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
<table>
<thead>
<tr>
<th>S. No.</th>
<th>Types of summary</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Generic and query-focused</td>
<td>whether general or query related data is required</td>
</tr>
<tr>
<td>2</td>
<td>Supervised and unsupervised</td>
<td>Availability of training data</td>
</tr>
<tr>
<td>3</td>
<td>Mono, multi and cross-lingual</td>
<td>Language</td>
</tr>
<tr>
<td>4</td>
<td>Web-based</td>
<td>For summarizing web pages</td>
</tr>
<tr>
<td>5</td>
<td>E-mail based</td>
<td>For summarizing e-mails</td>
</tr>
<tr>
<td>6</td>
<td>Personalized</td>
<td>Information specific to a user’s need</td>
</tr>
<tr>
<td>7</td>
<td>Sentiment-based</td>
<td>Opinions are detected</td>
</tr>
</tbody>
</table>
[English is the dominant language in the writing and publishing of scientific research in the form of scientific articles.] [However, many non-natives users of English suffer the interference of their mother tongues when writing scientific papers in English.] [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.] [In order to ease these users’ problems, we developed a learning environment for scientific writing named AMADEUS (Amiable Article Development for User Support).] [AMADEUS consists of several interrelated tools reference, support, critic and tutoring tools and provides the context in which this dissertation is inserted.] [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.] [In order to do that we introduce the concept of adaptivity in computer systems and describe several user model shells.] [We also provide details about intelligent agents which were used to implement the user model for the AMADEUS environment.]
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 judiciary official said Tuesday.
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

<table>
<thead>
<tr>
<th>Method</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA-SingleDocSum</td>
<td>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.</td>
</tr>
<tr>
<td>DE</td>
<td>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.</td>
</tr>
<tr>
<td>UnifiedRank</td>
<td>UnifiedRank method proposed by X. Wan and presents a graph-based model to solve single and multi-document summarization problem simultaneously.</td>
</tr>
<tr>
<td>CRF</td>
<td>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.</td>
</tr>
<tr>
<td>Method</td>
<td>Contribution</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>QSC</td>
<td>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.</td>
</tr>
<tr>
<td>SVM</td>
<td>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.</td>
</tr>
<tr>
<td>UnifiedRank</td>
<td>UnifiedRank method proposed by X. Wan and presents a graph-based model to solve single and multi-document summarization problem simultaneously.</td>
</tr>
<tr>
<td>FEOM</td>
<td>Song et al. have proposed fuzzy evolutionary optimization modeling (FEOM) technique and showed its application to extractive summarization.</td>
</tr>
<tr>
<td>Manifold Ranking</td>
<td>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.</td>
</tr>
</tbody>
</table>
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 \( x = \{ x_1, x_2 \ldots x_n \} \) of ‘n’ decision variables that optimizes \( M \) number of objective functions \( \{ f_1(x), f_2(x) \ldots f_M(x) \} \) simultaneously while satisfying some constraints if any.

• Mathematically, it is formulated as

\[
\text{min} \quad F(x) = \{ f_1(x), f_2(x) \ldots f_M(x) \}^T
\]

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

In real-world, we have to simultaneously optimize two or more than two objective functions which leads to more than one solution.
Minimize $f_1 = x^2$

Minimize $f_2 = (x-2)^2$

- The solution $x = 0$ is optimum w.r.t. $f_1$ but not so good with respect to $f_2$.
- The solution $x = 2$ is optimum w.r.t. function $f_2$ and not so good with respect to $f_1$.
- Optimal range: $0 \leq x \leq 2$ which provides a set of solutions.
A solution \( \text{sol} \) in M-dimensional objective space is represented as

\[
\text{sol} = \{ f_1(\text{sol}), f_2(\text{sol}) \ldots f_M(\text{sol}) \}
\]

where \( f_i(\text{sol}), 1 \leq i \leq M \) is the value of \( i \)th objective function.

- Representation of 5 solutions:
  - \( \text{sol1} = \{1, 1\} \)
  - \( \text{sol2} = \{1, 2\} \)
  - \( \text{sol3} = \{3, 1\} \)
  - \( \text{sol4} = \{2, 3\} \)
  - \( \text{sol5} = \{4, 2\} \)
Solutions Relationship: Dominance (for minimization problem) (2/4)

A solution $\text{sol}_i = \{f_1(\text{sol}_i), f_2(\text{sol}_i), \ldots, f_M(\text{sol}_i)\}$ dominates another solution $\text{sol}_j = \{f_1(\text{sol}_j), f_2(\text{sol}_j), \ldots, f_M(\text{sol}_j)\}$ denoted as $\text{sol}_i < \text{sol}_j$ iff

1. $f_m(\text{sol}_i) \leq f_m(\text{sol}_j) \quad \forall m \in \{1, 2, \ldots, M\}$
2. $f_m(\text{sol}_i) < f_m(\text{sol}_j) \quad \exists m \in \{1, 2, \ldots, M\}$

$\text{sol}_i$ and $\text{sol}_j$ are non-dominated represented as $\text{sol}_i \preceq \text{sol}_j$ iff neither $\text{sol}_i < \text{sol}_j$ nor $\text{sol}_j < \text{sol}_i$. 
Solutions Relationship (3/4)

In the Figure

- $\text{sol1} < \{\text{sol2, sol3, sol4, sol5}\}$
- $\text{sol2} < \{\text{sol4, sol5}\}$
- $\text{sol3} < \{\text{sol5}\}$
- $\text{Sol2} \preceq \text{sol3}$
- $\text{Sol3} \preceq \text{sol4}$
- $\text{sol4} \preceq \text{sol5}$
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. $\forall sol_i, sol_j \in F_k: sol_i \leq sol_j \quad 1 \leq k \leq K$
2. $\forall sol \in F_k, \exists sol' \in F_{k-1}: sol' < sol \quad 2 \leq k \leq K$
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 \( sol_1, sol_2, sol_3, sol_4, sol_5 \) are

0, 1, 1, 2 and 3, respectively.
An Example: Non-dominated Sorting

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

- For solution $sol_2$
  - $S_{sol_2} = \{sol_4, sol_5\}$
  - $n_{sol_2} = 1$
  - $sol_2 \succ \{sol_1\}$

- For solution $sol_3$
  - $S_{sol_3} = \{sol_5\}$
  - $n_{sol_3} = 1$
  - $sol_3 \succ \{sol_1\}$

- For solution $sol_4$
  - $S_{sol_4} = \{}$
  - $n_{sol_4} = 2$
  - $sol_4 \succ \{sol_1, sol_2\}$

- For solution $sol_5$
  - $S_{sol_5} = \{}$
  - $n_{sol_5} = 3$
  - $sol_5 \succ \{sol_1, sol_2, sol_3\}$

Fig.: Solutions in objective space
An Example: Non-dominated Sorting

\[
\begin{align*}
    n_{sol_1} &= 0 & S_{sol_1} &= \{sol_2, sol_3, sol_4, sol_5\} \\
    n_{sol_2} &= 1 & S_{sol_2} &= \{sol_4, sol_5\} \\
    n_{sol_3} &= 1 & S_{sol_3} &= \{sol_5\} \\
    n_{sol_4} &= 2 & S_{sol_4} &= \{\} \\
    n_{sol_5} &= 3 & S_{sol_5} &= \{\}
\end{align*}
\]

\[n_{sol_1} = 0\]

- \(sol_{1_{\text{rank}}} = 1\)
- \(F_1 = \{sol_1\}\)

Fig.: Solutions in objective space
An Example: Non-dominated Sorting

\[ n_{sol_1} = 0 \quad S_{sol_1} = \{sol_2, sol_3, sol_4, sol_5\} \]
\[ n_{sol_2} = 0 \quad S_{sol_2} = \{sol_4, sol_5\} \]
\[ n_{sol_3} = 1 \quad S_{sol_3} = \{sol_5\} \]
\[ n_{sol_4} = 2 \quad S_{sol_4} = \{\} \]
\[ n_{sol_5} = 3 \quad S_{sol_5} = \{\} \]
An Example: Non-dominated Sorting

\begin{align*}
  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{align*}

\begin{align*}
  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{align*}

\[ n_{sol_2} = 0, \ n_{sol_3} = 0 \]

- \( sol_{2\text{rank}} = sol_{3\text{rank}} = 2 \)
- \( F_2 = \{sol_2, sol_3\} \)
An Example: Non-dominated Sorting

<table>
<thead>
<tr>
<th>$n_{sol_1}$</th>
<th>$n_{sol_2}$</th>
<th>$n_{sol_3}$</th>
<th>$n_{sol_4}$</th>
<th>$n_{sol_5}$</th>
<th>$S_{sol_1}$</th>
<th>$S_{sol_2}$</th>
<th>$S_{sol_3}$</th>
<th>$S_{sol_4}$</th>
<th>$S_{sol_5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>${sol_2, sol_3, sol_4, sol_5}$</td>
<td>${sol_4, sol_5}$</td>
<td>${sol_5}$</td>
<td>${}$</td>
<td>${}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_{sol_1}$</th>
<th>$n_{sol_2}$</th>
<th>$n_{sol_3}$</th>
<th>$n_{sol_4}$</th>
<th>$n_{sol_5}$</th>
<th>$S_{sol_1}$</th>
<th>$S_{sol_2}$</th>
<th>$S_{sol_3}$</th>
<th>$S_{sol_4}$</th>
<th>$S_{sol_5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>${sol_2, sol_3, sol_4, sol_5}$</td>
<td>${sol_4, sol_5}$</td>
<td>${sol_5}$</td>
<td>${}$</td>
<td>${}$</td>
</tr>
</tbody>
</table>
An Example: Non-dominated Sorting

\[
\begin{align*}
    n_{sol_1} &= 0 & S_{sol_1} &= \{sol_2, sol_3, sol_4, sol_5\} \\
    n_{sol_2} &= 0 & S_{sol_2} &= \{sol_4, sol_5\} \\
    n_{sol_3} &= 0 & S_{sol_3} &= \{sol_5\} \\
    n_{sol_4} &= 0 & S_{sol_4} &= \{\} \\
    n_{sol_5} &= 0 & S_{sol_5} &= \{\}
\end{align*}
\]

\[
\begin{align*}
    n_{sol_4} &= 0, & n_{sol_5} &= 0 \\
    \text{sol}_{4_{\text{rank}}} &= \text{sol}_{5_{\text{rank}}} &= 3 \\
    F_3 &= \{\text{sol}_4, \text{sol}_5\}
\end{align*}
\]
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

**PBM Index: Internal Validity Index**

$$PBM = \left( \frac{1}{K} \sum_{k=1}^{K} \frac{E_k}{E} \right)^2$$

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

$$D_k = \max_{i,j=1, i \neq j} \| c_i - c_j \|$$

**Silhouette Score: Internal Validity Index**

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

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

Maximize
External Cluster Validity Indices

Minkowski Score: Rand Index

\[ RI = \frac{TP + TN}{TP + TN + FP + FN} \]

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

- 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

Image Source: Internet
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 ($\eta_0, \sigma, S, T$)

1: Initialize learning constant $\eta_0$ and neighborhood size $\sigma$; maximum iteration count $T$ for SOM Training; Initialize each map unit by assigning a weight vector randomly chosen from training data $S$.

2: while $t \neq T$ do
3:    Adjust Learning rate: $\eta = \eta_0 \ast (1 - \frac{t}{T})$.
4:    Randomly select a training sample $x \in S$
5:    Find winning map unit: $u^i = \text{arg min}_{1 \leq u \leq D} \| x - w^u \|_2$
6:    Find the neighboring neurons: $U = \{u | 1 \leq u \leq D, \| z^u - z^{u^i} \|_2 < \sigma \}$
7:    Update all neighboring neurons: $w^u = w^u + \eta \ast \exp(-\| z^u - z^{u^i} \|_2) \ast (x - w^u)$
8: return The weight vectors corresponding to map units, $w^u, u = 1, 2 \ldots, D$

Any Random Sample Xi

Update the weight vectors of winning neuron and neighboring neurons

Find winning and neighboring neurons using neighborhood relationship
SOM: A numerical example

- \( n = 4, m = 2 \)
- Training samples
  - \( i_1: (1, 1, 0, 0) \)
  - \( i_2: (0, 0, 0, 1) \)
  - \( i_3: (1, 0, 0, 0) \)
  - \( i_4: (0, 0, 1, 1) \)

Let \( \text{neighborhood} = 0 \)

- Only update weights associated with winning output unit (cluster) at each iteration
- Giving an updated weight matrix:
  \[
  \text{new} - \text{unit 2 - weights} = [0.8 \ 0.4 \ 0.7 \ 0.3] + 0.6([1 \ 1 \ 0 \ 0] - [0.8 \ 0.4 \ 0.7 \ 0.3]) =
  
  = [0.92 \ 0.76 \ 0.28 \ 0.12]
  
  Unit 1: [0.2 \ 0.6 \ 0.5 \ 0.9]
  Unit 2: [0.92 \ 0.76 \ 0.28 \ 0.12]

- Training sample: \( i_1 \)
  - Unit 1 weights
    \[
    d^2 = (2-1)^2 + (5-0)^2 + (9-0)^2 = 1.86
    \]
  - Unit 2 weights
    \[
    d^2 = (8-1)^2 + (4-1)^2 + (2-0)^2 = 0.98
    \]
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:

<table>
<thead>
<tr>
<th>Word</th>
<th>Cosine distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>norway</td>
<td>0.760124</td>
</tr>
<tr>
<td>denmark</td>
<td>0.715460</td>
</tr>
<tr>
<td>finland</td>
<td>0.620022</td>
</tr>
<tr>
<td>switzerland</td>
<td>0.588132</td>
</tr>
<tr>
<td>belgium</td>
<td>0.585835</td>
</tr>
<tr>
<td>netherlands</td>
<td>0.574631</td>
</tr>
<tr>
<td>iceland</td>
<td>0.562368</td>
</tr>
<tr>
<td>estonia</td>
<td>0.547621</td>
</tr>
<tr>
<td>slovenia</td>
<td>0.531408</td>
</tr>
</tbody>
</table>
● 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 an 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 move in a pack and attack a prey in a planned way.

![Leadership hierarchy diagram]

- Scouts, Sentinels, elders, hunters and caretakers in the pack
- Follow alpha and beta
- Help alpha in decision making
- Advisor to alpha
- Discipliner for the pack
- The Best candidate to be alpha in case alpha pass away or become very old
- Responsible for decision making
- The best in managing pack
- May not strongest member in the pack

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**

\[
D = |C. \mathbf{x}_p(t) - \mathbf{x}(t)| \quad \text{and} \quad \mathbf{x}(t + 1) = | \mathbf{x}_p(t) - A. D | 
\]

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:

\[
A = 2a.r_1 - a \quad \text{and} \quad C = 2.r_2 
\]

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**:

\[ D_\alpha = |C_1 \cdot x_\alpha(t) - x(t)| \quad \text{and} \quad D_\beta = |C_2 \cdot x_\beta(t) - x(t)| \quad \text{and} \quad D_\delta = |C_3 \cdot x_\delta(t) - x(t)| \]

\[ x(t+1) = \left( x_1(t) + x_2(t) + x_3(t) \right)/3 \]

where, $x(t+1)$ is the updated position of a wolf at $(t+1)$th iteration with respect to positions of $\alpha$, $\beta$ and $\delta$.

Thus, in this way wolves attack the prey.
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:

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

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

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

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

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

\[ x^{\text{stream}}(t + 1) = x^{\text{stream}}(t) + R \times C \times (x^{\text{sea}}(t) - x^{\text{stream}}(t)) \]
• 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

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

where \( d_{\text{max}} \) 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

$$x_{\text{steam new}} = lb + r_1 \times (up - lb)$$

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, \( \{S_1, S_2, \ldots, S_K\} \) in an automatic way which satisfies the following:
    - \( S_i = \{s_{i1}, s_{i2}, \ldots, s_{inp_i}\} \), \( np_i \): number of sentences in cluster \( i \), \( s_{ij} \): \( j \)th sentence of cluster \( i \).
    - \( \bigcup_{i=1}^{K} S_i = N \) and \( S_i \cap S_j = \emptyset \) for all \( i<>j \).
    - Several cluster validity indices, \( V_{al_1}, V_{al_2}, \ldots, V_{al_M} \) 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 as a reproduction operator: used to generate good quality solutions.

![Flow chart of proposed architecture, ESDS SMODE, where, $g_{max}$ is the user-defined maximum number of generations.](image-url)
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

\[
\text{if } \text{rand}() \leq \text{CR}, \text{ then } y'_{i} = x_i^{\text{current}} + F \times (x_i^{1} - x_i^{2}), \quad \text{Otherwise } y_{i} = x_i^{\text{current}}
\]

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

- Repairing of solution \(y'\)

\[
\text{if } y'_{i} < x_{i}^{L}, \text{ then } y''_{i} = x_{i}^{L} \quad \text{elseif } y'_{i} > x_{i}^{U}, \text{ then } y''_{i} = x_{i}^{U}, \quad \text{Otherwise, } y''_{i} = y'_{i}
\]
After repairing the solution generated by crossover operation, we have applied the concept of polynomial mutation which generates highly diverse solution.

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

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.
  - \[< c_1, c_2, c_3, 0, 0, 0, 0, 0 > \Longrightarrow < c_1, c_2, c_3, c_4, 0, 0, 0, 0 >\]

- **Delete mutation**: decreasing the number of clusters for ith solution by 1.
  - \[< c_1, c_2, c_3, 0, 0, 0, 0, 0 > \Longrightarrow < c_1, c_2, 0, 0, 0, 0, 0, 0 >\]
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.

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

\[
m = \arg \min_{i=1}^{N} \sum_{j=1}^{N} \sum_{i \neq j}^{N} \frac{\text{dist}_{\text{wmd}}(s_i, s_j)}{O}
\]

Where N is the number of sentences in the document, O is the total number of sentence pairs and is given as \((N \times (N - 1)/2\), \(s_i\) 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: \( z_k = \text{distwmd}(c_k, s_m) \), where \( 1 \leq K \leq N \), \( c_k \) 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) \( \uparrow \)

\[
L_{s_k}^i = \left( 1 - \exp\left( \frac{-l(s_i^k) - \mu(l)}{\text{std}(l)} \right) \right) / \left( 1 + \exp\left( \frac{-l(s_i^k) - \mu(l)}{\text{std}(l)} \right) \right)
\]

  - Position of the sentence in the document (F2) \( \uparrow \)

  - Similarity with title (F3) \( \uparrow \)
  - Anti-redundancy (F4) \( \downarrow \)

\[
sim\_title_{s_i^k} = \text{distwmd}(s_i^k, \text{title})
\]

\[
antred_{s_i^k} = \sum_{i,j=1, i \neq j} |c_k| \text{distwmd}(s_i^k, s_j^k)
\]

\[
p_i = \sqrt{\left( \frac{1}{q_i} \right)}
\]
Finally, sentence score is calculated by assigning different weights to various factors (defined above) as:

\[ \text{sentence}\_\text{score}_{si} = \alpha \times F1 + \beta \times F2 + \gamma \times F3 + \delta \times F4 \]

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).
Algorithm 1: ESDS_SMODE

Data: Single Text Document

Result: The fittest solution and corresponding summary generated

1. Initialize population_size (|P|), max_iteration;
2. Initialize population \( P = \{x^1, x^2, \ldots, x^{|P|}\} \) and calculate objective functions values for each solution;
3. Initial SOM Training data \( S=P \);
4. \( P' \leftarrow [\ ] \) //empty population to store new solutions;
5. for \( l=1 \) to max_iteration do
   6. Do SOM Training using its training data \( S \)
   7. for \( i=each \ solution \ in \ P \) do
      8. generate mating pool \( Q \), using neighborhood relationship of trained SOM
      9. generate new solution using \( Q \), crossover and mutation
      10. calculate new solution's objective functions values;
      11. Append new solution into \( P' \);
   end
   12. New_\( P = \) Merge \( P \) and \( P' \);
   13. Apply non-dominated sorting (and crowding distance operator if needed) on \( New_\( P \) to select the best \( |P| \) solutions
   14. Update SOM Training data as \( S = P' \setminus P \);
15. end
16. return the fittest solution;
17. Apply sentence extraction module on the fittest solution;
Method-2: ESDS_MGWO

- $\alpha$ is considered as the fittest solution.
- Archive (fixed length) which contains the non-dominated solutions of the Pareto optimal set.
- $\alpha$, $\beta$ and $\delta$ solutions are selected from archive using Roulette Wheel selection (RWL) such that they are not same.
- If $|\text{Archive\_size}| > \text{fixed\_length}$, 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 $\alpha/\beta/\delta$ wolf based on it’s fitness functional values.
- Finally, we have to report the generated summary corresponding to the fittest wolf (solution), i.e. $\alpha$. 
**Pseudo Code: ESDS_MGWO**

**Algorithm 2: ESDS_MGWO**

- **Data:** Single Text Document
- **Result:** The fittest solution, $\alpha$ and corresponding summary generated

1. Initialize grey wolf population $P = \{x^1, x^2, ..., x^|P|\}$ and calculate objective functional values for each wolf
2. Initialize $\bar{A}$, $\bar{C}$ and $\bar{a}$ and number of max iteration.
3. Apply non-dominated sorting on $P$ and initialize Archive with Pareto optimal solutions.
4. Select leaders $\alpha, \beta, \delta$ from archive such that $\alpha \neq \beta \neq \delta$.
5. For $l = 1$ to max iteration do
   - For each wolf in population do
     - Update wolf positions which will be considered as new wolf;
   - End
6. Update $\bar{A}$, $\bar{C}$ and $\bar{a}$;
7. Calculate objective functional values for each new wolf (wolf after updating positions);
8. Apply non-dominated sorting on new wolves and update Archive with Rank-1 solutions;
9. Apply non-dominated sorting on Archive and obtain the set of Rank-1 non-dominated solutions to form updated Archive.
10. If $Archive.size > threshold\_Archive\_size$ then
    - Remove just enough wolves using Roulette wheel selection mechanism;
11. End
12. If any wolf resides outside of hypercube then
    - Update grid to cover new solutions;
13. End
14. Select leaders $\alpha, \beta, \delta$ from archive such that $\alpha \neq \beta \neq \delta$;
15. Return $\alpha$ wolf;
16. Apply sentence extraction module on $\alpha$ 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.
Algorithm 3: ESDS_MWCA

Data: Single Text Document
Result: The fittest solution sea and corresponding summary generated
1 Initialize population_size (|P|), $N_r$, $d_{max}$ and max_iterations;
2 Initialize population $P = \langle \bar{x}^1, \bar{x}^2, \ldots, \bar{x}^{|P|} \rangle$ and calculate objective functional values for each solution as discussed in Sections
3 $[\text{front}_1, \text{front}_2, \ldots, \text{front}_k] =$ Apply non-dominated sorting on $P$;
4 for $\text{front}=1$ to $k$
5 | sort front in descending order of crowding distance;
6 end
7 Appoint sea, rivers;
8 Designate streams to rivers and sea
9 for $l=1$ to max_iteration do
10 | for $i=1$ to population_size do
11 | | move stream to river, river to sea and stream to sea
12 | end
13 for river in population do
14 | if $\text{dist}_{sea}(river, sea) < d_{max}$ or rand < 0.1 then
15 | | new streams are generated and objective function values are calculated;
16 | end
17 end
18 for stream in population do
19 | if $\text{dist}_{sea}(stream, sea) < d_{max}$ then
20 | | new streams are generated and objective functions values are calculated;
21 | end
22 end
23 $[\text{front}_1, \text{front}_2, \ldots, \text{front}_k] =$ Apply non-dominated sorting on new population;
24 for $\text{front}=1$ to $k$
25 | sort front in descending order of crowding distance;
26 end
27 Update sea, rivers;
28 end
29 return sea;
30 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.

<table>
<thead>
<tr>
<th></th>
<th>DUC2001</th>
<th>DUC2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Topics</td>
<td>30</td>
<td>59</td>
</tr>
<tr>
<td>#Documents</td>
<td>309</td>
<td>567</td>
</tr>
<tr>
<td>Source</td>
<td>TREC</td>
<td>TREC</td>
</tr>
<tr>
<td>length of summary (in words)</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Evaluation Measure

\[
ROUGE - N = \frac{\sum_{S \in \text{Summary}_{ref}} \sum_{N\text{-gram} \in S} \text{Count}_{\text{match}}(N\text{-gram})}{\sum_{S \in \text{Summary}_{ref}} \sum_{N\text{-gram} \in S} \text{Count}(N\text{-gram})}
\]

Where \( N \) represents the length of n-gram, \( \text{Count}_{\text{match}}(N\text{-gram}) \) is the maximum number of overlapping \( N \)−grams between reference summary and system summary, \( \text{Count}(N\text{-gram}) \) 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

<table>
<thead>
<tr>
<th>Method</th>
<th>DUC2001</th>
<th>DUC2002</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average ROUGE-2</td>
<td>Average ROUGE-1</td>
</tr>
<tr>
<td>ESDS_SMODE</td>
<td>0.21450</td>
<td>0.45214</td>
</tr>
<tr>
<td>ESDS_MGWO</td>
<td>0.15228</td>
<td>0.37108</td>
</tr>
<tr>
<td>ESDS_MWCA</td>
<td>0.14997</td>
<td>0.36702</td>
</tr>
<tr>
<td>MA-SingleDocSum</td>
<td>0.20142</td>
<td>0.44862</td>
</tr>
<tr>
<td>DE</td>
<td>0.18523</td>
<td>0.47856</td>
</tr>
<tr>
<td>UnifiedRank</td>
<td>0.17646</td>
<td>0.45377</td>
</tr>
<tr>
<td>FEOM</td>
<td>0.18549</td>
<td>0.47728</td>
</tr>
<tr>
<td>NetSum</td>
<td>0.17697</td>
<td>0.46427</td>
</tr>
<tr>
<td>CRF</td>
<td>0.17327</td>
<td>0.45512</td>
</tr>
<tr>
<td>QSC</td>
<td>0.18523</td>
<td>0.44852</td>
</tr>
<tr>
<td>SVM</td>
<td>0.17018</td>
<td>0.44628</td>
</tr>
<tr>
<td>Manifold Ranking</td>
<td>0.16635</td>
<td>0.43359</td>
</tr>
</tbody>
</table>

Fig.: ROUGE Scores of different methods on DUC2001 and DUC2002 data sets
An example of Good Summary: ESDS_SMODE

<table>
<thead>
<tr>
<th>Line No.</th>
<th>Actual Summary</th>
<th>Predicted Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>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.</td>
<td>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.</td>
</tr>
<tr>
<td>2</td>
<td>The Scandinavian Star was about 35 miles off Isla Mujeres, a Mexican island north of the tourist resort Cancun in the Yucatan Peninsula.</td>
<td>---------No Matching Line---------</td>
</tr>
<tr>
<td>3</td>
<td>Petty Officer Brian Lincoln of the U.S. Coast Guard’s Miami office said no injuries were reported.</td>
<td>No injuries were reported, said Petty Officer Brian Lincoln, of the U.S. Coast Guard’s Miami office.</td>
</tr>
<tr>
<td>4</td>
<td>&quot;All we know is that the ship’s captain&quot; is completely out of fire-fighting equipment and has requested additional equipment such as foam, water and CO2, &quot; said Lincoln.</td>
<td>&quot;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, &quot; said Lincoln, referring to carbon dioxide.</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>ESDS_SMODE</th>
<th>ESDS_MGWO</th>
<th>ESDS_MWCA</th>
<th>MA-SingleDocSum</th>
<th>DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pop. Size</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>30</td>
<td>200</td>
</tr>
<tr>
<td>NFEs</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>1600</td>
<td>1000</td>
</tr>
</tbody>
</table>

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:

\[ GD = \left( \frac{\sum_{i=1}^{|Q|} \text{dist}_{wmd}^p(s^i)}{|Q|} \right)^{\frac{1}{p}} \]

where \( \text{dist}_{wmd}(s^i, s^k) \) 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)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>DUC2001 GD</th>
<th>CPU time (In Sec.)</th>
<th>DUC2002 GD</th>
<th>CPU time (In Sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESDS_SMODE</td>
<td>0.47461</td>
<td>78.9915</td>
<td>0.44624</td>
<td>50.3384</td>
</tr>
<tr>
<td>ESDS_MGWO</td>
<td>0.48860</td>
<td>64.5141</td>
<td>0.45299</td>
<td>43.5634</td>
</tr>
<tr>
<td>ESDS_MWCA</td>
<td>0.50086</td>
<td>20.6499</td>
<td>0.45903</td>
<td>11.9936</td>
</tr>
</tbody>
</table>
Pareto Fronts: ESDS_SMODE
Pareto Fronts: ESDS_MGWO
Pareto Fronts: ESDS_MWCA
Diversity Plots

![Diversity Plots](image)
Improvement Obtained (1/2)

Improvement obtained (IO) is calculated as:

\[ IO = \frac{\text{ProposedMethod} - \text{OtherMethod}}{\text{OtherMethod}} \times 100 \]

Fig.: Improvements obtained by our proposed approach over other methods based on ROUGE−2 score
Table: Improvements obtained by our proposed approach over other methods using ROUGE−1 score on DUC2002 dataset

<table>
<thead>
<tr>
<th>Methods</th>
<th>DUC2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESDS_MGWO</td>
<td>17.37</td>
</tr>
<tr>
<td>ESDS_MWCA</td>
<td>17.50</td>
</tr>
<tr>
<td>MA-SingleDocSum</td>
<td>1.73</td>
</tr>
<tr>
<td>DE</td>
<td>5.19</td>
</tr>
<tr>
<td>UnifiedRank</td>
<td>1.30</td>
</tr>
<tr>
<td>FEOM</td>
<td>5.46</td>
</tr>
<tr>
<td>NetSum</td>
<td>9.24</td>
</tr>
<tr>
<td>CRF</td>
<td>11.61</td>
</tr>
<tr>
<td>QSC</td>
<td>9.48</td>
</tr>
<tr>
<td>SVM</td>
<td>13.60</td>
</tr>
<tr>
<td>Manifold Ranking</td>
<td>16.05</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Methods</th>
<th>DUC2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed approach</td>
<td>5.84</td>
</tr>
<tr>
<td>ESDS_MGWO</td>
<td>28.96</td>
</tr>
<tr>
<td>ESDS_MWCA</td>
<td>30.29</td>
</tr>
<tr>
<td>MA-SingleDocSum</td>
<td>6.67</td>
</tr>
<tr>
<td>UnifiedRank</td>
<td>5.46</td>
</tr>
<tr>
<td>FEOM</td>
<td>0.27</td>
</tr>
<tr>
<td>NetSum</td>
<td>3.08</td>
</tr>
<tr>
<td>CRF</td>
<td>5.15</td>
</tr>
<tr>
<td>QSC</td>
<td>6.70</td>
</tr>
<tr>
<td>SVM</td>
<td>7.23</td>
</tr>
<tr>
<td>Manifold Ranking</td>
<td>10.37</td>
</tr>
</tbody>
</table>
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.

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

- Readability Factor (↑):

\[
R = \sum_{i=2}^{N_p} \text{sim}(s_i, s_{i-1})
\]
Objective Functions Used (2/2)

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

\[
COH = \frac{\log(C_s \times 9 + 1)}{\log(M \times 9 + 1)}
\]

Where,

\[
C_s = \frac{\sum_{s_i, s_j \in \text{Summary}} \text{sim}(s_i, s_j)}{O_s}
\]

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 \(< \alpha, \beta, \gamma, \delta, \lambda, \varphi >\), then \(< ob1 \times \alpha, ob2 \times \beta, ob3 \times \gamma, ob4 \times \delta, ob5 \times \lambda, ob6 \times \varphi >\) 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:

\[
\sum_{s_i \in \text{Summary}} l_i \leq S_{\text{max}},
\]

where, \(l_i\) measures the length of sentence in terms of number of words, \(S_{\text{max}}\) is the maximum number of words allowed in generated summary.
Genetic Operators: Mutation and Crossover

- **Mutation:**

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

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

- **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 ($\sigma_0$)=2, initial learning rate ($\sigma_0$)=0.6, training iteration in SOM=$|P|$, topology=rectangular 2D grid; grid size=$5 \times 8$.
- 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

<table>
<thead>
<tr>
<th>Method</th>
<th>DUC’2001</th>
<th>DUC’2002</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ROUGE-2</td>
<td>ROUGE-1</td>
</tr>
<tr>
<td>Approach 1 (NGD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With SOM</td>
<td>0.26949</td>
<td>0.47699</td>
</tr>
<tr>
<td>Without SOM</td>
<td>0.26742</td>
<td>0.47521</td>
</tr>
<tr>
<td>Approach 2 (NGD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With SOM</td>
<td>0.26774</td>
<td>0.47291</td>
</tr>
<tr>
<td>Without SOM</td>
<td>0.26265</td>
<td>0.46762</td>
</tr>
<tr>
<td>Approach 1 (CS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With SOM</td>
<td>0.26459</td>
<td>0.47554</td>
</tr>
<tr>
<td>Without SOM</td>
<td>0.25282</td>
<td>0.46289</td>
</tr>
<tr>
<td>Approach 2 (CS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With SOM</td>
<td>0.26209</td>
<td>0.47398</td>
</tr>
<tr>
<td>Without SOM</td>
<td>0.26629</td>
<td>0.47862</td>
</tr>
<tr>
<td>Approach 1 (WMD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With SOM</td>
<td>0.29238*</td>
<td>0.50236*</td>
</tr>
<tr>
<td>Without SOM</td>
<td>0.28930</td>
<td>0.49486</td>
</tr>
<tr>
<td>Approach 2 (WMD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With SOM</td>
<td>0.28462</td>
<td>0.49863</td>
</tr>
<tr>
<td>Without SOM</td>
<td>0.28190</td>
<td>0.48877</td>
</tr>
</tbody>
</table>

**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 crew members. [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

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.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Improvements obtained by Proposed approach (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ROUGE-2</td>
</tr>
<tr>
<td>MA-SingleDocSum</td>
<td>45.16</td>
</tr>
<tr>
<td>DE</td>
<td>75.76</td>
</tr>
<tr>
<td>UnifiedRank</td>
<td>65.69</td>
</tr>
<tr>
<td>FEOM</td>
<td>57.85</td>
</tr>
<tr>
<td>NetSum</td>
<td>65.21</td>
</tr>
<tr>
<td>CRF</td>
<td>68.74</td>
</tr>
<tr>
<td>QSC</td>
<td>57.85</td>
</tr>
<tr>
<td>SVM</td>
<td>57.63</td>
</tr>
<tr>
<td>Manifold Ranking</td>
<td>71.81</td>
</tr>
<tr>
<td>NN-SE</td>
<td>xx</td>
</tr>
<tr>
<td>Summarunner</td>
<td>xx</td>
</tr>
</tbody>
</table>
Conclusion

- A self-organized multi-objective binary differential evolution technique is proposed for summary extraction.
- Three similarity/dissimilarity criteria are used to measure 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
- Apply this approach for query based single document
Publications


- **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)

Publications


- **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)*


Publications


References


Anu Queries???
Thank You!!