature*.

Existing MOO algorithms

Evolutionary

- VEGA
- MOGA
- NSGA
- PESA
- SPEA
- NSGA-II
- MODE

Nature inspired

- MPSO
- MACO

Physical property-inspired

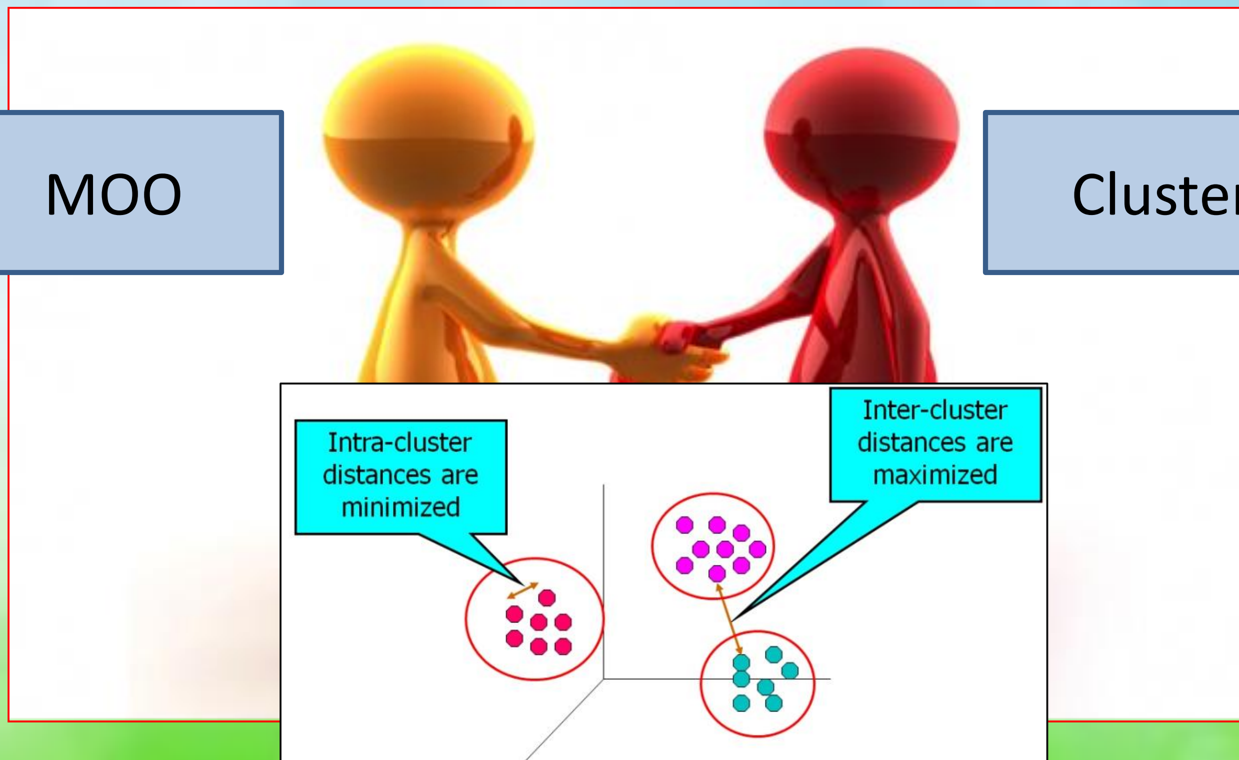
- MOSA
- AMOSA

1^{st} + 2^{nd} Key concepts → MOO based clustering

Two different paradigms in one frame!!!

MOO

Clustering



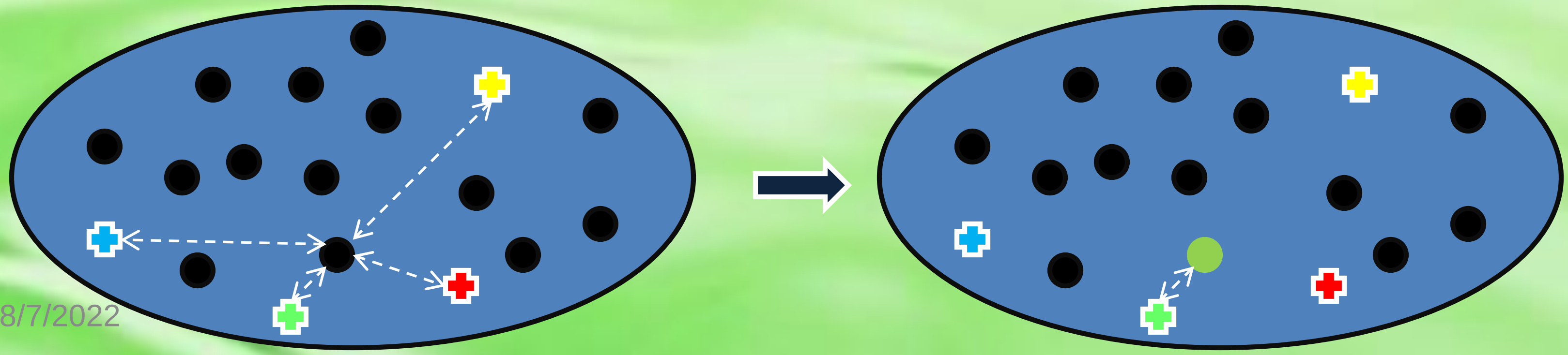
Multi objective optimization based clustering technique

3rd key concept

Existing distance measures →

- **Euclidean**
- **City block distance**
- **Point symmetry**
- **Line symmetry**
- **Cosine distance**
- **Symmetric conditional probability (SCP) metric**
- **Point-wise mutual information (PMI)**
- **Some Gene Ontology (GO) based similarity measures**

Proximity== Distance , determines **Similarity**



How to determine goodness of a clustering technique?

cluster validity measures

External	Internal
Need actual class labels	No class labels required

External — Need actual class labels

- **Adjusted rand index**
- **Classification accuracy**
- **F measure**
- **Minkowski score**
- **Jaccard index**

Internal — No class labels required

- **Xie-Beni index**
- **PBM index**
- **FCM index**
- **Silhouette index**
- **DB index**
- **Dunn index**
- **Mean squared residue**
- **Row variance**

Multiobjective Clustering Techniques

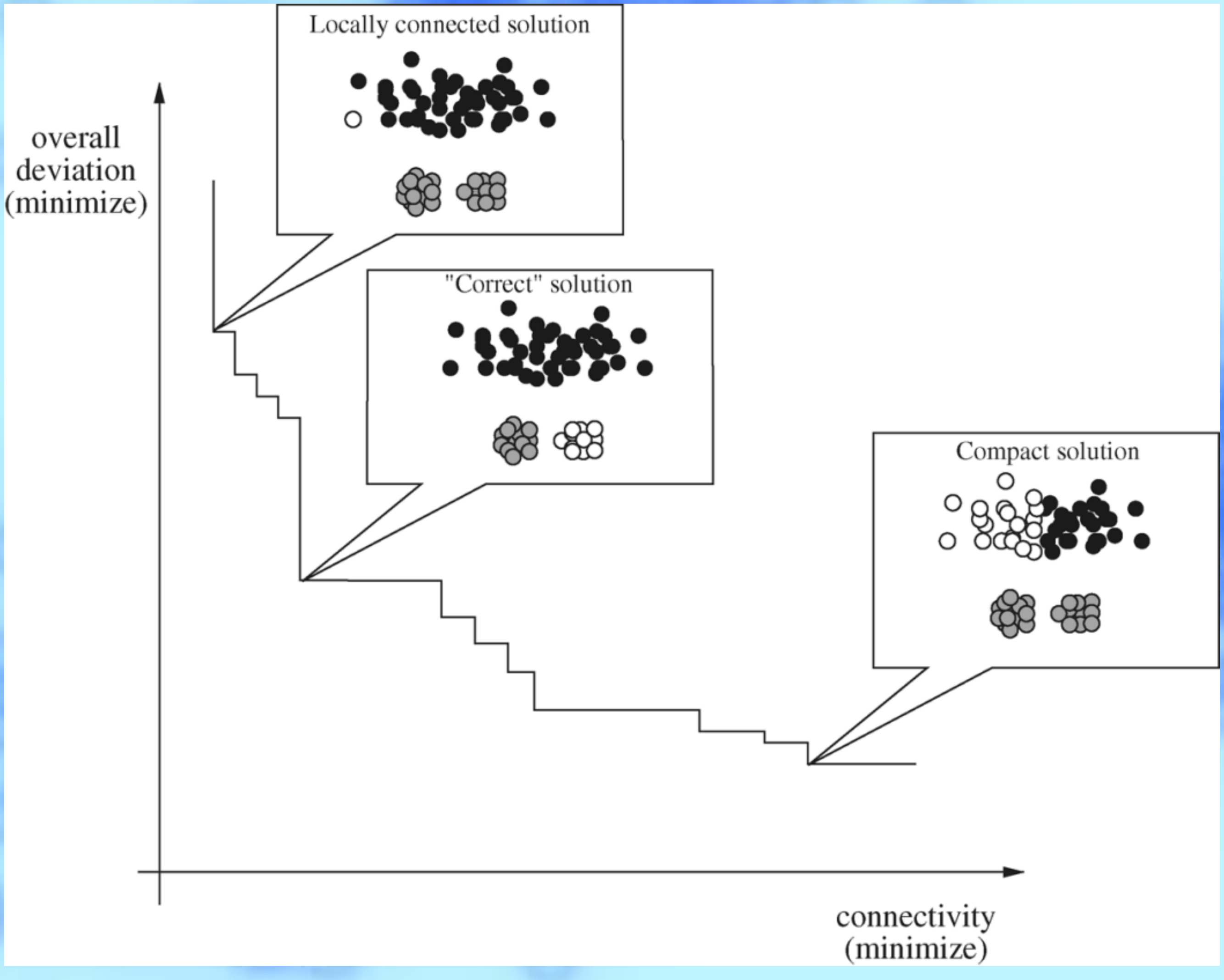
- **Often for many data sets no unambiguous partitioning exists**
  - **Multiple solutions representing different partitionings needed**
- **Existing clustering algorithms optimize only a single measure of cluster quality**
  - **Not applicable for different kinds of data sets with different characteristics**
  - **Need to optimize simultaneously different cluster quality measures capturing different data properties**

Pose Clustering as a MOO Problem

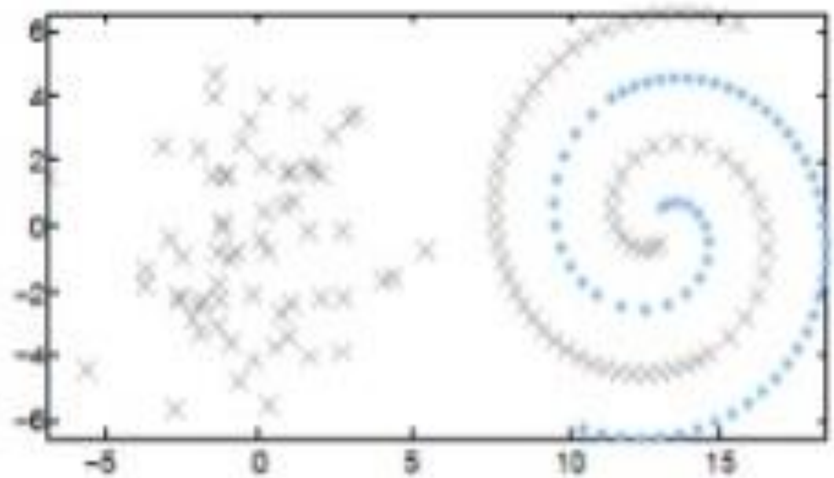
S. Saha and S. Bandyopadhyay (2010): ``A symmetry based multiobjective clustering technique for automatic evolution of clusters", Pattern Recognition. Volume 43, Issue 3, Pages 738-751 (impact factor: 2.607), H-index: 67.


Original data set
Two different clustering solutions

Example of Multiple Clustering Solutions for a Data Set



- Multi Objective Clustering
  - Decompose a data-set into similar groups maximizing multiple objectives in parallel.



Output for multiobjective clustering

A Multi-center Based Multiobjective Clustering Technique

- **Existing clustering techniques make strict assumptions about the shape, size, convexity and overlap of the data**
  - **K-means requires the clusters to be hyperspherical and of almost equal size**
  - **K-means known to get stuck at local optima**
  - **Hierarchical methods like single linkage clustering techniques require the clusters to be non-overlapping.**
  - **Algorithms like expectation maximization methods make assumptions about the data distributions.**
  - **Metaheuristic techniques like genetic algorithms, simulated annealing, etc often optimize measures that cannot capture clusters of any shape, size and convexity.**

A Multi-center Based Multiobjective Clustering Technique

- Aim : to develop a general purpose clustering method
    - avoid getting stuck at local optima
    - detecting clusters that need not necessarily be convex, may have overlaps, and can be of any size and density
    - even the number of clusters determined automatically

- A New MOO clustering technique (VGenClustMOO)
    - Uses AMOSA as underlying optimization method
    - Can determine the appropriate K and the appropriate partitioning
        - data sets with any kind of clusters
            - either symmetrical overlapping/well-separated

A Multi-center Based Multiobjective Clustering Technique

- **A new MOO clustering technique is proposed**
  - **can determine the appropriate number of clusters and the appropriate partitioning from data sets having any kind of clusters (either of hyperspherical shaped/connected/symmetric)**
  - **multiple cluster centers are used to represent a particular cluster**
  - **a cluster is divided into several nonoverlapping small hyperspherical clusters**

S. Saha and S. Bandyopadhyay, "Application of a Multiseed Based Clustering Technique for Automatic Satellite Image Segmentation", IEEE Geoscience and Remote Sensing Letters, Volume 7, Issue 2, Pages 306-308, April 2010.

Why Multiple Centers for Each Cluster?

- **Two possible representations:**
  - **Center based:**

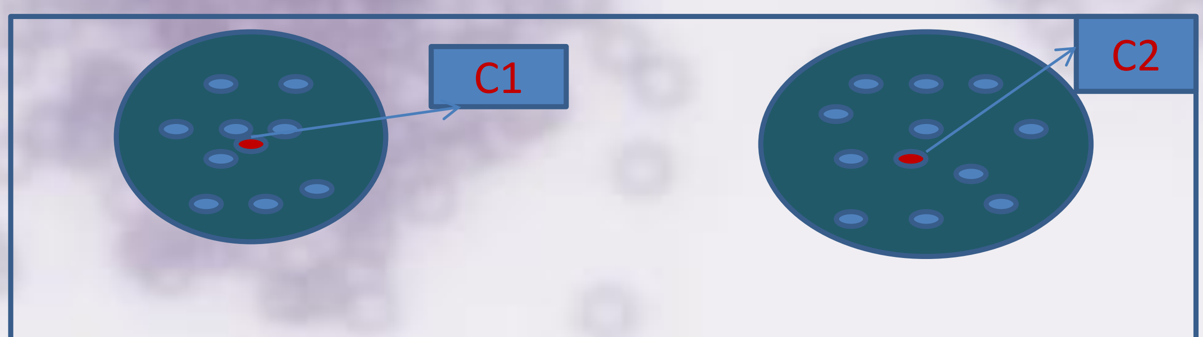
    Prefers hyperspherical shaped compact clusters

- **Point Based Encoding**
  - **String length increases with number of points**

C1	C2

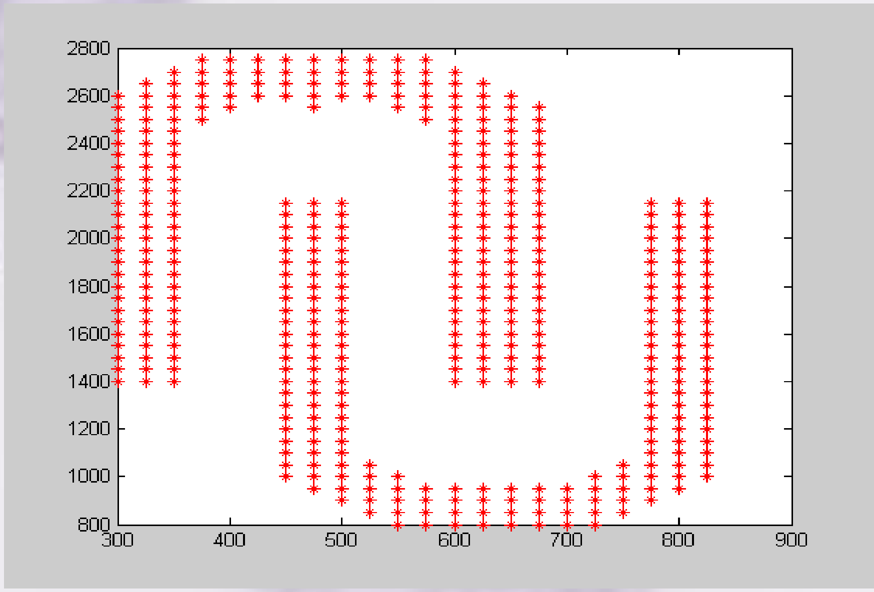
X1 X2X3 X4 X5 X6............X22

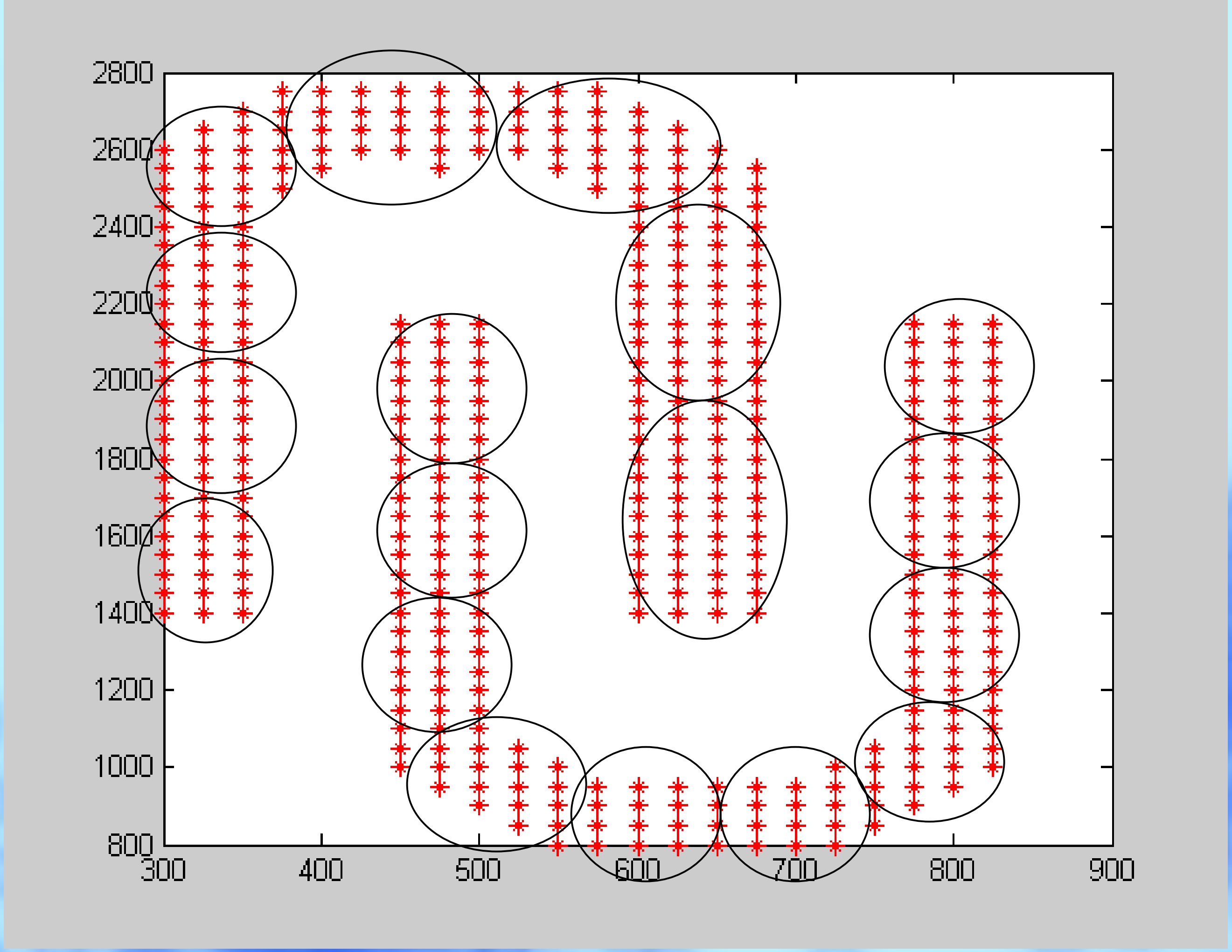
1111111111222222222222



Why Multiple Centers for Each Cluster?

- **Connected structures**
  - No symmetricity exists
  - Center based encoding fails to capture symmetry property
  - Can be decomposed into several small symmetrical shaped clusters
  - Centers of these local subclusters can be used to represent the whole cluster



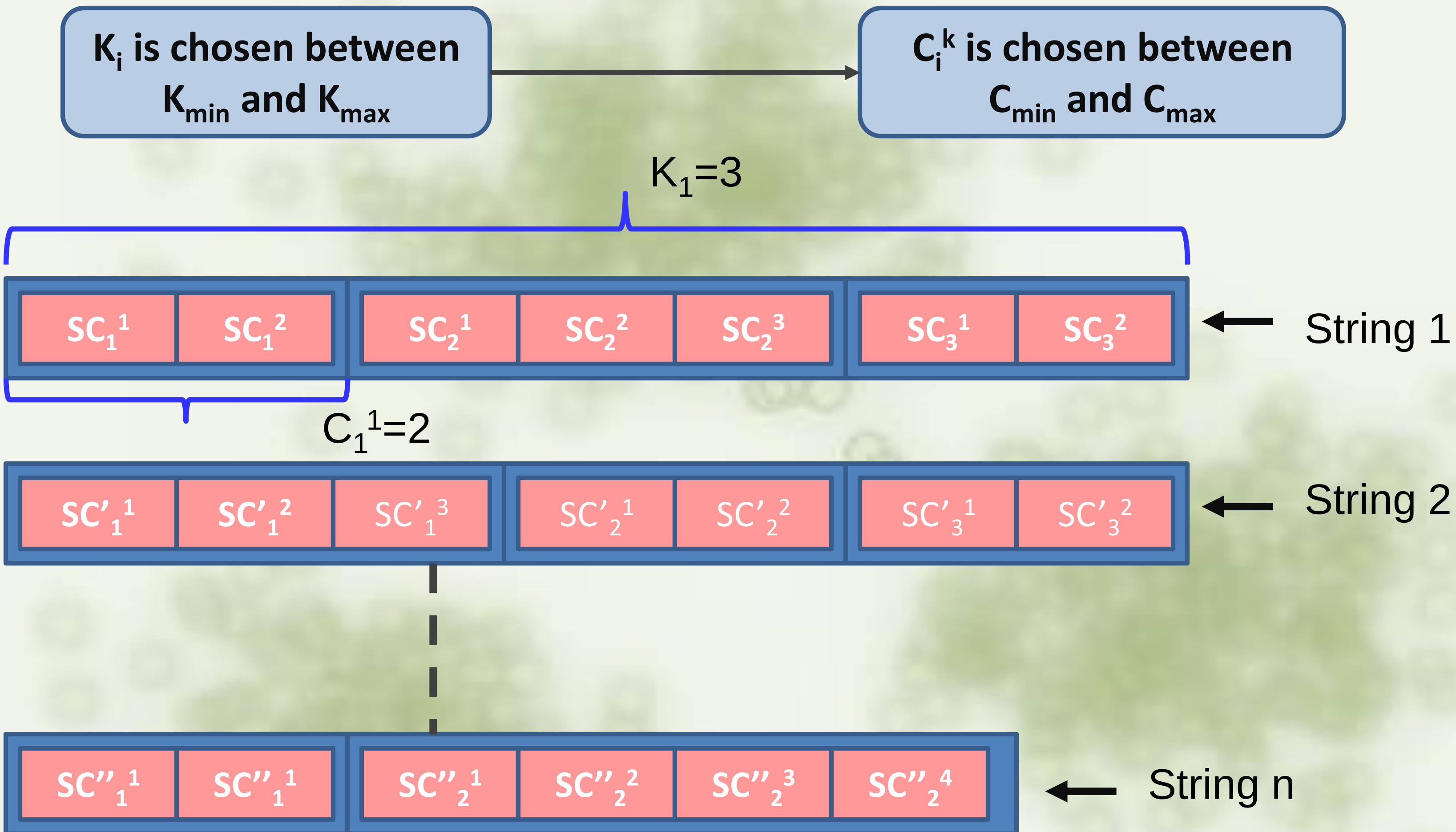


Representing a given cluster by some sub-cluster centers. Here first cluster is represented by 8 and second cluster is represented by 10 small hyperspherical clusters

- K_i : number of clusters
- C^i_k : number of centers for k^{th} cluster of i^{th} string
- d: dimension
- Length of each string : $\sum_{k=1}^{K_i} C_k^i \times d$
- K_i chosen randomly
  - K_i assumed to lie in between [K_{min}, K_{max}]
  - C^i_k assumed to lie in between [C_{min}, C_{max}]

- initialization procedure : partly random and partly based on two different single-objective algorithms in order to obtain a good initial spread of solutions
  - 1/3 of the solutions of the archive initialized after running single linkage for different values of number of clusters (K)
  - Another 1/3 initialized after running K-means
  - Last 1/3 initialized randomly

Population initialization



Assignment of Points

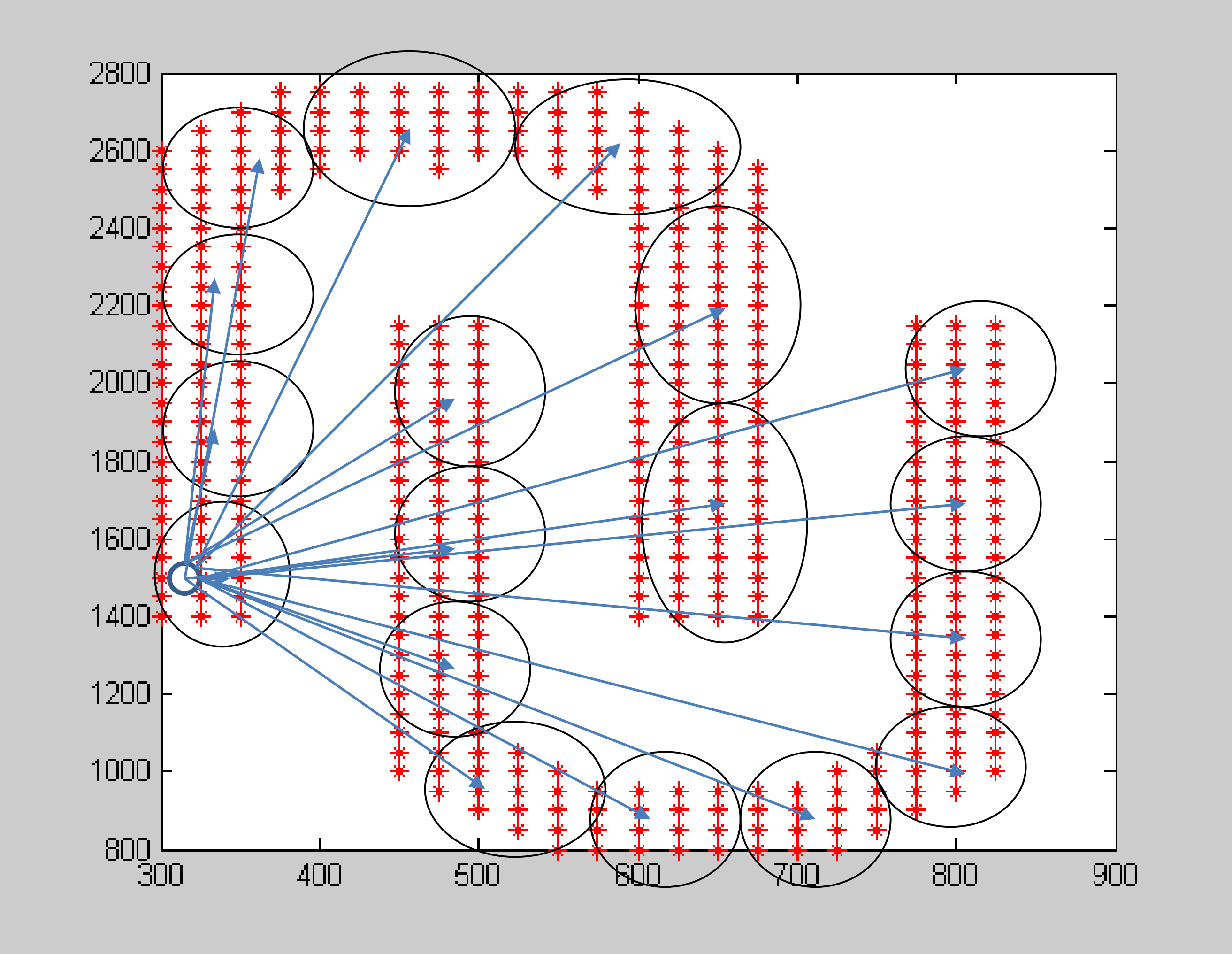
- For assignment of points to individual cluster
  - considering *total* subclusters as separate clusters

  - using minimum Euclidean distance criteria
  - Point x is assigned to cluster k where

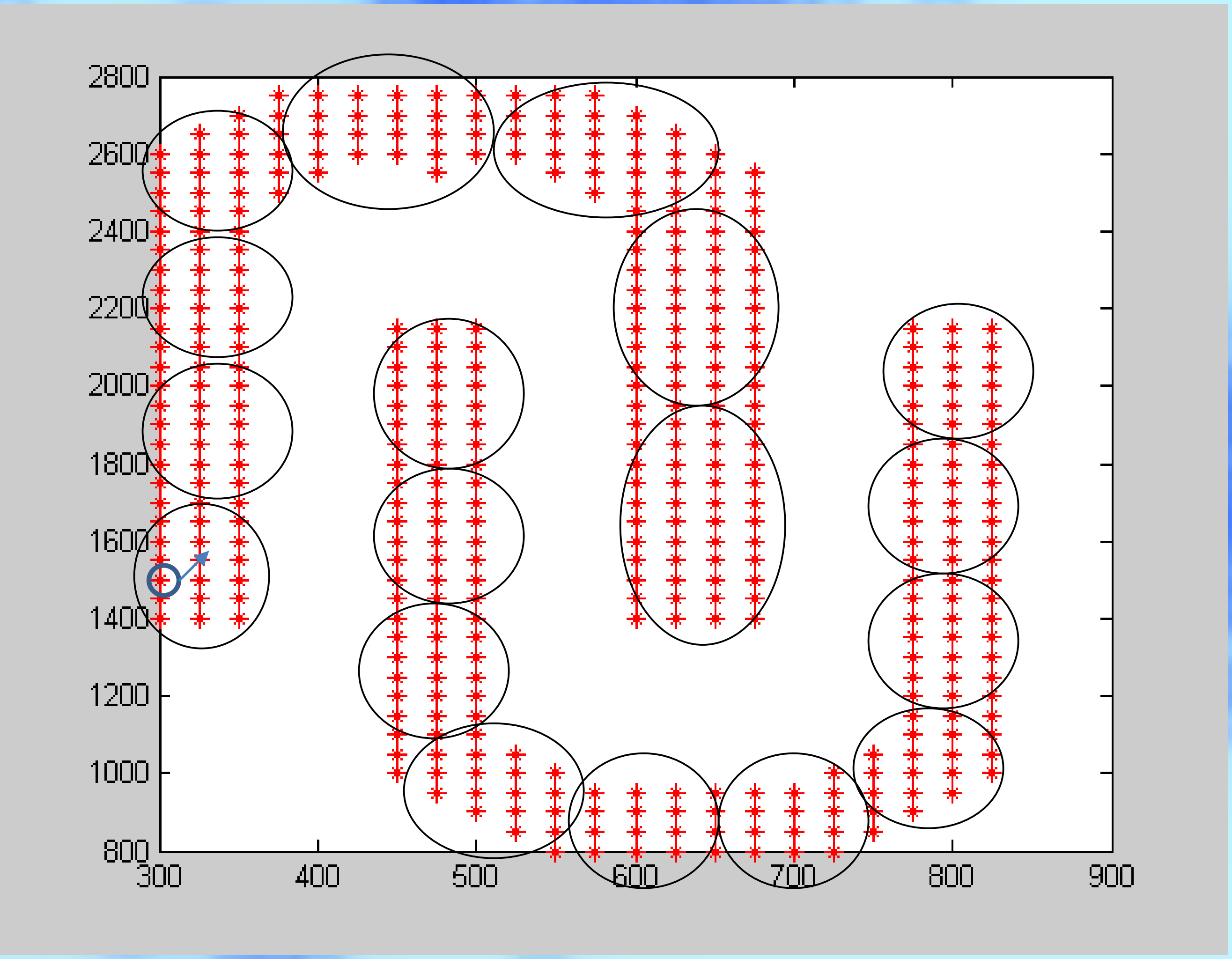
$$total = \sum_{k=1}^{K_i} C_k^i$$

$$k = \arg\min_{k=1}^{total} d(c_i, x)$$

Example: Calculate distance of a point from different sub-cluster centers



- **Assigned to the cluster with respect to which its distance is the minimum**



Fitness Value Computation

- For fitness computation
  - *total* number of subclusters merged to form K_i clusters

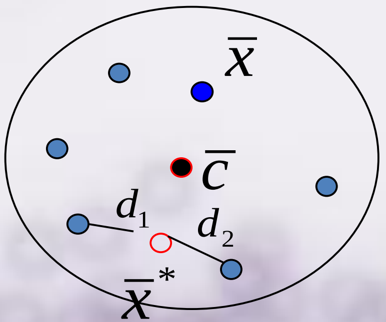
- Objective functions used
  - three cluster validity indices
- I-index :
  - cluster validity index based on Euclidean distance
  - detect appropriate K and the appropriate partitioning
    - from data sets having compact well-separated hyperspherical clusters

- Sym-index:
  - d_{PS} based cluster validity index
  - detect appropriate K and proper partitioning
    - from data sets having symmetrical shaped clusters irrespective of their geometrical shape, size and convexity

- **Connectivity-index:**
  - **newly developed**
  - **detect appropriate K and partitioning**
    - **data sets with nonoverlapping clusters irrespective of its shape, size and convexity**
  - **uses “connectedness" property of the points**

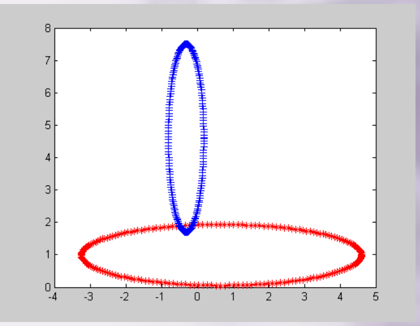
d_{PS} : A New Point Symmetry Distance

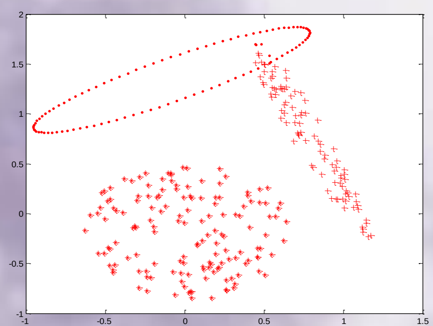
- $d_{ps}(x, c)$ - New PS distance of x wrt center c.
  - x': Symmetrical point of x wrt centre c
    - 2c−x (reflected point)
  - di, i=1, 2,….knear
    - dist. of knear nearest neighbors of x'

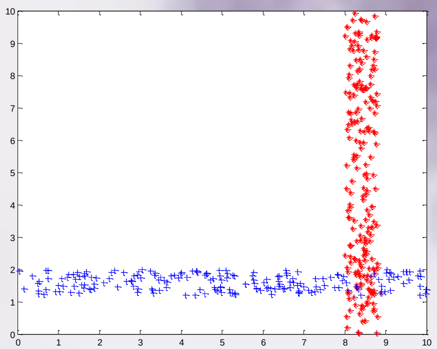


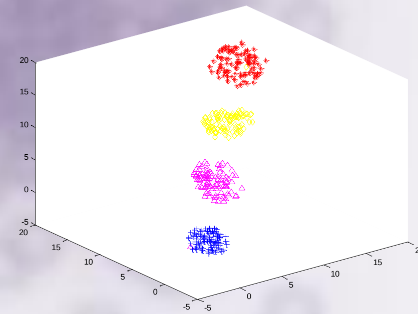
$$d_{ps}(\bar{x},\bar{c}) = d_{sym}(\bar{x},\bar{c}) * d_e(\bar{x},\bar{c}),$$

$$= \frac{\sum_{i=1}^{knear} d_i}{knear} * d_e(\bar{x},\bar{c})$$









S. Bandyopadhyay and S. Saha (2007): ``GAPS: A New Symmetry Based Genetic Clustering Technique", Pattern Recognition. Volume 40, Issue 12, Pages 3430-3451 (impact factor: 3.279), H-index: 67.

Cluster Validity Measure

- *Sym*-index
- **The newly developed point symmetry based distance is used to develop a new cluster validity index, Sym-index**
- **Consider a partition of the data set X={x_j, j=1…n }**
- **Let K cluster centers be denoted by c_i (1<=i<=K)**
- **n_i, i=1..n, number of points in cluster i.**
- **K : number of clusters**

$$\varepsilon_K = \sum_{i=1}^{K} E_i,$$

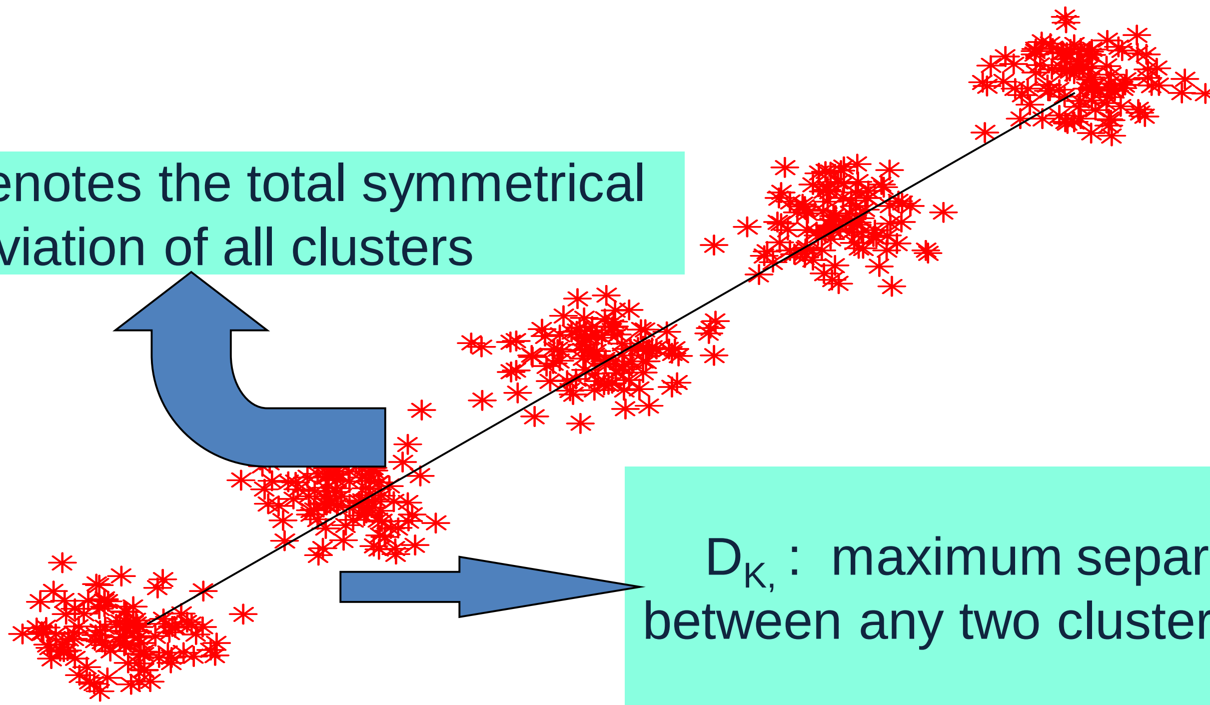
$$E_i = \sum_{j=1}^{n_i} d_{ps}^{*}(\bar{x}_j^i, \bar{c}_i)$$

$$D_K = \max_{i,j=1}^{K} \left\| \bar{c}_i, \bar{c}_j \right\|$$

$$Sym(K) = \left(\frac{1}{K} \times \frac{1}{\varepsilon_K} \times D_K \right)$$

- **Objective: maximize this index in order to maintain the actual number of clusters**
- ***Three factors complementary in nature, so they are expected to compete and balance each other critically for determining the proper partitioning.***

Here E_K denotes the total symmetrical deviation of all clusters



$D_{K,}$: maximum separation between any two cluster centers

S. Saha and S. Bandyopadhyay. *Application of a New Symmetry Based Cluster Validity Index for Satellite Image Segmentation.* IEEE Geoscience and Remote Sensing Letters, Vol. 5, No. 2, April 2008, Pages 166-170 (citations: 42)

I-index

- I-index , is an Euclidean distance based cluster validity index and is defined as:

$$I(k) = \left(\frac{1}{K} \times \frac{E_1}{E_K} \times D_K\right)^p$$

$$E_K = \sum_{k=1}^{K} \sum_{j=1}^{n_k} d_e(\overline{c_k}, \overline{x_j^k})$$

$$D_K = \max_{i,j=1}^{K} d_e(\overline{c_i}, \overline{c_j})$$

Connectivity Index

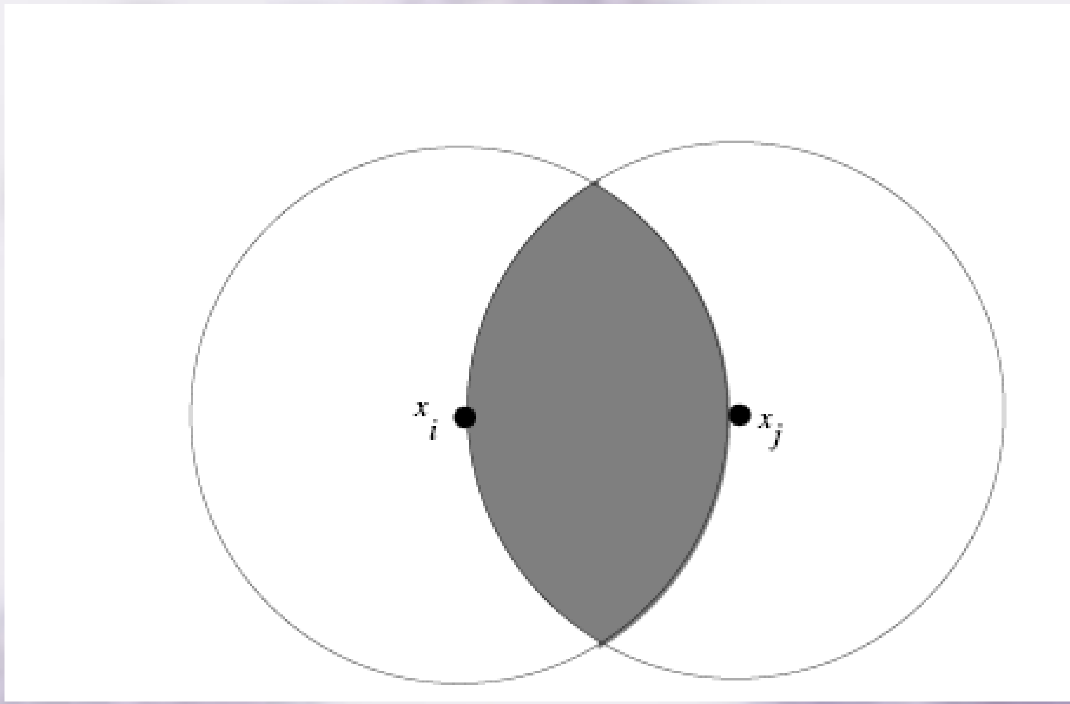
- **Relative Neighborhood Graph(RNG):**
- **Let X be a set of points represented as X = {x1; x2; : : : ; xn}.**
- **Two points xi and xj are said to be relative neighbors if the following condition holds:**

$$d(x_i, x_j) \le \max[d(x_i, x_k), d(x_j, x_k)], \forall x_k \in X, k \ne i, j$$

- **The region of influence of two points xi and xj in the RNG, denoted by R_{RNG}(xi; xj), formed as follows:**

$$R_{RNG}(x_i, x_j) = \{x : \max[d(x_i, x), d(x_j, x)] < d(x_i, x_j), i \ne j\}$$

Sriparna Saha, Sanghamitra Bandyopadhyay, Some connectivity based cluster validity indices, Applied Soft Computing, Volume 12, Issue 5, May 2012, Pages 1555–1565, ISSN 1568-4946 (impact factor: 2.415), H-index: 65.



Xi and Xj are relative neighbors

if there are no points that lie within the shaded region

Relative Neighborhood Graph (continued…)

- **Now the RNG is constructed as follows:**
- **For i=1 to n**
  - **For j= i+1 to n**
    - **Determine the region of influence R(xi,xj)**
    - **If R(xi,xj)$\cap${X-{xi,xj}}=$\varnothing$**
      - **Add the edge connecting xi,xj**
      - **Edge weight=Euclidean distance between xi,xj**
    - **End if**
  - **End for**
  - **End for**

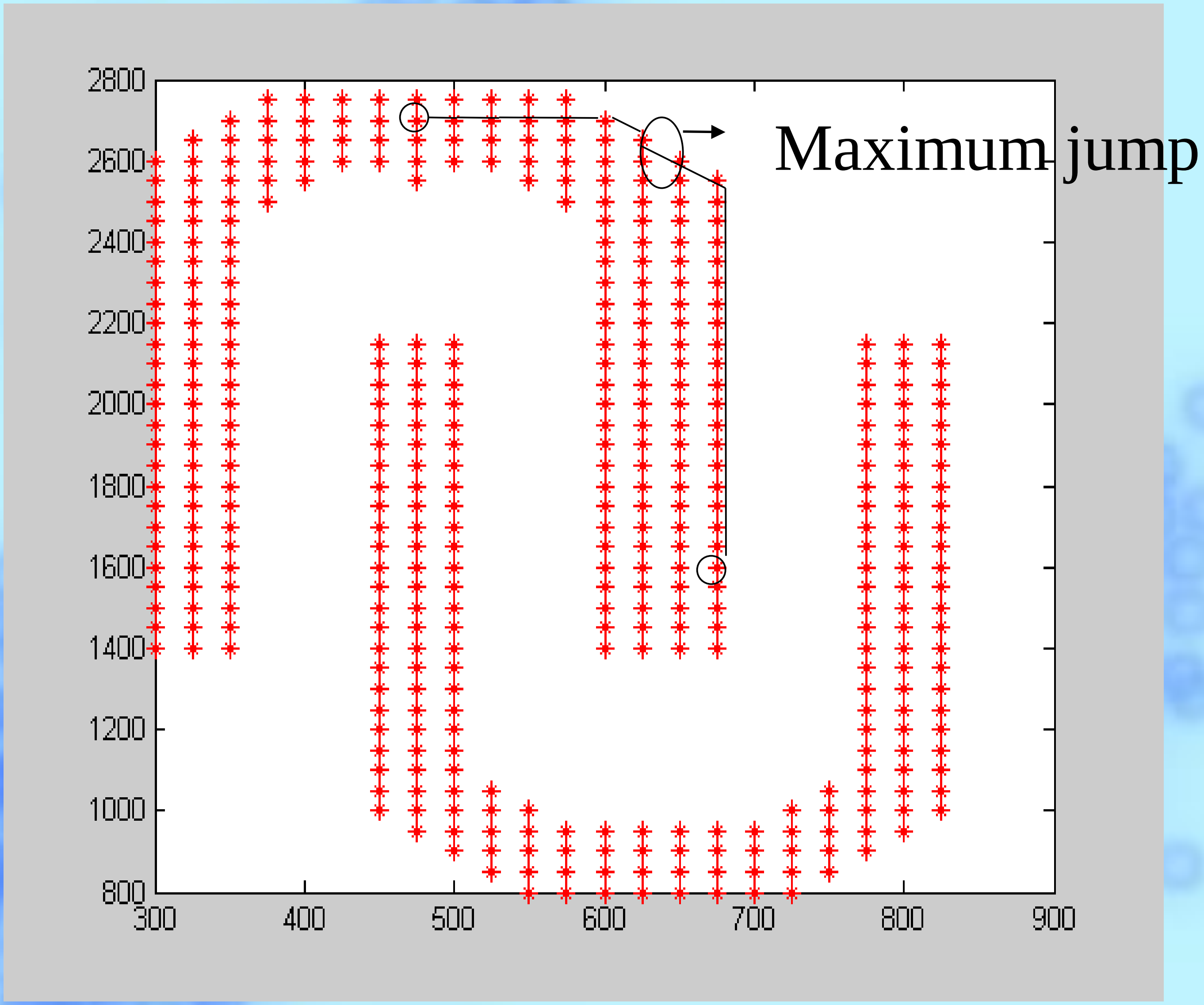
Connectivity Index

- **Find RNG of the whole data set**
- **Find all paths between medoid of a particular cluster and each point in that cluster along the RNG**
  - **Suppose there are total p paths between x and y**
  - **number of edges along the ith path : $nedge^i$, for i = 1, . . . , p**
  - **edges along the ith path denoted as $ed^i_1, \ldots, ed^i_{nedge^i}$ and the corresponding edge weights are $w(ed^i_1), \ldots, w(ed^i_{nedge^i})$,**
  - **shortest distance between x and y is defined as follows:**

$$d_{short}(p,q) = \min_{i=1}^{p} \max_{j=1}^{nedge_i} w(ed_j^i).$$

  - **Shortest distance : maximum jump along the path**
    - **Not the sum of the distances along the path**

Sriparna Saha, Sanghamitra Bandyopadhyay, Some connectivity based cluster validity indices, Applied Soft Computing, Volume 12, Issue 5, May 2012, Pages 1555–1565, ISSN 1568-4946 (impact factor: 2.415), H-index: 65



Maximum jump
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
300
400
500
600
700
800
900

Connectivity Index

Connectivity-index=

$$\frac{\sum_{i=1}^{K}\sum_{j=1}^{ni} \text{shortest\_distance}(\mathrm{m}_{\mathrm{i}}, \mathrm{x}^{i}_{j})}{n \times min\_sep}$$

– Here

min_sep=min$_{i,j=1 \text{ to } K,}$shortest_distance(m$_i$,m$_j$)

- **n_i= total number of points present in the ith cluster**
- **m_i: medoid of the ith cluster**
- **n=total number of data points**
- **Objective: minimize connectivity-index**

Sriparna Saha, Sanghamitra Bandyopadhyay, Some connectivity based cluster validity indices, Applied Soft Computing, Volume 12, Issue 5, May 2012, Pages 1555–1565, ISSN 1568–4946 (impact factor: 2.415), H-index: 65

Perturbation

- Any of the three types of perturbation operators can be applied

  –Modifying a cluster center;

  –Modifying each sub-cluster center by some small amount

  –Inserting a new sub-cluster center

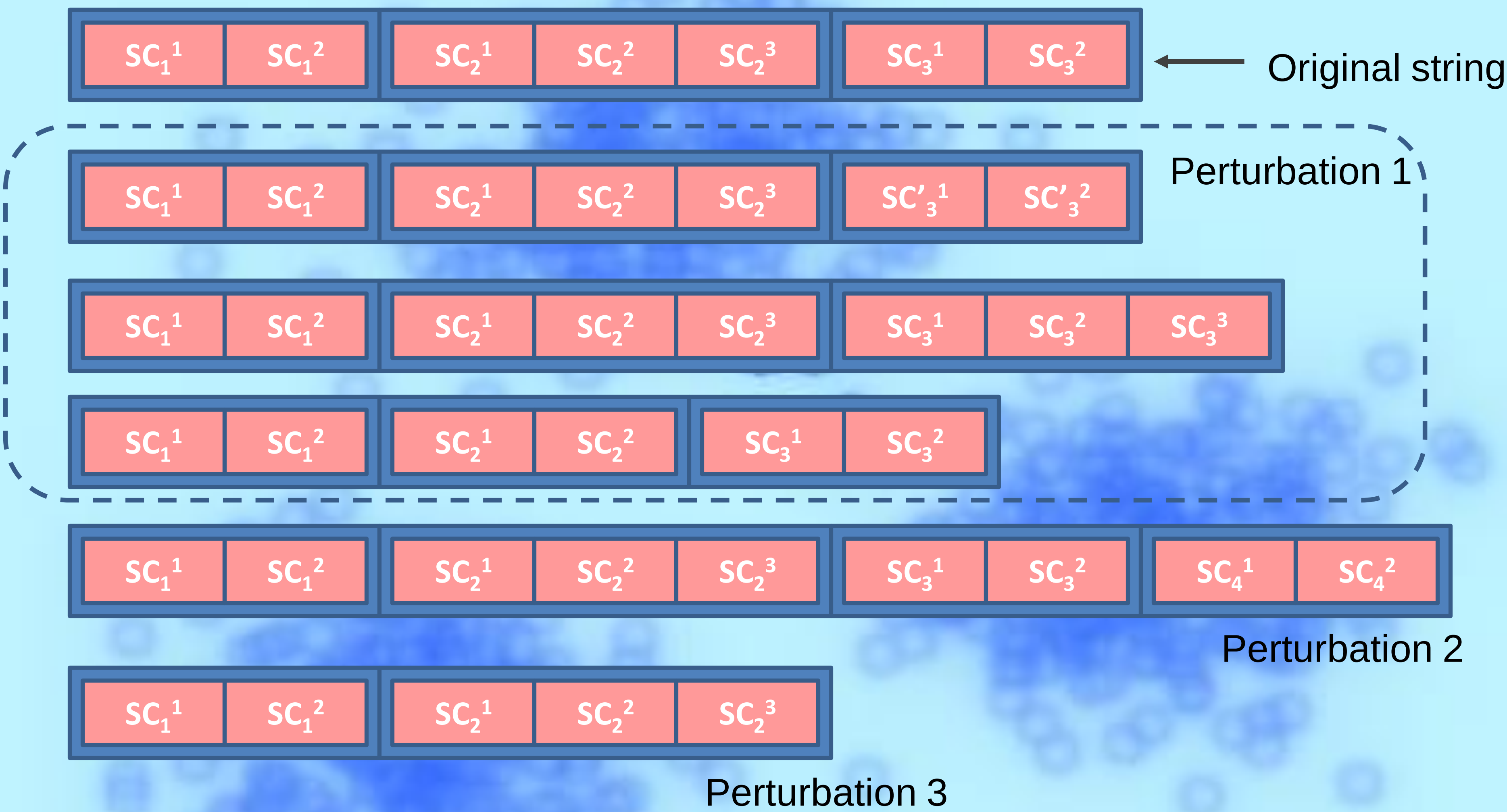
  –Deleting a sub-cluster center

  –increasing the number of whole clusters by 1

  – decreasing the number of whole clusters by 1

S. Saha and S. Bandyopadhyay, "A Generalized Automatic Clustering Algorithm in a Multiobjective Framework". *Applied Soft Computing Journal, Volume 13 (2013), Pages 89-108* (impact factor: 2.612).

Example:



S. Saha and S. Bandyopadhyay, "A Generalized Automatic Clustering Algorithm in a Multiobjective Framework". ***Applied Soft Computing Journal, Volume 13 (2013), Pages 89-108*** **(impact factor: 2.612).**

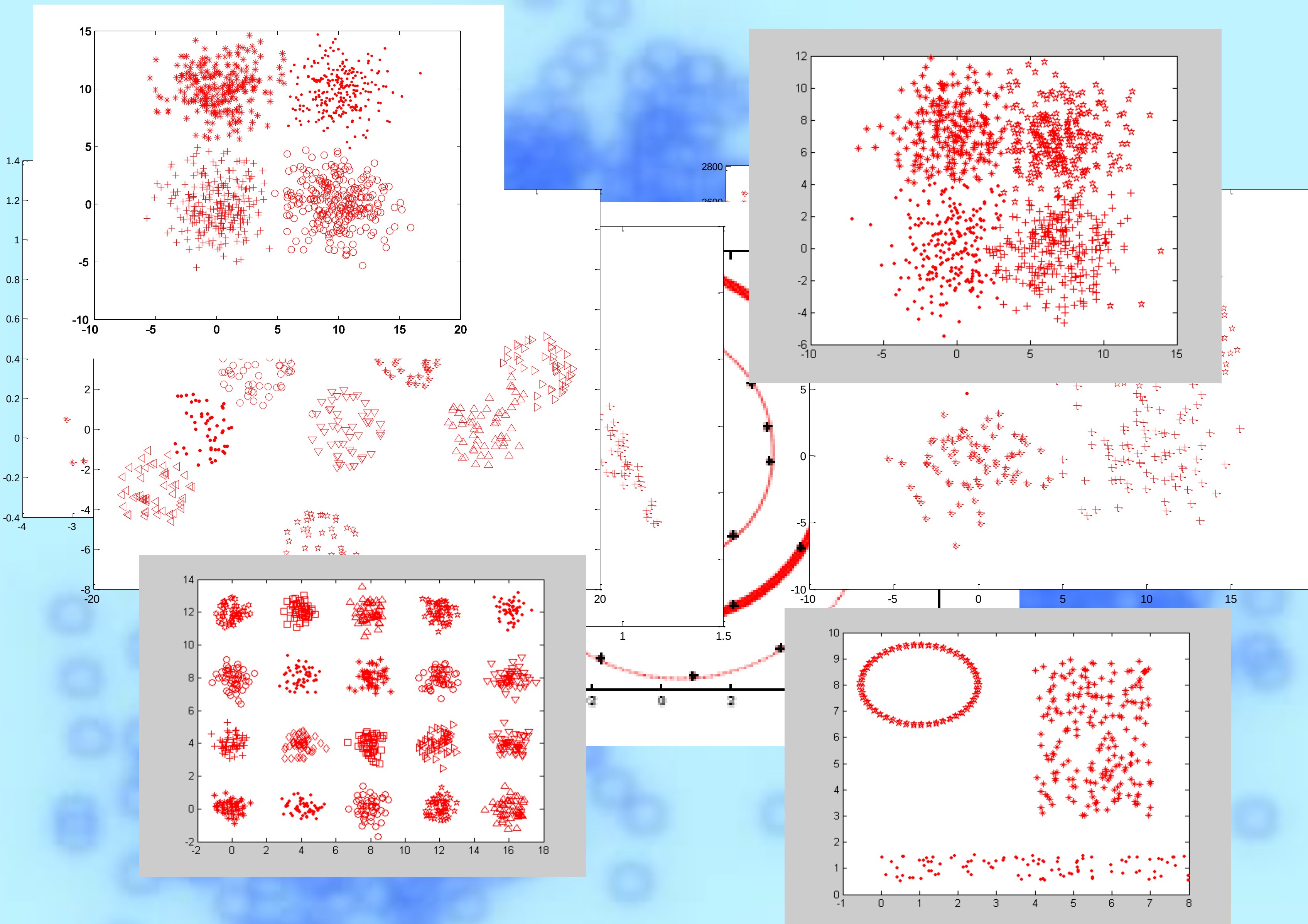
Experimental Results

- **Parameters of VGenClustMOO :**
  - **SL=100 HL=50, *iter*=50, Tmax=100, Tmin=0.00001.**
- **Comparisons done with:**
  - **MOCK, another multiobjective clustering technique**
  - **VAMOSA, a newly developed multiobjective clustering technique optimizing *Sym*-index and XB-index as two objective functions**
  - **VGAPS, a single objective genetic clustering technique with point symmetry based distance**
- **Minkowski score**
  - **calculated after application of each algorithm**
    - **MS: measure of the quality of a solution given the true clustering**
    - **MS: the optimum score is 0, with lower scores being "better"**

$$MS(T, S) = \sqrt{\frac{n_{01} + n_{10}}{n_{11} + n_{10}}}.$$

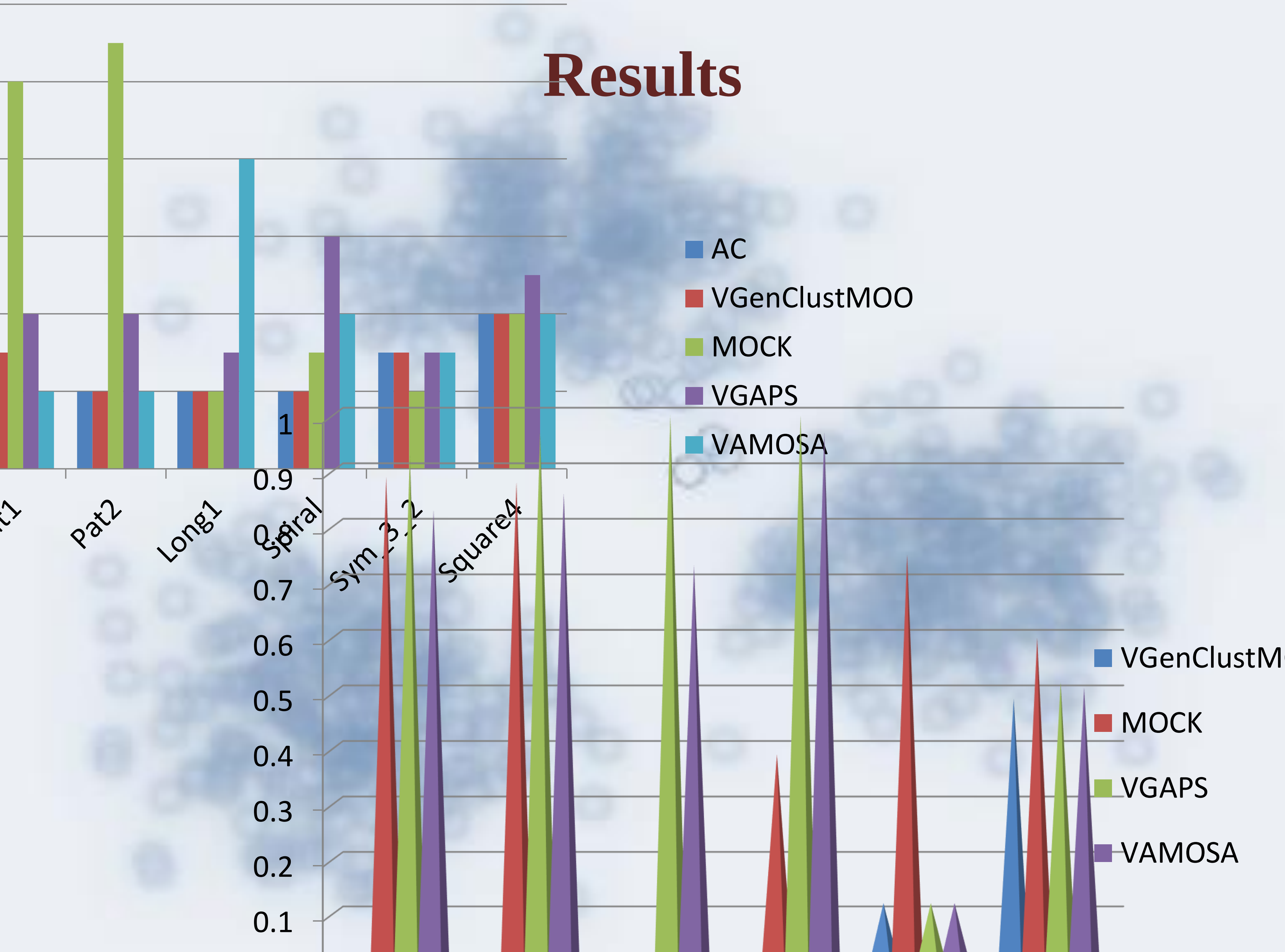
  - **The solution with minimum Minkowski Score**
    - **selected as the best solution**

S. Saha and S. Bandyopadhyay, "A Generalized Automatic Clustering Algorithm in a Multiobjective Framework". *Applied Soft Computing Journal, Volume 13 (2013), Pages 89-108* (impact factor: 2.612).

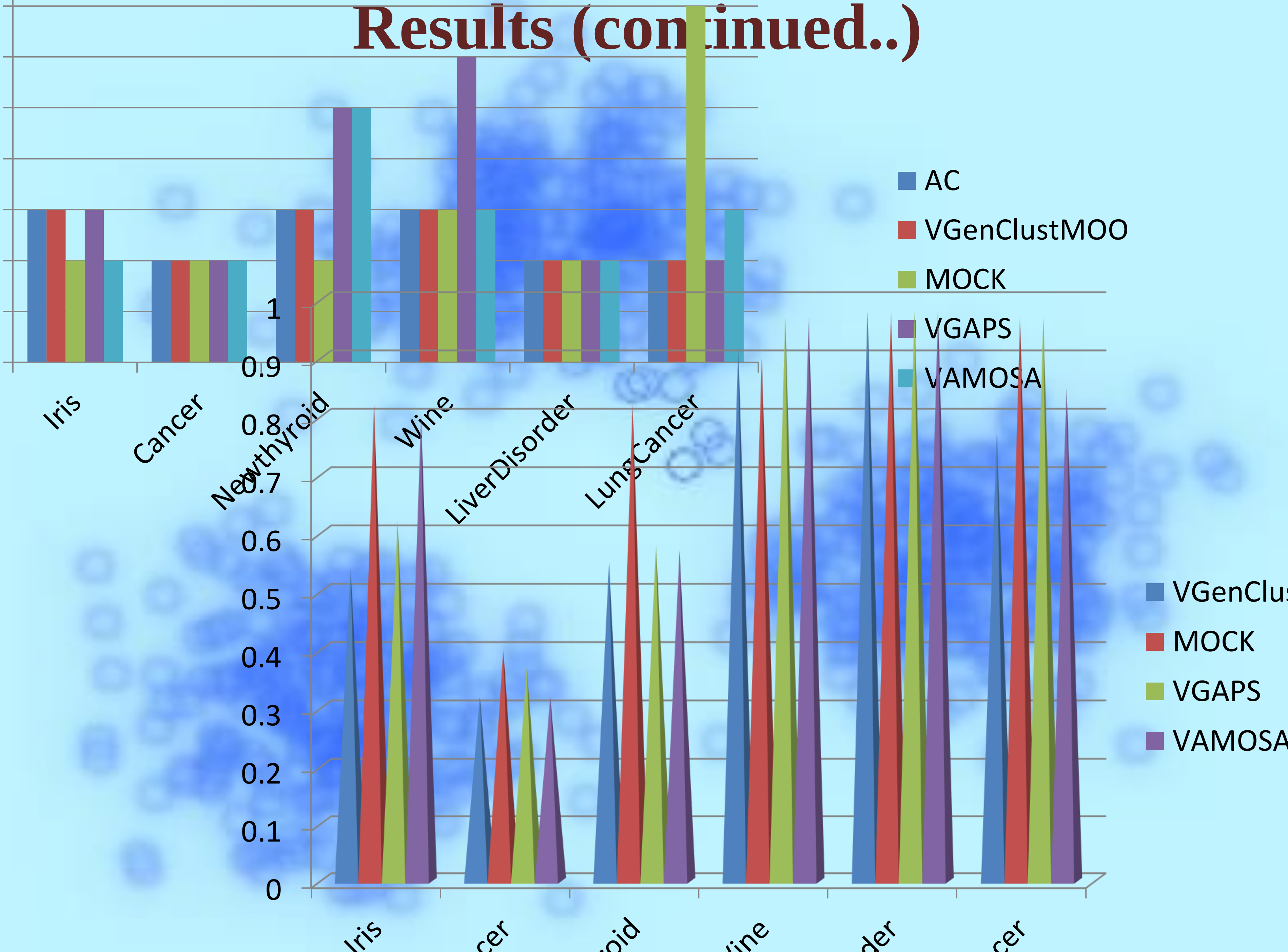


S. Saha and S. Bandyopadhyay, "A Generalized Automatic Clustering Algorithm in a Multiobjective Framework". ***Applied Soft Computing Journal, Volume 13 (2013), Pages 89-108*** **(impact factor: 2.612).**

Results


AC
VGenClustMOO
MOCK
VGAPS
VAMOSA
Pat2
Long1
Spiral
Sym_3_2
Square4
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
VGenClustM
MOCK
VGAPS
VAMOSA

Results (continued..)



Number of Sub-cluster Centers per Cluster

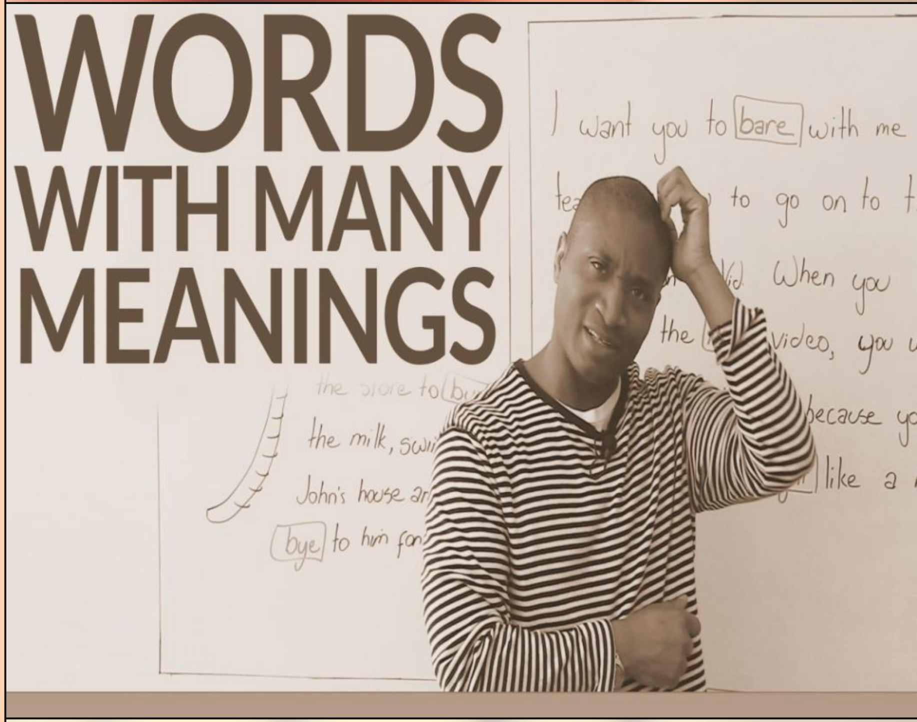
Data set	Actual no of Clusters	Obtained no of clusters	Sub-cluster centers per cluster
Mixed _3 _2	3	3	10,7,9
Sym_3 _2	3	3	4,10,3
AD _5 _2	5	5	9,10,10,9,4
AD _10 _2	10	10	7, 8, 10, 6, 10, 5, 7, 5, 6, 6
Long1	2	2	10,10
Sizes5	4	4	9,4,5
Spiral	2	2	14,16
Square4	4	4	3,3,2,5
Twenty	20	20	4 for 20 clusters
Iris	3	3	3,4,3
Cancer	2	2	3,4
Newthyroid	3	3	2,2,3

Solving search result clustering (SRC) and word sense induction (WSI) in multi-objective optimization framework

- Problem definition
- Literature survey and Motivation
- Working methodology
- Experiments and results

Word sense induction (WSI)

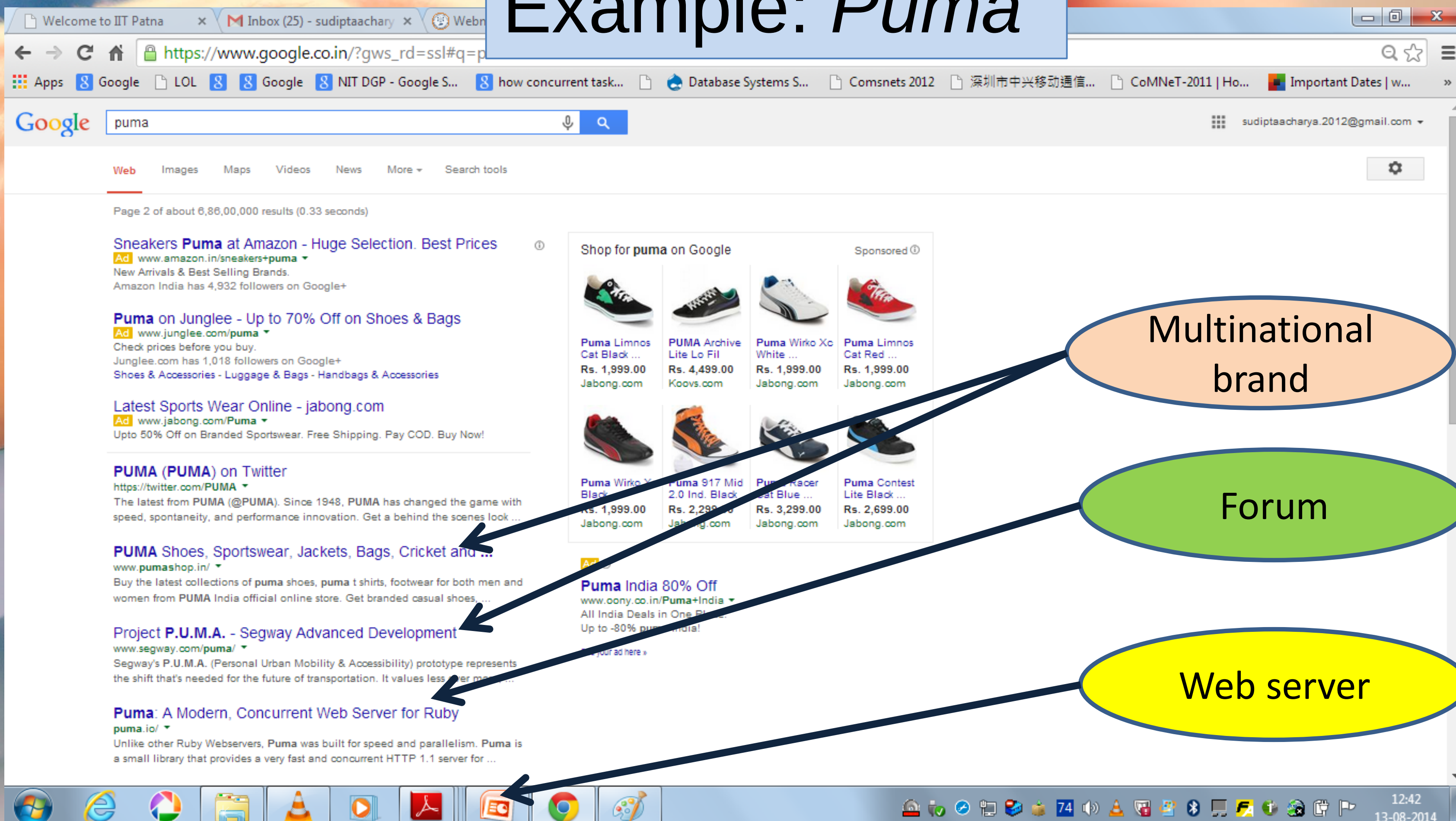
- One word → Several senses
  - Example:
    - I can hear ***bass*** sounds.→ Low frequency tones
    - They like grilled ***bass***.→ Type of fish
- **WSI** → Task of automatically inducing the different senses of a given word



End user application of WSI

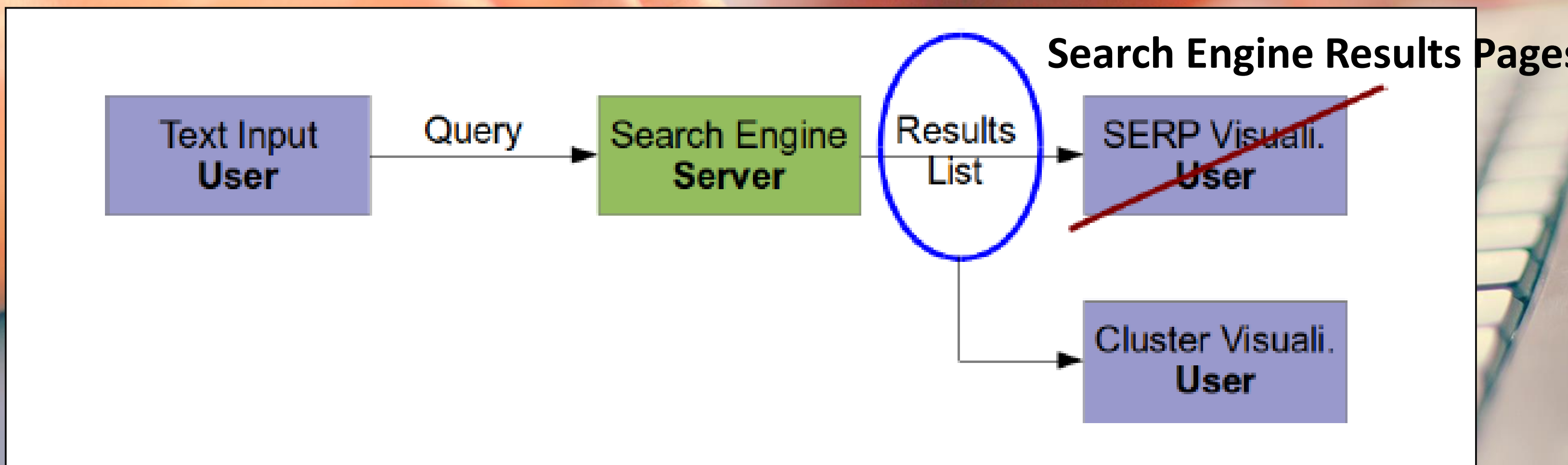
→ Web Search Result Clustering

Example: *Puma*



Working flow diagram of SRC technique

- Organizing results returned by a search engine in response to a keyword/query



Organize retrieved Web pages from a search Engine

Search Result Clustering (SRC)

Example SRC algorithms: Yippy, Carrot ,iBoogie, Vipaccess.

Our observations

❖SOO-based existing SRC techniques
❖No existing applications of MOO-based clustering to solve SRC
❖Study on smaller datasets

Our contribution

❖Developed a novel **MOO-based** clustering algorithm to solve SRC task namely ***MOO-clus***
❖AMOSA is used as underlying MOO strategy
❖Applied on large benchmark datasets

Chosen objective functions

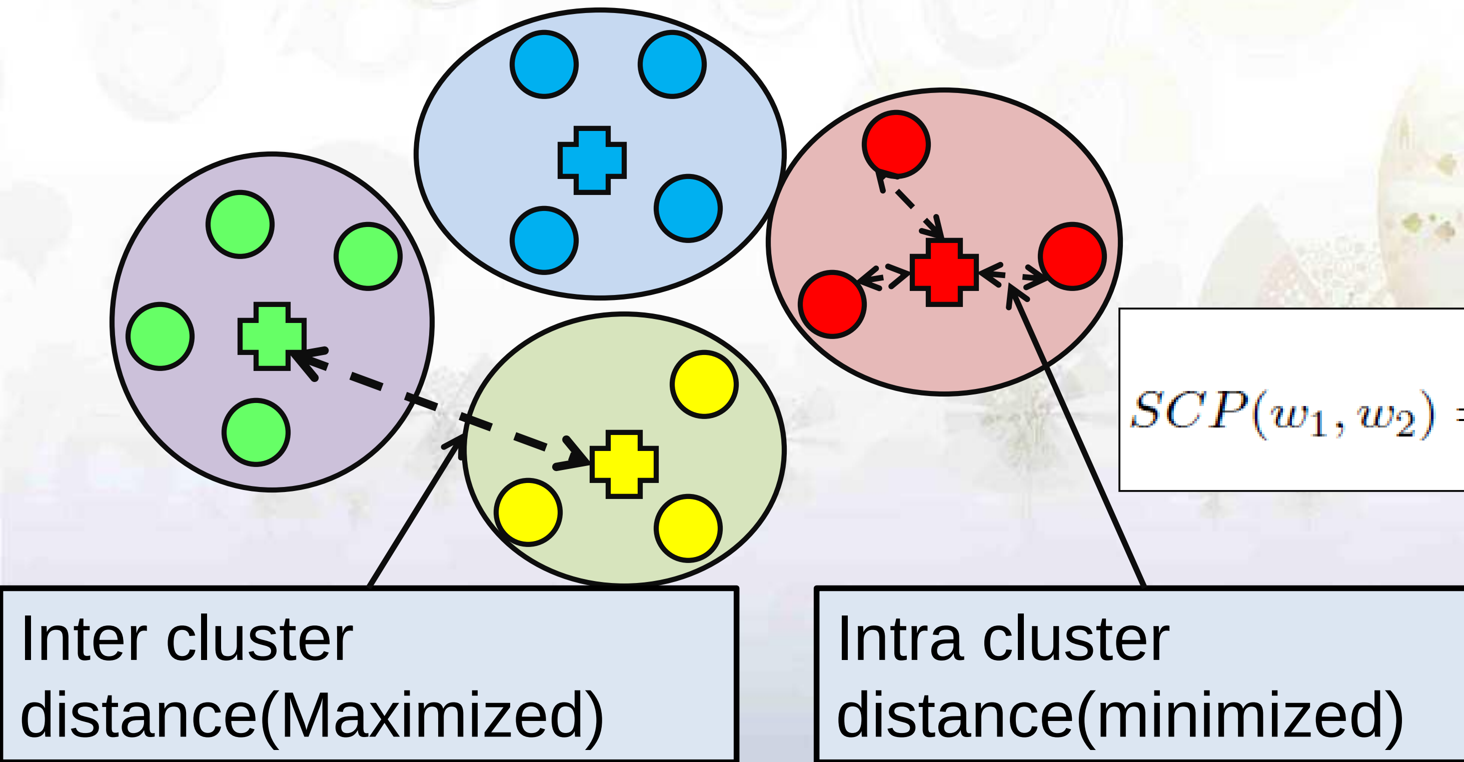
- Cluster Compactness→ maximize

$$COM = \sum_{k=1}^{K} \sum_{x_i \in \pi_k} S(x_i, m_{\pi_k})$$

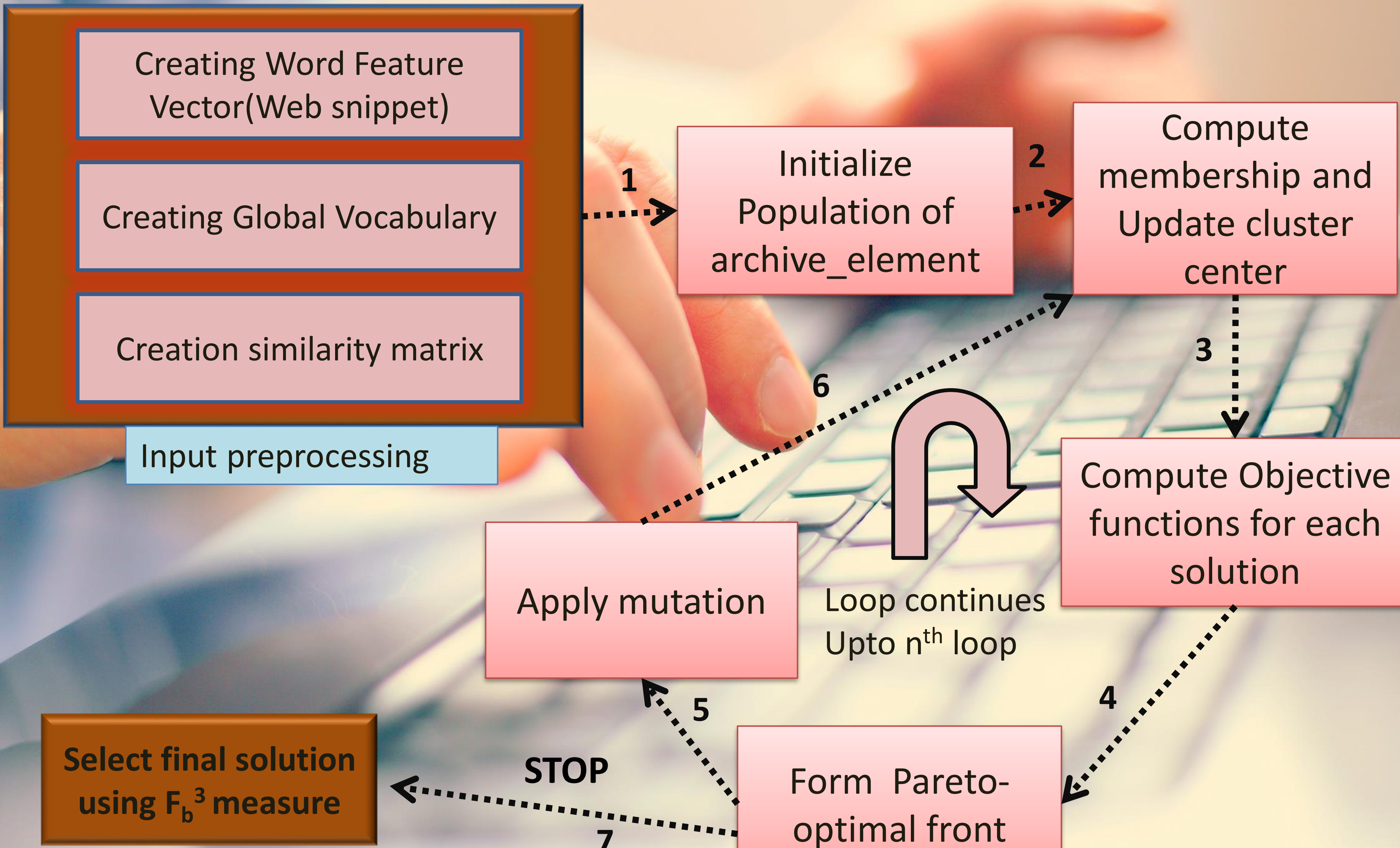
- Cluster separation→ minimize

$$SEP = \sum_{k=1}^{K} \sum_{o=k+1}^{K} S(m_{\pi_k}, m_{\pi_o})$$

$$SCP(w_1, w_2) = \frac{P(w_1, w_2)^2}{P(w_1) \times P(w_2)}$$



Steps of Proposed Algorithm



Working methodology

Creating Word Feature Vector(Web snippet)

- Word feature vector
- Based on tf-idf method most significant topical words are selected

Query 1:

OUTPUT: 4 snippets/documents:
d1→W1:1 W2:3 W3:2
d2→W4:2 W1:5 W5:1
d3→W6:3 W3:3 W1:1 W7:3
d4→W3:3 W8:2 W9:2 W2:2

Back

Working methodology

Global Vocabulary

1. W1
2. W2
3. W3
4. W4
5. W5
6. W6
7. W7
8. W8
9. W9

Back

Working methodology

Similarity matrix(Symmetric Conditional Probability)

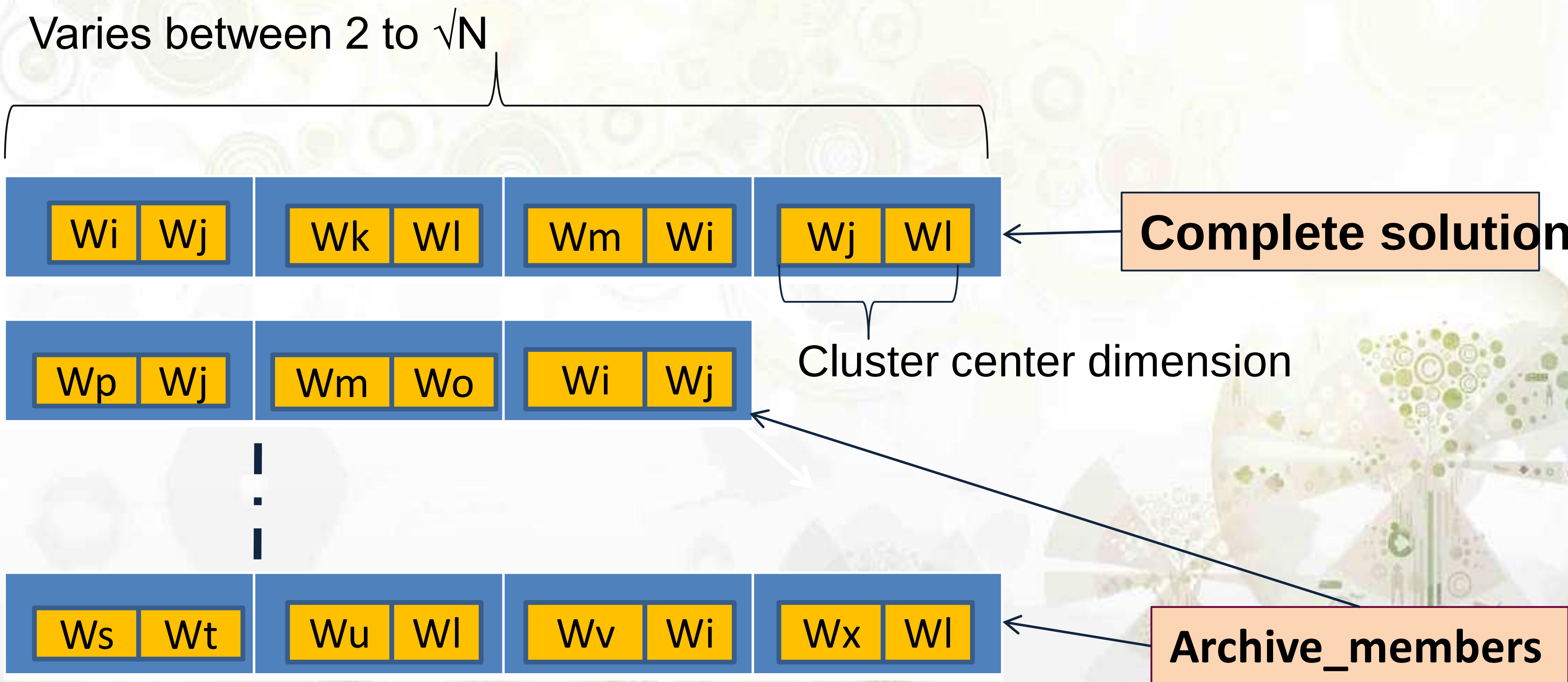
	W1	W2	W3	W4	W5	W6	W7	W8	W9
W1	1.0	0.8	0.3	0.0	0.2	0.1	0.3	0.0	0.6
W2	0.8	1.0	..	..	..	..	..	..	..
W3	0.3	..	..	..	..	..	..	..	..
W4	0.0	..	..	..	..	..	..	..	..
W5	0.2	..	..	..	..	..	..	..	..
W6	0.1	..	..	..	..	..	..	..	..
W7	0.3	..	..	..	..	..	..	..	..
W8	0.0	..	..	..	..	..	..	..	..
W9	0.6	..	..	..	..	..	..	..	..

$$SCP(w_1, w_2) = \frac{P(w_1, w_2)^2}{P(w_1) \times P(w_2)}$$

Back

Working methodology

String encoding of clustering solution and archive initialization



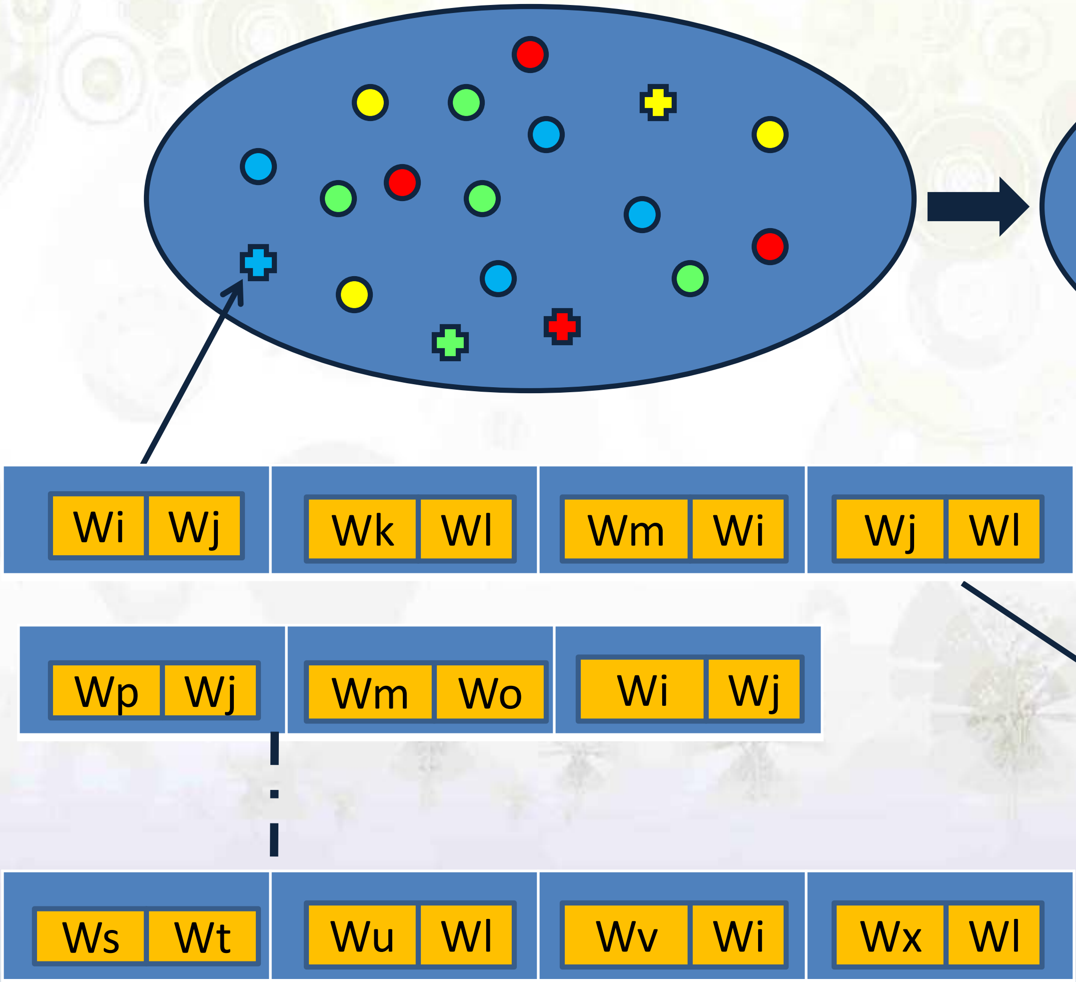
Where, $K_{max} = \sqrt{N}$

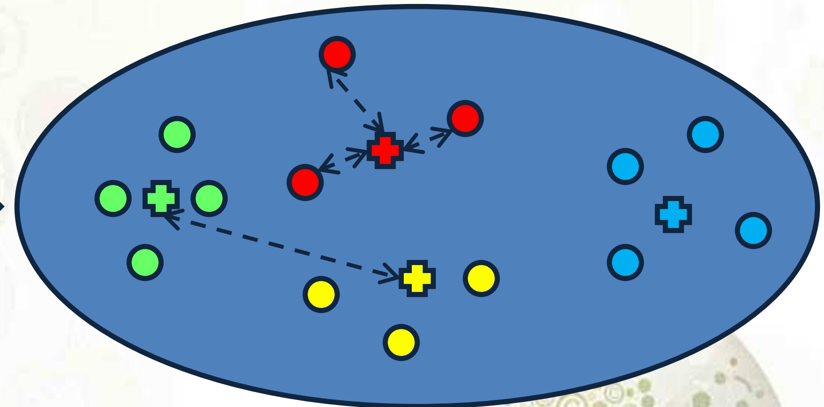
$N_d = (rand() mod(K_{max} - 1)) + 2$

Back

Working methodology

Membership calculation





	d1	d2	..	..	dm
C_1	0	1	0	1	0
C_2	1	0	0	0	0
C_3	0	0	1	0	0
C_4	0	0	0	0	1

Back

Working methodology

- For calculating membership SCP matrix is used to calculate similarity between
  - Each snippet and encoded cluster centers
- To find the SCP measure, each sample is first assigned to the maximum similarity cluster centers respectively. i.e.

$$i = \arg\max_{k=1 \ldots p} S(d_j, m_k)$$

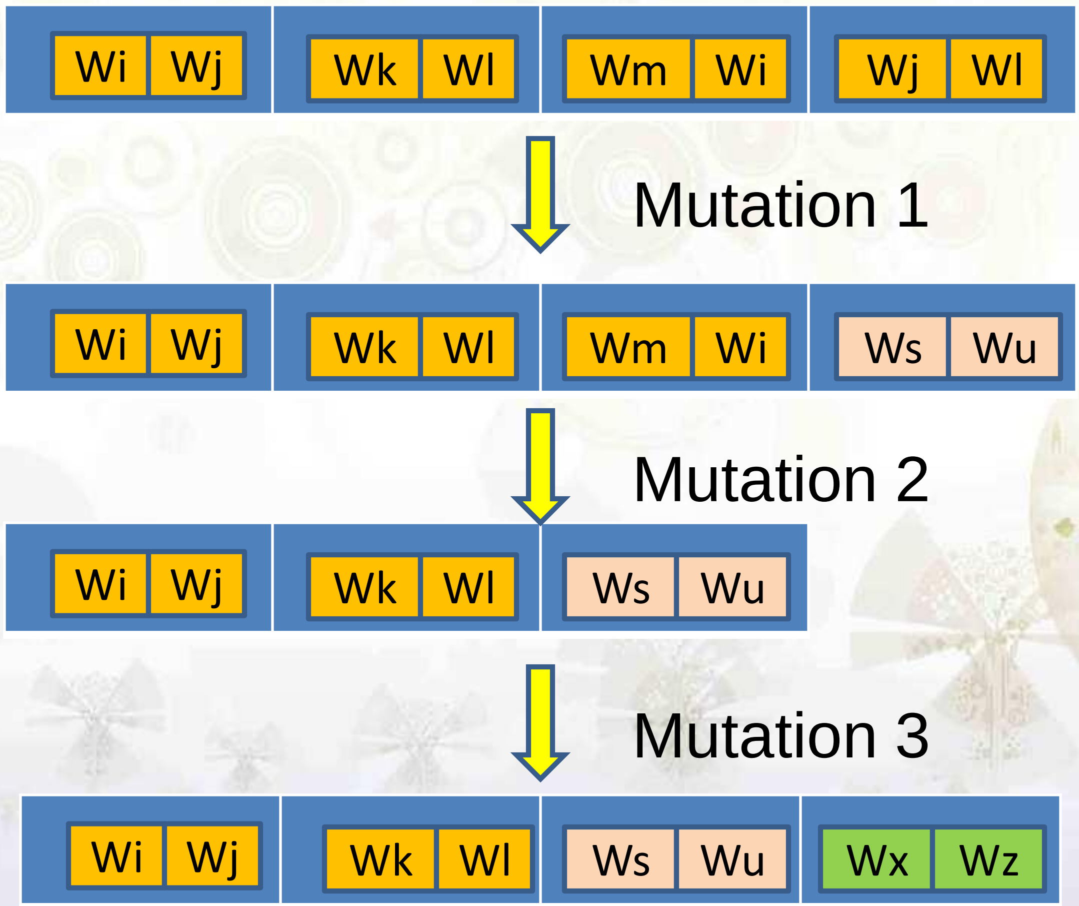
Where, p=Number of encoded clusters

$S(d_j, m_k)$ = denotes similarity measurement between the point *d_j and cluster centroid m_k defined in equation below*

$$S(d_i, d_j) = \frac{1}{\|d_i\| \|d_j\|} \sum_{r=1}^{\|d_i\|} \sum_{b=1}^{\|d_j\|} SCP(w_i^r, w_j^b)$$

Back

Search Operator: Mutation



Back

Experiment and results

Our chosen datasets:

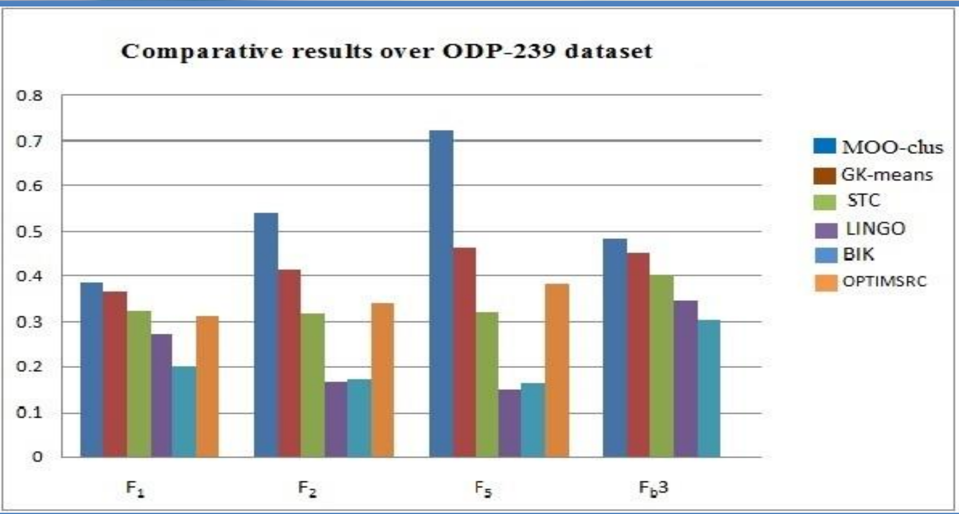
Dataset	# of queries	# of Snippets
ODP-239 (Carpineto and Romano, 2010)	**239**	**25580**
MORESQUE (Navigli and Crisafulli, 2010)	**114**	**11402**

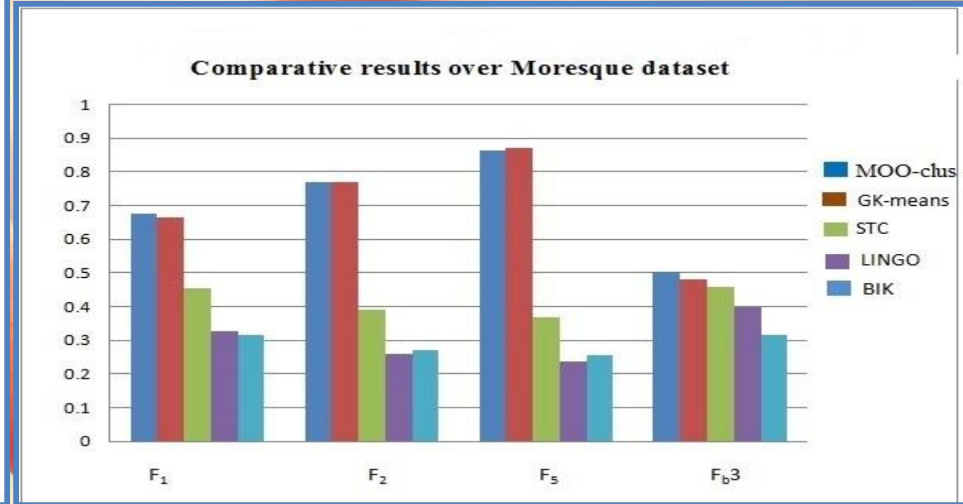
Our chosen evaluation metrics:

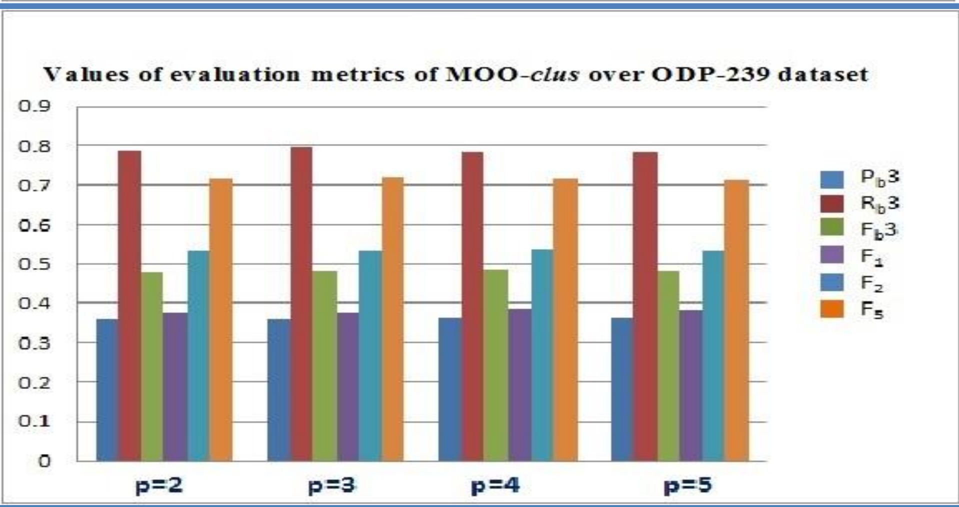
F_{β} (for β = 1, 2, 5) , F_{b}^{3}

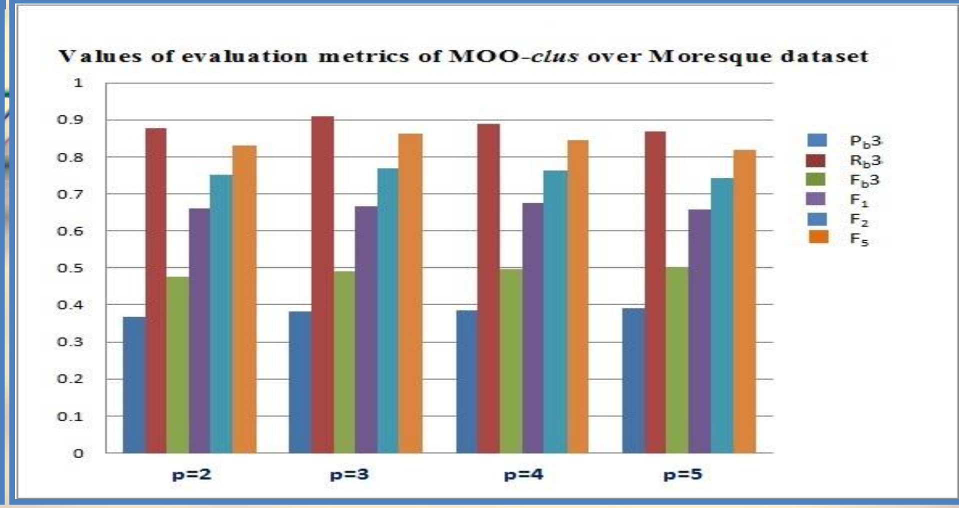
- Weighted average of Precision and recall
- Used to measure cluster quality i.e., Cluster Completeness, Rag-bag, Homogeneity, and cluster size vs. cluster number constraints.

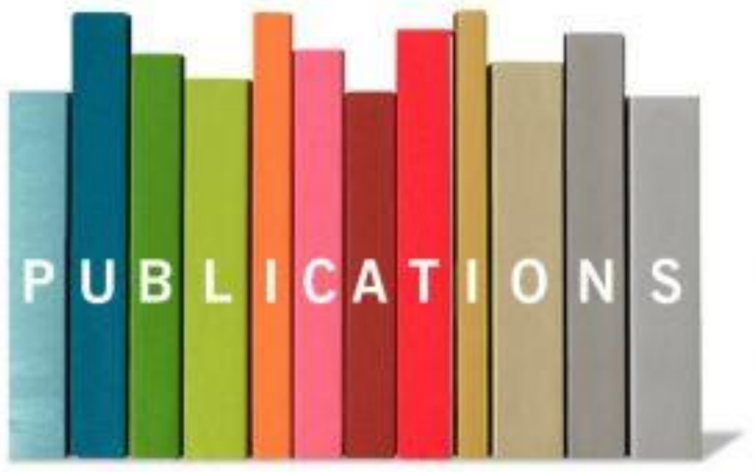
Experiment and results











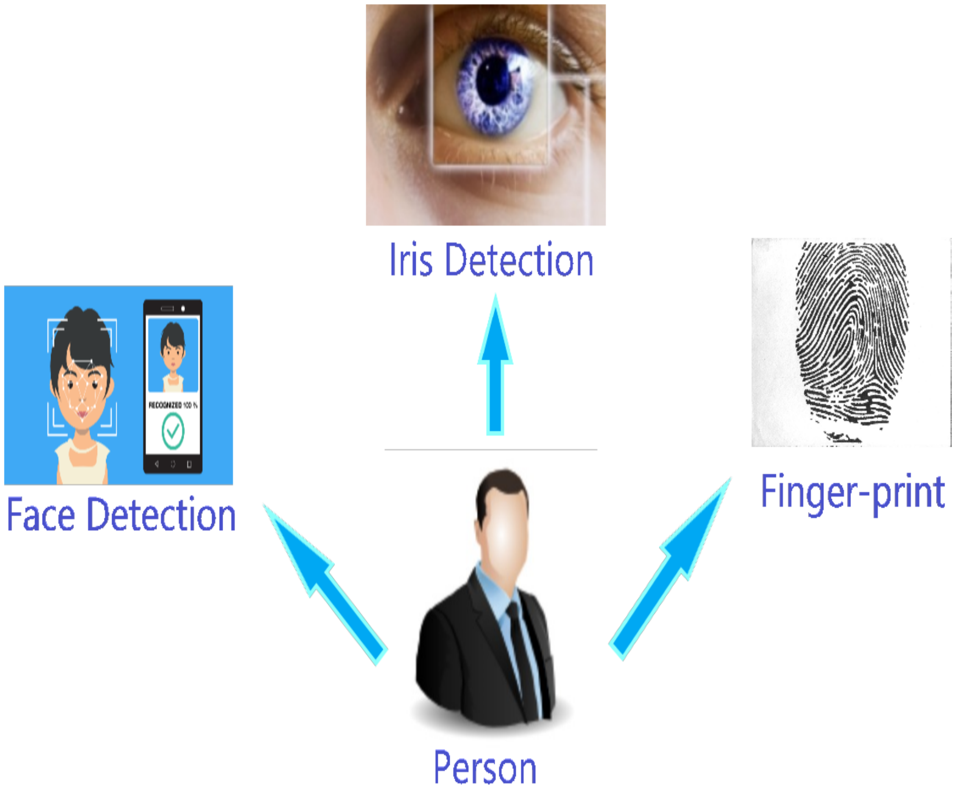
S. Acharya, S. Saha, J. G. Moreno and G. Dias, “Multi-Objective Search Results Clustering“, In Proceedings of **COLING 2014**, the 25th International Conference on Computational Linguistics, Dublin, Ireland, pg 99-108, August 2014**.(Core Ranking: A)**

S. Acharya, A. Ekbal, S. Saha, P. Santhanam, J. G. Moreno, G. Dias, “Multi-Objective Word Sense Induction based on Content and Interlink Connections", in the proceedings of 21st International Conference on Applications of Natural Language to Information Systems**(NLDB, 2016**) Vol. 9612, p. 366, Springer, June 22-24**.(Core Ranking: B)**

What is Multi-view Learning?

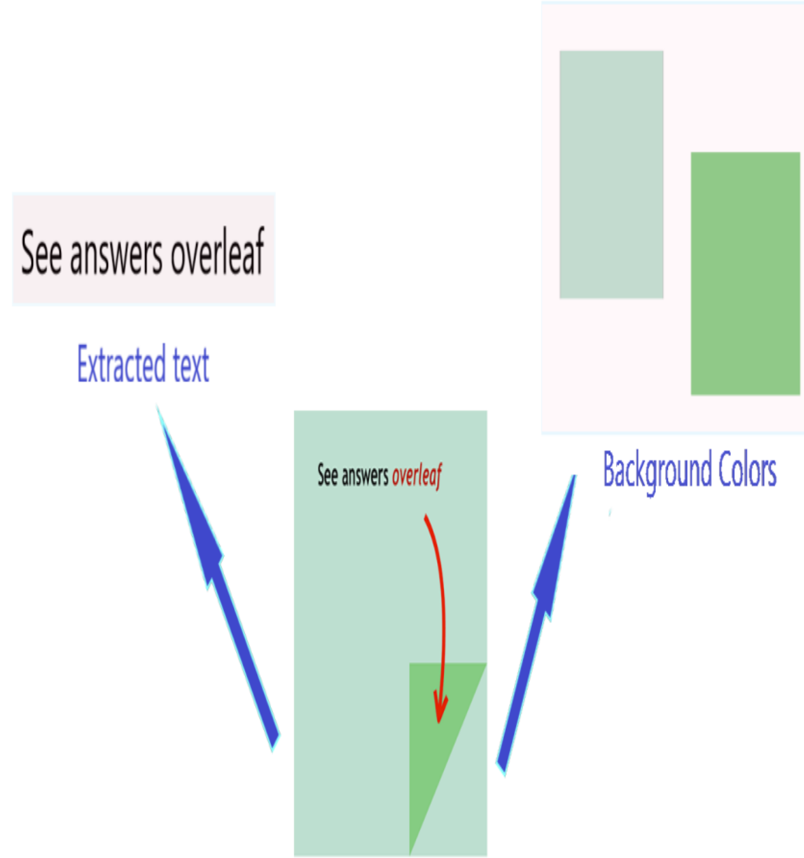
Multi-view Data Sets

- Data sets can be represented using different views.
- View/Modality
  - a certain type of information and/or the representation format in which information is stored
  - obtained from multiple sources or different feature subsets
- Learning from multimodal (or multi-view) data by considering the diversity of different views.



A person can be identified by face, fingerprint, signature or iris with information obtained from multiple sources.

Multi-view data: a) a web document can be represented by its url and words on the page, b) a web image can be depicted by its surrounding text separate to the visual information, c) video clips are combinations of audio signals and visual frames, d) multilingual documents have one view in each language.



An image can be represented by its color or texture features, which can be seen as different feature subsets of the image.

Literature Survey

Authors	Proposed Work
Xu et al. (2013)	A detailed survey on multi-view learning.
Bickel and Scheffer (2004)	Comparative study of multi-view versions of k-Means and EM with their single-view counterparts on text data. Reported a significant improvement in the performance of multi-view contrary to a single view.
Wahid et al. (2014)	Multi-view clustering of web documents using multiobjective genetic algorithm (NSGA II).
Angela Serra et al. (2015)	Proposed a multi-view approach in which the information from different data layers (genomic data) is integrated at the levels of the results of each single view clustering iterations.
Handl *et al*.(2006)	Multiobjective optimization technique: PESA-II, Objective functions : cluster compactness , cluster separability;
S. Acharya et al. (2016)	Proposed a new multiobjective algorithm for classifying RNA datasets
Saha *et al*.(2009)	Genetic algorithm, objective functions: cluster impurity and symmetrical compactness, point symmetry distance
Mukhopadhyay et al., 2015	A survey on existing multiobjective evolutionary clustering techniques
Mukhopadhyay et al. 2012	Aims find the best set of validity indices that should be optimized simultaneously to obtain good clustering results; A novel interactive genetic algorithm based multiobjective approach is proposed that simultaneously finds the clustering solution as well as evolves the set of validity measures that are to be optimized simultaneously;

Motivation

- Most of the existing multi-view clustering techniques are SOO
- A single quality measure for partitioning optimized implicitly or explicitly

None of the existing multiobjective based multi-view approaches follows a partitional approach to clustering a data set described by multiple views

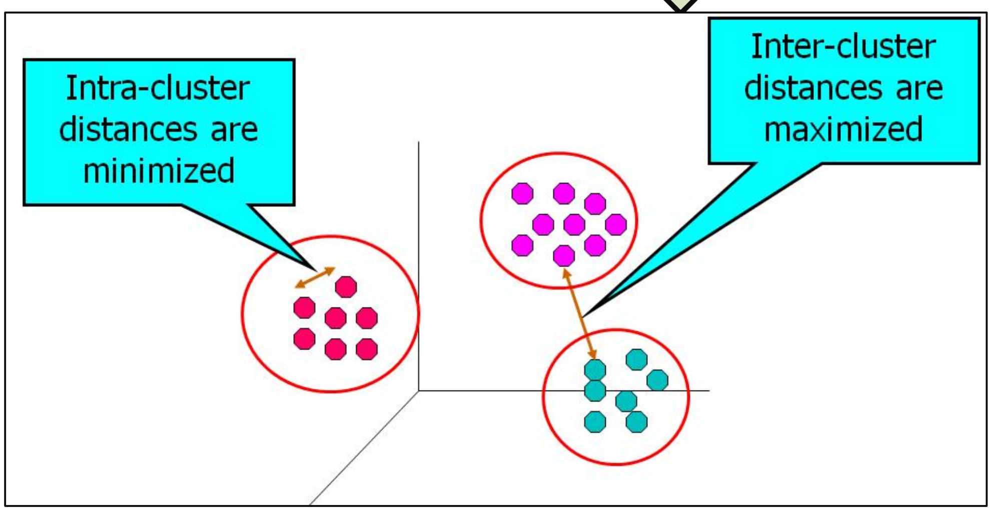
A way of automatically determining the optimal partitioning satisfying multiple views

- Development of some multiobjective multi-view approach
  - Automatic approach to determine the appropriate number of clusters

MOO based Multi-view Clustering

Three different paradigms in one frame!!

MOO + Clustering + Multi-views of data



+ Agreement between partitionings obtained using different views

Multi objective optimization based multi-view clustering technique

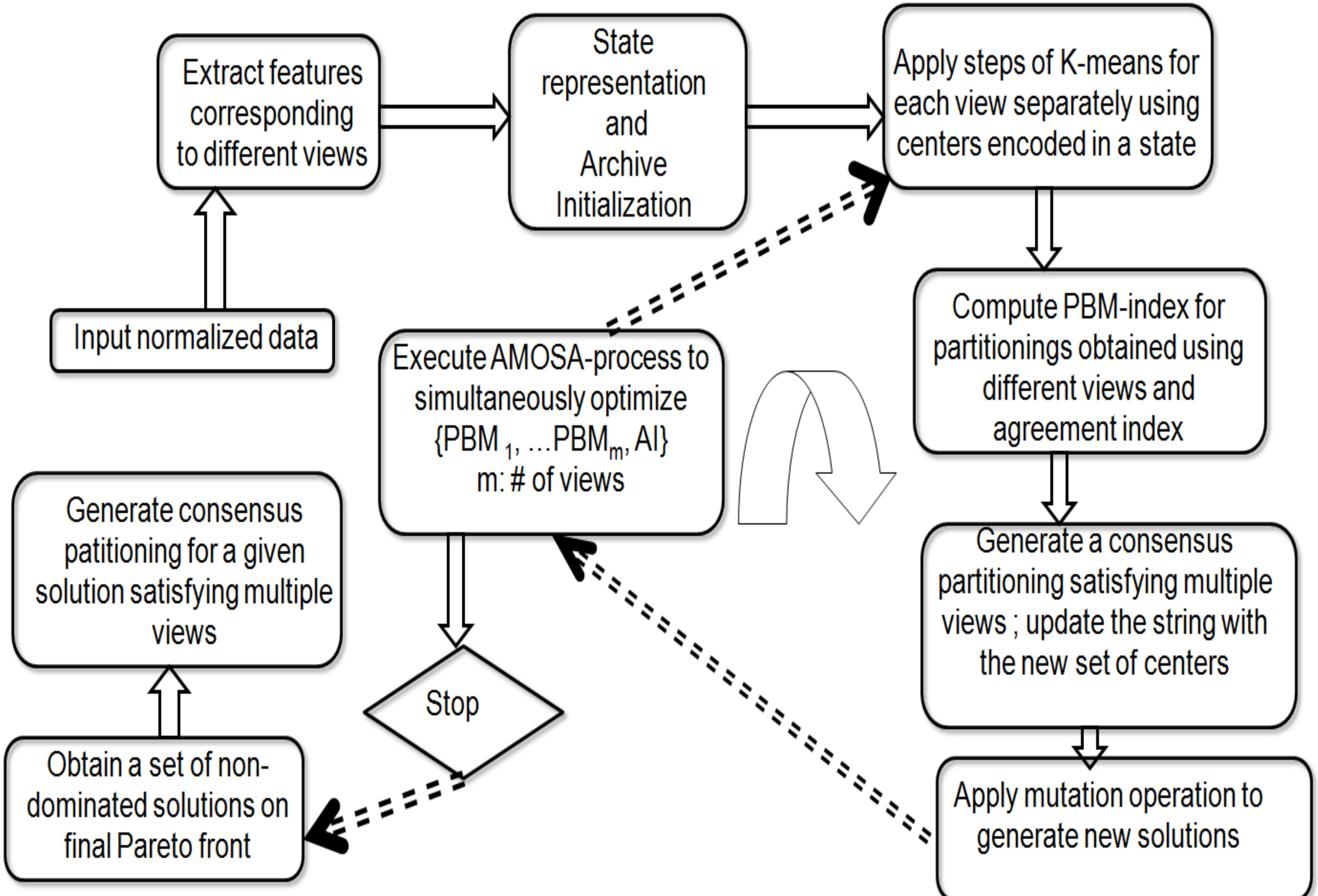
Problem Formulation

The multi-view clustering problem is formulated as a multiobjective optimization problem.

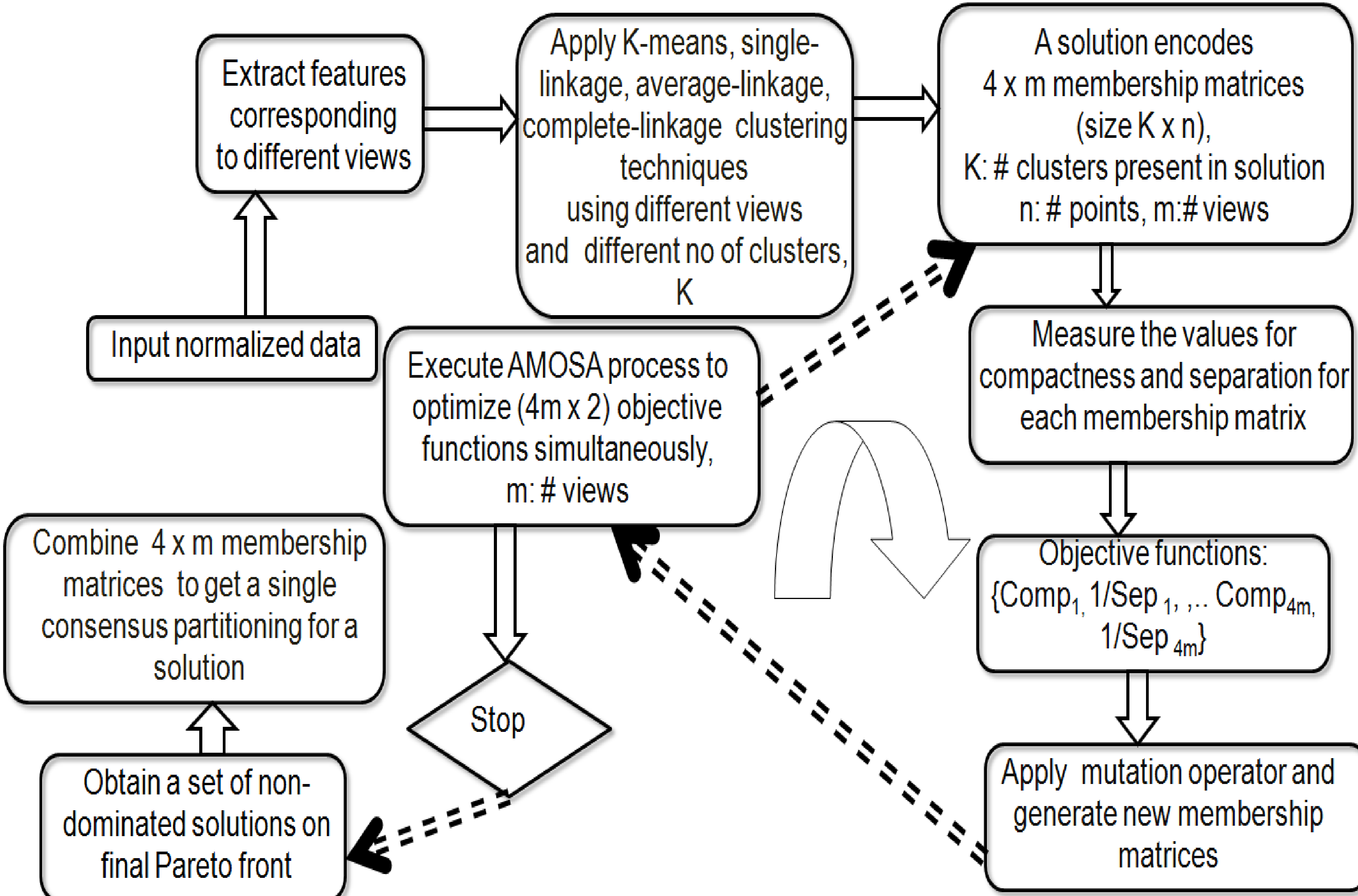
- **Given:**
  - **A data set of n samples S = {x1, x2, . . . , xn},**
  - **each described by d different features**
  - **and in total m different views,**
  - **and a set of objective functions**
    - **$CV_1, CV_2, \ldots, CV_m$, AI,**
    - **CV_i :cluster validity index computed on partitioning obtained using view I**
    - **AI: agreement index**
- **Find:**
  - **A consensus partitioning (U) satisfying all views**
    - **The set of data-points, S, is divided into K clusters, $\{U_1, U_2, \ldots, U_K\}$**
    - **$U_i = \{x^i_1, x^i_2, \ldots, x^i_{n_i}\}$; n_i: number of points in cluster i; x^i_j : jth point of cluster i.**
    - **$\cup^K_{i=1} U_i = S$ and $U_i \cap U_j = \emptyset$ for all $i \neq j$.**

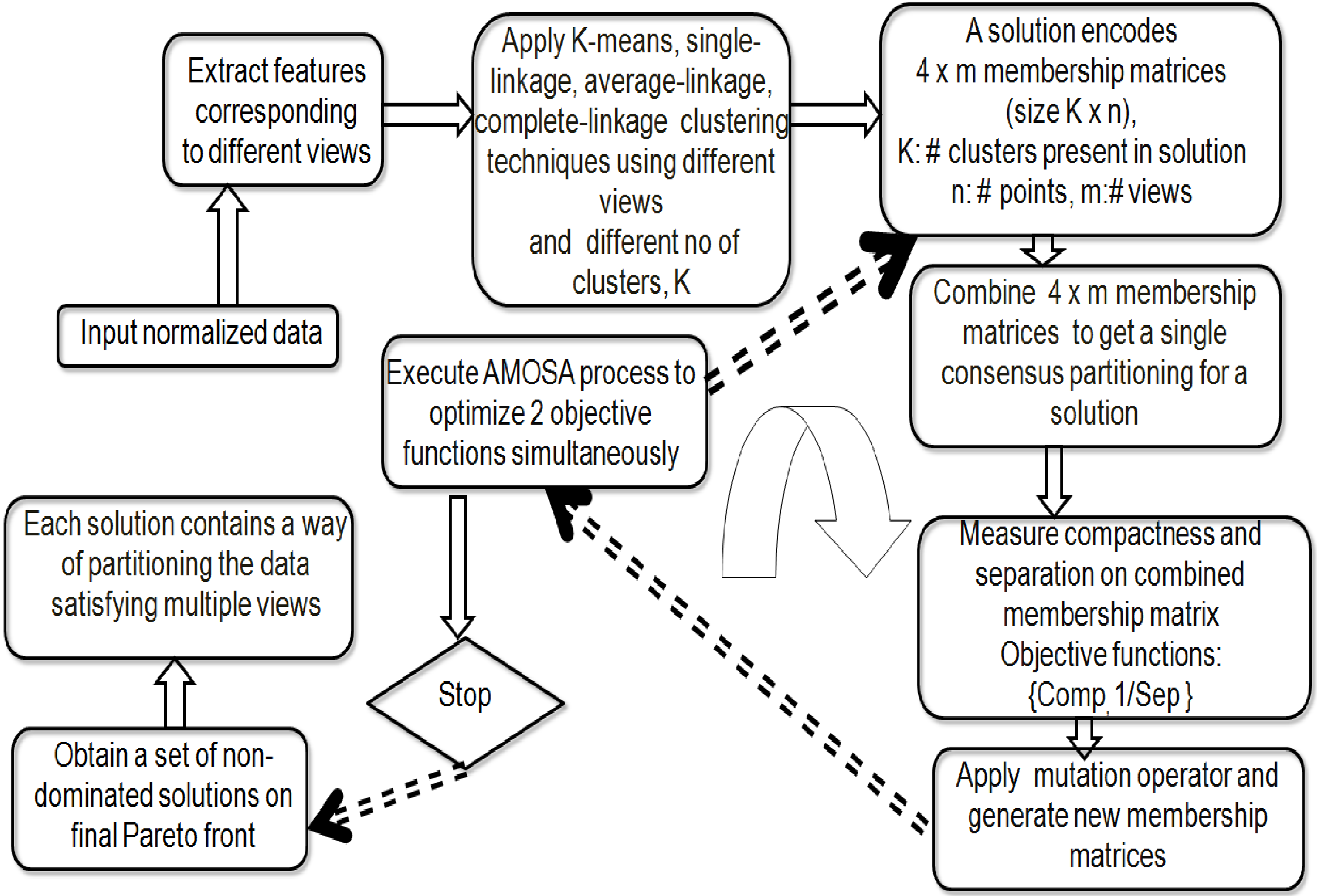
  - **which simultaneously optimizes the objective functions.**
  - **The simultaneous optimization of these objectives provides a Pareto optimal front.**

Flow-chart



S. Saha , S. Mitra, S. Kramer (2018): ``Exploring Multiobjective Optimization for Multi-view Clustering", **ACM Transactions on Knowledge Discovery from Data**, Vol. 12(4), Pages 44:1-44:30.


Input normalized data
Extract features corresponding to different views
Apply K-means, single-linkage, average-linkage, complete-linkage clustering techniques using different views and different no of clusters, K
A solution encodes 4 x m membership matrices (size K x n), K: # clusters present in solution n: # points, m:# views
Measure the values for compactness and separation for each membership matrix
Objective functions: {Comp$_1$, 1/Sep$_1$, ,.. Comp$_{4m}$, 1/Sep$_{4m}$}
Apply mutation operator and generate new membership matrices
Execute AMOSA process to optimize (4m x 2) objective functions simultaneously, m: # views
Stop
Obtain a set of non-dominated solutions on final Pareto front
Combine 4 x m membership matrices to get a single consensus partitioning for a solution



Extract features corresponding to different views
Apply K-means, single-linkage, average-linkage, complete-linkage clustering techniques using different views and different no of clusters, K
A solution encodes 4 x m membership matrices (size K x n), K: # clusters present in solution n: # points, m:# views
Input normalized data
Execute AMOSA process to optimize 2 objective functions simultaneously
Combine 4 x m membership matrices to get a single consensus partitioning for a solution
Each solution contains a way of partitioning the data satisfying multiple views
Measure compactness and separation on combined membership matrix Objective functions: {Comp, 1/Sep}
Stop
Obtain a set of non-dominated solutions on final Pareto front
Apply mutation operator and generate new membership matrices

String Representation

- Let us assume that each sample point has *m* different views, $V_1, V_2, \ldots, V_m$.
- Distribution of features in *m* different views is available,

$$\sum_{i=1}^{m} |Vi| = d$$

- Archive member i represents the centroids of K_i clusters
  - length l_i , where $l_i = d * K_i$.
  - $K_i = (rand()mod(K_{max} - 1)) + K_{min}$

Attributes of view 1 for cluster center 1

Attributes of view2 for cluster center 1

Attributes of view1 for cluster center 2

Attributes of view2 for cluster center 2

$C_1^1(1)$	$C_1^1(2)$	$C_1^1(3)$	$C_1^2(1)$	$C_1^2(2)$	$C_2^1(1)$	$C_2^1(2)$	$C_2^1(3)$	$C_2^2(1)$	$C_2^2(2)$

S. Saha , S. Mitra, S. Kramer (2018): ``Exploring Multiobjective Optimization for Multi-view Clustering", **ACM Transactions on Knowledge Discovery from Data**, Vol. 12(4), Pages 44:1-44:30

Formation of Clusters

- **K :number of clusters present in a solution**
- **Centers of a string:**

$$\{\overline{C}_1^d, \overline{C}_2^d, \ldots, \overline{C}_K^d\} = \{C_1^1, C_2^1, \ldots, C_d^1, \ldots, C_1^K, C_2^K, \ldots, C_d^K\};$$

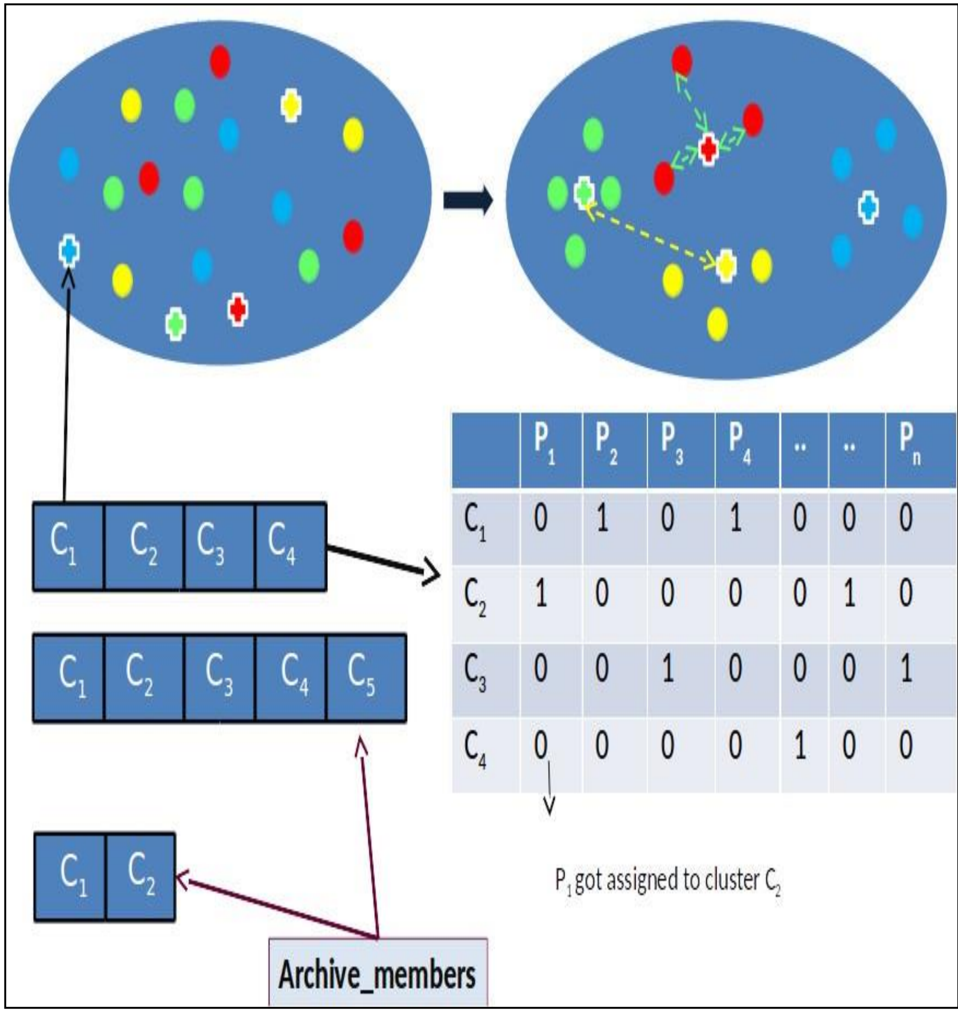
corresponding subset of cluster centers :

- **Formation of clusters:**

$$\pi_j^v = \{\forall \overline{x}_i^v \in \overline{X}^v : d(\overline{x}_i^v, \overline{C}_j^v) < d(\overline{x}_i^v, \overline{C}_l^v), l \neq j\}$$

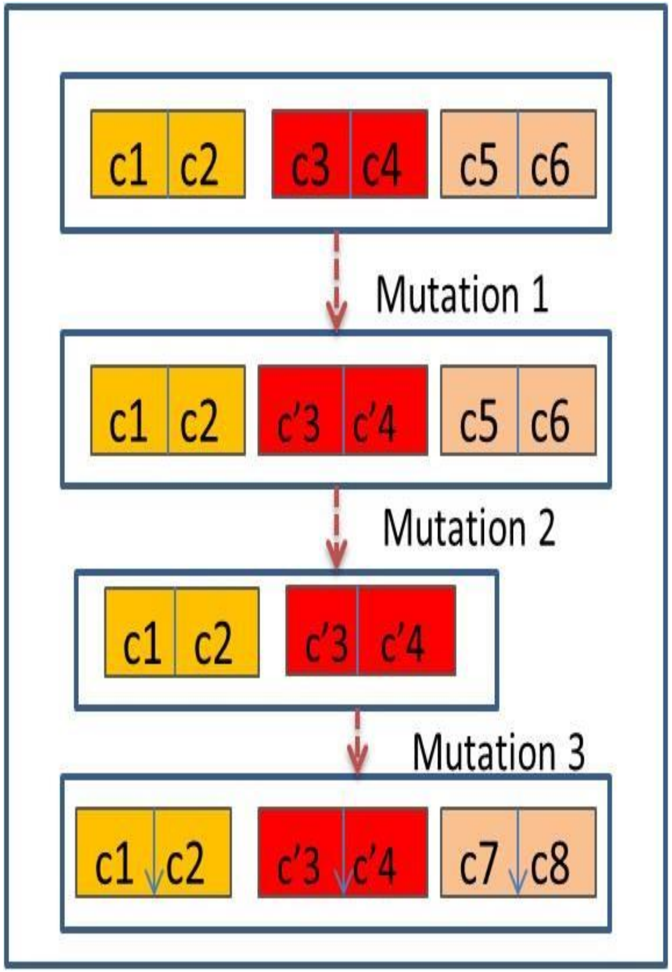
- **New cluster centers :**

$$\overline{c}_j^v = \frac{\sum_{\overline{x}^v \in \pi_j^v} \overline{x}^v}{n_j^v}$$



S. Saha , S. Mitra, S. Kramer (2018): ``Exploring Multiobjective Optimization for Multi-view Clustering", **ACM Transactions on Knowledge Discovery from Data**, Vol. 12(4), Pages 44:1-44:30

Mutation Operation



- Three types of mutation operations applied to generate new solution through AMOSA process:
  - In Mutation 1, some cluster centers present in a state are modified by some value.
  - In Mutation 2, total number of clusters present in a state is decreased by 1.
  - In Mutation 3, total number of clusters present in a state is increased by 1.

Objective Functions

- **PBM index:**

$$D_B = \max_{k<k'} d(C_k, C_{k'})$$

$$E_w = \sum_{k=1}^{K} \sum_{i \in U_k} d(x_i, C_k)$$

$$E_T = \sum_{i=1}^{N} d(x_i, C)$$

E_W= The sum of the distances of the points of each cluster to their center

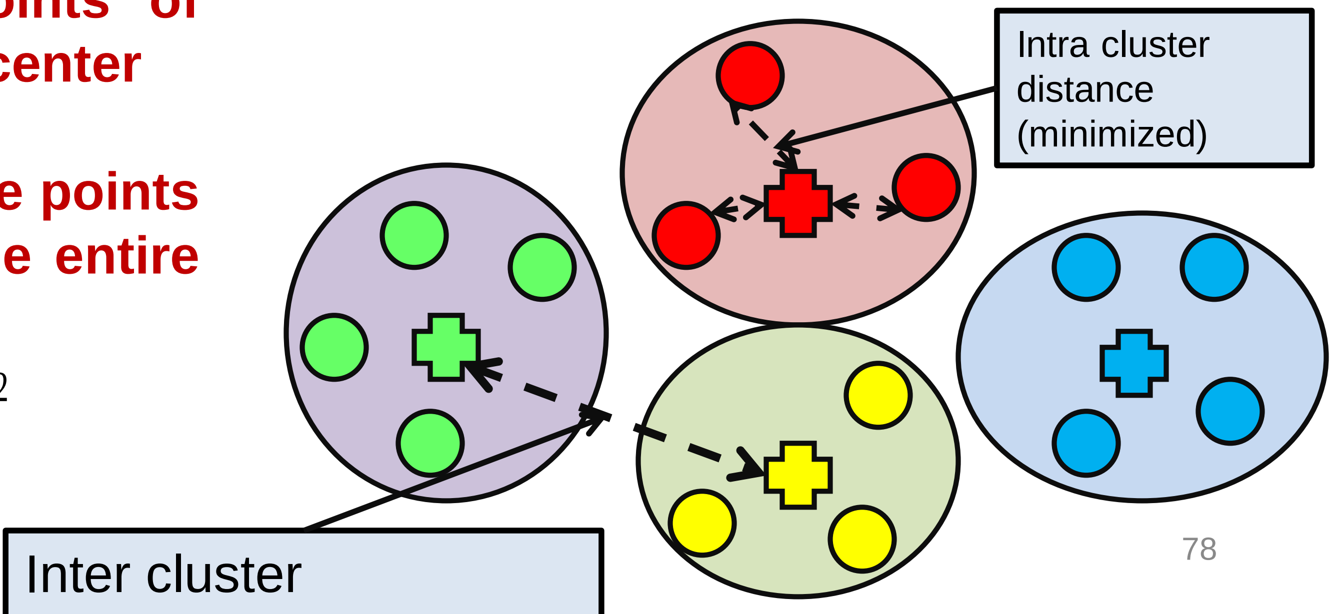
E_T= The sum of all the points to the center C of the entire dataset.

$$PBM = \left(\frac{1}{K} x \frac{E_T}{E_W} x D_B\right)^2$$

- **Xie-Beni (XB):**

$$XB = \frac{1}{N} \frac{\sum_{k=1}^{K} \sum_{i \in U_k} \| x_i - C_k \|^2}{\min\{d(C_i, C_j) \mid \forall i, j = 1 to K, i \neq j\}^2}$$

x_i = Point in cluster U_i , N = Total number of points in the dataset, C_k = center of cluster k, K = Total number of clusters



Objective Functions

- **Agreement Index**

A new objective function *Agreement Index (AI)* is calculated ; $n_a = \sum_{i=1}^{n} \sum_{j=1}^{n} I_{A_{ij}^{v1}, A_{ij}^{v2}}$, here

$$
\begin{aligned}
I_{A_{ij}^{v1}, A_{ij}^{v2}} &= 1 \quad \text{if } \; A_{ij}^{v1} = A_{ij}^{v2} \\
&= 0 \quad \text{otherwise}
\end{aligned}
$$

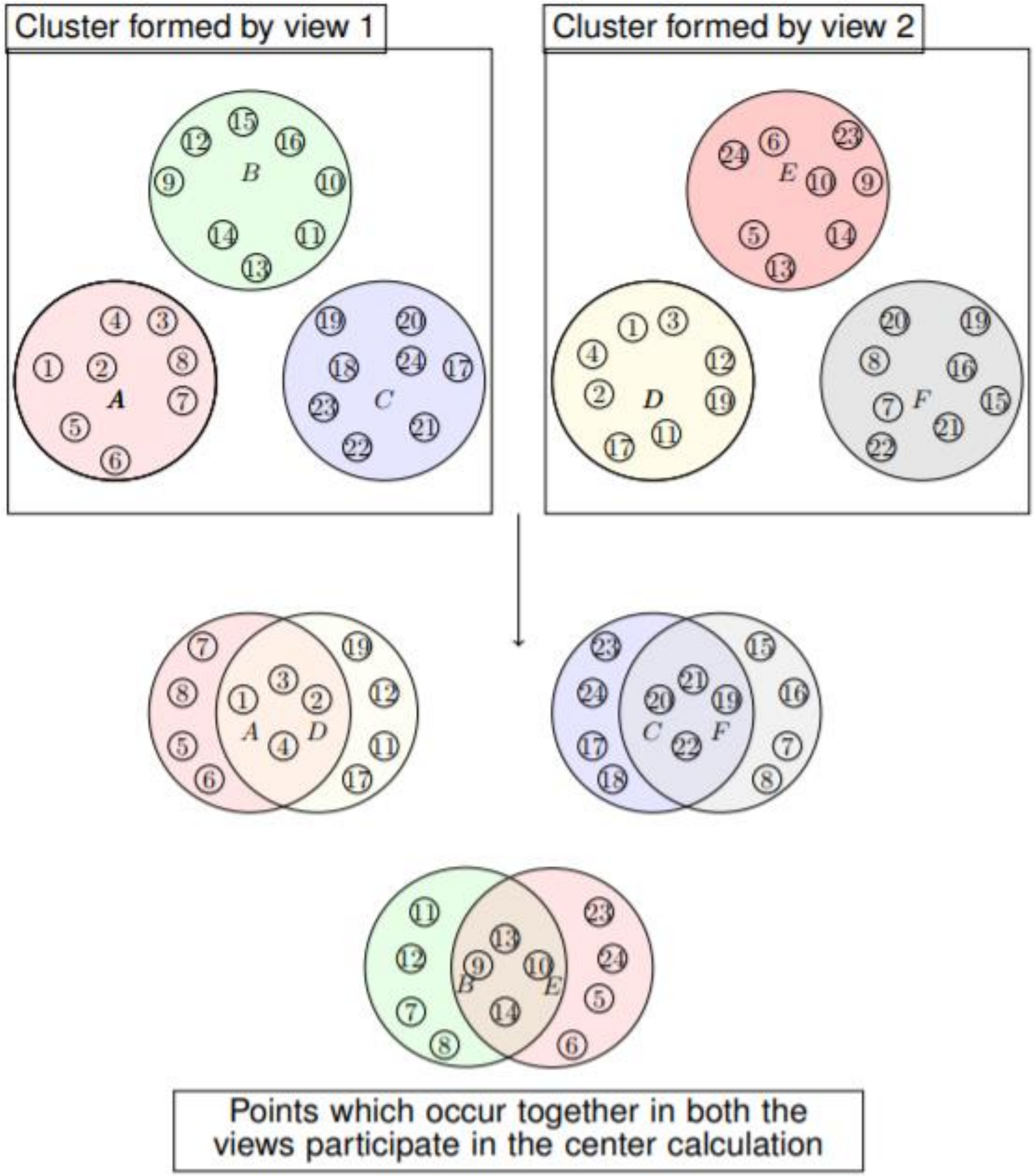
$$
\begin{aligned}
A_{ij}^{v} &= 1 \text{ if } \overline{x}_i \text{ and } \overline{x}_j \text{ belong to the same cluster} \\
&= 0 \quad \text{otherwise} \\
&= 1 \quad \text{if } i = j
\end{aligned}
$$

$$n_d = n^2 - n_a$$

$$AI_{v1,v2} = \frac{n_a + 1}{n_d + 1}$$

$$AI = \frac{\sum_{i=1}^{m} \sum_{j=1, j \neq i}^{m} 2 \times AI_{v_i, v_j}}{m \times (m-1)},$$

Consensus Partitioning



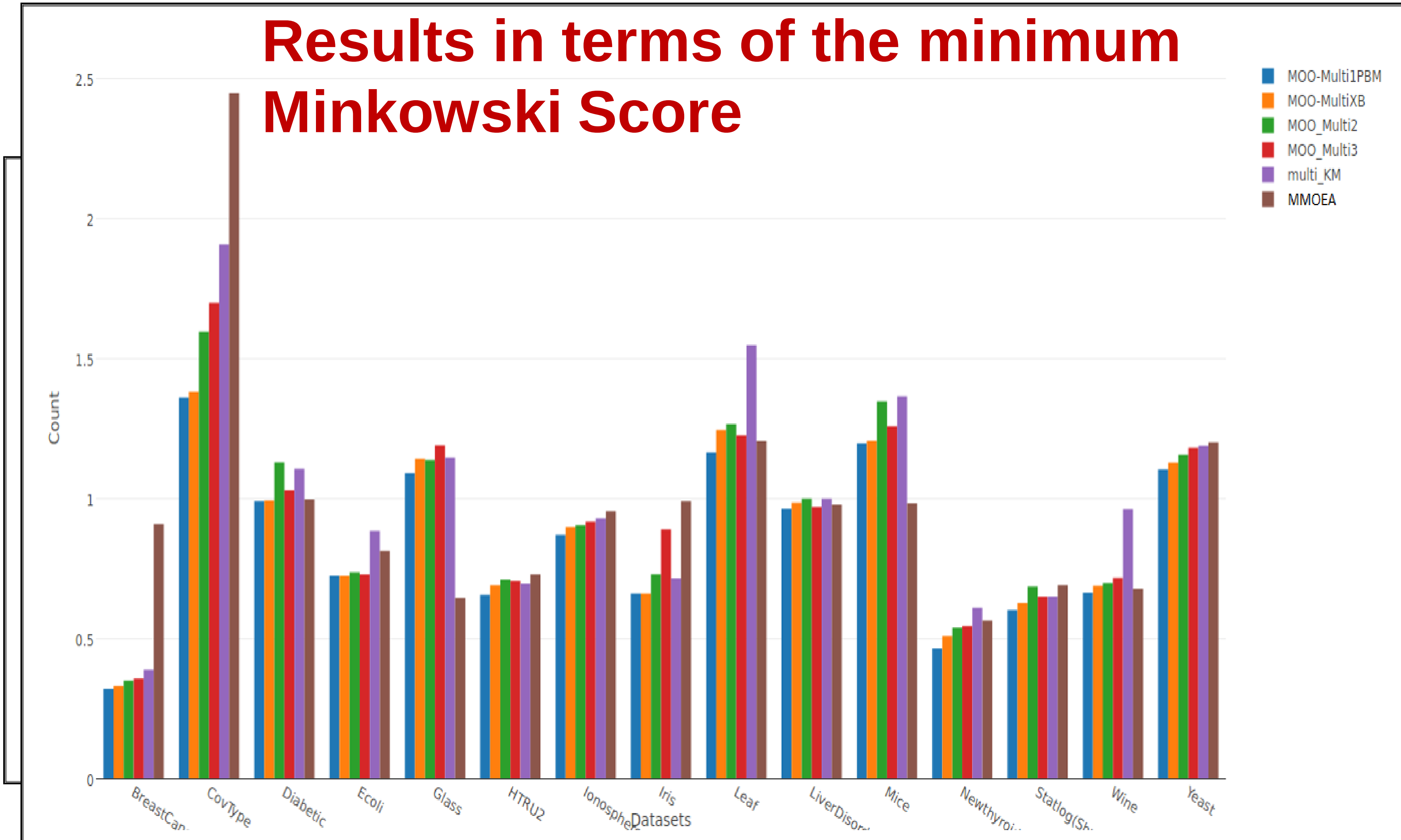
Experimental Results

Description of UCI Datasets

- **First batch of experiments**: **standard data from UCI Machine Learning Repository .**
- **Two views for each data set**
  - **First view : original set of attributes**
  - **Second view : apply Principal Component Analysis to the data set and select few attributes**
- **Results taken for three approaches :** proposed, multi-view KM, multiobjective ensemble clustering
- **Parameters:**
  - T_{min} = 0.01, T_{max} = 100, α = 0.85, HL = 50, SL = 70, itr = 20, K_{max} = √ #of samples and
  - K_{min} = 2

Datasets	AC	Instances	Actual no. of features	Features used in View1	No. of PCA features/ Features used in view2	Total features
Iris	3	150	4	4	2	6
Newthyroid	3	215	5	5	2	7
Liver Disorder	2	345	7	6	4	10
Glass	6	214	10	9	4	13
Breast Cancer	2	699	10	9	5	14
Wine	3	178	13	13	6	19
Ionosphere	2	351	34	34	10	44
Leaf	40	340	16	14	7	21
Yeast	10	1484	8	8	4	12
Ecoli	8	336	8	7	4	11
Mice Protein Expression	8	1080	82	80	40	120
Diabetic Retinopathy Debrecen	2	1151	20	19	4	23
HTRU2	2	17898	9	8	4	12
Statlog (Shuttle)	7	58000	9	9	5	14
CovType	7	581012	54	54	26	80

Experimental Results



S. Saha , S. Mitra, S. Kramer (2018): ``Exploring Multiobjective Optimization for Multi-view Clustering", **ACM Transactions on Knowledge Discovery from Data**, Vol. 12(4), Pages 44:1-44:30

View Generation from SRC Data Sets

- **Two views :**
- **Syntactic view : represents the syntactic information of a document given a particular query :**
  - **The terms in the documents are first extracted**
  - **A document-term matrix is created to be used as view1**
  - **TF-IDF (a common weighting scheme) values are used to fill this matrix**
- **Semantic view: captures semantic contents of the web snippets :**
  - **A content-based similarity measure, Symmetric Conditional Probability (SCP) to calculate the similarity between two web snippets.**
  - **Given two word feature vectors, d_i and d_j, the corresponding similarity :**

$$S(d_i, d_j) = \frac{1}{\|d_i\|\|d_j\|} \sum_{r=1}^{\|d_i\|} \sum_{b=1}^{\|d_j\|} SCP(w_i^r, w_j^b)$$

$$SCP(W_1, W_2) = \frac{P(W_1, W_2)^2}{P(W_1) \times P(W_2)},$$

Comparison of F1 Measure

Datasets	MOO-multi1		MOO-clus		SOO-SRC			
	Min	Max	Min	Max	GK-means	STC	LINGO	BIK
MORESQUE	0.6689	**0.6987**	0.658	0.675	0.655	0.455	0.326	0.317
ODP-239	0.4059	**0.4294**	0.379	0.384	0.366	0.324	0.273	0.2

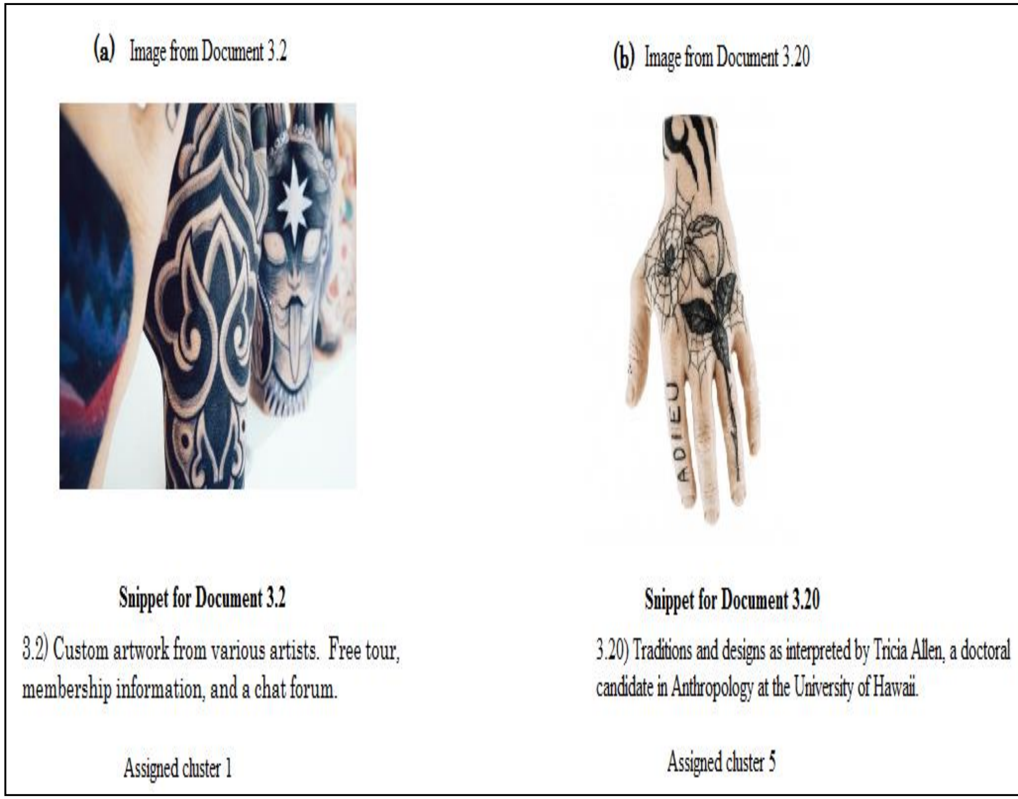
Evaluation results in terms of F 1 (should be maximized) over MORESQUE and ODP239 data sets: Comparison of the proposed approach with state-of-the-art single view based approaches.

S. Saha , S. Mitra, S. Kramer (2018): ``Exploring Multiobjective Optimization for Multi-view Clustering", **ACM Transactions on Knowledge Discovery from Data**, Vol. 12(4), Pages 44:1-44:

LINK

Visually Augmented Multiobjective based Multi-view Search Results Clustering

1. **Image view: pre-trained 19-layer VGG net (Simonyan and Zisserman, 2014) to extract high-level features directly from the image.**



2. **Textual view: combine the benefits of both Word2Vec (Mikolov et al., 2013) and TF-IDF.**

Example where text content differs but image information is same.

Generation of Different Views

- **Textual view:**
  - This view represents both **syntactic and semantic** information.
  - **Using word embedding** each word in the vocabulary is represented by a vector of dimension 1 × 100.
  - Document similarity matrix **using word embedding :**

$$S_{emb}(\bar{d}_i, \bar{d}_j) = \frac{1}{\|d_i\|\|d_j\|} \sum_{r=1}^{\|d_i\|} \sum_{b=1}^{\|d_j\|} CosSim(w_i^r, w_j^b)$$

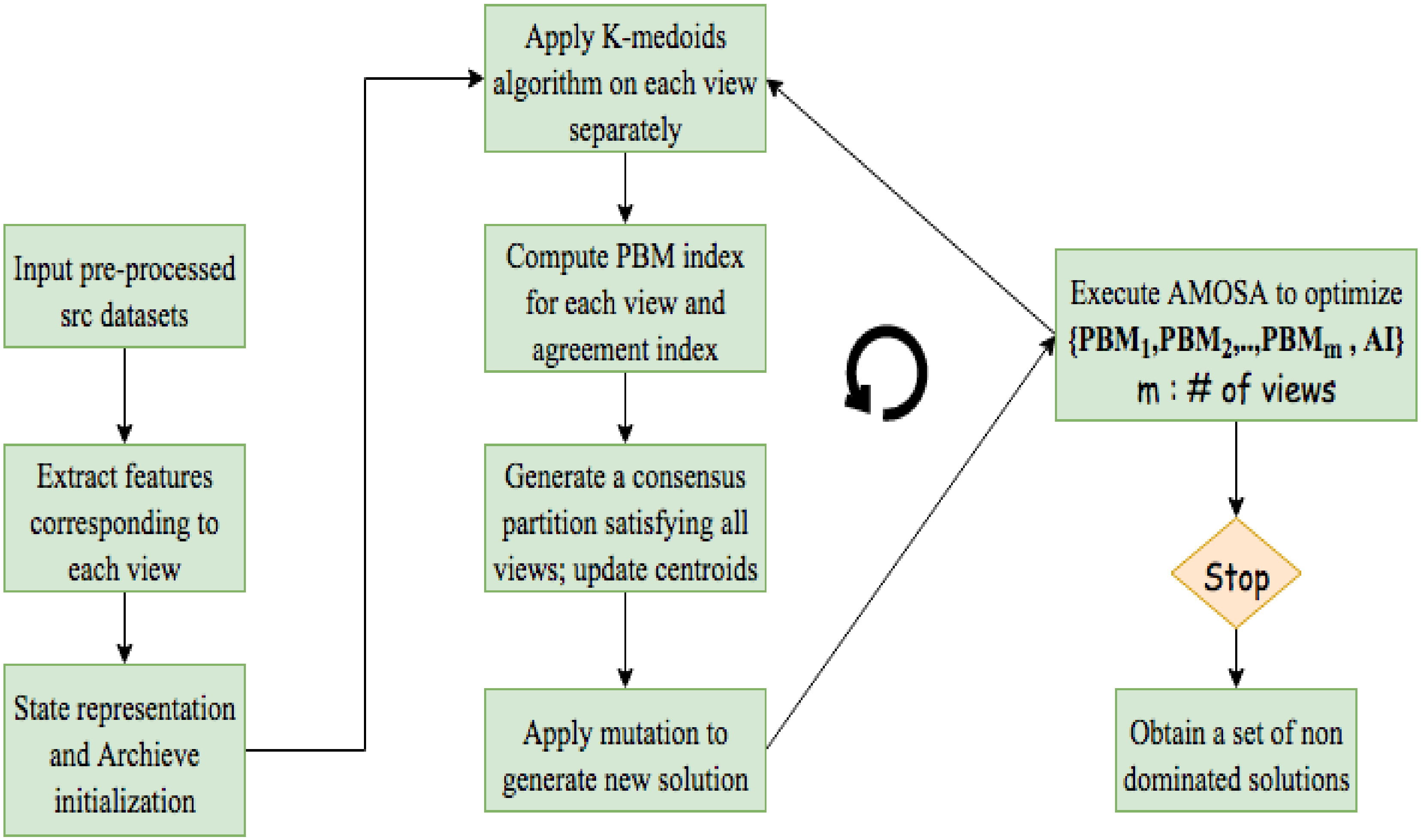
  - **Cosine similarity calculated between TF-IDF vectors of two documents to generate document × document similarity matrix $S_{tf\ idf}$.**

- **Image view:**
  - **Images fed to the pre-trained VGG19 network**
    - **A 4096 dimensional feature vector**

$$S_{image}(\bar{d}_i, \bar{d}_j) = \frac{1}{\|d_i\|\|d_j\|} \sum_{r=1}^{\|d_i\|} \sum_{b=1}^{\|d_j\|} CosSim(img_i^r, img_j^b)$$

$$S_{text}(\bar{d}_i, \bar{d}_j) = S_{embd}(\bar{d}_i, \bar{d}_j) \times S_{tfidf}(\bar{d}_i, \bar{d}_j), i, j = 1, \ldots, n$$

Flowchart



S. Mitra, M. Hasanuzzaman, S. Saha (2018): Incorporating Deep Visual Features into Multiobjective based Multi-view Search Result Clustering, In the proceedings of 27th International Conference on Computational Linguistics (COLING 2018) to be held in Santa Fe, New-Mexico, USA from August 20th--26th 2018 (accepted) (Core Ranking: A).

Results

		MOO-Multiview-PBM (word2vec*tf-idf)		MOO-Multiview-PBM (word2vec)		MOO-Clus		SOO-SRC			
		Min	*Max*	*Min*	*Max*	*Min*	*Max*	*GK-means*	*STC*	*LINGO*	*BIK*
MORESQUE	F_{b^3}	0.506	**0.564**†	0.510	0.557	0.477	0.502	0.482	0.460	0.399	0.315
	F1	0.698	**0.742**†	0.682	0.728	0.658	0.675	0.655	0.455	0.326	0.317
ODP-239	F_{b^3}	0.491	**0.549**†	0.482	0.531	0.478	0.484	0.452	0.403	0.346	0.307
	F1	0.438	**0.474**†	0.431	0.462	0.379	0.384	0.366	0.324	0.273	0.2

MORESQUE	114	10 / 10 / 10	11400	11400	10834	566	8/ 4 / 18
ODP-239	239	6.7 / 2 / 38	25580	25580	19513	3067	6 / 3 / 25

Evaluation results in terms of F1 and Fb3 over MORESQUE and ODP239 data sets: Comparison of the proposed approach with state-of-the-art approaches.

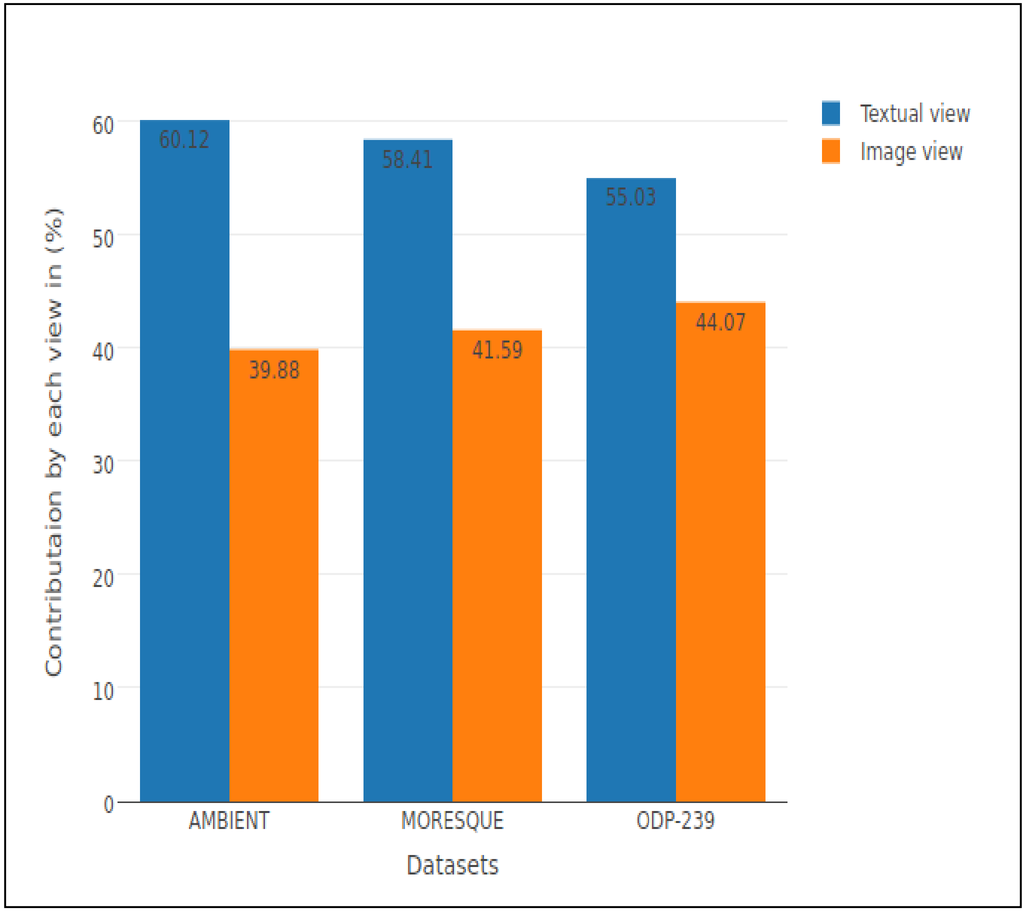
First part represents SRC gold standard data sets. Second part represents total number of relevant images extracted for each query and number of active query links present in each data set.

S. Mitra, M. Hasanuzzaman, **S. Saha** (2018): Incorporating Deep Visual Features into Multiobjective based Multi-view Search Result Clustering, **In the proceedings of 27th International Conference on Computational Linguistics (COLING 2018)** to be held in Santa Fe, New-Mexico, USA from August 20th--26th 2018 (accepted) (Core Ranking: A).

Results on Individual Views

		MOO-image		MOO-word2vec		MOO-word2vec*tfidf	
		Min	Max	Min	Max	Min	Max
MORE SQUE	Fb3	0.427	0.4684	0.479	0.5174	0.489	0.5267
	F1	0.613	0.657	0.664	0.6812	0.6793	0.6973
ODP-239	Fb3	0.4429	0.4725	0.4803	0.4831	0.4821	0.4921
	F1	0.342	0.376	0.3814	0.3901	0.4031	0.4083

Results and Discussion



The following bar graph shows contribution by each view in the clustering process for each dataset.

Quantified the influence of individual views in obtaining the final partitioning. The degree is expressed as follows :

$$Degree_v = \frac{\|A^v \cap A^{\cup v}\|}{\|A^{\cup v}\|}$$

Error Analysis

S. Mitra, M. Hasanuzzaman, **S. Saha** (2018): Incorporating Deep Visual Features into Multiobjective based Multi-view Search Result Clustering, **In the proceedings of 27th International Conference on Computational Linguistics (COLING 2018)** to be held in Santa Fe, New-Mexico, USA from August 20th--26th 2018 (accepted) (Core Ranking: A)

Advancements in State-of-the-art

- First attempt to develop a multiobjective based automated clustering technique for multi-view data sets
  - Automatic determination of number of clusters
  - Novel use of image and text features for search result clustering
  - No labeled data is used

- Proposal of a generic approach
  - Can be applied on any multi-modal data sets

- Use of MOO helps in getting a set of alternative solutions on Pareto optimal front
  - Depending on domain requirement a single solution can be selected
- Future works:
  - Multi-view clustering for scientific documents
    - Views correspond to texts, images, captions of tables + figures, reference sections
  - Cyber-bully detection
  - Multi-view subspace clustering
  - Investigation on objective functions
  - Investigation on selection of a single solution from final Pareto front

Thank You!