Extractive Single Document Summarization using Multi-objective Optimization

Dr. Sriparna Saha
Associate Professor
Department of Computer Science and Engineering
Indian Institute of Technology Patna
Email: sriparna.saha@gmail.com/sriparna@iitp.ac.in
Webpage: www.iitp.ac.in/~sriparna

Roadmap

- Summarization?
- Categories of Summarization
- Types of Summarization
- Literature Survey and Related Background
- Proposed Methods
- Data sets
- Experimental Results
- Result discussion
- Summary

What is Summarization??

- Task of automatically creating a compressed version of the text document that should be concise, relevant, non-redundant and representative of the main idea of the text.

- A text that is produced from one or more texts that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that.

Why summarization?

- **Internet has provided large collection of text on a variety of topics**
- **large number of electronic documents are available online**

↓ Problems

- **Users get so exhausted reading large amount**
- **Users face difficulty in finding relevant information**

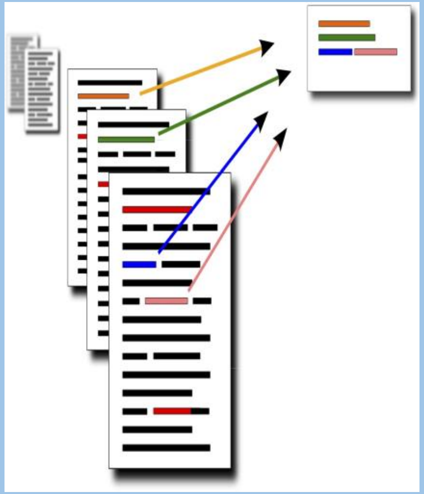
↓ Solution

- **Automatic text summarization system is needed that compress information into shorter length that must follow coverage of information, non-redundancy, information significance and Cohesion in the text**

Categories of Summarization

Single

Multi-document



Types of Summarization (1/2)

Extractive

- selecting a few relevant sentences from the original document
- Relevance of sentences is decided using sentence scoring features like sentence position, similarity with the title etc.

Abstractive

- Abstract summary which includes words and phrases different from the ones occurring in the source document
- Required natural language processing

Others Types of Summaries (2/2)

S. No.	Types of summary	Factors
1	Generic and query-focused	whether general or query related data is required
2	Supervised and unsupervised	Availability of training data
3	Mono, multi and cross-lingual	Language
4	Web-based	For summarizing web pages
5	E-mail based	For summarizing e-mails
6	Personalized	Information specific to a user’s need
7	Sentiment-based	Opinions are detected

Example of Extractive Summarization (sentence based)

[English is the dominant language in the writing and publishing of scientific research in the form of scientific articles.]$_1$ [However, many non-natives users of English suffer the interference of their mother tongues when writing scientific papers in English.]$_2$ [These users face problems concerning rules of grammar and style, and/or feel unable to generate standard expressions and clauses, and the longer linguistic compositions which are conventional in this genre.]$_3$ [In order to ease these users' problems, we developed a learning environment for scientific writing named AMADEUS (Amiable Article Development for User Support).]$_4$ [AMADEUS consists of several interrelated tools reference, support, critic and tutoring tools and provides the context in which this dissertation is inserted.]5 [The main goal of this research is to implement AMADEUS as an agent -based architecture with collaborative agents communicating with a special agent embodying a dynamic user model.]$_6$ [In order to do that we introduce the concept of adaptivity in computer systems and describe several user model shells.] $_7$ [We also provide details about intelligent agents which were used to implement the user model for the AMADEUS environment.]$_8$

→

English is the dominant language in the writing and publishing of scientific research in the form of scientific articles. In order to ease these users' problems, we developed a learning environment for scientific writing named AMADEUS (Amiable Article Development for User Support). The main goal of this research is to implement AMADEUS as an agent -based architecture with collaborative agents communicating with a special agent embodying a dynamic user model. We also provide details about intelligent agents which were used to implement the user model for the AMADEUS environment.

Example of Abstractive Summarization

A detained iranian-american academic accused of acting against national security has been released from a tehran prison after a hefty bail was posted, a to p judiciary official said tuesday.

→

iranian-american academic held in tehran released on bail.

Different Quality Measures for Summarization

- Sentence Similarity with the title
- Anti-redundancy
- Position of the sentence in the document
- Length of the sentence
- Readability
- Coverage
- Cohesion

Literature Survey & Related Background

Existing Summarization System

Method	Contribution
MA-SingleDocSum	Mendoza et al. proposed this method and developed an automatic summarization technique using population-based meta-heuristic algorithm, namely, Memetic algorithm as the optimization technique. It considers single document summarization as a binary optimization problem. and optimizes the weighted sum of different aspects of the summary like readability etc.
DE	Aliguliyev proposed an automatic document summarization technique using differential evolution (DE) approach. It is a sentence clustering-based approach. It first clusters the sentences of the document; then extracts sentences from different clusters. It optimizes a single cluster validity index.
UnifiedRank	UnifiedRank method proposed by X. Wan and presents a graph-based model to solve single and multi-document summarization problem simultaneously.
CRF	CRF was proposed by Shen et al. Authors of this paper have treated extractive single document summarization as a sequence labeling problem where the approach assigns a label of 1 or zero to sentences.

Method	**Contribution**
QSC	QSC method was proposed by Dunlavy et al. where Query-based single document summarization system was proposed which makes use of K-means clustering followed by Hidden Markov Model (HMM). HMM selects sentences from each cluster based on some probability value.
SVM	In Yeh et al., authors have proposed two approaches: Modified Corpus-Based Approach and LSA-based text relationship map. First one is based on the trainable classifier which used various features like sentence position etc. to represent the sentence.The second approach uses latent semantic analysis for summarization task.
UnifiedRank	UnifiedRank method proposed by X. Wan and presents a graph-based model to solve single and multi-document summarization problem simultaneously.
FEOM	Song et al. have proposed fuzzy evolutionary optimization modeling (FEOM) technique and showed its application to extractive summarization.
Manifold Ranking	This method was proposed by Wan et al. In this method, a topic based multi-document summarization system is developed which utilizes the manifold ranking process to assign a score to each sentence. It considers the relationship between sentences in the document and the given topic.

Drawbacks of existing meta-heuristic techniques

- Several ESDS algorithms have been developed (MA-SingleDocSum, FEOM, DE) utilizing the search capabilities of some meta-heuristic based optimization techniques, namely genetic algorithm, differential evolution etc. and shown good results in summarization task.
- These approaches suffer from the following drawbacks:
  - Unable to automatically detect the number of clusters
  - None of the existing ESDS techniques captures the semantic similarity present in the sentences
  - Low convergence rate and ROUGE-score
  - Formulated the summarization problem in the framework of single objective optimization

Solution to drawbacks

- Needs to develop an automatic text summarization system using multi-objective optimization (sentence clustering)
- Able to detect the number of clusters automatically.
- Makes use of several sentence scoring features to select the sentences
- Able to achieve better ROUGE score as comparison to state-of-the-art techniques

Multi-objective Optimization (1/2)

- Multi-objective optimization (MOO) problem aims at finding a vector x = { x_1, x_2 … x_n} of ‘n’ decision variables that optimizes M number of objective functions {$f_1(x)$, $f_2(x)$ … $f_M(x)$} simultaneously while satisfying some constraints if any.
- Mathematically, it is formulated as

$$\textit{min} \qquad F(x) = \{f_1(x), f_2(x) \ldots f_M(x)\}^T$$

such that $x = \{ x_1 , x_2 \ldots x_n\}^T \in \Omega$, where x is a decision vector in n-dimensional decision space Ω.

Multi-objective Optimization (2/2)

- **Example:** Find out tickets in the train with minimum cost and minimum travel time with some constraint
- Here:
  - Optimizing Criteria:
    - Minimizing the ticket cost
    - Minimizing the travel time
  - Constraints:
    - Not more than 2 stoppage between source and destination
    - Should have pantry car
  - Decision variables
    - The available trains

Conflicting behaviour

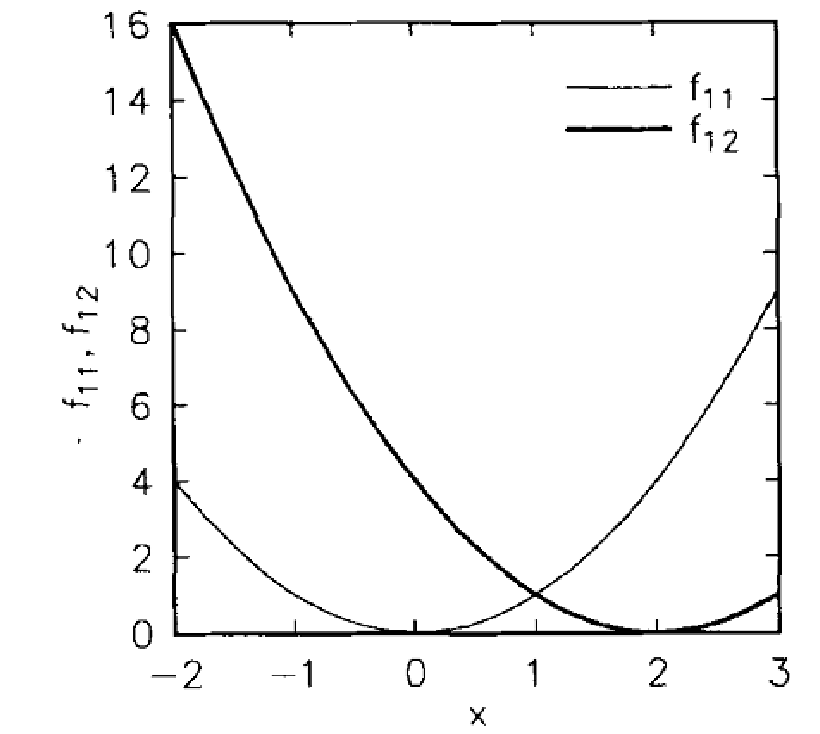
In real-world, we have to simultaneously optimize two or more than two objective functions which leads to more than one solution.

MOO: A Numerical Example

Minimize $f1 = x^2$

Minimize $f2 = (x-2)^2$

- The solution x = 0 is optimum w.r.t. f1 but not so good with respect to f2.
- the solution x = 2 is optimum w.r.t. function f2 and not so good with respect to f1 .
- Optimal range: 0 <= x <= 2 which provides a set of solutions.



Solutions Relationship(1/4)

- A solution *sol* in M-dimensional objective space is represented as

$$sol = \{f_1(sol), f_2\,(sol) \dots f_M\,(sol)\}$$

where $f_i(sol)$, $1 \leq i \leq M$ is the value of ith objective function

- Representation of 5 solutions
    - sol1 = {1, 1}
    - sol2 = {1, 2}
    - sol3 = {3, 1}
    - sol4 = {2, 3}
    - sol5 = {4, 2}

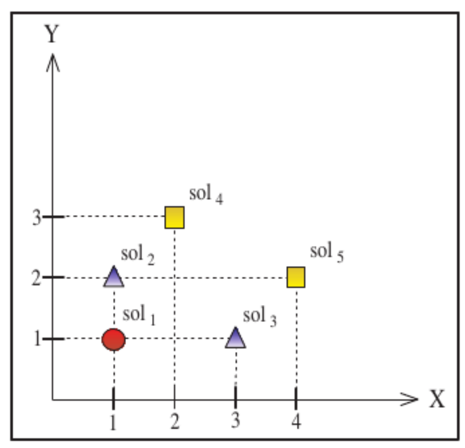


Fig: Solutions in 2-dimensional objective space

Solutions Relationship: Dominance (for minimization problem) (2/4)

A solution $sol_i = \{f_1(sol_i), f_2(sol_i), \ldots, f_M(sol_i)\}$ dominates another solution $sol_j = \{f_1(sol_j), f_2(sol_j), \ldots, f_M(sol_j)\}$ denoted as $sol_i \prec sol_j$ iff

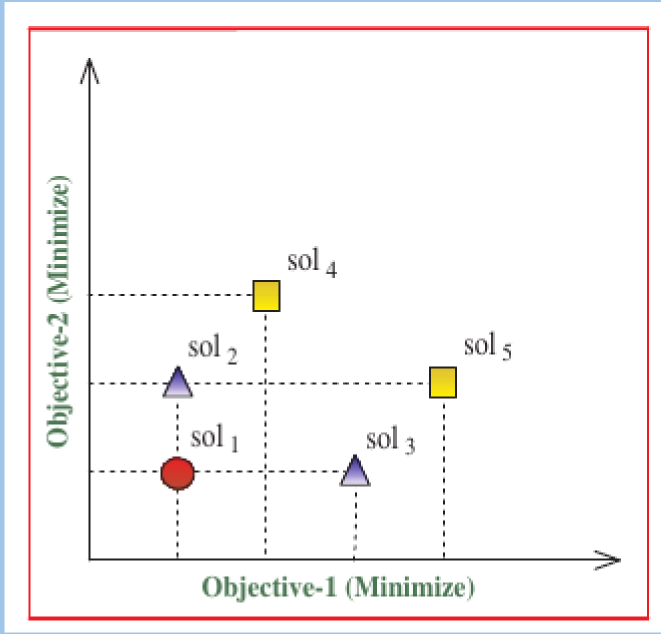
1. $f_m(sol_i) \leq f_m(sol_j) \qquad \forall m \in \{1, 2, \ldots, M\}$
2. $f_m(sol_i) < f_m(sol_j) \qquad \exists m \in \{1, 2, \ldots, M\}$

sol_i and sol_j are non-dominated represented as $sol_i \underline{\prec} sol_j$ iff neither $sol_i \prec sol_j$ nor $sol_j \prec sol_i$

Solutions Relationship (3/4)

In the Figure

- sol1 $\prec$ {sol2 , sol3 , sol4 , sol5 }
- sol2 $\prec$ {sol4 , sol5 }
- sol3 $\prec$ {sol5 }
- Sol2 $\preceq$ sol3
- Sol3 $\preceq$ sol4
- sol4 $\preceq$ sol5



Solutions Relationship: Non-dominated Sorting (4/4)

Non-Dominated Sorting is to divide the population P in K ($1 \leq K \leq N$) fronts. Let the set of these K fronts in decreasing order of their dominance (increasing order of non-domination level) be $F = \{F_1, F_2, \ldots, F_K\}$. The division of the solutions in fronts is such that

1. $\forall sol_i, sol_j \in F_k : soli \preceq solj \qquad 1 \leq k \leq K$
2. $\forall sol \in F_k, \ \exists\ sol' \in F_{k-1} : sol' \prec sol \qquad 2 \leq k \leq K$

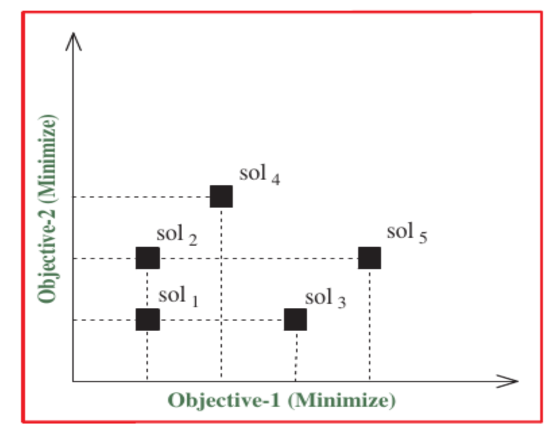


Fig.: Solutions in objective space

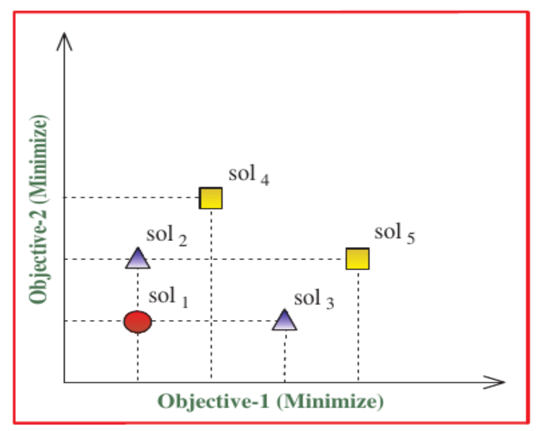


Fig.: Non-dominated Fronts

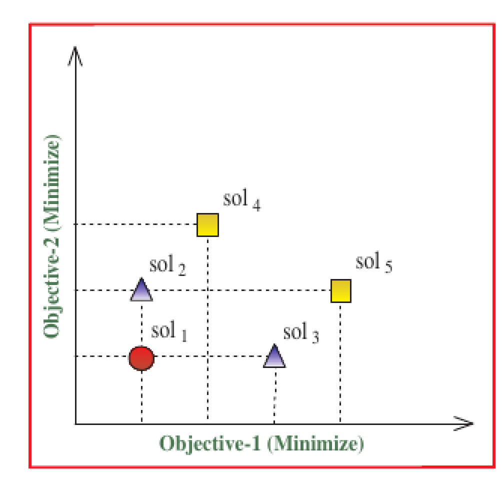
An Example: Non-dominated Sorting

- To obtain K fronts, two things need to maintain
  - Domination count
  - Dominance relationship between solutions
- Domination Count (DC): Domination count of a solution *sol'* in population P is the number of solutions in P which dominates solution *sol'* .

In the Figure,

DC of sol1, sol2, sol3, sol4, sol5 are

0, 1, 1, 2 and 3, respectively.



An Example: Non-dominated Sorting

1. For solution sol_1
   - $S_{sol_1} = \{sol_2, sol_3, sol_4, sol_5\}$
   - $n_{sol_1} = 0$
2. For solution sol_2
   - $S_{sol_2} = \{sol_4, sol_5\}$
   - $n_{sol_2} = 1$ $\quad sol_2 \succ \{sol_1\}$
3. For solution sol_3
   - $S_{sol_3} = \{sol_5\}$
   - $n_{sol_3} = 1$ $\quad sol_3 \succ \{sol_1\}$
4. For solution sol_4
   - $S_{sol_4} = \{\}$
   - $n_{sol_4} = 2$ $\quad sol_4 \succ \{sol_1, sol_2\}$
5. For solution sol_5
   - $S_{sol_5} = \{\}$
   - $n_{sol_5} = 3$ $\quad sol_5 \succ \{sol_1, sol_2, sol_3\}$

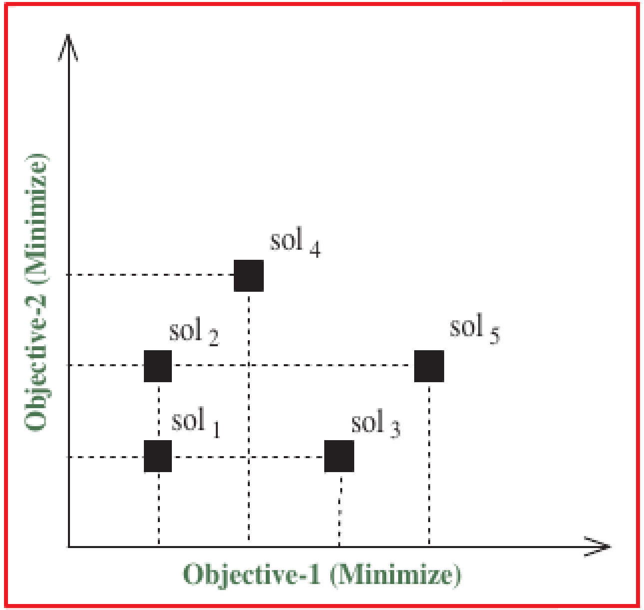


Fig.: Solutions in objective space

An Example: Non-dominated Sorting

$n_{sol_1} = 0$	$S_{sol_1} = \{sol_2, sol_3, sol_4, sol_5\}$
$n_{sol_2} = 1$	$S_{sol_2} = \{sol_4, sol_5\}$
$n_{sol_3} = 1$	$S_{sol_3} = \{sol_5\}$
$n_{sol_4} = 2$	$S_{sol_4} = \{\}$
$n_{sol_5} = 3$	$S_{sol_5} = \{\}$

$\mathbf{n}_{sol_1} = \mathbf{0}$

- $\mathrm{sol}_{1_{\mathrm{rank}}} = 1$
- $F_1 = \{\mathrm{sol}_1\}$

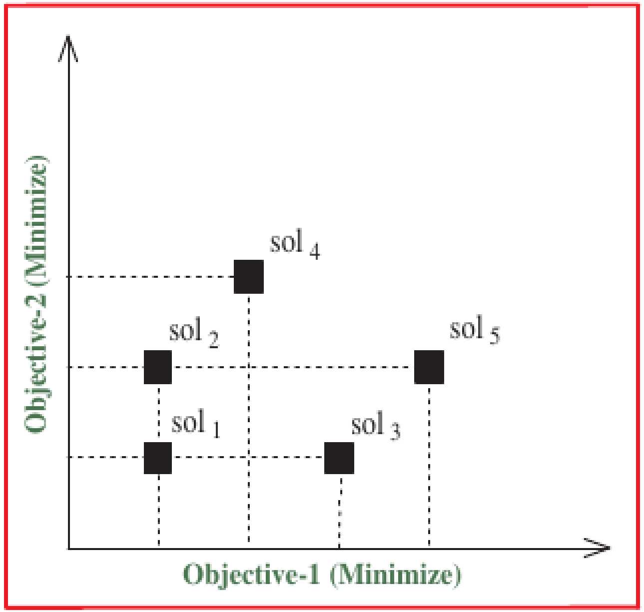
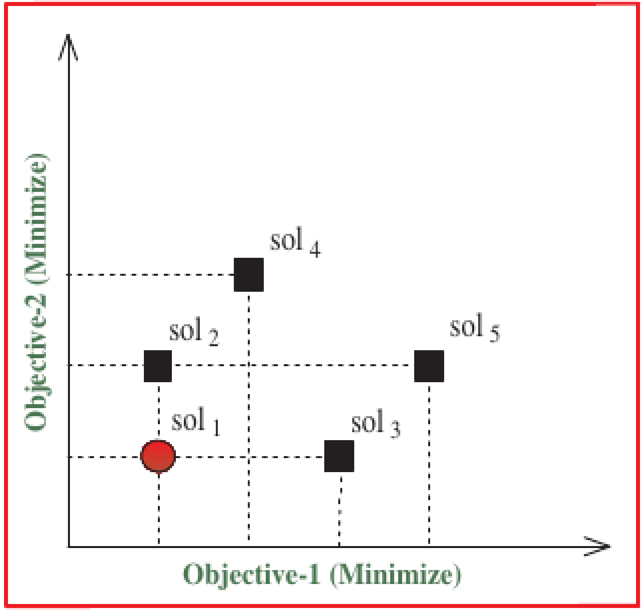


Fig.: Solutions in objective space

An Example: Non-dominated Sorting

$$
\begin{aligned}
n_{sol_1} &= 0 & S_{sol_1} &= \{sol_2, sol_3, sol_4, sol_5\} \\
n_{sol_2} &= 0 & S_{sol_2} &= \{sol_4, sol_5\} \\
n_{sol_3} &= 1 & S_{sol_3} &= \{sol_5\} \\
n_{sol_4} &= 2 & S_{sol_4} &= \{\} \\
n_{sol_5} &= 3 & S_{sol_5} &= \{\}
\end{aligned}
$$

$$
\begin{aligned}
n_{sol_1} &= 0 & S_{sol_1} &= \{sol_2, sol_3, sol_4, sol_5\} \\
n_{sol_2} &= 0 & S_{sol_2} &= \{sol_4, sol_5\} \\
n_{sol_3} &= 0 & S_{sol_3} &= \{sol_5\} \\
n_{sol_4} &= 2 & S_{sol_4} &= \{\} \\
n_{sol_5} &= 3 & S_{sol_5} &= \{\}
\end{aligned}
$$



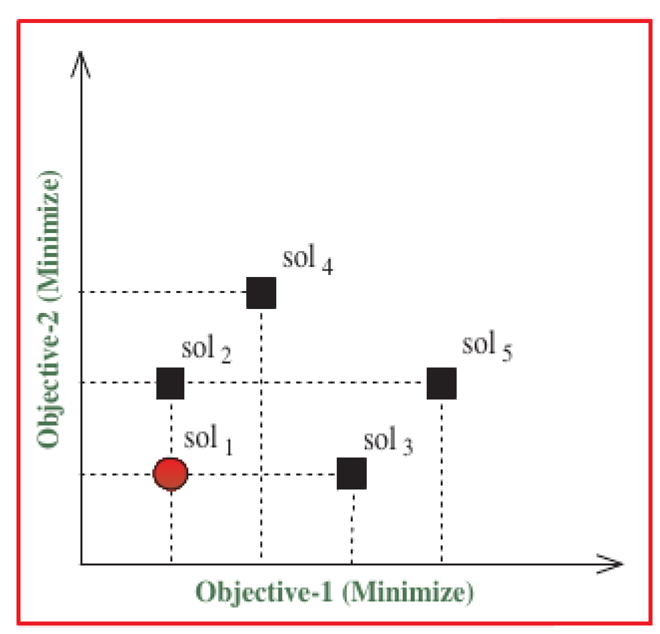
An Example: Non-dominated Sorting

$$
\begin{aligned}
n_{sol_1} &= 0 & S_{sol_1} &= \{sol_2, sol_3, sol_4, sol_5\} \\
n_{sol_2} &= 0 & S_{sol_2} &= \{sol_4, sol_5\} \\
n_{sol_3} &= 0 & S_{sol_3} &= \{sol_5\} \\
n_{sol_4} &= 1 & S_{sol_4} &= \{\} \\
n_{sol_5} &= 3 & S_{sol_5} &= \{\}
\end{aligned}
$$

$$
\begin{aligned}
n_{sol_1} &= 0 & S_{sol_1} &= \{sol_2, sol_3, sol_4, sol_5\} \\
n_{sol_2} &= 0 & S_{sol_2} &= \{sol_4, sol_5\} \\
n_{sol_3} &= 0 & S_{sol_3} &= \{sol_5\} \\
n_{sol_4} &= 1 & S_{sol_4} &= \{\} \\
n_{sol_5} &= 2 & S_{sol_5} &= \{\}
\end{aligned}
$$

$\mathbf{n}_{sol_2} = \mathbf{0},\ \mathbf{n}_{sol_3} = \mathbf{0}$

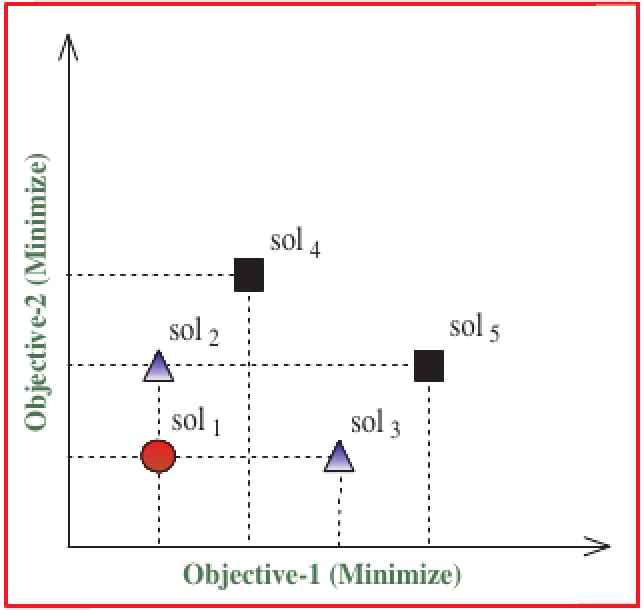
- $\mathrm{sol2}_{\mathrm{rank}} = \mathrm{sol3}_{\mathrm{rank}} = 2$
- $F_2 = \{\mathrm{sol}_2, \mathrm{sol}_3\}$



An Example: Non-dominated Sorting

$$
\begin{aligned}
n_{sol_1} &= 0 & S_{sol_1} &= \{sol_2, sol_3, sol_4, sol_5\} \\
n_{sol_2} &= 0 & S_{sol_2} &= \{sol_4, sol_5\} \\
n_{sol_3} &= 0 & S_{sol_3} &= \{sol_5\} \\
n_{sol_4} &= 0 & S_{sol_4} &= \{\} \\
n_{sol_5} &= 2 & S_{sol_5} &= \{\}
\end{aligned}
$$

$$
\begin{aligned}
n_{sol_1} &= 0 & S_{sol_1} &= \{sol_2, sol_3, sol_4, sol_5\} \\
n_{sol_2} &= 0 & S_{sol_2} &= \{sol_4, sol_5\} \\
n_{sol_3} &= 0 & S_{sol_3} &= \{sol_5\} \\
n_{sol_4} &= 0 & S_{sol_4} &= \{\} \\
n_{sol_5} &= 1 & S_{sol_5} &= \{\}
\end{aligned}
$$

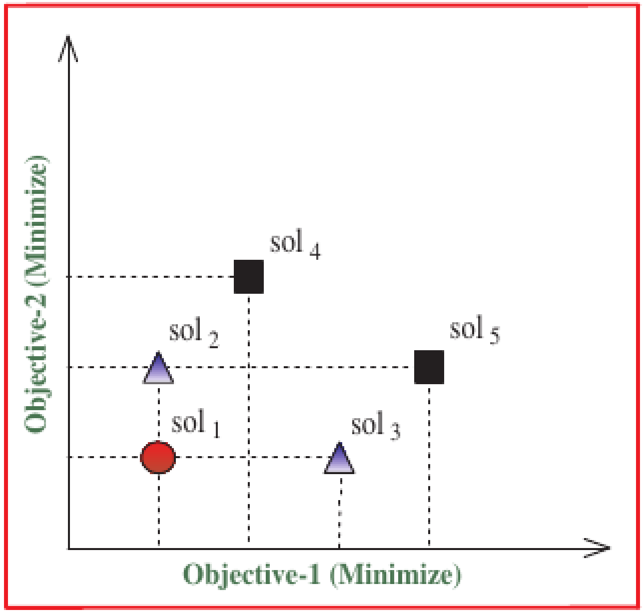


An Example: Non-dominated Sorting

$n_{sol_1} = 0$	$S_{sol_1} = \{sol_2, sol_3, sol_4, sol_5\}$
$n_{sol_2} = 0$	$S_{sol_2} = \{sol_4, sol_5\}$
$n_{sol_3} = 0$	$S_{sol_3} = \{sol_5\}$
$n_{sol_4} = 0$	$S_{sol_4} = \{\}$
$n_{sol_5} = 0$	$S_{sol_5} = \{\}$

$\mathbf{n}_{sol_4} = \mathbf{0}$, $\mathbf{n}_{sol_5} = \mathbf{0}$

- $\text{sol}_{4_{\text{rank}}} = \text{sol}_{5_{\text{rank}}} = 3$
- $F_3 = \{\text{sol}_4, \text{sol}_5\}$



An Example: Non-dominated Sorting

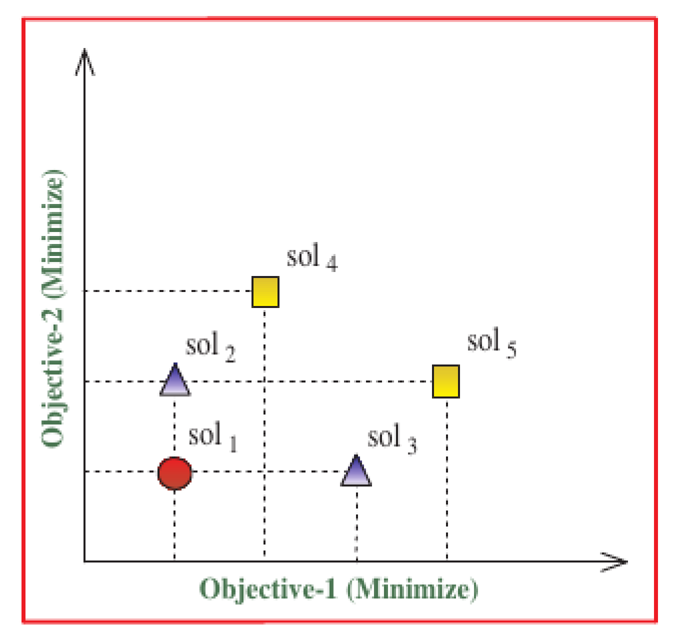
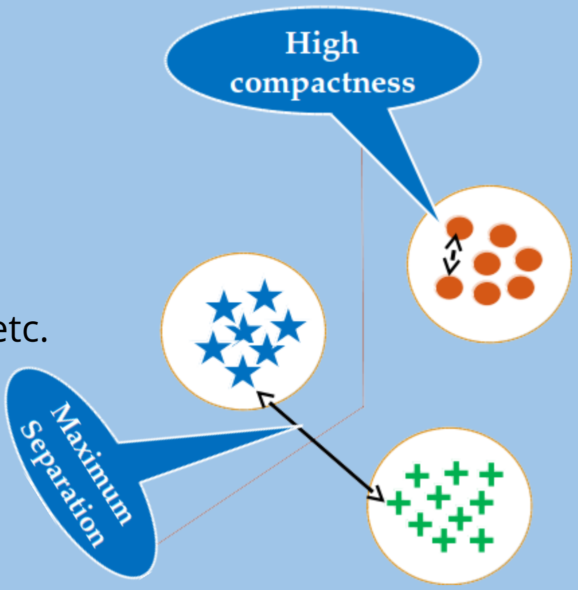


Fig.: Solutions in Pareto Fronts

Clustering

- Grouping of similar elements into various groups
- Main Objective:
  - High compactness
  - Maximize Separation
- Examples:
  - K-means, K-medoids, Hierarchical
- How to measure goodness of partitioning:
  - Using Cluster Validity Indices
    - External: Adjusted rand index, Minkowski Score etc.
    - Internal: Silhouette index, PBM index etc.



Cluster Validity Indices

- Used to validate the quality of clusters
- **External:** Used to measure the extent to which cluster labels match externally supplied class labels.
    - Examples: Rand Index, Minkowski score
- **Internal:** Used to measure the goodness of a clustering structure without using any external information.
    - Example: PBM Index, Xie-beni index, Silhouette index

Internal Cluster Validity Indices

Maximize

PBM Index: Internal Validity Index

$$PBM = \left(\frac{1}{K} X \frac{E_1}{E_K} X D_K \right)^2$$

$$E_K = \sum_{k=1}^{K} E_k, \text{ and } E_k = \sum_{j=1}^{n} \mu_{kj} \parallel x_j - c_k \parallel$$

$$D_K = \max_{i,j=1, i\neq j}^{K} \parallel c_i - c_j \parallel$$

Maximize

Silhouette Score : Internal Validity Index

$$S = \frac{b - a}{max(b, a)}$$

$$XB = \frac{\sum_{k=1}^{K} \sum_{s \in S_k} dist_{wmd}(s, c_k)}{N \times \min_{i, i \neq j} dist_{wmd}(c_i, c_j)}$$

Maximize

External Cluster Validity Indices

Minkowski Score : Rand Index

$$RI = \frac{TP + TN}{TP + TN + FP + FN}$$

Maximum

Minimum

Minkowski Score : External Validity Index

$$MS(AL, CL) = \sqrt{\frac{n_{01} + n_{10}}{n_{11} + n_{10}}}$$

Multi-objective Clustering (in relation with Summarization)

- Nowadays, sentence based extractive summarization techniques are popularly used in producing summary.
  - First perform sentence clustering
  - Rank the clusters
  - Extract sentences from top rank clusters using some sentence scoring features until we get desirable length of summary.
- Multiple cluster quality measures capturing different data properties are required to be optimized simultaneously.
- Problem of sentence clustering is framed as a MOO-based clustering problem where sentence clusters are identified in an automatic way.
- Some of the example of MOO clustering: MOCK, SMEA_Clust etc.

Self-Organizing Map

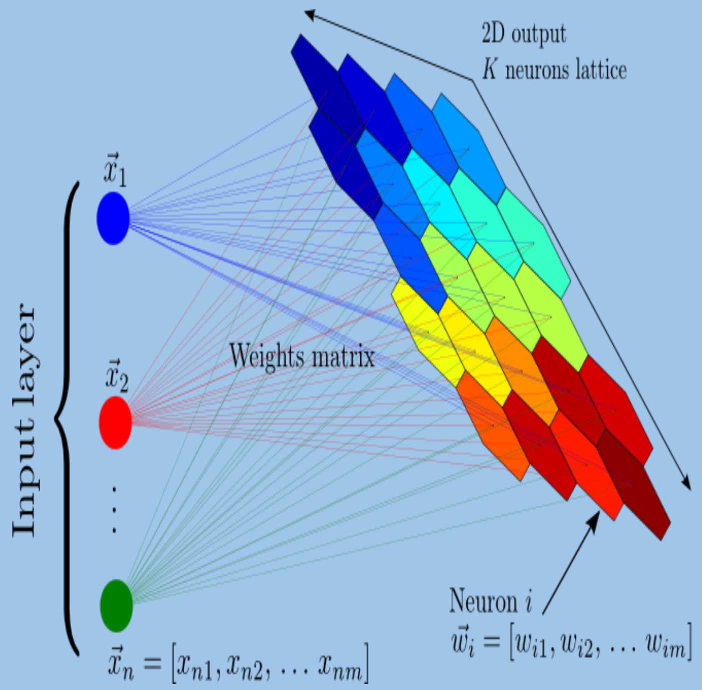


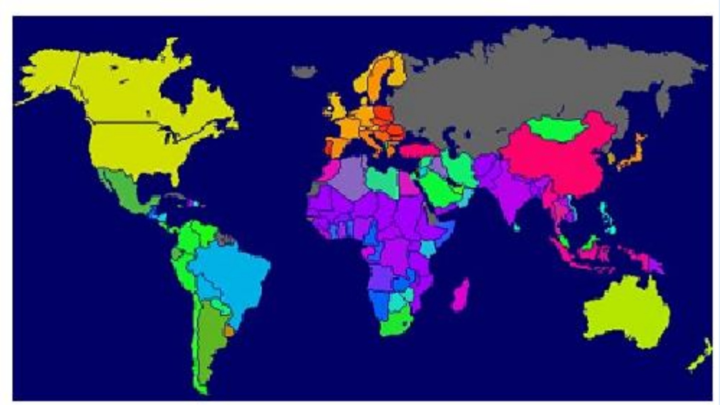
Image Source: Internet

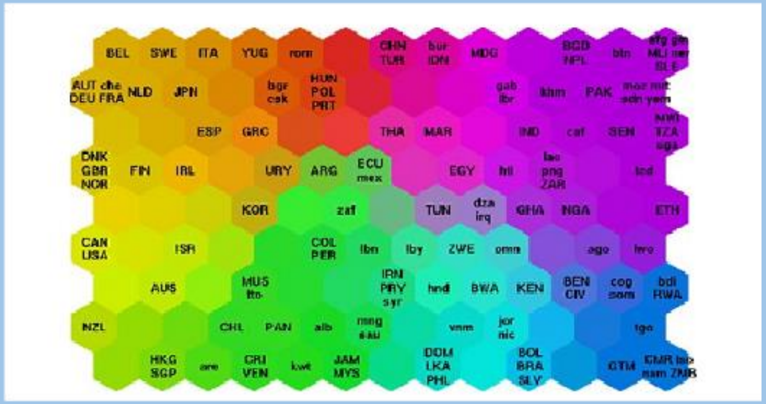
- Special type of Artificial Neural Network
- Associated with each node is a weight vector of the same dimension as the input data vectors, and a position in the map space.
- Arrangement of nodes is two-dimensional regular spacing in a hexagonal or rectangular grid.
- Maps High dimensional Map to low dimensional usually 2-D in a topographic order
- Makes use of Unsupervised and Does not include any hidden layer
- Used for: Data visualization, Clustering

Example: SOM

First figure represents map of the World quality-of-life. ***Yellows*** and ***oranges*** **wealthy nations**, while ***purples*** and ***blues*** **the poorer nations**. From this view, it can be difficult to visualize the relationships between countries.

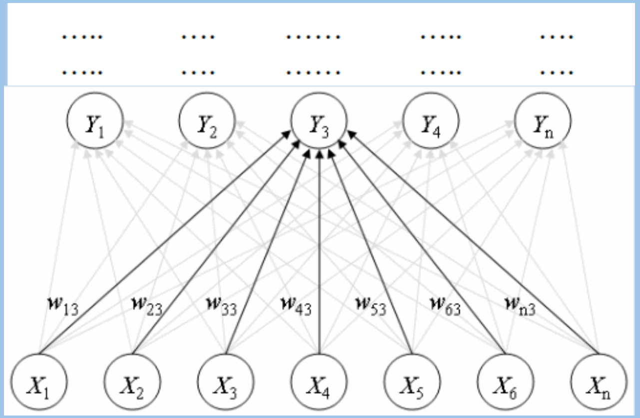
Second Figure (After applying SOM), we can see the United States, Canada, and Western European countries, on the left side of the network, being the wealthiest countries. The poorest countries (like NPL, BGD), then, can be found on the opposite side of the map (at the point farthest away from the richest countries), represented by the purples and blues.



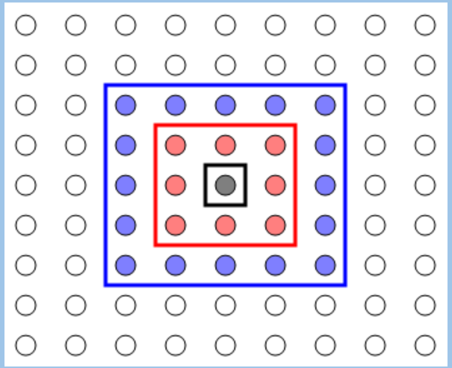


Algorithm 1 SOM Framework(η_0, σ, S, T)

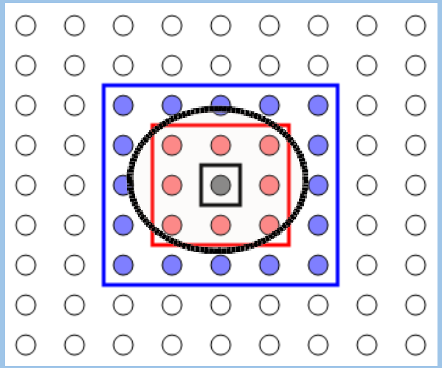
Initialize learning constant η_0 and neighborhood size σ ; maximum iteration count T for SOM Training; Initialize each map unit by assigning a weight vector randomly chosen from training data S.
while $t \neq T$ **do** ▷ t is current iteration no.
    Adjust Learning rate: $\eta = \eta_0 * (1 - \frac{t}{T})$.
    Randomly select a training sample x $\in$ S
    Find winning map unit: $u' = arg\min_{1 \leq u \leq D} \wedge \| x - w^u \|_2$
    Find the neighboring neurons: $U = \{u | 1 \leq u \leq D \| z^u - z^{u'} \|_2 < \sigma\}$
    Update all neighboring neurons: $w^{u'} = w^{u'} + \eta * \exp(- \| z^u - z^{u'} \|_2) * (x - w^{u'})$
return The weight vectors corresponding to map units, $w^u, u = 1, 2 \ldots, D$



Any Random Sample Xi



Find winning and neighboring neurons using neighborhood relationship



Update the weight vectors of winning neuron and neighboring neurons

SOM: A numerical example

- n = 4, m = 2

- Training samples

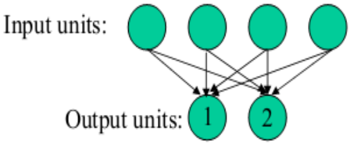
  i1: (1, 1, 0, 0)

  i2: (0, 0, 0, 1)

  i3: (1, 0, 0, 0)

  i4: (0, 0, 1, 1)

Network Architecture



$$\begin{array}{l} \text{Unit 1:} \\ \text{Unit 2:} \end{array} \begin{bmatrix} .2 & .6 & .5 & .9 \\ .8 & .4 & .7 & .3 \end{bmatrix}$$

- Training sample: i1
  - Unit 1 weights
    - $d^2 = (.2\text{-}1)^2 + (.6\text{-}1)^2 + (.5\text{-}0)^2 + (.9\text{-}0)^2 = 1.86$
  - Unit 2 weights
    - $d^2 = (.8\text{-}1)^2 + (.4\text{-}1)^2 + (.7\text{-}0)^2 + (.3\text{-}0)^2 = .98$
  - Unit 2 wins
  - Weights on winning unit are updated

Let neighborhood = 0

- Only update weights associated with winning output unit (cluster) at each iteration
- Giving an updated weight matrix:

$$new-unit2-weights = [.8 \quad .4 \quad .7 \quad .3] + 0.6([1 \; 1 \; 0 \; 0] - [.8 \quad .4 \quad .7 \quad .3]) =$$

$$= [.92 \;\; .76 \;\; .28 \;\; .12]$$

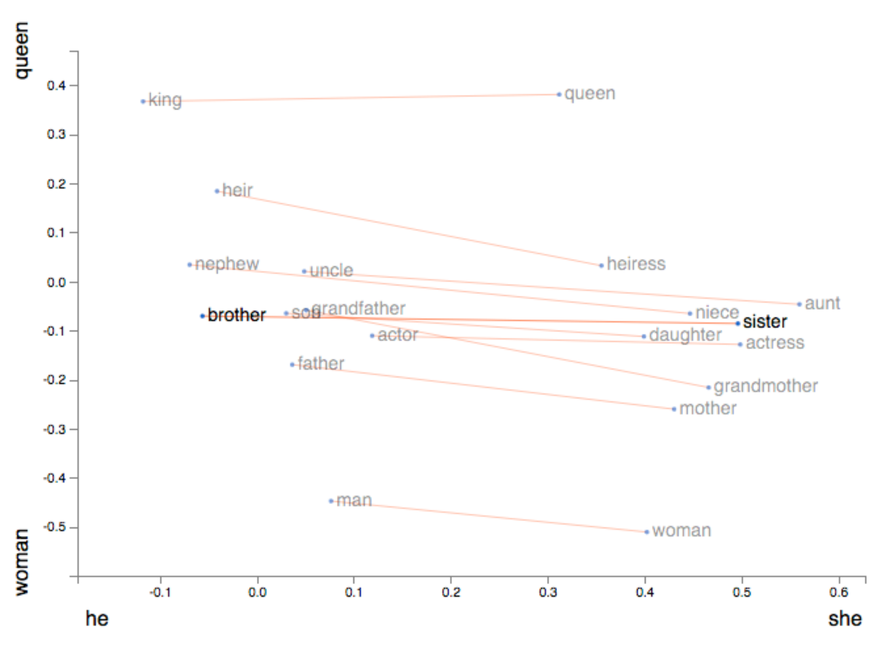
$$\begin{array}{l} \text{Unit 1:} \\ \text{Unit 2:} \end{array} \begin{bmatrix} .2 & .6 & .5 & .9 \\ .92 & .76 & .28 & .12 \end{bmatrix}$$

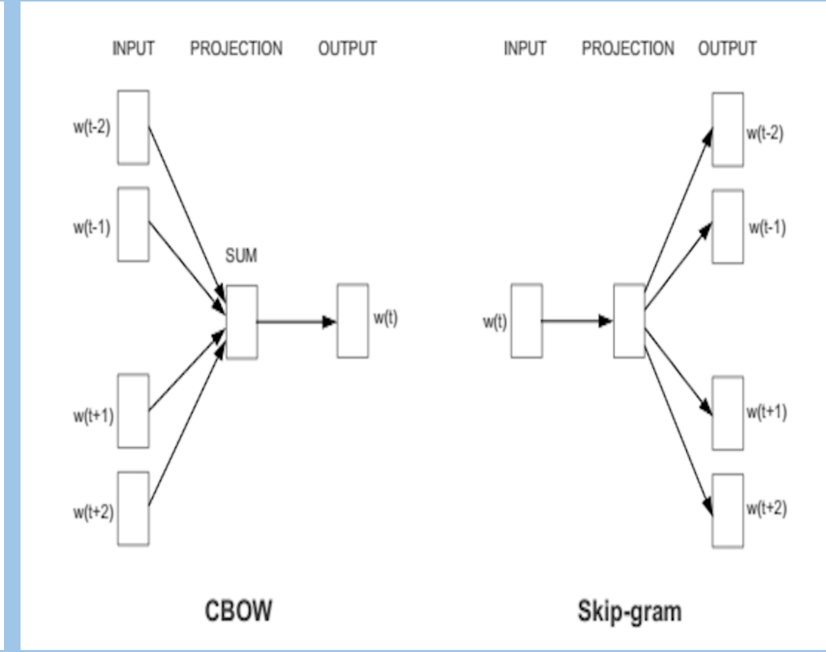
Word2vec Model

- two-layer neural net that processes text and word embeddings (texts converted into numbers).
- Able to associate words with other words (e.g. “man” is to “boy” what “woman” is to “girl”), or cluster documents and classify them by topic.
- In other words, able to capture semantics between words. Here’s a list of words associated with “Sweden” using Word2vec, in order of proximity:

Word	Cosine distance
norway	0.760124
denmark	0.715460
finland	0.620022
switzerland	0.588132
belgium	0.585835
netherlands	0.574631
iceland	0.562368
estonia	0.547621
slovenia	0.531408

- It’s training done in two ways: CBOW and Skip-gram i.e. either using context to predict a target word or using a word to predict a target context, which is called skip-gram.





Word Mover Distance

“Amount of distance that the embedded words of one text needs to travel to reach the embedded words of another text.”

- Makes use of word embedding.
- If two sentences are similar, then WMD will be 0.

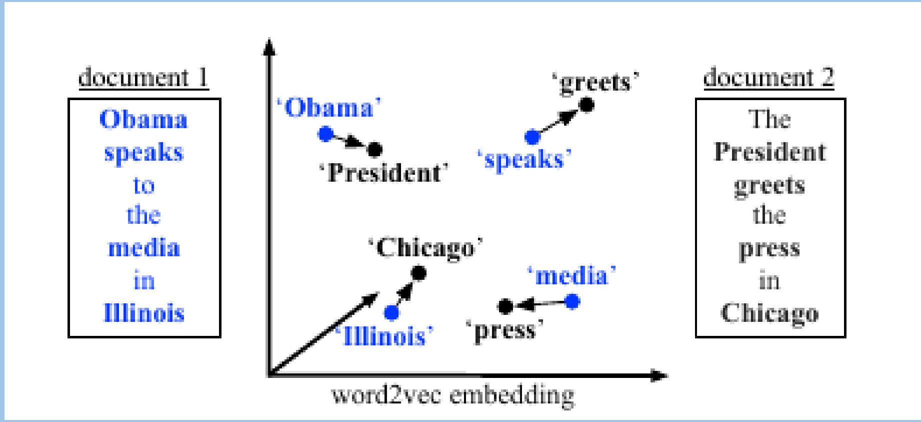
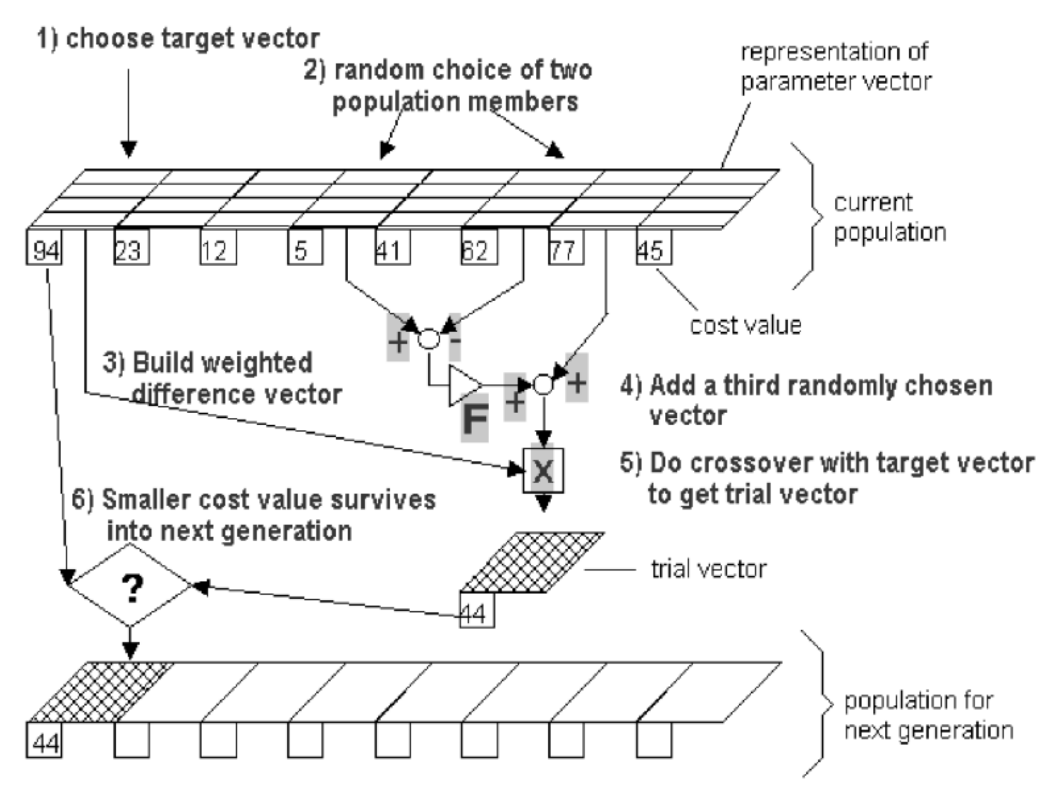


Figure: An illustration of the word mover’s distance. All non-stop words (bold) of both documents are embedded into a word2vec space. The distance between the two documents is the minimum cumulative distance that all words in document 1 need to travel to exactly match document 2. (Best viewed in color.)

Optimization Techniques: Differential Evolution, Grey Wolf Optimizer, Water Cycle Algorithm

Differential Evolution

- **Differential Evolution** is a Optimization algorithm, and is an instance of an **Evolutionary** Algorithm.
- involves maintaining a population of candidate solutions
- Crossover, mutation and selection takes over the number of iteration.
- Fig. shows the flow of single - objective DE.



Grey Wolf Optimizer

- Algorithm is based on **leadership hierarchy** and **hunting procedure** of grey wolves in nature.
- Wolves usually moves in a pack and attack a prey in a planner way.

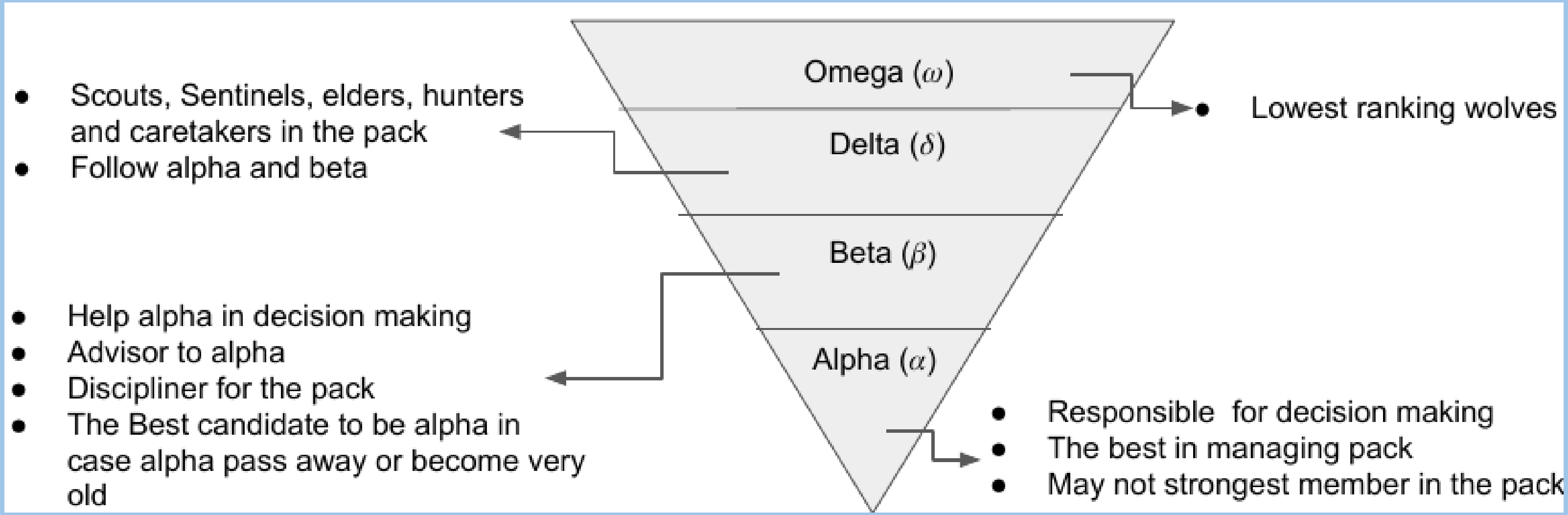


Figure: Leadership hierarchy

- **Hunting technique** of wolves:
    - Chasing and approaching the prey
    - harassing and encircling the prey until it stops
    - Attacking the prey
- During hunting, wolves **update** their **positions towards the prey**

$$\mathbf{D = |C.\, x_p\,(t) - x(t)| \qquad and \qquad x(t+1) = |\, x_p(t) - A.\, D|}$$

where, $x(t)$ and $x_p(t)$ are the position vectors of grey wolf and prey, t indicates current iteration number. Vector A and C are expressed as:

$$\mathbf{A = 2a.r_1 - a \qquad and \qquad C = 2.r_2}$$

where, components of a linearly decrease from 2 to 0 as the iteration passes, r_1 and r_2 are random vectors in [0, 1]

- **Exploration vs. Exploitation:** If |A| > 1, wolf diverges from the prey, while for |A| < 1, wolf converges towards the prey.
- Following equations are applied for **hunting mechanism**:

$$D_\alpha=|C_1 . x_\alpha(t) - x(t)| \text{ and } D_\beta = | C_2 . x_\beta(t) - x(t)| \text{ and } D_\delta=|C_3 . x_\delta(t) - x(t)|$$

$$x(t+1)= (x_1(t) + x_2(t) + x_3(t))/3$$

where, x(t+1) is the updated position of a wolf at (t+1)th iteration with respect to positions of α, β and δ.

Thus, in this way wolves attack the prey.

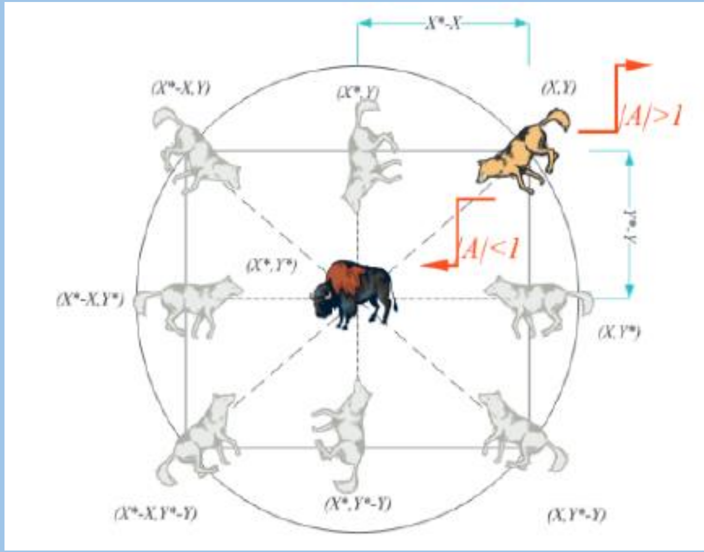
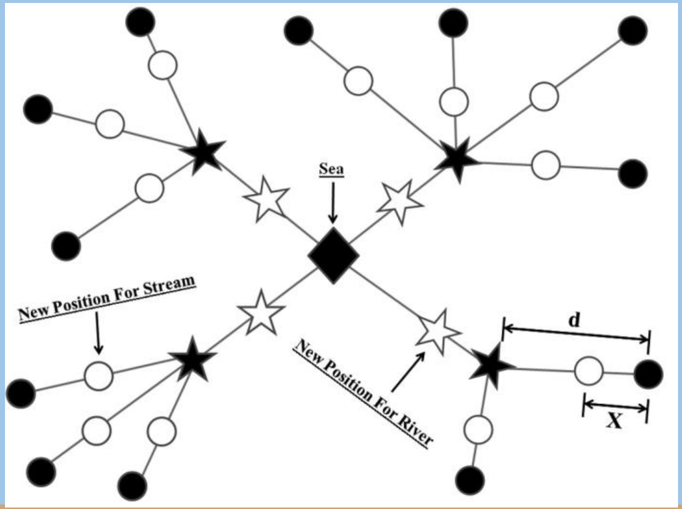


Fig.: Updation mechanism of wolf’s position

Water Cycle Algorithm

- A meta-heuristic algorithm that mimics the water cycle process in nature, i.e., the flow of rivers and streams to sea and flow of streams to rivers.
- The fittest solution is considered as the sea. The second to N_{sr} solutions are considered as rivers and remaining as streams. Here N_{sr} =1 sea +N number_of_rivers



To show the flow of streams to rivers, following equation is considered:

$$x^{stream}(t+1) = x^{stream}(t) + R \times C \times (x^{river}(t) - x^{stream}(t))$$

where $1 < C < 2$ and R lies between [0, 1], $x^{stream}(t + 1)$ represents the updated position of stream x^{stream} at time (t + 1), $x^{river}(t)$ shows the position of river at time t.

Equation to update the position of river in case river flows to sea

$$x^{river}(t+1) = x^{river}(t) + R \times C \times (x^{sea}(t) - x^{river}(t))$$

Equation to update the position of stream in case stream flows to sea

$$x^{stream}(t+1) = x^{stream}(t) + R \times C \times (x^{sea}(t) - x^{stream}(t))$$

- If solution given by stream (after updating position) is better than its connecting river, then stream and river exchange their positions. Similar steps can be executed between stream and sea, river and sea.
- After updating position, evaporation condition is checked to generate new solutions i.e. to check whether stream/rivers are close to sea within a radius to make the evaporation process to occur

$$\| x^{sea} - x^{river} \| < d_{max} \quad \text{or} \quad rand() < 0.1$$

where d_{max} is a small number close to zero and linearly decreases over the course of iteration.

- After evaporation, new streams are formed at different locations. due to raining process. This step is like exploration.

- The new stream generated can be shown as

$$x^{steam}_{new} = lb + r_1 \times (up - lb)$$

where r_1 is the random number between [0, 1], lb and ub are the lower and upper bounds given by the problem.

- Thus, these steps are executed over the fixed number of iterations to search for the optimal solution, i.e., the sea.

Proposed Methods

Problem Definition

- We have formulated the ESDS problem as a sentence clustering problem using multi-objective optimization
- Qualities of sentence clusters are measured using two validity indices, PBM and Xie-Beni index.
- In case of summarization, the problem of sentence clustering is formulated
  - Find a set of optimal sentence-clusters, {S 1 , S 2 , . . . , S K } in an automatic way which satisfies the following:
    - $S_i = \{s^i_1 , s^i_2 , . . . , s^i_{npi} \}$, np_i : number of sentences in cluster i, s^i_j : jth sentence of cluster i.
    - $\cup K^{|Si|}_{i=1}$ = N and $S_i \cap S_j = \emptyset$ for all i<>j.
    - Several cluster validity indices, $V\ al_1$, $V\ al_2$, . . . , $V\ al_M$ computed on this partitioning have attained their optimum values.

Proposed Methods

Three methods are proposed based on different multi-objective optimization techniques for summarization task:

- Development of Self-organized multi-objective differential evolution (MODE) based sentence clustering technique
- Development of multi-objective water cycle algorithm (MWCA) and multi-objective grey wolf optimizer (MGWO) based sentence clustering techniques.

NOTE: Differential Evolution, water cycle algorithm and grey wolf optimizer are the optimization algorithms. The developed algorithms for summarization task corresponding to these techniques are called as ESDS_SMODE, ESDS_MWCA and ESDS_MGWO.

Key-points of the proposed algorithms

- A semantic-based scheme is used to represent a sentence in the form of a vector.
- In order to properly calculate the similarity/dissimilarity between two sentences, Word Mover Distance (WMD) is used which also utilizes word2vec model.
- A multi-objective clustering technique is developed to cluster the sentences present in a document.
- Two well-known cluster validity indices, are deployed as the optimization criteria.
- Capable of automatic determination of the number of sentence clusters from a given document.
- Makes use of several sentence scoring features to select some informative sentences from each cluster.

Method-1: ESDS_SMODE

- Uses Differential Evolution as the underlying optimization technique
- Self-organizing Map is used a reproduction operator: used to generate good quality solutions.

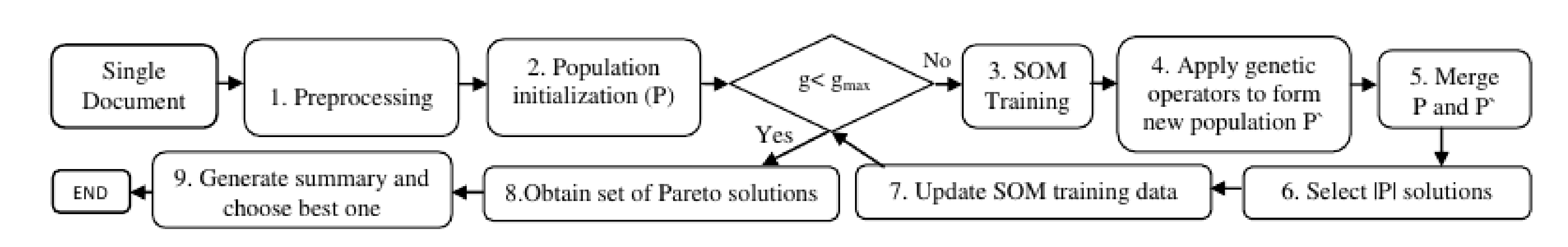


Figure: Flow chart of proposed architecture, ESDS SMODE, where, *gmax* is the user-defined maximum number of generations.

Population Initialization and Objective function Calculation

- Population comprises of set of solutions/chromosomes
- Each solution encodes cluster centers (representative sentences of the documents)
- Each solution has varied number of clusters between [1, N] and associated with two objective functions, PBM and Xie-Beni, where, N is the total number of sentences in the document.

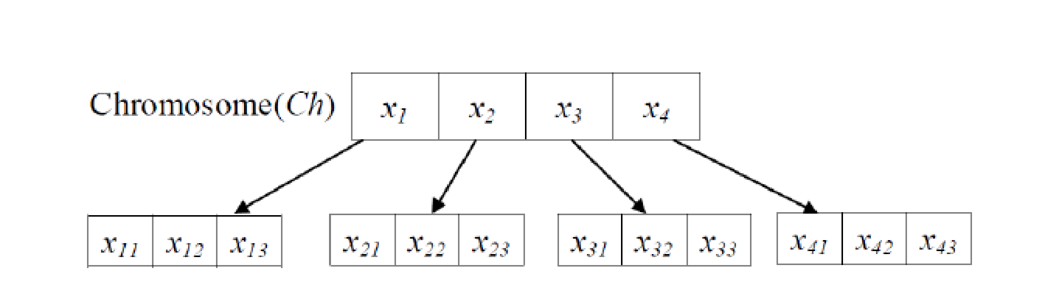


Figure: Chromosome representation; {x1, x2 , x3 , x4 } are the cluster centers where each center is in 3-dimensional vector space.

Genetic operators

- Mating Pool Construction
- Crossover
- Mutation

Mating Pool Construction

- Mating pool is constructed after considering the neighborhood solutions of the current solution retrieved using SOM.
- Only neighboring solutions can mate to generate new solutions.

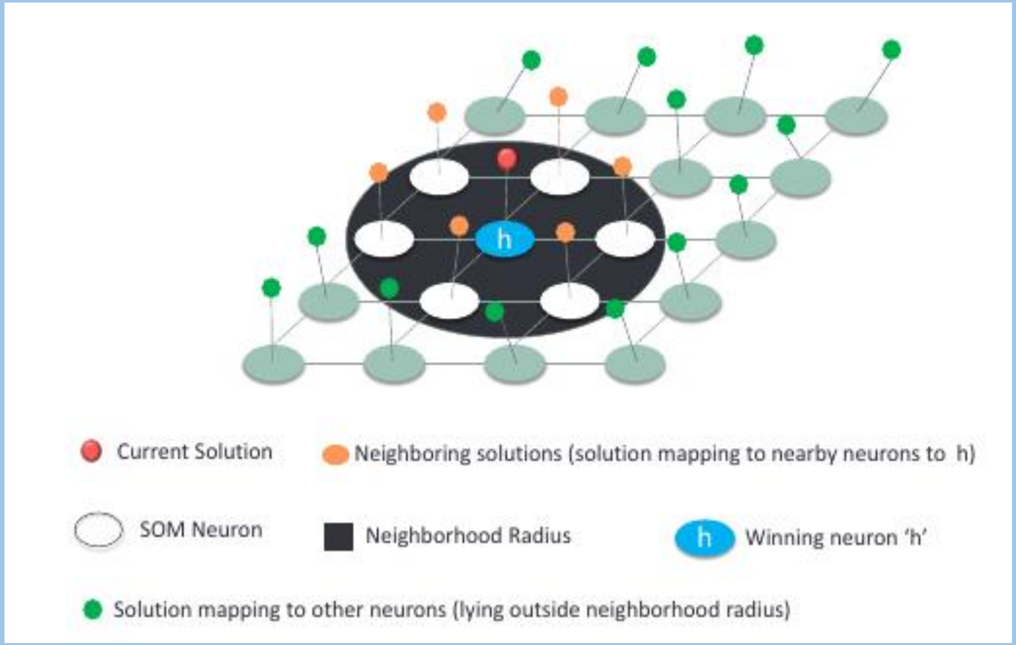


Figure: Mating pool construction for current solution

Crossover

- Random two solutions are selected from the mating pool

  if rand() ≤ CR, then $y'_i = x_i^{current} + F \times (x_i^1 - x_i^2)$, Otherwise $y_i = x_i^{current}$

  Here, rand() is the random probability lying between [0, 1], CR is the crossover probability.

- Repairing of solution y'

  if $y'_i < x^L_i$, then $y''_i = x^L_i$ elseif $y'_i > x^U_i$, then $y''_i = x^U_i$, Otherwise,

$y''_i = y'_i$

Mutation Operator

- After repairing the solution generated by crossover operation, we have applied the concept of polynomial mutation which generates highly diverse solution.

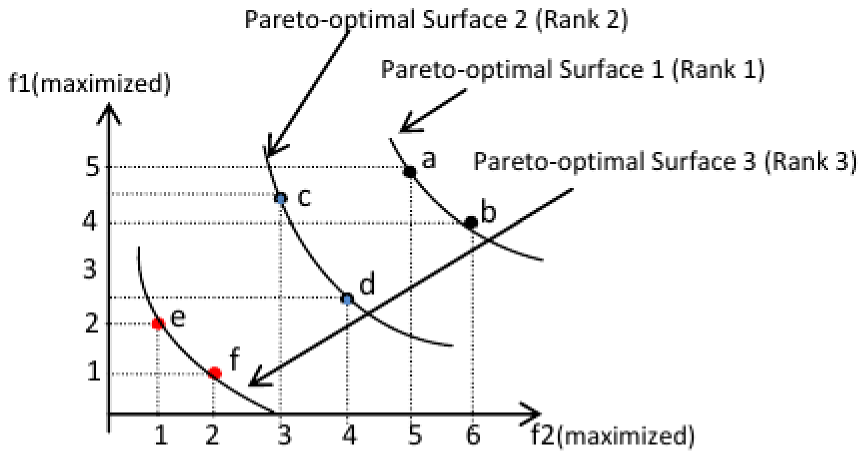
$$\begin{cases} \text{if rand()} \le p_m, \text{then } y_i''' = y_i'' + \delta_i \times (x_i^U - x_i^L), \\ \text{Otherwise } y_i''' = y_i'' \end{cases}$$

$$\begin{cases} \text{if } r_1 < 0.5, \text{then } \left[2r_1 + (1 - 2r_1) \left(\frac{x_i^U - y_i''}{x_i^U - x_i^L} \right)^{\eta_m + 1} \right]^{\frac{1}{\eta_m + 1}} - 1 \\ \text{else, } 1 - \left[2 - 2r_1 + (1 - 2r_1) \left(\frac{y_i'' - x_i^L}{x_i^U - x_i^L} \right)^{\eta_m + 1} \right]^{\frac{1}{\eta_m + 1}} \end{cases}$$

- In order to detect the number of clusters in a document, two more types of mutation are used
  - Insert mutation: increasing the number of clusters present in the ith solution by 1.
    - < c 1 , c 2 , c 3 , 0, 0, 0, 0, 0 > ====> < c 1 , c 2 , c 3 , c 4 , 0, 0, 0, 0 >
  - Delete mutation: decreasing the number of clusters for ith solution by 1.
    - < c 1 , c 2 , c 3 , 0, 0, 0, 0, 0 > ====> < c 1 , c 2 , 0, 0, 0, 0, 0, 0 >

Environmental Selection

- New generated ‘N’ solutions form a new population (P’) which are combined with old population (P) containing ‘N’ solutions
- Non-dominated sorting and crowding distance operator of NSGA-II algorithm is applied to select the top N solutions.



Rank 1: Solutions *a* and *b* are non-dominating to each other because in terms of objective *f1*, solution *a* is better. While in terms of *f2*, solution *b* is better.
Rank 2: Solutions *c* and *d* are non-dominating but dominated by at least one solution of Rank 1 solutions. For example, here, solution *c* dominated by solution *a* because in terms of *f1* and *f2*, solution *a* is better than *c*.
Rank 3: Solutions *e* and *f* are non-dominating but dominated by at least one solution of Rank 1 and Rank 2 solutions. For example, here, solution *e* is dominated by solution *c* and *a*.

Fig.: Representation of non-dominated solutions and dominated relationship.

Summary generation(1/3)

- At the end of the optimization algorithm, set of solutions are obtained.
- ROUGE score of all the solutions are calculated with respect to Gold summary and the solution having the best ROUGE-1 recall score will be considered as the best solution.
- To generate summary for ith solution, following steps are followed:
- First document center is identified

$$m = arg\min \sum_{i=1}^{N} \sum_{j=1, i \neq j}^{N} \frac{dist_{wmd}(s_i, s_j)}{O}$$

Where N is the number of sentences in the document, O is the total number of sentence pairs and is given as (N ×(N −1)/2), s_i is the ith sentence, m is the document center index (m^{th} sentence in the document).

Summary generation(2/3)

- Clusters present in the ith solution are ranked
  - The WMD of each cluster center present in the ith solution to document center is calculated as follows: z_k = distwmd (c_k , s_m), where $1 \le K \le N$, c_k is the kth cluster center. Finally clusters are ranked in descending order based on these z_k scores.
- Calculate sentence score in each cluster:
  - Length of the sentence (F1) ↑

$$L_{s_i^k} = \left(1 - \exp\left(\frac{-l(s_i^k) - \mu(l)}{std(l)}\right)\right) / \left(1 + \exp\left(\frac{-l(s_i^k) - \mu(l)}{std(l)}\right)\right)$$

  - Position of the sentence in the document (F2) ↑

$$p_i = \sqrt{\left(\frac{1}{q_i}\right)}$$

  - Similarity with title (F3) ↑

$$sim\_title_{s_i^k} = dist_{wmd}(s_i^k, title)$$

  - Anti-redundancy (F4) ↓

$$antred_{s_i^k} = \sum_{i,j=1,i\neq j}^{|c_k|} dist_{wmd}(s_i^k, s_j^k)$$

Summary generation(3/3)

- Finally, sentence score is calculated by assigning different weights to various factors (defined above) as:

$$sentence\_score_{s_i} = \alpha \times F1 + \beta \times F2 + \gamma \times F3 + \delta \times F4$$

- Arrange sentences in descending order present in a cluster according to their sentence scores.
- Now, to generate summary, clusters are considered rank-wise. Given a cluster, top ranked sentences are extracted sequentially until summary length reaches to some threshold (in terms of number of words).

Pseudo Code: ESDS_SMODE

Algorithm 1: ESDS_SMODE

Data: Single Text Document

Result: The fittest solution and corresponding summary generated

```
Initialize population_size (| P |), max_iteration ;
Initialize population P=< x^1, x^2 ... x^|P| > and calculate objective functions values for
  each solution
Initial SOM Training data S=P ;
P' ← [ ] //empty population to store new solutions ;
for l=1 to max_iteration do
    Do SOM Training using its training data S
    for i=each solution in P do
        generate mating pool, Q, using neighborhood relationship of trained SOM
        generate new solution using Q, crossover and mutation
        calculate new solution's objective functions values;
        Append new solution into P' ;
    end
    New_P = Merge P and P' ;
    Apply non-dominated sorting (and crowding distance operator if needed) on
      New_P to select the best | P | solutions
    Update SOM Training data as S= P'\ P ;
end
return the fittest solution;
Apply sentence extraction module on the fittest solution;
```

Method-2: ESDS_MGWO

- α is considered as the fittest solution.
- Archive (fixed length) which contains the non-dominated solutions of the Pareto optimal set.
- α, β and δ solutions are selected from archive using Roulette Wheel selection (RWL) such that they are not same.
- If |Archive_size|>fixed_length, RWL mechanism is used to drop out some solutions
- Population initialization, SOM training, mutation (insert and deletion) remains same as in ESDS_SMODE.
- The new wolf generated has a chance to become α/β/δ wolf based on it's fitness functional values.
- Finally, we have to report the generated summary corresponding to the fittest wolf (solution), i.e. α.

Pseudo Code: ESDS_MGWO

Algorithm 2: ESDS_MGWO

Data: Single Text Document

Result: The fittest solution, α and corresponding summary generated

Initialize grey wolf population P=$< \vec{x}^1, \vec{x}^2 \dots \vec{x}^{|P|} >$ and calculate objective functional values for each wolf

Initialize $\vec{A}$, $\vec{C}$ and $\vec{a}$ and number of max_iteration

Apply non-dominated sorting on P and initialize *Archive* with Pareto optimal solutions ;

Select leaders α, β, δ from archive such that $\alpha \neq \beta \neq \delta$;

for *l=1 to max_iteration* **do**

    for *each wolf in population* **do**

        Update wolf positions which will be considered as new wolf;

    end

    Update $\vec{A}$, $\vec{C}$ and $\vec{a}$;

    Calculate objective functional values for each new wolf (wolf after updating positions);

    Apply non-dominated sorting on new wolves and update *Archive* with Rank-1 solutions;

    Apply non-dominated sorting on Archive and obtain the set of Rank-1 non-dominated solutions to form updated Archive ;

    if *Archive.size > threshold_Archive_size* **then**

        remove just enough wolves using Roulette wheel selection mechanism;

    end

    if *any wolf resides outsides of hypercube* **then**

        Update grid to cover new solutions;

    end

    Select leaders α, β, δ from archive such that $\alpha \neq \beta \neq \delta$;

end

return α wolf;

Apply sentence extraction module on α wolf

Method-3: ESDS_MWCA

- Sea is considered as the fittest solution.
- Similar steps are executed as adopted in ESDS MGWO.
- Here, non-dominated sorting along with crowding distance algorithm are used to sort the solutions based on their rankings in the objective space. While, there was no role of crowding distance algorithm in ESDS_MGWO.
- Whenever a new stream is generated, normal, insertion and delete mutation operations are applied as done in ESDS MGWO.
- After number of iterations, the summary corresponding to the solution denoted as sea is reported.

Pseudo Code: ESDS_MWCA

Algorithm 3: ESDS_MWCA

Data: Single Text Document

Result: The fittest solution **sea** and corresponding summary generated

```
Initialize population_size (|P|), N_sr, d_max and max_iterations;
Initialize population P=< x^1, x^2 ... x^|P| > and calculate objective functional values for
  each solution as discussed in Sections
[front_1, front_2 .... front_k]=Apply non-dominated sorting on P;
for front=1 to k do
  | sort front in descending order of crowding distance;
end
Appoint sea, rivers;
Designate streams to rivers and sea
for l=1 to max_iteration do
  | for i=1 to population_size do
  |   | move stream to river, river to sea and stream to sea
  | end
  | for river in population do
  |   | if dist_cos(river, sea) < d_max or rand < 0.1 then
  |   |   | new streams are generated and objective function values are
  |   |   |   calculated;
  |   | end
  | end
  | for stream in population do
  |   | if dist_cos(stream, sea) < d_max then
  |   |   | new streams are generated and objective functions values are
  |   |   |   calculated;
  |   | end
  | end
  | [front_1, front_2 .... front_k]= Apply non-dominated sorting on new population;
  | for front=1 to k do
  |   | sort front in descending order of crowding distance;
  | end
  | Update sea, rivers;
end
return sea ;
Apply sentence extraction module on sea
```

Datasets Used

- Gold standard data from Document Understanding Conference for the years 2001 and 2002 are used.
- Contain 30 and 59 topics each with 309 and 567 news reports.

	DUC2001	DUC2002
#Topics	30	59
#Documents	309	567
Source	TREC	TREC
length of summary (in words)	100	100

Evaluation Measure

$$ROUGE - N = \frac{\sum_{S \in Summary_{ref}} \sum_{N-gram \in S} Count_{match}(N - gram)}{\sum_{S \in Summary_{ref}} \sum_{N-gram \in S} Count(N - gram)}$$

Where N represents the length of n-gram, $Count_{match}(N -gram)$ is the maximum number of overlapping N −grams between reference summary and system summary, Count(N −gram) is the total number of N − gram in the reference summary. In our experiment, N takes the values of 1 and 2 for ROUGE−1 and ROUGE−2, respectively.

Results

Method	DUC2001		DUC2002	
	Average ROUGE-2	Average ROUGE-1	Average ROUGE-2	Average ROUGE-1
ESDS_SMODE	**0.21450**	0.45214	**0.34132**	**0.49117**
ESDS_MGWO	0.15228	0.37108	0.18838	0.41849
ESDS_MWCA	0.14997	0.36702	0.18812	0.41800
MA-SingleDocSum	0.20142	0.44862	0.22840	0.48280
DE	0.18523	**0.47856**	0.12368	0.46694
UnifiedRank	0.17646	0.45377	0.21462	0.48487
FEOM	0.18549	0.47728	0.12490	0.46575
NetSum	0.17697	0.46427	0.11167	0.44963
CRF	0.17327	0.45512	0.10924	0.44006
QSC	0.18523	0.44852	0.18766	0.44865
SVM	0.17018	0.44628	0.10867	0.43235
Manifold Ranking	0.16635	0.43359	0.10677	0.42325

Fig.: ROUGE Scores of different methods on DUC2001 and DUC2002 data sets

An example of Good Summary: ESDS_SMODE

Data set: DUC 2001, Topic: d21d, Document No.: AP880316-0061		
Line No.	Actual Summary	Predicted Summary
1	An engine fire broke out early today on a cruise ship carrying more than 700 people in the Gulf of Mexico , but the blaze appeared to have been brought under control, according to officials .	An engine fire broke out early today on a cruise ship carrying more than 700 people in the Gulf of Mexico , but the blaze appeared to have been brought under control, according to officials and a nearby ship 's captain.
2	The Scandinavian Star was about 35 miles off Isla Mujares , a Mexican island north of the tourist resort Cancun in the Yucatan Peninsula.	-----------No Matching Line------------
3	Petty Officer Brian Lincoln of the U.S. Coast Guard 's Miami office said no injuries were reported.	No injuries were reported , said Petty Officer Brian Lincoln , of the U.S. Coast Guard 's Miami office.
4	``All we know is that the ship's captain`` is completely out of fire-fighting equipment and has requested additional equipment such as foam, water and CO2 , " said Lincoln.	``All we know is that the ship 's captain ... is completely out of fire-fighting equipment and has requested additional fire-fighting equipment such as foam, water and CO2, " said Lincoln, referring to carbon dioxide .

An example of Poor Summary: ESDS_SMODE

Actual Summary:

[Line-1] An earthquake struck in the Soviet Republic of Tadzhikistan at 5 A.M. local time today. *[Line-2]* It measured 5.4 on the Richter Scale. *[Line-3]* The quake unleashed a 50-foot wall of dirt and sodden mud that buried two small towns. *[Line-4]* The center of the quake was located about 25 miles southwest of Dushanbe, the capital of the republic. *[Line-5]* It is estimated that 1,000 are dead ; most of them peasants who were buried beneath mud as they slept. *[Line-6]* This is the second quake to hit the Soviet Union in less than two months . A destructive quake in Armenia on December 7 registered 6.9 and left 25,000 dead .

Predicted Summary:

[Line-1] A predawn earthquake in Soviet Central Asia unleashed a 50-foot-high wall of dirt and mud that buried a mountain village and swept through at least two others Monday, killing up to 1,000 people as they slept, officials said. *[Line-2]* He estimated the number of dead there alone at 600. *[Line-3]* He declined to give his name. *[Line-4]* According to Nasreddinov, President Mikhail S. Gorbachev telephoned Tadzhikistan officials to promise Moscow 's help. *[Line-5]* Damaged roads were hampering those efforts .

Population Size and Number of fitness evaluations

	ESDS_SMODE	ESDS_MGWO	ESDS_MWCA	MA-SingleDocSum	DE
Pop. Size	10	10	10	30	200
NFEs	110	110	110	1600	1000

Fig.: Population size and number of fitness evaluations used by different ESDS approaches.

Quality of Pareto Front Obtained(1/2)

- Generational Distance: It measures the convergence of Pareto optimal front obtained by our approach towards the true Pareto optimal front. Let Q be obtained and Q$*$ be actual Pareto optimal front, M be the number of objective functions. Then GD is denoted as:

$$GD = \frac{\left(\sum_{i=1}^{|Q|} dist_{wmd_i}^{p}\right)^{\frac{1}{p}}}{|Q|} \quad \text{where} \quad dist_{wmd_i} = \min_{s^i \in Q, s^k \in Q^*, k=1}^{|Q^*|} dist_{wmd}(s^i, s^k)$$

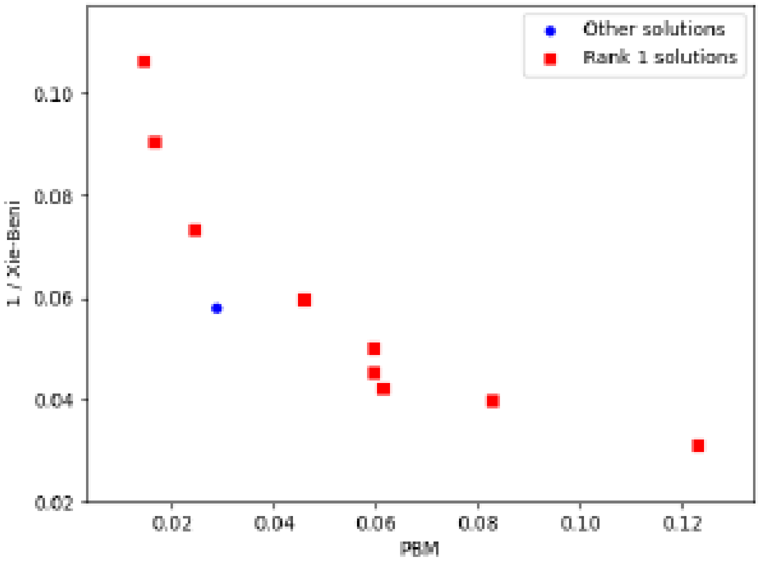
where, $dist_{wmd}(s^i, s^k)$ is the word mover distance between sentences s^i and s^k, the value of p is taken as 2.

- CPU Time: It is the average time taken by our algorithm to generate the final summary.

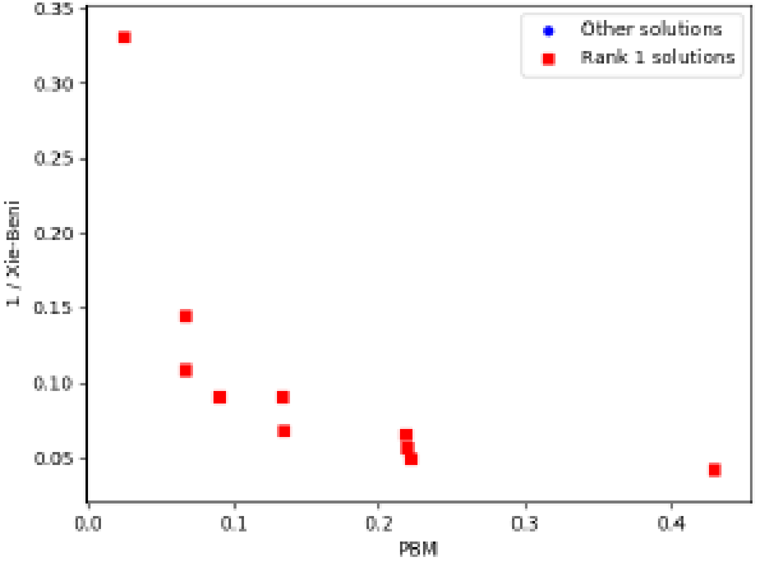
Quality of Pareto Front Obtained(2/2)

Algorithm	DUC2001		DUC2002	
	GD	CPU time (In Sec.)	GD	CPU time (In Sec.)
ESDS_SMODE	0.47461	78.9915	0.44624	50.3384
ESDS_MGWO	0.48860	64.5141	0.45299	43.5634
ESDS_MWCA	0.50086	20.6499	0.45903	11.9936

Pareto Fronts: ESDS_SMODE

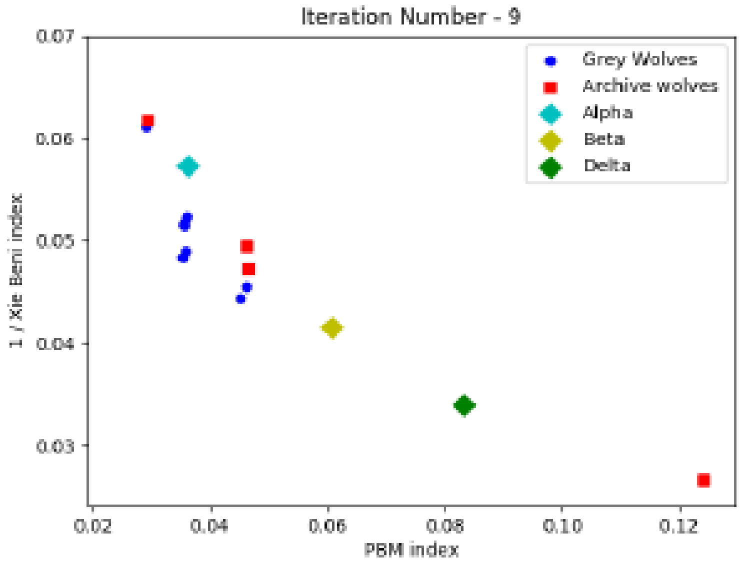


(a)

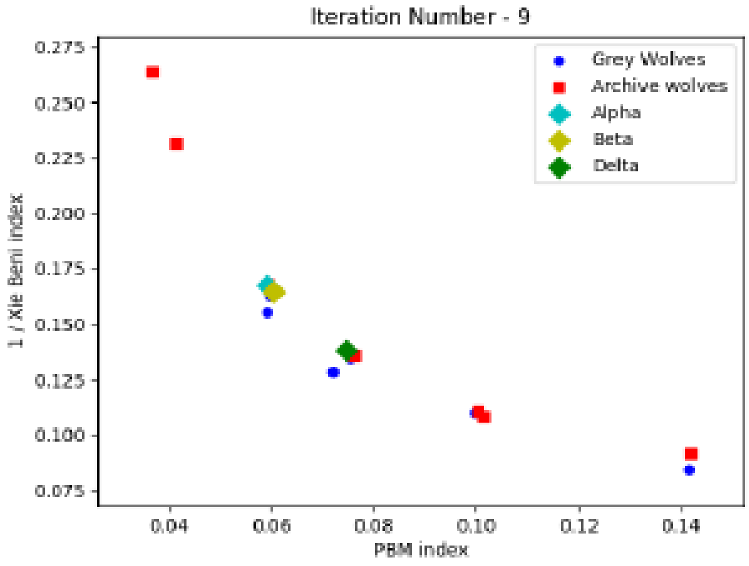


(b)

Pareto Fronts: ESDS_MGWO

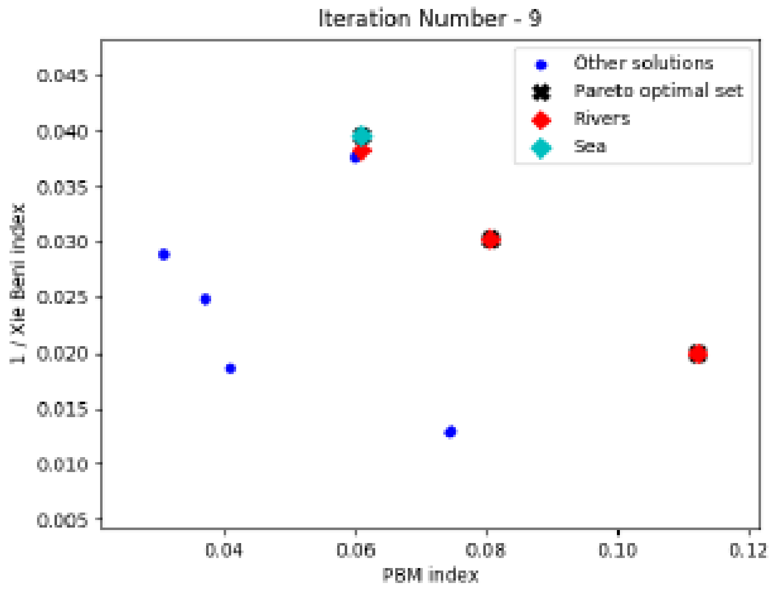


(a)

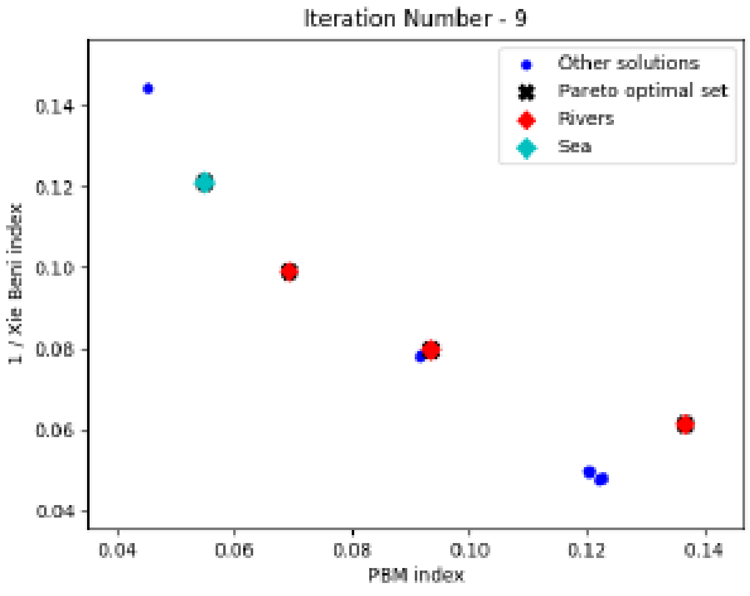


(b)

Pareto Fronts: ESDS_MWCA

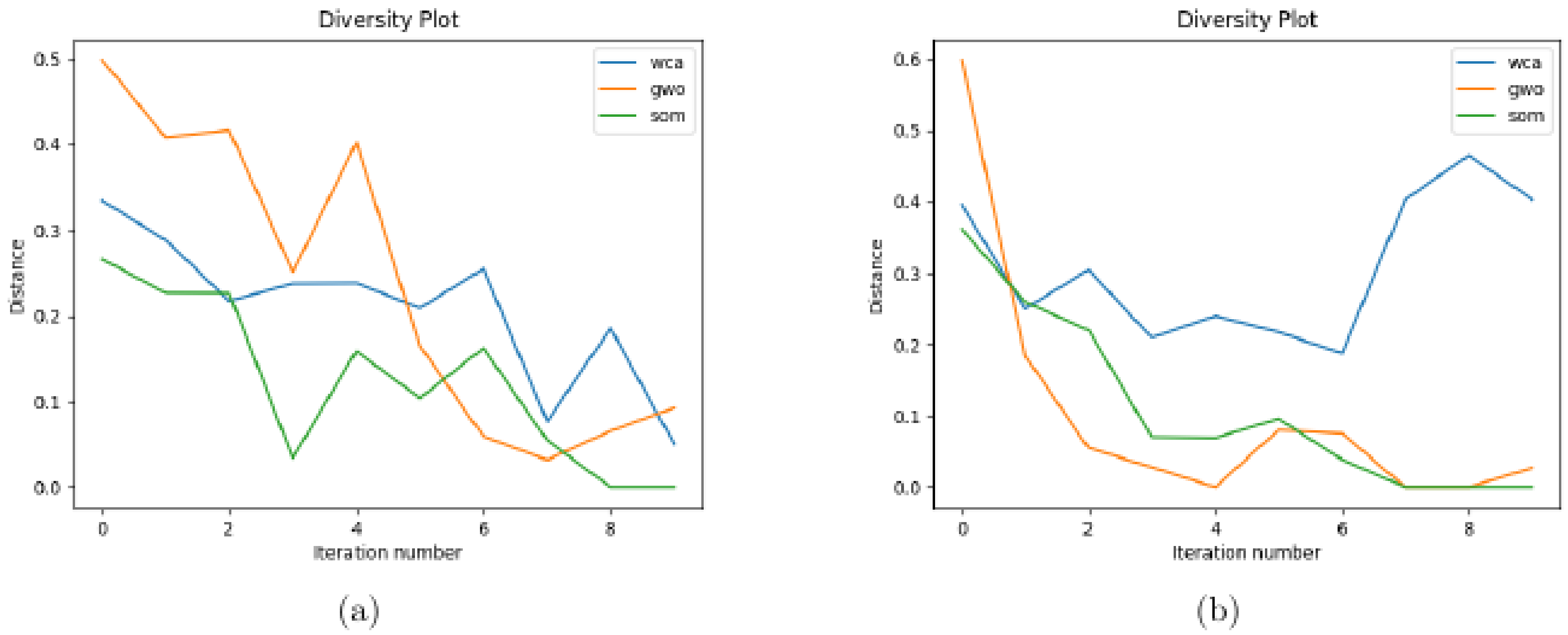


(a)



(b)

Diversity Plots



Improvement Obtained(1/2)

Improvement obtained (IO) is calculated as :

$$IO = \frac{ProposedMethod - OtherMethod}{OtherMethod} \times 100$$

Methods	Improvements obtained by Proposed approach (%)	
	DUC2001	DUC2002
ESDS_MGWO	40.86	81.19
ESDS_MWCA	43.03	81.44
MA-SingleDocSum	6.49	49.44
DE	15.80	175.97
UnifiedRank	21.56	59.03
FEOM	15.64	173.27
NetSum	21.21	205.65
CRF	23.80	212.45
QSC	15.80	81.88
SVM	26.04	214.09
Manifold Ranking	28.94	219.68

Fig.: Improvements obtained by our proposed approach over other methods based on ROUGE−2 score

Improvement Obtained(2/2)

Methods	DUC2002
ESDS_MGWO	17.37
ESDS_MWCA	17.50
MA-SingleDocSum	1.73
DE	5.19
UnifiedRank	1.30
FEOM	5.46
NetSum	9.24
CRF	11.61
QSC	9.48
SVM	13.60
Manifold Ranking	16.05

Table: Improvements obtained by our proposed approach over other methods using ROUGE−1 score on DUC2002 dataset

Methods	DUC2001
Proposed approach	5.84
ESDS_MGWO	28.96
ESDS_MWCA	30.29
MA-SingleDocSum	6.67
UnifiedRank	5.46
FEOM	0.27
NetSum	3.08
CRF	5.15
QSC	6.70
SVM	7.23
Manifold Ranking	10.37

Table: Improvements obtained by DE over other methods using ROUGE−1 score on DUC2001 dataset

Conclusion

- Three methods are proposed for summarization utilizing three search approaches: self-organized multi-objective differential evolution, multi-objective grey wolf optimizer and multi-objective water cycle algorithm.
- Two sentence-cluster quality measures are optimized simultaneously.
- ESDS_SMODE improves by 6.49% points for DUC2001, while, for DUC2002 dataset, our best approach improves by 49.44% points over the best approach, namely, MA-SingleDocSum.
- ROUGE-1: for DUC2002 dataset, our best approach improves by 1.30% points over UnifiedRank approach.

Future Work

- The effect on performance using other sentence representation schemes and different sentence similarity/dissimilarity measures
- Application to multi-document summarization, Microblog summarization.
- Automatic adaption of various parameters
- Application of this approach for query based single document text summarization

Another Proposed Approach based on Binary Differential Evolution

Key-points

- Consider extractive text-summarization as a binary optimization problem
- Multi-objective binary differential evolution (DE) based optimization strategy is employed to solve this.
- Six quality measures of summary are optimized simultaneously.
- Self-organizing Map based genetic operators are incorporated in the optimization process to improve the convergence similar to ESDS_SMODE approach..
- To measure the similarity/dissimilarity between sentences, different existing measures like normalized Google distance, word mover distance, and cosine similarity are explored.

- We have obtained 45% and 4% improvements, while for the DUC2002 dataset, improvements obtained by our approach are 26% and 6%, considering ROUGE−2 and ROUGE−1 scores, respectively.
- It was also shown that the best performance not only depends on the objective functions used but also on the correct choice of similarity/dissimilarity measure between sentences.

NOTE: It differs from ESDS_SMODE, in terms of type of optimization problem (here it is binary), #objectives functions, solution representation, crossover and mutation operator definition, sentence similarity/dissimilarity measures.

Objective Functions Used(1/2)

- Sentence Position (↑) Similarity with the title (↑), length of the sentence (↑) are similar as used in ESDS_SMODE.
- Coverage (↑): measures the extent to which sentences in the summary provide useful information about the document.

$$CoV = \sum_{\forall s_i \in Summary} \sum_{\forall s_j \in Doc, s_i \neq s_j} \frac{sim(s_i, s_j)}{N-1}$$

- Readability Factor(↑):

$$R = \sum_{i=2}^{N_p} sim(s_i, s_{i-1})$$

Objective Functions Used(2/2)

- Cohesion: measures the relatedness of the sentences in the summary.

$$COH = \frac{\log(C_s \times 9 + 1)}{\log(M * 9 + 1)}$$

Where,

$$C_s = \frac{\sum_{\forall s_i, s_j \in Summary} sim(s_i, s_j)}{O_s} \text{and, } O_s = \frac{N \times (N-1)}{2}$$

Developed methods

1. **Approach-1:** In this approach all objective functions are assigned some importance factors. For example, if fitness values of six objective functions are < ob1 , ob2 , ob3 , ob4, ob5 , ob6 > and weights assigned are <α, β, γ, δ, λ, φ >, then < ob1 ×α, ob2 ×β, ob3 ×γ, ob4 ×δ, ob5 ×λ, ob6 ×φ > are optimized simultaneously. The values of these weights are selected after conducting a thorough literature survey.
2. **Approach-2**: In this approach all objective functions are simultaneously optimized without assigning any weight values.

NOTE: Both approaches are developed with SOM and without SOM.

Solution Representation

- Each solution is represented as a binary vector.
- Example: if a document consists of 10 sentences then a valid solution can be represented as [1, 0, 0, 1, 1, 0, 1, 0, 0, 0].
- This solution indicates that first, fourth, fifth and seventh sentences of the original document should be in summary.
- Each solution associated with six objective functions values..
- Summary length constraint:

$$\sum_{s_i \in Summary} l_i \leq S_{max},$$

where, l_i measures the length of sentence in terms of number of words, S_{max} is the maximum number of words allowed in generated summary.

Genetic Operators: Mutation and Crossover

- Mutation:

$$P(x_j^t) = \frac{1}{1 + e^{\frac{2b \times [x_{r1,j}^t + F \times (x_{r2,j}^t - x_{r3,j}^t) - 0.5]}{1+2F}}}$$

$$y_j' = \begin{cases} 1, \text{ if rand}() \leq P(x_j^t) \\ 0, otherwise \end{cases}$$

- Crossover

$$y_j'' = \begin{cases} y_j', \text{ if rand}() \leq CR \\ x_j, \text{ Otherwise} \end{cases}$$

Parameters

- |P|= 40, mating Pool size=4, max. generations= 25, crossover probability (CR)=0.2, b=6, F=0.8.
- SOM parameters: initial neighborhood size (σ_0)=2, initial learning rate (σ0)=0.6, training iteration in SOM=|P|, topology=rectangular 2D grid; grid size=5 × 8.
- Importance factors/weight values assigned to different objective functions: $\alpha = 0.25$, $\beta = 0.25$, $\gamma = 0.10$, $\delta = 0.11$, $\lambda = 0.19$, $\varphi = 0.10$; System summary: length (in words)=100 words.
- Word Mover Distance makes use of pre-trained GoogleNews corpus to calculate the distance between two sentences.

Results

		DUC2001		DUC2002	
		ROUGE-2	ROUGE-1	ROUGE-2	ROUGE-1
Approach1 (NGD)	With SOM	**0.26949**	**0.47699**	**0.27846**	**0.50225**
	Without SOM	**0.26742**	**0.47521**	**0.27705**	**0.50191**
Approach2 (NGD)	With SOM	**0.26774**	**0.47291**	**0.27519**	**0.49899**
	Without SOM	**0.26265**	**0.46762**	**0.27654**	**0.50162**
Approach1 (CS)	With SOM	**0.26459**	**0.47554**	**0.27649**	**0.50624**
	Without SOM	**0.25282**	**0.46289**	**0.27292**	**0.50050**
Approach2 (CS)	With SOM	**0.26209**	**0.47398**	**0.25961**	**0.49159**
	Without SOM	**0.26629**	**0.47862**	**0.27319**	**0.50147**
Approach1 (WMD)	With SOM	***0.29238*** †	***0.50236*** †	***0.28846*** †	***0.51662*** †
	Without SOM	**0.28930**	**0.49486**	**0.28556**	**0.51441**
Approach2 (WMD)	With SOM	**0.28462**	**0.49863**	**0.28520**	**0.51538**
	Without SOM	**0.28190**	**0.48877**	**0.28656**	**0.51406**
MA-SingleDocSum [9]	-	0.20142	0.44862	0.22840	0.48280
DE [6]	-	0.18523	0.47856	0.12368	0.46694
UnifiedRank [48]	-	0.17646	0.45377	0.21462	0.48487
FEOM [32]	-	0.18549	0.47728	0.12490	0.46575
NetSum [29]	-	0.17697	0.46427	0.11167	0.44963
CRF [27]	-	0.17327	0.45512	0.10924	0.44006
QSC [26]	-	0.18523	0.44852	0.18766	0.44865
SVM [25]	-	0.17018	0.44628	0.10867	0.43235
Manifold Ranking [31]	-	0.16635	0.43359	0.10677	0.42325
NN-SE [33]	-	xx	xx	0.23	0.474
SummaRuNNer [34]	-	xx	xx	0.231 ± 0.008	0.466 ± 0.008

Table : ROUGE Scores of different methods on DUC2001 and DUC2002 data using Normalized Google Distance (NGD), Cosine Similarity (CS) and Word Mover Distance (WMD). Here, † denotes the best results; it also indicates that results are statistically significant at 5% significance level; xx indicates results are not available in reference paper.

Pareto fronts obtained

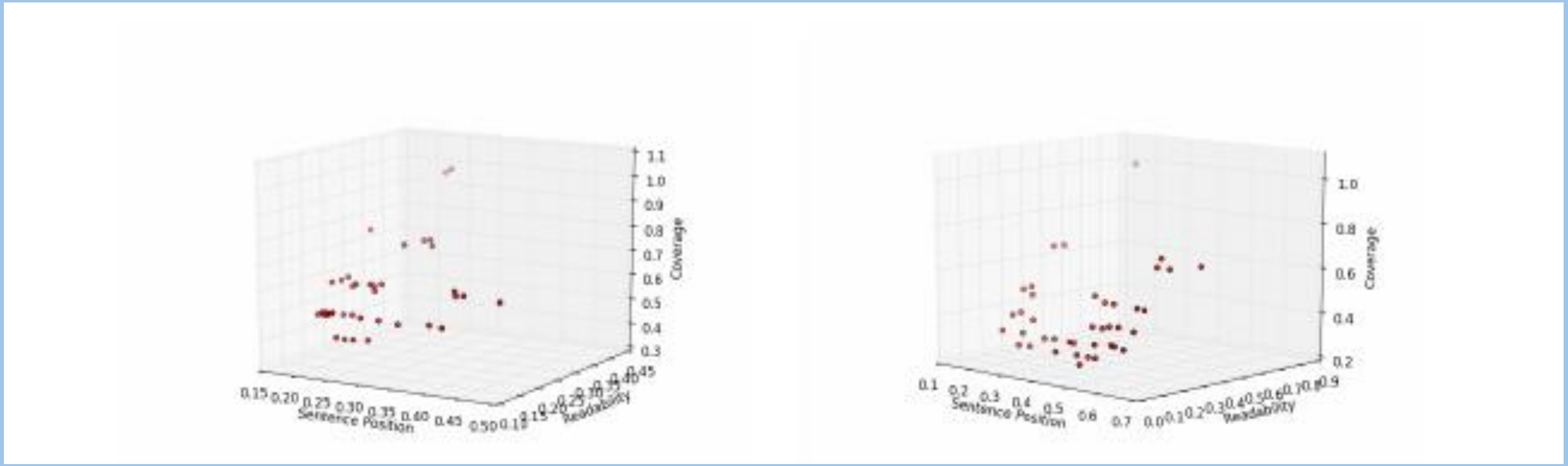


Fig: Figure 3: Pareto optimal fronts obtained after application of the proposed Approach-1 (WMD) with SOM based operators after first, and fourteen generations. Here, red color dots represent Pareto optimal solutions; three axes represent three objective functional values, namely, sentence position, readability, coverage.

An example of good summary

Reference summary: DUC2001 -> d21d -> AP881109-0149

[Line-1] The cruise ship Song of America was forced to return to port after an engine seized up and started a small fire, but no one was hurt . *[Line 2]* The ship left Miami on Sunday with about 1,300 passengers on a Caribbean cruise. Rick Steck, a spokesman for Royal Caribbean Cruise Line said the fire was quickly doused by crewmembers . *[Line 3]* The passengers, who had been brought on deck , were allowed to resume the evening's activities. *[Line 4]* The 705-foot ship turned around and returned to Cozumel on its remaining three engines to replenish firefighting supplies. The passengers stayed aboard , and the ship will return to Miami on Thursday or Friday .

Predicted summary:

[Line-1] The cruise ship Song of America was forced to return to port after an engine seized up and started a small fire, but no one was hurt, the ship 's owner said today. *[Line 2]* The ship left Miami on Sunday with about 1,300 passengers on a Caribbean cruise. *[Line 3]* The passengers were mustered on deck while crew members doused the blaze, but then allowed to resume the evening's activities, he said. *[Line 4]* The 705-foot ship turned around and returned to Cozumel on its remaining three engines to replenish firefighting supplies, Steck said.

An example of bad summary

Reference summary: **DUC2001 -> d60k ->** SJMN91-06106024

[Line-1] Rodney King spends his time seeing doctors and thinking about his injuries he fears may become permanent. ***[Line-2]*** He is staying with relatives and fears retribution by the police. *[Line-3]* His ex-wife says he's depressed and frightened; his attorney has hired guards to protect him. ***[Line-4]*** King suffers *headaches and numbness of the face* after five hours of plastic surgery to repair fractures of his cheek and eye bones, and has instituted an $83 million law suit against the city for excessive force. ***[Line-5]*** In another development, he 's now a suspect in a February 21 robbery and shooting, a result of the wide publicity.

Predicted summary:

[Line-1] Six weeks after his beating by Los Angeles police and seemingly forgotten in the political turmoil that has followed -- Rodney G. King fears retribution, spends most of his time seeing doctors, and thinks a lot about the headaches, scars and facial numbness he worries might become permanent. ***[Line-2]*** Lerman has filed an $83 million claim against the city on King's behalf. King's neat, blue home in Altadena has the curtains drawn, its phone number and those of other family members long changed .

Improvements obtained

Methods	Improvements obtained by Proposed approach (%)			
	DUC2001		DUC2002	
	ROUGE-2	ROUGE-1	ROUGE-2	ROUGE-1
MA-SingleDocSum	45.16	11.98	26.3	7.01
DE	75.76	15.86	170.17	22.06
UnifiedRank	65.69	10.71	34.41	6.55
FEOM	57.85	4.97	133.23	10.64
NetSum	65.21	8.21	158.32	14.9
CRF	68.74	10.38	164.07	17.4
QSC	57.85	12.01	53.72	15.15
SVM	57.63	5.26	130.96	10.92
Manifold Ranking	71.81	12.57	165.45	19.49
NN-SE	xx	xx	25.41	8.99
SummaRuNNer	xx	xx	22.33	13.21

Table: Improvements obtained by the proposed approach, Approach-1 (WMD) with SOM based operators over other methods concerning ROUGE scores. Here, xx indicates non-availability of results on the DUC2001 dataset.

Conclusion

- A self-organized multi-objective binary differential evolution technique is proposed for summary extraction.
- Three similarity/dissimilarity criteria are used to measures the same between two sentences.
- Six objectives are optimized simultaneously covering different aspects of summary.
- SOM-based approach with WMD as a distance measure has obtained 45% and 4% improvements over the best existing method considering ROUGE−2 and ROUGE−1 scores, respectively, for the DUC2001 dataset. While for the DUC2002 dataset, improvements obtained by our approach are 26% and 6%, considering ROUGE−2 and ROUGE−1.

Future Work

- As the performance of summarization system depends on types of similarity/dissimilarity measures used and also depends on the dataset, therefore, in future, we will try to make the similarity/dissimilarity measure selection automatic for different datasets. In future, we also want to extend the current approach for multi-document summarization.

Future Work

- The effect on performance using other sentence representation schemes and different sentence similarity/dissimilarity measures
- Application to multi-document summarization, Microblog summarization.
- Automatic adaption of various parameters u
- Apply this approach for query based single document

Publications

- **Saini, N.**, Saha, S., & Bhattacharyya, P. (**2018**). Automatic Scientific Document Clustering Using Self-organized Multi-objective Differential Evolution. ***Cognitive Computation***, *11*(2), 271-293. (**Impact factor: 4.87**)

- **Saini, N.**, Saha, S., Jangra, A., & Bhattacharyya, P. (**2018**). Extractive single document summarization using multi-objective optimization: Exploring self-organized differential evolution, grey wolf optimizer and water cycle algorithm. ***Knowledge-Based Systems***, *164*, 45-67. (**Impact factor: 5.10**)

- **Saini, N.**, Saha, S., Bhattacharyya, P., Tuteja, H. (August **2019**). Textual Entailment based Figure Summarization for Biomedical Articles, ***ACM Transactions on Multimedia Computing Communications and Applications***. (accepted) (**Impact Factor: 2.25**)

- **Saini, N.,** S., Saha, S., & Bhattacharyya, P. (**2019**) A Multi-objective Based Approach for Microblog Summarization, ***IEEE Transactions On Computational Social Systems.***

Publications

- **Saini, N.**, S., Saha, S., Chakraborty, D., & Bhattacharyya, P. **(2019)**. Extractive single document summarization using binary differential evolution: Optimization of difference sentence quality measures, **PLoS ONE** 14(11): e0223477. (**Impact factor: 2.76, h5 index: 176**)

- **Saini, N.**, S., Saha, S., Potnuru, V., Grover, R., & Bhattacharyya, P. **(2019)** Figure-Summarization: A Multi-objective optimization based approach, ***IEEE Intelligent Systems.*** (**Impact factor: 4.64**)

- **Saini, N.**, Saha, S., Soni, C., Bhattacharyya, P. (September **2019**). Automatic Evolution of Bi-clusters from Microarray Data using Self-Organized Multi-objective Evolutionary Algorithm, ***Applied Intelligence***. (**Impact factor: 2.88**)

- **Saini, N.,** Saha, S., Harsh, A., & Bhattacharyya, P. (**2018**). Sophisticated SOM based genetic operators in multi-objective clustering framework. ***Applied Intelligence***, *49*(5), 1803-1822**.** (**Impact factor: 2.88**)

Publications

- **Saini, N.,** Chourasia, S., Saha, S., & Bhattacharyya, P. (**2017**). A self-organizing map based multi-objective framework for automatic evolution of clusters. In ***International Conference on Neural Information Processing*** (ICONIP 2017) (pp. 672-682). Springer, Cham. (**Core ranking: A**).

- **Saini, N.,** Saha, S., & Bhattacharyya, P. (**2018**). Cascaded SOM: An improved technique for automatic email classification. In 2018 **International Joint Conference on Neural Networks** (IJCNN 2018) (pp. 1-8). IEEE. (**Core ranking: A**)

- **Saini, N.,** S., Saha, S., Bhattacharyya, P. (September **2019**). Incorporation of Neighborhood Concept in Enhancing SOM based Multi-label Classification. In ***International Conference on Pattern Recognition and Machine Intelligence*** (PReMI 2019). Springer.

- **Saini, N., S.,** Saha, S., Kumar, A., & Bhattacharyya, P. (September 2019). Multi-document Summarization using Adaptive Composite Differential Evolution. **In *International Conference on Neural Information Processing* (ICONIP 2019).** Springer. **(Core ranking: A).**

References

[1] R. M. Aliguliyev, A new sentence similarity measure and sentence based extractive technique for automatic text summarization, Expert Systems with Applications 36 (4) (2009) 7764–7772.

[2] H. Zhang, A. Zhou, S. Song, Q. Zhang, X. Z. Gao, J. Zhang, A self-organizing multiobjective evolutionary algorithm, IEEE Transactions on Evolutionary Computation 20 (5) (2016) 792–806. doi:10.1109/TEVC.2016.2521868.

[3] S. Mirjalili, S. Saremi, S. M. Mirjalili, L. d. S. Coelho, Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization, Expert Systems with Applications 47 (2016) 106–119.

[4] A. Sadollah, H. Eskandar, A. Bahreininejad, J. H. Kim, Water cycle algorithm for solving multi-objective optimization problems, Soft Computing 19 (9) (2015) 2587–2603.

[5] X. Wan, Towards a unified approach to simultaneous single-document and multi-document summarizations, in: Proceedings of the 23rd international conference on computational linguistics, Association for Computational Linguistics, 2010, pp. 1137–1145.

[6] K. Svore, L. Vanderwende, C. Burges, Enhancing single-document summarization by combining ranknet and third-party sources, in: Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL), 2007.

[7] D. Shen, J.-T. Sun, H. Li, Q. Yang, Z. Chen, Document summarization using conditional random fields., in: IJCAI, Vol. 7, 2007, pp. 2862–2867.

[8] D. M. Dunlavy, D. P. OLeary, J. M. Conroy, J. D. Schlesinger, Qcs: A system for querying, clustering and summarizing documents, Information processing & management 43 (6) (2007) 1588–1605.

[9] X. Wan, J. Yang, J. Xiao, Manifold-ranking based topic-focused multi-document summarization., in: IJCAI, Vol. 7, 2007, pp. 2903–2908.

[10] J.-Y. Yeh, H.-R. Ke, W.-P. Yang, I.-H. Meng, Text summarization using a trainable summarizer and latent semantic analysis, Information processing & management 41 (1) (2005) 75–95.

[11] W. Song, L. C. Choi, S. C. Park, X. F. Ding, Fuzzy evolutionary optimization modeling and its applications to unsupervised categorization and extractive summarization, Expert Systems with Applications 38 (8) (2011) 9112–9121.

[12] T. Kohonen, The self-organizing map, Neurocomputing 21 (1) (1998) 1–6.

[13] S. Saha, S. Bandyopadhyay, A generalized automatic clustering algorithm in a multiobjective framework, Applied Soft Computing 13 (1) (2013) 89–108.

Anu Queries???

Thank You!!