

# Optimizing Data-Driven Models for Summarization as Parallel Tasks

Aleš Zamuda

*Faculty of Electrical Engineering and Computer Science, University of Maribor  
Koroška cesta 46, 2000 Maribor, Slovenia*

Elena Lloret

*Department of Software and Computing Systems, University of Alicante  
Apdo. de correos 99, E-03080 Alicante, Spain*

---

## Abstract

This paper presents tackling of a hard optimization problem of computational linguistics, specifically automatic multi-document text summarization, using grid computing. The main challenge of multi-document summarization is to extract the most relevant and unique information effectively and efficiently from a set of topic-related documents, constrained to a specified length. In the Big Data/Text era, where the information increases exponentially, optimization becomes essential in selection of the most representative sentences for generating the best summaries. Therefore, a data-driven summarization model is proposed and optimized during a run of Differential Evolution (DE).

Different DE runs are distributed to a grid in parallel as optimization tasks, seeking high processing throughput despite the demanding complexity of the linguistic model, especially on longer multi-documents where DE improves results given more iterations. Namely, parallelization and the grid enable, running several independent DE runs at same time within fixed real-time budget. Such approach results in improving a Document Understanding Conference (DUC) benchmark recall metric over a previous setting.

*Keywords:* Text Summarization, Discrete Optimization, Distributed Computing, Data-Driven Model, Differential Evolution

---

*Email addresses:* ales.zamuda@um.si (Aleš Zamuda), elloret@dlsi.ua.es (Elena Lloret)

---

## 1. Introduction

In the era in which the information increases at an exponential rate, and it is impossible to digest it in an efficient and effective manner, the task of automatic text summarization can be of great help, both for humans and other computer-based tasks. On the one hand, providing summaries that condense all the relevant information for a specific topic instead of all the source documents to users, avoids them the costly task of having to read and determine which information should be kept [1]. On the other hand, using summaries can benefit other processes when it comes to managing large amounts of information at intermediate stages, thus decreasing the execution time, and consequently, providing the output faster, since, in this case, the text summarization process would act as a filtering step, keeping only the relevant information to process, and considering the remaining information as noisy information [2].

Formally, the task of text summarization can be defined as “the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user (or users) and task (or tasks)” [3]. This task poses great challenges to the research community, due partly to the inherent subjectivity associated with the process of determining “the most important” information. When and how can information be considered as relevant? This will very much depend on the person who will read the summary, or even the final goal of the summary itself, since the same piece of information could be considered relevant or irrelevant, depending on many factors. Different taxonomies have been proposed to classify the summaries according to different factors, such as its input, output, purpose, or language [4]. For instance, multi-document summarization specifically takes as input a number of documents — that could, potentially, be related to the same topic, or not — and has to produce an effective summary based on them. This type of input increases the complexity of the task, because it has to deal with redundancy, as well as the type of input documents — formal and informal text documents could also be input.

The usefulness of text summarization, either for users or for applications, also depends on the extent to which the summary can be obtained in real time, regardless of the number of input documents it has to cope with. The real time limitation to be considered may vary from one task specification

to another, and might even be different when planning the operation's scenario [5, 6]. Also, for real case scenarios, the optimum values might not be known beforehand, and running an optimization algorithm for a longer time setting would be useful [7, 5].

The aim of this paper is to elaborate on the possibility that the text summarization model can be run on a parallel computer grid, and the feedback obtained is applied for design of the text summarization approach. We motivate the fact that Natural Language Processing (NLP) tasks do not normally take care of the efficiency in terms of time. They just focus on solving the problem with high accuracy, which is good, but not enough to transfer the High-Performance Computing (HPC) technology to it by improving the NLP approaches with HPC support. In this context, we emphasize the need to investigate methods/approaches that take into account accuracy, as well as efficiency. To investigate this issue would be the main contribution of our research work, since, to the best of our knowledge, existing publications only focus on either accuracy or efficiency at a time, thus, not analyzing how both aspects could be integrated appropriately into a summarization task to produce effective parallelized automatic summaries. But just to note, we have limited our approach to extractive summarization, i.e. our approach cannot be applied for abstractive summarization, and it is also not applicable for summarization with word length less than the shortest sentence, because it selects whole sentences to generate a summary.

Obtaining an appropriate balance between accuracy and efficiency would be possible by including parallel computing and using an HPC grid to run these summarization tasks. We investigate that, during working on a design of a summarization algorithm, one first conducts a preliminary research for showing that this is feasible. The knowledge obtained through the proposed research would allow the design and development of summarization approaches that can be both effective and accurate, so they can be used further in real contexts, being directly transferable to society. The main novelty in the newly proposed approach in this paper is, therefore, the parallelization of tasks for optimization that are applied in the perspective of benchmarking, i.e. the individual benchmark tests are parallelized in such a manner that summarization parallel tasks are submitted and distributed to a computer grid individually and merged after tasks are complete. Also, the information identification and quantitative evaluation is obtained using Freeling [8], coreference resolution, semantic analysis, and concept matrix that take place in the pre-processing stage for summarization (in the corpus

preprocessing phase for task construction, before the parallel task is submitted — much before the loop for summarization optimization commences). The computational perspective defining static (precomputed) and changing (computed during evaluation call) parts of fitness function are also newly discussed, as this perspective is important in high-performance computing scenarios; namely, the parts of fitness function that do not need to be recomputed, are precomputed. Regarding the optimization part novelty, a new algorithm combination is presented, a constrained version of a self-adaptive Differential Evolution (DE) with binarization. Additionally, an important computationally-intensive correlation analysis is, finally, presented empirically, providing insight into the correlation among ROUGE metric values on the dedicated benchmark (DUC 2002) along with the number of fitness functions, testing the trend and payoff of prolonged optimization execution runs (quality improvement after extended optimization run time).

In the following, Section 2 provides more related work, covering optimization and Differential Evolution, then text summarization and task parallelization on HPC. Section 3 presents our proposed contributions and defines novelty of the paper, defining the multi-document extractive summarization approach via DE that is run in parallel on HPC. Section 4 reports and discusses the results of the proposed approach. Section 5 provides conclusions and outlines future work.

## 2. Related Work

This section presents related work, as covered in the following subsections, in which, first, text summarization is explained, then optimization is addressed, followed by covering task parallelization and HPC.

### 2.1. Text Summarization

The task of text summarization has been researched for more than 50 years. Despite this, it is still a very challenging task in NLP [9]. As was outlined in the previous Section, its difficulty is due partly to the fact that there is a lot of subjectivity in the process that is influenced by many cognitive aspects [10, 11]. Whereas the main objective of text summarization is to produce a summary automatically, i.e., with no human intervention, such a summary could be output in many different forms, also being influenced by a wide range of factors that should be considered during the process of summary generation. In this sense, there are different classical taxonomies proposed in the literature that lead to different types of summaries [12, 13].

A special type of summarization is sentence extraction from multiple documents, i.e. multi-document extractive summarization. Different approaches can be found in the literature for addressing multi-document summarization, which range from simple approaches using statistical techniques, such as tf-idf [14], to more recent ones that use neural computing [15, 16, 17]. However, efficiency is not normally taken into account, and although summaries can obtain good results in terms of their content, the associated drawback concerns the impossibility to apply those approaches in realtime scenarios, especially those approaches based on Neural Networks (NNs) that require a lot of training time. One approach in-between is to take into account optimization issues and integrate them within the summarization approach. For instance, Alguliev et al. [18] applied Differential Evolution to multi-document summarization using sentence extraction, being formalized as a discrete optimization problem. Further on, Alguliev et al. also extended their work, modeling the multi-document summarization tasks as different algorithmic problems, such as a quadratic boolean programming problem; a non-linear programming problem; or as a modified p-median problem [19, 20, 21, 22, 23].

The summarization approach presented in this paper is based on the multi-document sentence extraction summarization approach of Alguliev et al. [18]. Therefore, in the next subsection, the basic text summarization approach [18] is covered over three subsections, followed by a subsection on the DUC 2002 multi-document summarization set, and an explanation of the assessment through ROUGE peer evaluation.

### *2.1.1. Multi-Document Summary: Text Indexing*

This subsection presents how a summarization task is defined mathematically, based on the approach from [18]. Addressing it from an optimization perspective, summarization is optimization of a summarization text model (fitness function), constrained to word length. From the perspective of optimization problems, this model does not have a more general problem description, as it is a special case of a very complex feature selection with constraints, like, e.g. feature preselection for stability selection in Data Mining of Big Data from healthcare inpatients records [24].

To define the features that can be selected during Feature Selection (FS) for text summarization, the approach in [18] extracts a definition of the text summarization model from a given dataset. This dataset is a collection of documents, which are converted to a text representation and concatenated, composed of e.g. UTF-8 encoded character bytes. The input text is split into

sentences, so that if the sentences are extracted in the resulting summary, the summary can be printed by recomposing and concatenating the text written in these sentences. The sentences in the input text are indexed, as well as the words that these sentences contain, as explained in the following.

Within one text  $D$  of a multi-document set from the whole benchmark set, the sentences in all contained documents are indexed. A statement consists of words (terms), words consist of letters. Within the text, beginnings of statements are marked as text character indices  $D = \{s_1, s_2, s_3, \dots, n\}$ , and a dictionary of different words (terms) and their appearances in the text is composed of  $W = \{w_1, w_2, w_3, \dots\}$ . A symbol table (e.g. implemented using hash table) is then used to compare terms and build the dictionary.

### 2.1.2. Multi-Document Summary: Term Weights' Computation

Term weights' computation contains three steps. In the first step, for each  $i$ -th term ( $w_i$ ), during indexing, the number of occurrences in the text is gathered, and the number of occurrences ( $n_k$ ) of a term in some  $k$ -th statement. Then, in the second step, for each term  $w_i$  in the document, an inverse frequency of each term is calculated:

$$isf_{w_i} = \log\left(\frac{n}{n_k}\right), \quad (1)$$

where  $n$  denotes the number of statements in the document, and  $n_k$  the number of statements including a term  $w_i$ . In the third step, for each term in the document, a weight is calculated:

$$w_{i,k} = tf_{i,k} isf_k, \quad (2)$$

where  $tf_{i,k}$  is the number of occurrences (term frequency) of a term  $w_k$  in a statement  $s_i$ .

### 2.1.3. Multi-Document Summary: Summary Model Fitness Evaluation

Based on the weights, the similarity between two selected statements  $s_i = [w_{i,1}, w_{i,2}, \dots, w_{i,m}]$  and  $s_j = [w_{j,1}, w_{j,2}, \dots, w_{j,m}]$  is computed as in [18]:

$$sim(s_i, s_j) = \sum_{k=1}^m \frac{w_{i,k} w_{j,k}}{\sqrt{\sum_{k=1}^m w_{i,k} w_{i,k} \sum_{k=1}^m w_{j,k} w_{j,k}}}, \quad (3)$$

where  $w_{i,k}$  is a term weight defined in the previous subsection, and  $m$  is the number of all terms in the summarized text.

A sentence set  $\mathbf{x}$  is evaluated as a 0/1 knapsack-like problem, the content of which is to be maximized. To include a selection of statements in the extractive summary, an  $i$ -th sentence  $s_i$  is selected ( $x_i = 1$ ), or unselected ( $x_i = 0$ ). The selected statements from the sentence combination  $\mathbf{x}$  are printed in order as they appear in the text, to recompose an extractive summary text. To define a sentence selection  $\mathbf{x}$ , needed to provide the selected text summaries to be extracted from a multi-document set  $\mathbf{D}$  with  $N$  sentences, a set of sentences  $\mathbf{x}$  needs to be selected:

$$\mathbf{x} = \{x_j\}, \forall j \in \{1, 2, \dots, 1, N\}, \forall x_j \in \{0, 1\}. \quad (4)$$

For a sentence selection  $\mathbf{x}$ , content coverage  $V(\mathbf{x})$  is computed as a cross-sentence multiplied double sum of similarities [18]:

$$V(\mathbf{x}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N (sim(s_i, O) + sim(s_j, O))x_{i,j}, \quad (5)$$

where  $x_{i,j}$  denotes the inclusion of both statements,  $s_i$ , and  $s_j$ , and  $x_{i,j}$  is only 1 if  $x_i = x_j = 1$ , otherwise 0. The  $O = (o_1, o_2, \dots, o_M)$  represents an average term vector, which is an auxiliary weight used for all different indexed  $M$  terms  $i = \{1, 2, \dots, M\}$  based on the corresponding term weights  $w_{i,k}$ :

$$o_i = \frac{\sum_{j=1}^N w_{i,j}}{N}. \quad (6)$$

For a sentence combination  $\mathbf{x}$ , redundance  $R(\mathbf{x})$  is then calculated as a similarity sum among all statements, similar to content coverage  $V(\mathbf{x})$  by reusing the  $sim(s_i, s_j)$  calculations [18]:

$$R(\mathbf{x}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N sim(s_i, s_j)x_{i,j}, \quad (7)$$

where  $x_{i,j}$  denotes the inclusion of both statements,  $s_i$ , and  $s_j$ , and again,  $x_{i,j}$  is only 1 if  $x_i = x_j = 1$ , otherwise 0.

Finally, the fitness of a generated text model represents the ratio between content coverage ( $V(\mathbf{x})$ ) and redundancy ( $R(\mathbf{x})$ ), defined using negation as a minimization function [18]:

$$f(\mathbf{x}) = -\frac{V(\mathbf{x})}{R(\mathbf{x})}. \quad (8)$$

As the constraint of the summary length is  $L \pm \epsilon_w$  words, therefore, a constraint handling mechanism also needs to be used to address this during optimization.

The meaning and evaluation of a recomposed summary, hence, does not have the same features as a 0/1 knapsack problem with a fixed knapsack capacity and a linear relationship between the weights and profit values of unsorted items, as addressed for standard strongly correlated sets of unsorted data and efficiently solved approximately using NNs, like recently in [25]. These features are more complex and the weights (of sentences) are dynamic, according to which sentences are chosen (coverage and redundancy change without preservation of linear relationships among selected sentences due to distance non-preserving multiplications with term weights, see Eqs. (5) and (7)). Also, a summarization fitness function as modeled here does not merely change the dynamic capacity of a knapsack [26], but changes the values of knapsack items added up together at higher powers than linearly. The problem selected as the case study in this paper is not only an important challenge within NLP due to the real and big data processing, but also, moreover, technically, as a large scale, non-linear, constrained, and non-separable problem in the benchmarking domain [27].

#### *2.1.4. Multi-document Summarization Sets and Peer Evaluation*

The summarization task addressed in this research focus on the scenario of news digest. Everyday, a high number of news are published, and it is impossible to read them all and keep up-to-date. The use of summaries to help digest them fits perfectly in this context.

The existing collection of English newswire documents from the Document Understanding Conferences<sup>1</sup> (DUC) fits very well for experimenting within the aforementioned scenario. On the one hand, it provides pairs of documents-summary or cluster-summary for diverse summarization types (extractive, single-document, multi-document, etc.), and, on the other hand, there is a wide number of previous summarization systems to be compared with, and determine to what extent our approach is effective. In particular, we used the DUC 2002 dataset, which contains 567 documents grouped in 59 clusters (denoted as d061 to d120), where each cluster represents a set of topic-related documents (the average number of documents per cluster is

---

<sup>1</sup><https://duc.nist.gov/>



10). Besides this dataset, a recent new dataset, i.e., CNN/DailyMail<sup>2</sup> has also been made available for the research community [28]. This dataset contains more than 300,000 documents, and it provides summaries of about 50 words. However, this dataset can only be used for single-document summarization, which is not the type of summaries we are addressing in this research.

Concerning the evaluation of summaries, ROUGE [29] is one of the common standard, and most used tools. The idea behind ROUGE is that, if two texts have a similar meaning, they must also share similar words or phrases. As a consequence, it relies on n-gram co-occurrence, and the idea behind it is to compare the content of a peer summary with one or more model summaries, and compute the number of n-gram of words they all have in common. Different types of n-grams can be obtained, such as unigrams (ROUGE-1), bigrams (ROUGE-2), the longest common subsequence (ROUGE-L), or bigrams with a maximum distance of four words in-between (ROUGE-SU4), and, based on them, values for recall, precision and F-measure can be obtained, thus determining the summary accuracy in terms of content (the higher recall, precision and F-measure values, better). Among all the metrics, recall values are then usually reported for peer evaluation of generated model summaries. Other metrics also exist, like AutoSummENG [30] which is an automatic character n-gram based evaluation method with high correlation with human judgments, or SumTriver [31], which does not need to have human summaries for the evaluation; however, they are not often used by the research community, thus it is difficult to find results for comparison purposes. Other ways to evaluate the summaries would be to consider standard similarity metrics, such as Simmetrics<sup>3</sup>. A comprehensive literature review on summarization evaluation methods and metrics can be found in [32].

## 2.2. Optimization and Differential Evolution

Differential Evolution (DE) [33] is a floating-point encoding Evolutionary Algorithm [34, 35] for global optimization over continuous spaces. The DE algorithm was introduced in 1995 by Storn and Price [33] and, since then, has formed the basis for a set of successful algorithms for optimization domains, such as continuous, discrete, mixed-integer, or other search spaces and features [36]. The whole encompassing research field around

---

<sup>2</sup><https://github.com/abisee/cnn-dailymail>

<sup>3</sup><https://github.com/Simmetrics/simmetrics>

DE was surveyed most recently in [37], and even since then, several other domain- and feature-specific surveys, studies, and comparisons have also followed [38, 39, 40, 41]. Theoretical insight and insights to inner workings and behaviors of DE during consecutive generations has been studied in works like [42, 43, 44, 45, 46, 47, 48].

The main performance advantages of DE over other Evolutionary Algorithms [49, 50, 51, 52, 53, 54] lie in floating-point encoding, and a good combination of evolutionary operators, mutation step size adaptation, and elitist selection. The DE algorithm has a main evolution loop in which a population of vectors is computed for each generation of the evolution loop. During one generation  $g$ , for each vector  $\mathbf{x}_i$ ,  $\forall i \in \{1, 2, \dots, NP\}$  in the current population, DE employs evolutionary operators, namely mutation, crossover, and selection, to produce a trial vector (offspring) and to select one of the vectors with the best fitness value.  $NP$  denotes population size and  $g \in \{1, 2, \dots, G\}$ , the current generation number.

Mutation creates a mutant vector  $\mathbf{v}_{i,g+1}$  for each corresponding population vector. Among many proposed, one of the most popular DE mutation strategies [55, 33] are the ‘*rand/1*’:

$$\mathbf{v}_{i,g+1} = \mathbf{x}_{r_1,g} + F(\mathbf{x}_{r_2,g} - \mathbf{x}_{r_3,g}) \quad (9)$$

and the ‘*best/1*’:

$$\mathbf{v}_{i,g+1} = \mathbf{x}_{best,g} + F(\mathbf{x}_{r_1,g} - \mathbf{x}_{r_2,g}), \quad (10)$$

where the indexes  $r_1$ ,  $r_2$ , and  $r_3$  represent the random and mutually different integers generated within the range  $\{1, 2, \dots, NP\}$  and are also different from index  $i$ . The  $\mathbf{x}_{best,g}$  denotes the currently best vector.  $F$  is an amplification factor of the difference vector within the range  $[0, 2]$ , but usually less than 1. The vector at index  $r_1$  is a base vector. The term  $\mathbf{x}_{r_2,g} - \mathbf{x}_{r_3,g}$  denotes a difference vector which after multiplication with  $F$ , is named an amplified difference vector. The simple DE mutation ‘*rand/1*’ is by far most the widely used [56], however, a form of ‘*best/1*’ mutation has also been signified beneficial, especially in more restrictive evaluation scenarios [57, 58, 45, 59, 6].

After mutation, the mutant vector  $\mathbf{v}_{i,g+1}$  is taken into a recombination process with the target vector  $\mathbf{x}_{i,g}$  to create a trial vector  $\mathbf{u}_{i,g+1} = \{u_{i,1,g+1}, u_{i,2,g+1}, \dots, u_{i,D,g+1}\}$ . The binary crossover operates as follows:

$$u_{i,j,g+1} = \begin{cases} v_{i,j,g+1} & \text{if } rand(0, 1) \leq CR \text{ or } j = j_{rand} \\ x_{i,j,g} & \text{otherwise} \end{cases}, \quad (11)$$

where  $j \in \{1, 2, \dots, D\}$  denotes the  $j$ -th search parameter of  $D$ -dimensional search space,  $rand(0, 1) \in [0, 1]$  denotes a uniformly distributed random number, and  $j_{rand}$  denotes a uniform randomly chosen index of the search parameter, which is always exchanged to prevent cloning of target vectors.  $CR$  denotes the crossover rate [60]. Finally, the selection operator propagates the fittest individual in the new generation (for a minimization problem):

$$\mathbf{x}_{i,g+1} = \begin{cases} \mathbf{u}_{i,g+1} & \text{if } f(\mathbf{u}_{i,g+1}) < f(\mathbf{x}_{i,g}) \\ \mathbf{x}_{i,g} & \text{otherwise} \end{cases}. \quad (12)$$

### 2.2.1. Self-adaptation of Control Parameters

As studied in [49, 45], the original DE algorithm [33] keeps all three control parameters fixed during the optimization process. As observed from the initial experiments in [49], the necessity was confirmed for changing control parameters during the optimization process. In the jDE algorithm [49], which extends the original DE algorithm, a self-adaptive control mechanism is introduced to change the control parameters  $F$  and  $CR$  during the evolutionary process after their initial setting at  $F = 0.5$  and  $CR = 0.9$ . New control parameters  $F_{i,G+1}$  and  $CR_{i,G+1}$  are calculated for jDE as in [49]:

$$F_{i,G+1} = \begin{cases} F_1 + rand_1 \times F_u & \text{if } rand_2 < \tau_F \\ F_{i,G} & \text{otherwise} \end{cases} \quad (13)$$

$$CR_{i,G+1} = \begin{cases} rand_3 & \text{if } rand_4 < \tau_{CR} \\ CR_{i,G} & \text{otherwise.} \end{cases} \quad (14)$$

This produces control parameters  $F$  and  $CR$  within a new vector. The  $rand_j \in [0, 1]$ ,  $j \in \{1, 2, 3, 4\}$  are uniform random values. The  $\tau_F$  and  $\tau_{CR}$  (generally denoted in literature as  $\tau_1$  and  $\tau_2$ , respectively) represent the probabilities of adjusting control parameters  $F$  and  $CR$ , respectively. The  $\tau_F$ ,  $\tau_{CR}$ ,  $F_1$ ,  $F_u$  are taken as fixed values, as proposed in [49], but in [45], a thorough study was conducted on a full benchmark using HPC, showing that adjusting these values can have significant improvement effects.

### 2.2.2. Constraints' Definition and Handling

As explained and compared experimentally in [59], several methods have been proposed for defining constraints, and addressing them in Genetic Algorithms for parameter optimization problems [61]. A summary of constraint

handling techniques can be found in [62, 63, 64], which also contain information on many stochastic techniques. An application of the constraint definition is to use it to guide the search towards feasible areas of the search space. Given constraints, there are different methods for consideration when defining feasibility and possible violation estimation of a solution, varying with the ease of defining such methods on the one hand, and guiding feedback amount suitability for constraint violation handling on the other [59].

A solution  $\mathbf{x}$  is regarded as *feasible* with regard to inequality constraints  $g_i(\mathbf{x}) \leq \mathbf{0}$  and equality constraints  $h_j(\mathbf{x}) = \mathbf{0}$  if

$$g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, q, \quad (15)$$

$$|h_j(\mathbf{x})| - \epsilon \leq 0, \quad j = q + 1, \dots, m, \quad (16)$$

where equality constraints are transformed into inequalities. The mean value of all constraints' violations  $\bar{\nu}$  is defined as:

$$\bar{\nu} = \frac{(\sum_{i=1}^q G_i(\mathbf{x})) + (\sum_{j=q+1}^m H_j(\mathbf{x}))}{m}, \quad (17)$$

where the sum of all constraint violations is zero for feasible solutions, and positive when at least one constraint is violated:

$$G_i(\mathbf{x}) = \begin{cases} g_i(\mathbf{x}), & g_i(\mathbf{x}) > 0, \\ 0, & g_i(\mathbf{x}) \leq 0, \end{cases} \quad (18)$$

$$H_j(\mathbf{x}) = \begin{cases} |h_j(\mathbf{x})|, & |h_j(\mathbf{x})| - \epsilon > 0, \\ 0, & |h_j(\mathbf{x})| - \epsilon \leq 0. \end{cases} \quad (19)$$

The  $\epsilon$ -jDE algorithm [65] follows the jDE-2 algorithm [66], and emphasizes constraints as follows. It compares two solutions, say  $i$  and  $j$ , during the selection operation:

$$\mathbf{x}_{i,g+1} = \begin{cases} \mathbf{x}_{j,g} & \text{if } (\bar{\nu}_{i,g} > \bar{\nu}_{j,g}), \\ \mathbf{x}_{j,g} & \text{else if } (\bar{\nu}_{j,g} = 0) \wedge (f(\mathbf{x}_{i,g}) > f(\mathbf{x}_{j,g})), \\ \mathbf{x}_{i,g} & \text{otherwise.} \end{cases} \quad (20)$$

The algorithm distinguishes between feasible ( $\bar{\nu} = 0$ ) and infeasible individuals: Any feasible solution being better than any infeasible one. Namely, the jDE-2 algorithm had difficulties when solving constrained optimization problems with equality constraints. Since Takahama and Sakai in [67] pointed out

that, for problems with equality constraints, the  $\epsilon$  level should be controlled properly in order to obtain high quality solutions, the  $\epsilon$ -jDE algorithm, therefore, uses  $\epsilon$  level controlling, where the  $\epsilon$  level constraint violation precedes the Objective Function when the aggregated constraints' value is small. The  $\epsilon$ -jDE also uses an adaptive and self-adaptive control of  $\epsilon$  level, but thereby introduces some additional control parameters. The  $\epsilon$  level is updated until the number of generations  $g$  reaches the control generation  $g_c$ . After the number of generations exceeds  $g_c$ , the  $\epsilon$  level is set at 0 to obtain solutions with minimum constraint violations [65].

### 2.3. Task Parallelization and High-Performance Computing

High-Performance Computing (HPC) architectures involve the use of many interconnected processing elements to reduce the time to solution of a given problem, such as processing of huge amounts of Big Data, as defined in [68]. As presented in [68], many powerful HPC systems are heterogeneous, in the sense that they combine general-purpose CPUs with *accelerators* such as, Graphics Processing Units (GPUs), or Field Programmable Gate Arrays (FPGAs) [69]. Several HPC approaches exist [70, 71, 72, 73] developed to improve the performance of advanced and data intensive modeling and simulation applications. The parallel computing paradigm may be used on multi-core CPUs, many-core processing units (such as, GPUs [74]), re-configurable hardware platforms (such as FPGAs), or over distributed infrastructure (such as, cluster, Grid, or Cloud). To access the HPC, tools like ARC from ATLAS [75] can be utilized, to submit workload tasks as, e.g. Bash scripts and other executable and data files. Regarding widely used parallel programming frameworks [76] for heterogeneous systems, these include OpenACC [77], OpenCL [78], OpenMP [79], and NVIDIA CUDA [80]. OpenMP is a set of compiler directives, library routines, and environment variables for programming shared-memory parallel computing systems. Furthermore, OpenMP has been extended to support programming of heterogeneous systems that contain CPUs and accelerators. OpenCL supports portable programming of hardware provided by various vendors, while CUDA runs only on NVIDIA hardware. CUDA C/C++ compiler, libraries, and run-time software enable programmers to develop and accelerate data-intensive applications on a GPU. With regard to distributed parallel computing, the available frameworks include the Message Passing Interface (MPI) [81], MapReduce/Hadoop [82] or Apache Spark [83]. For Big Data processing, an important platform is also Data Flow [84], specifically, Maxeller platform [85].

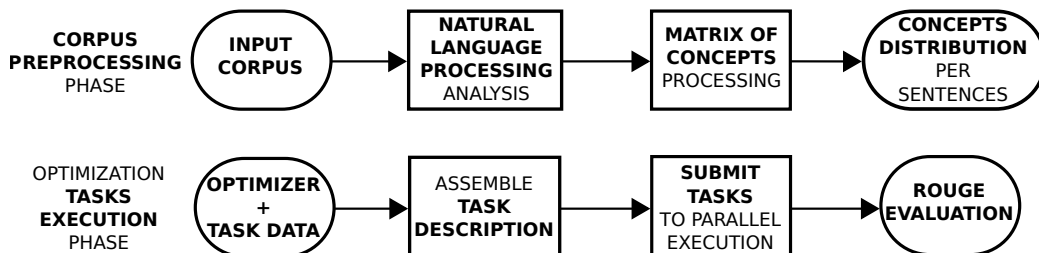


Figure 1: Architecture of the Multi-document Extractive Summarization Approach. Tasks are prepared in the first phase, then executed as the second phase in parallel.

### 3. Multi-document Extractive Summarization Approach via Differential Evolution and Sentence Optimization

The architecture, with phases, is provided in Figure 1. The two distinct phases are preprocessing of the corpus and execution of optimization tasks. The first phase needs to be completed before submitting a task in the next phase.

The proposed multi-document extractive summarization approach has been designed and developed following the standard stages for the automatic summarization process described in [86]: i) Information identification; ii) Information interpretation; and iii) Summary generation. The first stage consists of determining the particular subject the document is about. It is usually approached by assigning each unit (words, sentences, phrases, etc.) a score which is indicative of its importance. In the end, the top score units are extracted up to a desired length. This is far from trivial, since NLP techniques are necessary to understand the meaning of the documents' content. In the information interpretation stage, information in the form of topics identified as important are fused, represented in new terms, and expressed using a new formulation, which includes concepts or words not found in the original text. This stage is what distinguishes extractive from abstractive summarization. Finally, the last stage only makes sense if abstractive summaries are generated. In these cases, natural language generation techniques (text planning, sentence planning, and sentence realization) are needed to produce the final text of the summary.

Since, in this research work, we are focusing on an extractive summarization approach, i.e., the most important sentences would be extracted without any modification in vocabulary or format, the last two stages will not be necessary. Despite this, the extractive task is far from trivial, and it

still requires exploring novel methods and techniques to address it properly, as far as the results obtained by the generated summaries is concerned, as well as the processing time required to generate it, especially when dealing with a high number of input documents. The main novelty in the newly proposed approach in this paper is, therefore, the parallelization of tasks for optimization that are applied in the perspective of benchmarking, i.e. the individual benchmark tests are parallelized in such manner that summarization parallel tasks are submitted individually to a computer grid and merged after tasks are complete. Also, the information identification and quantitative evaluation using Freeling [8], coreference resolution, semantic analysis, and concept matrix in the pre-processing stage for summarization before the task execution for summary optimization (that loops DE generations, etc.) commences. The computational perspective defining static (precomputed) and changing (computed during evaluation call) parts of fitness function are also discussed, as this perspective is important in high-performance computing scenarios; namely, the parts of fitness function that do not need to be re-computed are precomputed. Regarding the optimization part novelty, a new algorithm combination is presented, a constrained version of a self-adaptive DE with binarization. In the following subsections, we explain in detail how the process has been designed, by first elaborating on the information identification and information interpretation and then presenting the distribution of parallel tasks to a grid.

### *3.1. Information identification and Quantitative evaluation*

In this stage, a linguistic analysis process is applied to the input. Using NLP tools, such as Freeling linguistic analyzer [8], the linguistic process consists of performing sentence splitting, tokenization, semantic analysis and coreference resolution. Sentence splitting aims at delimiting when a sentence starts and ends. Tokenization identifies the beginning and end of an element (token) in a specific language. A token could be a word, but also a punctuation mark. Semantic analysis will allow us to obtain the specific meaning of a word within its context, together with its semantic relations, such as synonymy, hypernymy, meronymy, etc. Coreference resolution, and, specifically, pronominal anaphora resolution will allow us to determine whether a pronoun and a word are referring to the same item within the text. Once this process is finished, the text will be represented, associating each of its words to its corresponding linguistic meaning.

Although the linguistic process may resemble the compilation process

of a programming language [87], specially the tokenization part included in the lexical analysis of a compiler, understanding and interpreting natural language automatically gets even harder, since contrary to programming languages, natural languages include ambiguity, so a deeper linguistic knowledge and appropriate linguistic resources have to be used, such a Freeling [8].

### 3.1.1. Sentence Splitting and Tokenization

Sentence and word text processing is different between ROUGE and Freeling. Therefore, we did a couple of experiments with different word whitespace configurations. We got the same results with texts containing punctuation marks as with those texts from the same source that did not contain any punctuation marks. On the contrary, we observed that spaces do influence the ROUGE results. Therefore, since the difference in some punctuation marks with the Freeling recomposed texts compared to the SPL (i.e., the format the source documents have), the following processing was used to create SPL from Freeling recomposed text (deleting spaces in front of " ", " ", ")", "( ", and splitting multi-word phrases " \_ ") using command `sed -e 's/ \././g' -e 's/ ,/,/g' -e 's/ -/_/g' -e 's/( /(g' -e 's/ )/)/g'`.

### 3.1.2. Semantic Analysis and Coreference Resolution

Before applying an evolutionary algorithm, we carried out a linguistic process over the DUC datasets in order that semantic understanding of texts could also be obtained. In addition, input documents were cleansed by removing XML tags, and a sentence segmentation step was then applied to detect sentence boundaries. Further on, documents were passed through a coreference resolution system (i.e. JavaRap<sup>4</sup>), so that pronoun references were replaced by their correct antecedent in the document. Once the references were removed from documents, a semantic analysis process was carried out. This is the most relevant step in the proposed linguistic process, since it involved the detection of concepts semantically-related in the documents. WordNet 3.0<sup>5</sup> was employed for this semantic analysis. WordNet is a lexico-semantic English resource that groups content words (i.e., nouns, verbs, adjectives and adverbs) into sets of synonyms called synsets, providing information about the semantic relationships between them. In our process,

---

<sup>4</sup><http://ilo.nus.edu.sg/for-industry/online-software-licences/g-online-software-licences/java-rap-resolution-of-anaphora-program/>

<sup>5</sup><https://wordnet.princeton.edu/>



we used such information to detect sets of synonyms in the documents, so in this manner, we could work with concepts instead of literal terms. For example, *detonation* and *explosion* are different words but their WordNet’s synsets are the same (*07323181*), so they belong to the same concept. Specifically, WordNet was used through the Freeling analyzer<sup>6</sup>, which, among other linguistic information, also provides the WordNet synset corresponding to a word. The advantage of using WordNet through Freeling tool, is that we can configure the word disambiguation method used for determining the meaning of a word in a particular context (this is especially useful in dealing with ambiguity). We set this option to the most frequent sense, that corresponds to the first WordNet synset of a term, but also taking also into consideration its part-of-speech (i.e., the grammatical class of a term). This guarantees that the meaning of that word in the document is its most probable meaning. In this manner, if two words have the same first synset, it will be considered that they are synonyms and they will be represented under the same concept.

It is worth mentioning that more sophisticated disambiguation methods could have been employed; however, they are not matured enough yet [88], and, therefore, they could introduce errors that would be propagated to further stages of the approach, and, therefore, be detrimental for the whole approach.

### 3.1.3. Concept Matrix

The information from the previous stages is used to build a concept Matrix (M), where the rows represent the concepts<sup>7</sup> of the document —extracted through the previous linguistic analysis— and the columns represent the sentences of the document/s. Each cell  $M_{i,j}$  contains the frequency of appearance of each concept. For those cases where the concept is not included in the sentence, a 0 is assigned to the cell. Once the Matrix has been filled in, the next stage of the process is to determine which concepts are the most relevant ones in order to extract the sentences with the highest relevance.

Specifically, we used concept matrix and stopword omitting for information identification in multi-document summarization on DUC2002 (using a max. 200 words per summary). The sentences were split using Freeling, and, therefore differed from the SPL sentences provided in the source documents;

---

<sup>6</sup><http://nlp.lsi.upc.edu/freeling/>

<sup>7</sup>Concepts provide a higher level of abstraction compared to words, since they can group together sets of synonyms

however, they could be used in the ROUGE, since the human models are free text and different from the provided SPL input. The wordcount and number of sentences for the optimizer were, therefore, reconstructed from the Freeling sentence analysis.

The words were classified into concepts (concept matrix) using the Freeling tool, and statistics were gathered on how many times some concepts appeared in a certain sentence. The Freeling was also used for sentence splitting, not only for word splitting and word classification. Also, no punctuation signs were treated as concepts, and, therefore, were omitted in the concept-matrix. The punctuation marks were not considered as stopwords (i.e. to be omitted), but they do not affect the concept matrix computations, because WordNet was used, which only identifies as concepts those words that are nouns, verbs, adjectives, or adverbs.

#### *3.1.4. Obtained Identified Multi-Document Abstractions to be Optimized*

Coherent with the definitions of the above subsections in this subsection, we then obtained the following number of sentences (i.e. features to select from), total number of words in the multi-documents for each DUC benchmark. The number of sentences and terms for DUC 2002 is listed in Table 1. In SPL-based input from DUC (using the original sentence splitting tool), 22,114 sentences are given. In Freeling format, 16,878 sentences were outputted for the DUC 2002. This gives a 23,68% difference in number of reduced number of sentences when using the Freeling. Namely, using Freeling sentence splitting approach, we used it for input, as well as for the reconstruction of terms into sentences, to match the correct number of words in sentences that were used in the input to the Concepts matrix that used WordNet term classification.

#### *3.2. Information interpretation: Sentence Selection Optimization*

The phasing of a summarization task using the multi-document sets is as follows (see algorithm pseudocode in Fig. 2):

1. All single multi-document sentences are packed together in one set to be summarized,
2. Each of these sets is sent as input to the summarization algorithm 20 times independently (i.e. for 20 independent runs),
3. In a summarization optimization run, each sentence in the set is received as indexed, and also each term in a sentence, then, for each term, a term dictionary and weight are built, and for sentences  $i =$

Figure 2: Algorithm Constraint-adjusting Binary Self-adaptive Differential Evolution for Data-Driven Models Extractive Text Summarization Optimization (CaBiSEETS)

1: **algorithm** CaBiSEETS( $RNi$ , MDOC,  $L$ ,  $GMAX$ ,  $NP$ ,  $F_l$ ,  $F_u$ ,  $\tau_F$ ,  $\tau_{CR}$ )

**Require:** variable parameters —  $RNi$  (current random seed number, run), MDOC (multi-document input to be summarized, already prepared by Freeling decomposition of text to sentences and terms),  $L$  (summary target length); DE constants —  $GMAX$  (number of generations to run DE),  $NP$  (DE population size),  $F_l$ ,  $F_u$ , (bounds of  $F$  and  $CR$  control-parameters' self-adaptation);  $\tau_F$ ,  $\tau_{CR}$  (frequencies of  $F$  and  $CR$  control-parameters' self-adaptation).

**Ensure:**  $\mathbf{x}$  – list of sentence indexes ( $D$  binary variables setting on/off switch for each of  $N$  sentences of a MDOC) for an extractive summary.

2: Read the input files  $MDOC$  previously processed by Freeling, precompute the ISF and weights for all  $M$  terms and pan-sentence similarities between all  $N$  sentences:

$$3: \forall n \in \{1, 2, \dots, N\}, \forall m \in \{1, 2, \dots, M\}: isf_m = \log\left(\frac{N}{|\{\forall n : m \in s_n\}|}\right);$$

$$4: \forall n \in \{1, 2, \dots, N\}, \forall m \in \{1, 2, \dots, M\}: w_{n,m} = tf_{n,m} isf_m;$$

$$5: \forall n \in \{1, 2, \dots, N\}: o_n = \frac{\sum_{j=1}^n w_{i,j}}{n};$$

$$6: \forall k, \forall l \in \{1, 2, \dots, N\}: sim(s_k, s_l) = \sum_{m=1}^M \frac{w_{k,m} w_{l,m}}{\sqrt{\sum_{m=1}^M w_{k,m} w_{k,m} \sum_{m=1}^M w_{l,m} w_{l,m}}};$$

7: initialize random number generator function  $\mathcal{U}$  seed to  $RNi$ ; set dimension  $D$  to  $N$ ;

8: uniform randomly generate DE initial population  $\mathbf{x}_{i,0} := \mathcal{U}[-5, 5]$ ,  $\forall i \in \{1, 2, \dots, NP\}$ ;

9: evaluate initial population and set constraint level  $\epsilon$  at  $\bar{v}$  of 30% least fit DE vector;

10: **for** DE generation loop  $g = 1$  to  $g = 10000$  **do**

11:   **for** DE iteration loop  $i$  (all individuals  $\mathbf{x}_{i,g}$  in a population) **do**

12:     DE new trial vector  $\mathbf{x}_{i,g}$  computation:

13:       choose mutually and from  $i$  distinct indices  $r_1, r_2, r_3 \in \mathcal{U}\{1, 2, \dots, NP\}$ ;

$$14: \quad F_{i,g+1} = \begin{cases} F_l + \mathcal{U}[0, 1] \times F_u & \text{if } \mathcal{U}[0, 1] < \tau_F \\ F_{i,g} & \text{otherwise} \end{cases};$$

$$15: \quad CR_{i,g+1} = \begin{cases} \mathcal{U}[0, 1] & \text{if } \mathcal{U}[0, 1] < \tau_{CR} \\ CR_{i,g} & \text{otherwise} \end{cases};$$

$$16: \quad \mathbf{v}_{i,g+1} = \mathbf{x}_{r_1,g} + F_{i,g+1}(\mathbf{x}_{r_2,g} - \mathbf{x}_{r_3,g})$$

$$17: \quad \forall j \in \{1, \dots, D\}: u_{i,j,g+1} = \begin{cases} v_{i,j,g+1} & \text{if } \mathcal{U}[0, 1] \leq CR_{i,g+1} \text{ or } j = j_{rand} \\ x_{i,j,g} & \text{otherwise} \end{cases};$$

18:     DE fitness evaluation (summary model calculation) for  $i$ -th vector  $\mathbf{u}_{i,g+1}$ :

$$19: \quad \mathbf{u}_{i,g+1} \text{ as } \mathbf{x} := \{x_1, x_2, \dots, x_D\}, \forall j: x_j := \mathcal{U}[0, 1] < \left|\left(\frac{2}{\pi} \arctan\left(\frac{\pi}{2} u_{i,j,g+1}\right)\right)\right|;$$

$$20: \quad V(\mathbf{x}) = \sum_{k=1}^{N-1} \sum_{l=k+1}^N (sim(s_k, O) + sim(s_l, O)x_k x_l);$$

$$21: \quad R(\mathbf{x}) = \sum_{k=1}^{N-1} \sum_{l=k+1}^N (sim(s_k, s_l)x_k x_l);$$

$$22: \quad f(\mathbf{u}_{i,g+1}) = f(\mathbf{x}) = -\frac{V(\mathbf{x})}{R(\mathbf{x})};$$

$$23: \quad \mathbf{x}_{i,g+1} = \begin{cases} \mathbf{u}_{i,g+1} & \text{if } \bar{v}_{u,g} < \epsilon \text{ and } \bar{v}_{i,g} < \epsilon \text{ and } f(\mathbf{u}_{i,g+1}) < f(\mathbf{x}_{i,g}) \\ \mathbf{x}_{i,g} & \text{if } \bar{v}_{u,g} < \epsilon \text{ and } \bar{v}_{i,g} < \epsilon \text{ and } f(\mathbf{u}_{i,g+1}) \geq f(\mathbf{x}_{i,g}) \\ \mathbf{u}_{i,g+1} & \text{if } (\bar{v}_{u,g} \geq \epsilon \text{ or } \bar{v}_{i,g} \geq \epsilon) \text{ and } \bar{v}_{u,g} < \bar{v}_{i,g} \\ \mathbf{x}_{i,g} & \text{if } (\bar{v}_{u,g} \geq \epsilon \text{ or } \bar{v}_{i,g} \geq \epsilon) \text{ and } \bar{v}_{u,g} > \bar{v}_{i,g} \\ \mathbf{u}_{i,g+1} & \text{if } \bar{v}_{u,g} \geq \epsilon \text{ and } \bar{v}_{i,g} = \bar{v}_{u,g} \text{ and } f(\mathbf{u}_{i,g+1}) \leq f(\mathbf{x}_{i,g}) \\ \mathbf{x}_{i,g} & \text{if } \bar{v}_{u,g} \geq \epsilon \text{ and } \bar{v}_{i,g} = \bar{v}_{u,g} \text{ and } f(\mathbf{u}_{i,g+1}) > f(\mathbf{x}_{i,g}) \end{cases};$$

24:     **end for**

25:     update constraint level  $\epsilon$  to 0 if  $g > 0.2GMAX$ , else to population's 30%-worst  $\bar{v}$ ;

26: **end for**

27: **return** the best individual obtained among  $\mathbf{x}_{i,g}$ ;

Table 1: DUC 2002 multi-document (MDOC), number of terms per one MDOC (Freeling wordcount). SntD stands for number of sentences in an MDOC, as used for input to the summarization, as given from the DUC repository for DUC2002 (`docs.with.sentence.breaks`, counting number of lines containing "`<s docid=`"). SntF stands for the number of sentences in an MDOC, as used for input to the summarization, as given from the Freeling decomposition of each MDOC. SntR stands for the number of sentences in an MDOC, as used for input to the ROUGE evaluation, based on the recomposition of the Freeling-decomposed sentences. After that column, for the latter case, the number of terms in total follow, for a certain MDOC.

MDOC	#SntD	#SntF	#SntR	#Terms	MDOC	#SntD	#SntF	#SntR	#Terms
d061j	238	202	202	4039	d092c	375	227	227	5026
d062j	158	118	118	2998	d093c	343	211	211	4350
d063j	319	261	261	5449	d094c	253	201	201	4740
d064j	254	192	192	4759	d095c	320	244	244	5278
d065j	372	328	328	6557	d096c	353	212	212	4478
d066j	250	200	200	4650	d097e	316	229	229	5323
d067f	168	121	121	3170	d098e	579	498	498	10245
d068f	182	134	134	2749	d099e	540	420	420	8416
d069f	450	360	360	8519	d100e	712	691	691	11693
d070f	250	163	163	3506	d101e	620	354	354	8798
d071f	204	126	126	2354	d102e	802	770	770	13831
d072f	483	480	480	9141	d103g	299	194	194	4414
d073b	306	168	168	3816	d104g	421	332	332	6265
d074b	318	176	176	4007	d105g	355	246	246	6888
d075b	362	274	274	6512	d106g	313	213	213	4285
d076b	421	321	321	7146	d107g	336	228	228	5355
d077b	432	323	323	6883	d108g	296	185	185	4092
d078b	359	337	337	5475	d109h	275	139	139	4189
d079a	398	343	343	7146	d110h	598	433	433	9341
d080a	653	548	548	12249	d111h	292	210	210	4781
d081a	403	291	291	6763	d112h	306	178	178	4868
d082a	284	216	216	5450	d113h	183	112	112	2903
d083a	275	209	209	4754	d114h	485	380	380	7629
d084a	460	395	395	8211	d115i	338	264	264	5611
d085d	307	232	232	4905	d116i	430	363	363	7727
d086d	514	395	395	8673	d117i	344	322	322	6214
d087d	363	293	293	6514	d118i	467	389	389	7904
d089d	480	376	376	8022	d119i	319	218	218	5035
d090d	344	247	247	5327	d120i	477	358	358	8755
d091c	360	286	286	5307	TOTAL:	22114	16878	16878	480197

$1, 2, \dots, D$ , the  $isf_{w_i}$ ,  $sim(i, O)$ , and  $sim(i, j)$  are calculated for all  $i, j \in \{1, 2, \dots, N\}$ , then

4. The DE generation loop is started, repeating  $GMAX$  times over the population of  $NP$  vectors representing candidate model summaries, yielding the obtained optimized summary model sentence selection,
5. Printing of the statistics and ROUGE evaluations.

After information identification (pre-processing), a system summary is computed, i.e. evolved by optimization. We chose BinaryDE [89] as the Evolutionary Algorithm, like in [18], but adding to the BinaryDE the  $\epsilon$ -jDE self-adaptation of control parameters [65, 53, 45], which evolves *sentence in-*

*clusion vectors* (i.e. a 0/1 set of flags if a sentence is to be included in the final summary). The genotype-phenotype mapping from floating points to binary values is done the same as in BinaryDE [89]. Also, we applied inside DE, that the epsilon constraint definition was such that for the upper limit, the sentence length constraint bound value was strictly violated, but for the lower bound, a difference was allowed for an epsilon between the maximum and minimum sentence. During evolution, the DE evaluates the trial vectors using the evaluation as defined in the previous subsection (3.1). The best vector sentences obtained by BinaryDE (each sentence marked “1”) is output as a reconstructed sentence text (in Sentence-Per-Line, SPL format), prepared by Freeling decomposition using each sentence’s corresponding sentence number. The obtained system summary (evolved by DE) sentences are mapped using its Freeling printed SPL format to unstemmed original SPL sentences, and compared to human model sentences using jackknifing on ROUGE 1.5.5 with Porter stemmer enabled. The algorithm listed using pseudocode in Fig. 2 is executed as a call to `CaBiSDETS(RNi=$RNi, MDOC=$MDOC, L = 200, GMAX = 10000, NP = 500, Fl = 0, Fu = 0.9,  $\tau_F = 0.1$ ,  $\tau_{CR} = 0.1$ )` using different *RNi* settings, where the `$RNi` seed and `$MDOC` filename are chosen in lines 4 and 5 of the script in Fig. 4. The pseudocode in Fig. 2 shows the use of input files and summary model preparation in lines 2–6, DE initializations in lines 7–9, then the DE loop follows in lines 10–26, which return the result in line 27. The inner DE loop of the algorithm in Fig. 2 runs trial vector computation in lines 12–17 and summary fitness evaluation in lines 18–22 (binarization happens in line 19), then it calculates the propagated DE vector in line 23, and updates the constraint handling level in line 25. After this, the algorithm is executed, and it determines the list of sentences that are in the optimized summary set of extracted sentences.

### 3.3. Distributing Parallel Tasks to a Grid

This subsection provides the definition of executing parallel tasks for summarization, where the benchmark for summarization is first split to different parallel tasks, and these tasks are then executed independently and, finally, merged.

The following two subsections provide two approaches to run the tasks in parallel, i.e. to run each task in the dataset in parallel in the same batch on one machine, or, to send these tasks as independent batches to be scheduled on different machines. Then, gathering of data from the parallel runs is explained.

```

1  #!/bin/bash
2  mkdir -p volatile
3
4  for RNi in {1..20}; do
5    for MDOC in `ls -1 input/duc2002.FreelingSentences_recomposed.TEXT0/\
6      | sed -e 's/\.txt//g'`; do
7      mkdir -p output/unstemmedmodels/$MDOC
8
9      ./summarizer & --constraintHandlingEpsilonAdaptivePopulations --useBinaryDE\
10     --RNi $RNi --GMAX 10000 --NP 500 --summarylength 200\
11     --inputfile input/duc2002docs_CONCEPT_COUNTS_CONCEPT_ID/${MDOC}.txt\
12     --withStatementMarkers --MDOCmatrix\
13     --useConceptMatrix input/duc2002docs.matrixConcepts.NoStopW_OK/${MDOC}.txt.
14     matrix\
15     input/DUC-2002.FreelingSentences_WORDCOUNT_PER_SENTENCE/${MDOC}.txt.wordcount\
16     --printSummaryStatementIDsONLY --printOptimizationNewBests\
17     --printParameterVectors --printOptimizationNewBests\
18     --printOptimizationNewBestsOutputROUGE\
19     | tee volatile/output-smdoc-$MDOC-$RNi;
20 done # MDOC
21 done # RNi
22 wait # all tasks to finish

```

Figure 3: Script listing for running in parallel for summarization tasks in DUC2002 using a single multi-processor machine.

### 3.3.1. Running each task in parallel on one machine

Running all tasks in parallel, the optimization execution file `summarizer` is called on a multi-processor Bash-operated machine using following script listed in Figure 3. The script generates output directories and uses input directories from a preprocessed DUC 2002 datasets. Lines 4 and 5 represent the loops over different multi-documents for all independent runs. The lines 9–17 run the specific code of the summarization algorithm, using the CaBiS-DETS algorithm with adaptive constraint handling (line 9), settings for DE (line 10), input text file (line 11), concepts matrix format settings and files (lines 11–14). Lines 2 and 7 prepare the output structure, while line 18 saves the results, and the lines 15–17 set the output print format for the optimizer executable.

### 3.3.2. Sending tasks as independent batches to be scheduled on different machines

Given the algorithm code from `code` directory put in package together with the corresponding input files as one task input file, the tasks are sent as independent batches for scheduling on different machines as shown in Figure 4. While here the optimization executable inputs and outputs (on

lines 12–21) and setup of output structure are same as in Figure 3, the difference is how the loops over different multi-documents for all independent runs are provided. Namely, these loops generate task shell Bash script (line 23), packaged together with input files (lines 26–32) and prepare the task package (lines 34–40) to be submitted and retrieved as seen in Figure 5 using ARC [75] commands, `arcsub ${task}.xrs1` (line 7) for every parallel task submitted and `arcget -a` command (line 13) after all tasks complete to retrieve the summaries (line 12 as proof-of-concept that could be changed to e.g. some loop checking using `[ $(arcstat -s Finished | wc -l) == 1180 ]`) for these 1180 tasks.

### 3.3.3. Merging of distributed tasks results

For sake of clarity, merging of ROUGE results on DUC 2002 dataset is listed in script in Figure 6. This shell script uses any of the outputs provided (lines 3-11) from the files structure generated from previous two parallelization approaches in Figures 4 and 3 and then prints the average ROUGE values (lines 13-16) or plots the results (lines 18-23).

Table 2: Average ROUGE 1, 2, L, and SU recall values through generations (logscale), for the best obtained solution during optimization with CaBiSDETS algorithm, using parameters  $NP = 500$  and  $GMAX=10000$ . Final generation yielded results are in bold.

Metric / Generation	ROUGE 1	ROUGE 2	ROUGE L	ROUGE SU4	Fitness
13	0.22588	0.0380907	0.184536	0.0573557	-80.1165
147	0.272849	0.054693	0.240132	0.0882645	-164.708
1008	0.287191	0.0580012	0.258406	0.0967963	-251.043
10000	<b>0.318715</b>	<b>0.0655228</b>	<b>0.290371</b>	<b>0.107963</b>	-511.336

## 4. Results and Discussion

This section presents results by reporting the quantitative evaluation of the obtained summary models using comparison to human peer summaries. The optimization algorithm parameter settings used were: Population size was set at  $NP=500$  and the maximum number of generations was set at  $GMAX=10000$ .

Figure 7 shows the text summarization model and ROUGE metric recall values’ convergence during generations, averaged over 20 runs. As observed, all the values improved with adding more generations and, therefore, it is

```

1  #!/bin/bash
2  mkdir -p xrsl
3
4  for RNi in {1..20}; do
5    for MDOC in `ls -l input/duc2002.FreelingSentences_recomposed.TEXT0/\
6    | sed -e 's/\.txt//g'`; do
7
8      task="$MDOC-seed-${RNi}"
9
10     echo -e "#!/bin/bash\ntar xvf summarizer-code-and-inputdata-${task}.tgz;
11     mkdir -p volatile;
12     ./summarizer --constraintHandlingEpsilonAdaptivePopulations --useBinaryDE\
13     --RNi $RNi --GMAX 10000 --NP 500 --summarylength 200\
14     --inputfile input/duc2002docs_CONCEPT_COUNTS_CONCEPT_ID/${MDOC}.txt\
15     --withStatementMarkers --MDOCmatrix\
16     --useConceptMatrix input/duc2002docs_matrixConcepts_NoStopW_OK/${MDOC}.txt.
17     matrix\
18     input/DUC-2002.FreelingSentences_WORDCOUNT_PER_SENTENCE/${MDOC}.txt.wordcount\
19     --printSummaryStatementIDsONLY --printOptimizationNewBests\
20     --printParameterVectors --printOptimizationNewBests\
21     --printOptimizationNewBestsOutputROUGE\
22     | tee volatile/output-smdoc-${MDOC}-${RNi};
23     tar cvzf ${task}.tgz volatile/output-smdoc-${MDOC}-${RNi};
24     echo Experiment successfully finished: ${task}" > ${task}.sh
25     chmod 755 ${task}.sh
26
27     cd code;
28     tar cvzf ../summarizer-code-and-inputdata-${task}.tgz .\
29     ../input/duc2002.FreelingSentences_recomposed.TEXT0/${MDOC}\
30     ../input/duc2002docs_CONCEPT_COUNTS_CONCEPT_ID/${MDOC}\
31     ../input/duc2002docs_matrixConcepts_NoStopW_OK/${MDOC}\
32     ../input/DUC-2002.FreelingSentences_WORDCOUNT_PER_SENTENCE/${MDOC}
33
34     cd ..
35
36     cat experiment.xrsl | sed -e "s/experiment.tgz/${task}.tgz/g"\
37     | sed -e "s/jobname=experiment/jobname=${task}/g"\
38     | sed -e "s/executable=task.sh/executable=${task}.sh/g"\
39     | sed -e "s/workload.tgz/workload-${task}.tgz/g"\
40     > ${task}.xrsl
41
42     mv ${task}.xrsl workload-${task}.tgz ${task}.sh xrsl
43 done # MDOC
44 done # RNi

```

Figure 4: Script listing for preparing parallel summarization tasks in DUC2002 using a grid to call the algorithm from Fig. 2.

reasonable to run the summarization processing for more generations. The ROUGE values, as well as the fitness of the model, were improving throughout the evolution. As an example, values over generation numbers at natural values 13, 147, 1008, and the last generation, the ROUGE values and fitness are seen in Table 2. The results in Figure 7 show an increasing trend as far as



```

1 #!/bin/bash
2 for RNi in {1..20}; do
3   for MDOC in `ls -1 input/duc2002-FreelingSentences-recomposed.TEXT0/\
4     | sed -e 's/\.txt//g'`; do
5
6     task="$MDOC-seed-${RNi}"
7     cd xrsl; arbsub ${task}.xrsl -j jobs.dat; cd ..
8
9   done # MDOC
10 done # RNi
11
12 read -p "Finished?" # wait for parallel tasks completion
13 arcget -a -j jobs.dat
14 for F in *.tgz; do tar xf $F; done

```

Figure 5: Submitting, retrieving, and merging results of parallel tasks execution on a grid (runs 1180 CaBiSDETS algorithm executions from Fig. 2 as Parallel Tasks, possibly merely simultaneously if allocated 1180 nodes at once by Slurm).

```

1 #!/bin/bash
2
3 for RUN in {1..20}; do
4   for DOC in $(ls -1 *-$RUN|sed -e 's/output-smdoc-//g' -e 's/-'$RUN'$//g'); do
5     export DOC; cat output-smdoc-$DOC-$RUN |\
6     awk '/ROUGE/{print ENVIRON["DOC"], $5, $(NF-3), $(NF-2), $(NF-1), $NF}' ;
7     done | tee ../plots/rougeCross4-$RUN; done
8
9 for RUN in {1..20}; do for i in {061..120}; do
10   cat rougeCross4-$RUN|grep d${i} > plot4-$RUN-$i; tail -n 1 plot4-$RUN-$i
11 done; done
12
13 for RUN in {1..20}; do
14   for i in {061..120}; do tail -n 1 plot4-$RUN-$i; done;
15 done > COMBINED
16 for i in {2..6}; do cat COMBINED | avg $i; done # prints the average ROUGE values
17
18 echo -n "plot " # generate plotting commands for Gnuplot
19 for RUN in {1..20}; do for i in {61..120}; do
20   cat rougeCross4-$RUN|grep ${i}..txt > plot-$RUN-$i; echo -n "'plot-$RUN-"$i;
21   echo -n "' u 3 w l lc "$i" t 'Run #'$RUN" for model ";
22   cat plot-$RUN-$i | awk 'NR==1 {printf("%s", $1)}'; echo -n ", ";
23 done; done

```

Figure 6: Merging of ROUGE results on DUC 2002 dataset from optimization tasks run in parallel.

the ROUGE values obtained, which indicates that the more generations, the better the model is for selecting relevant information. As the language model for multi-documents are usually not presented and might vary for different computations, we, hence, also included the obtained count of words for each

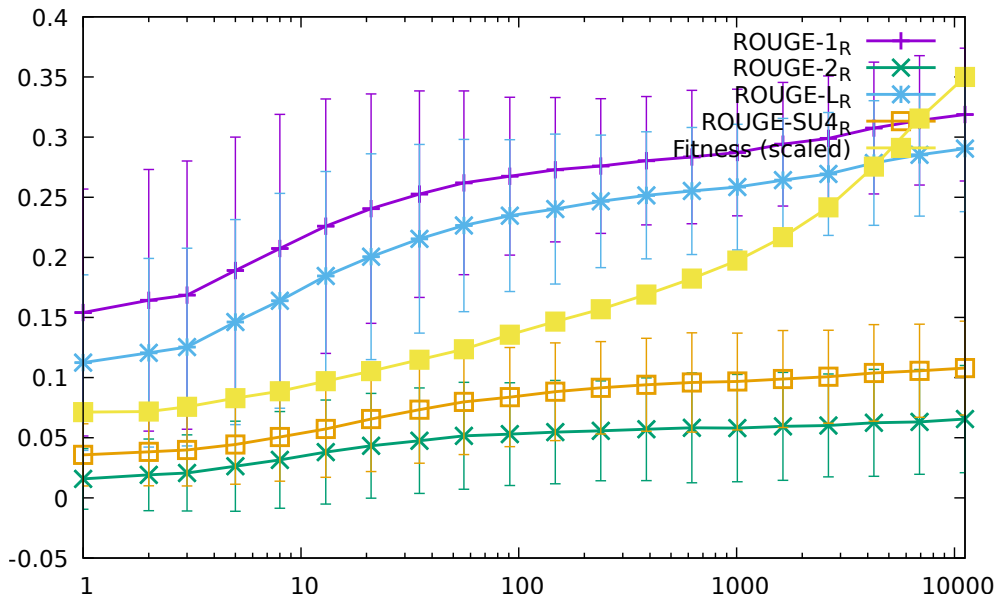


Figure 7: Average ROUGE 1, 2, L, and SU recall values through generations (logscale) for the best obtained solution during optimization with CaBiSDETS, using parameters  $NP = 500$  and  $GMAX=10000$ . Also, fitness obtained on average, is plotted along this evolution, as a scaled graph starting at 0.05 and divided by the best fitness value (-511.336). A vertical line on each data point draws the standard deviation at this point.

sentence in each multi-document set, in detail Figures 8 and 9. Furthermore, for the demonstrative purpose of describing the research, three outputs with different approaches for a sample summarization task (document d061j from DUC 2002) are demonstrated in Figure 10.

In terms of accuracy performance, the resulting summaries obtained acceptable results for the recall value compared to other summarization approaches. In this manner, with 10,000 generations, results overperformed the COMPENDIUM system, which is a very powerful summarization system in NLP [90]. Moreover, the results also improved with respect to a very competitive baseline (i.e. *Lead*) when the model was run with at least 1,000 generations. This baseline extracts the first sentences of the documents, and, taking into account the structure of a news document, in which its most important information is stated at the very beginning, the first paragraph contains the answer to the five Ws questions (i.e., *Who*, *What*, *When*, *Where*, and *Why*), so it can be appropriate to be considered as a summary.

		<b>ROUGE-1</b>	<b>ROUGE-2</b>	<b>ROUGE-L</b>	<b>ROUGE-SU4</b>
Our (10000 genera- tions)	model	0.31872 (0.05518)	0.065523 (0.04456)	0.29037 (0.05241)	0.10796 (0.03898)
COMPENDIUM		0.30340	0.05357	0.26554	0.09282
Lead baseline		0.28684	0.05283	-	-

Table 3: Comparison in terms of summary information appropriateness (recall values of ROUGE-1, ROUGE-2, ROUGE-L and ROUGE-SU4). Results in paranthesis for our approach report the standard deviation of the value considering different independent runs of DE over the multi-document sets.

Enabling factors of the analysis conducted were also the setup and definition of the approach, which included that the information identification and quantitative evaluation use Freeling [8], coreference resolution, semantic analysis, and concept matrix in the pre-processing stage before summary optimization (task) commenced; and, in the active part, a constrained version was used of a self-adaptive DE with binarization as an optimizer. As the parts of fitness function that do not need to be recomputed during optimization, are precomputed, the computational perspective defining static (pre-computed) and changing (computed during evaluation call) parts of fitness function yielded important and efficient high-performance computing scenarios. Under these scenarios, the parallelization of summarization tasks submitted, distributed to a computer grid and merged after tasks were complete, an important computationally-intensive correlation analysis was able to be reported, empirically providing insight into the correlation and trend among ROUGE metric values on the dedicated benchmark (DUC 2002) along with the number of fitness functions. Thereby, testing the trend and pay-off of prolonged optimization execution runs showed summarization quality improvement after extended optimization run time, that was possible due to task parallelization on grid.

Therefore, the experiments conducted and the results obtained confirm the feasibility of using parallelization and high-performance computing in summarization tasks, which renders it beneficial to be applied in such costly and very time-consuming tasks.

## 5. Conclusions and Future Work

This paper presented tackling of a hard optimization problem of computational linguistics, specifically automatic multi-document text summarization. Feasibility and analysis were demonstrated for the problem, and parallelization was explained for the case of High-Performance Computing. A Differential Evolution algorithm was applied to optimize the text summarization model.

Different DE runs were distributed to a grid in parallel as optimization tasks, seeking high processing throughput despite the demanding complexity of the linguistic model, especially on longer multi-documents, where DE improves results given more iterations. The impact of a larger number of function evaluations was demonstrated to be beneficial to improve summarization performance along with evolution steps in terms of ROUGE metrics.

As prospects for future research, we see further improvement of the optimization approach, including multi-modal and multi-objective approaches. A multi-objective approach, using multiple criteria functions, could be an interesting challenge for DE, and, although it is difficult to define more objective functions for the current problem (coverage vs. redundancy), multi-objective optimization seems likely for deeper knowledge insight, but criteria still need to be defined. Also, more comparisons with other techniques could be done based on different aspects of evaluation, like qualitative human-based evaluations.

## Acknowledgments

This paper is based upon work from COST Action IC1406 High-Performance Modelling and Simulation for Big Data Applications (cHiPSet), supported by COST (European Cooperation in Science and Technology). This paper is also based upon work from COST Actions CA15140 “Improving Applicability of Nature-Inspired Optimisation by Joining Theory and Practice (ImAppNIO)”, and CA18231 “Multi3Generation: Multi-task, Multilingual, Multi-modal Language Generation”, both supported by COST. The author AZ acknowledges the financial support from the Slovenian Research Agency (Research Core Funding No. P2-0041). AZ also acknowledges EU support under Project No. 5442-24/2017/6 (HPC – RIVR). AZ also acknowledges the EU Interreg Alpine Space project SmartVillages and Erasmus TSM grant. The author EL acknowledges the financial support by the Generalitat Valenciana through the Research Project PROMETEU/2018/089, and by the Spanish Government through the INTEGER project (RTI2018-094649-B-I00), and network RED iGLN (TIN2017-90773-REDT).





1. Automatic summary (our method, for peer 1, run 1, cluster d061j):  
 An estimated 100,000 people live in the province, including 70,000 in the city of Barahona, about 125 miles west of Santo Domingo.  
 On Saturday, Hurricane Florence was downgraded to a tropical storm and storm 's remnants pushed inland from the U. S. Gulf Coast.  
 The National Weather Service reported heavy damage to Kingston 's airport and aircraft parked on The National Weather Service 's fields.  
 Public buildings in Cancun were used as shelters, said Cecilia Lavalle, a spokesman for Quintana Roo state government in Chetumal, 155 miles southeast of Cozumel.  
 Thursday does n't happen very often.  
 If clouds form, the heat of condensation in the clouds occasionally provides ' positive feedback ' to the convergence pattern.  
 The sun puts energy into the water, the top of the oceans and lowest part of the atmosphere.  
 Hurricane Gilbert swept toward Jamaica yesterday with 100-mile-an-hour winds, and officials issued warnings to residents on the southern coasts of the Dominican Republic, Haiti and Cuba.

2. Human (d061jb):  
 Gilbert reached Jamaica after skirting southern Puerto Rico, Haiti and the Dominican Republic.  
 Hurricane Gilbert, one of the strongest storms ever, slammed into the Yucatan Peninsula Wednesday and leveled thatched homes, tore off roofs, uprooted trees and cut off the Caribbean resorts of Cancun and Cozumel.  
 The storm killed 19 people in Jamaica and five in the Dominican Republic before moving west to Mexico.  
 Prime Minister Edward Seaga of Jamaica said Wednesday the storm destroyed an estimated 100,000 of Jamaica's 500,000 homes when it throttled the island Monday.  
 More than 120,000 people on the northeast Yucatan coast were evacuated, the Yucatan state government said. Shelters had little or no food, water or blankets and power was out.  
 The Mexican National Weather Service reported winds gusting as high as 218 mph earlier Wednesday with sustained winds of 179 mph.  
 The National Hurricane Center said a hurricane watch was in effect on the Texas coast from Brownsville to Port Arthur and along the coast of northeast Mexico from Tampico north.  
 The National Hurricane Center said Gilbert was the most intense storm on record in terms of barometric pressure.  
 Tropical Storm Gilbert formed in the eastern Caribbean and strengthened into a hurricane Saturday night.

3. Lead baseline (d061j):  
 AP880912-0137 AP-NR-09-12-88 1555EDT u i AM-HurricaneGilbert Bjt 09-12 0681 AM-Hurricane Gilbert, Bjt,0702 Hurricane Hits Jamaica With 115 mph Winds; Communications Disrupted By LLOYD WILLIAMS Associated Press Writer KINGSTON, Jamaica AP Hurricane Gilbert slammed into Kingston on Monday with torrential rains and 115 mph winds that ripped roofs off homes and buildings, uprooted trees and downed power lines.  
 WSJ880912-0064 Hurricane Gilbert Heading for Jamaica With 100 MPH Winds LATAM SANTO DOMINGO, Dominican Republic AP Hurricane Gilbert swept toward Jamaica yesterday with 100-mile-an-hour winds, and officials issued warnings to residents on the southern coasts of the Dominican Republic, Haiti and Cuba.

4. COMPENDIUM (d061j):  
 Hurricane Gilbert swept toward the Dominican Republic Sunday, and the Civil Defense alerted its heavily populated south coast to prepare for high winds, heavy rains and high seas.  
 The storm was approaching from the southeast with sustained winds of 75 mph gusting to 92 mph.  
 There is no need for alarm," Civil Defense Director Eugenio Cabral said in a television alert shortly before midnight Saturday.  
 Cabral said residents of the province of Barahona should closely follow Gilbert's movement.  
 An estimated 100,000 people live in the province, including 70,000 in the city of Barahona, about 125 miles west of Santo Domingo.

Figure 10: Sample summaries generated for document cluster d061j, using different approaches.

## References

- [1] K. McKeown, R. J. Passonneau, D. K. Elson, A. Nenkova, J. Hirschberg, Do Summaries Help? A Task-Based Evaluation of Multi-Document Summarization, in: In Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1–8.
- [2] K. Kaczmarek-Majer, O. Hryniewicz, Application of linguistic summarization methods in time series forecasting, *Information Sciences* 478 (2019) 580 – 594.
- [3] I. Mani, D. House, G. Klein, L. Hirschman, T. Firmin, B. Sundheim, The tipster summac text summarization evaluation, in: Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics, EACL '99, Association for Computational Linguistics, Stroudsburg, PA, USA, 1999, pp. 77–85.
- [4] E. Lloret, M. Palomar, Text summarisation in progress: a literature review, *Artif. Intell. Rev.* 37 (2012) 1–41.
- [5] A. Zamuda, J. Brest, On Tenfold Execution Time in Real World Optimization Problems with Differential Evolution in Perspective of Algorithm Design, in: 2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP), IEEE, pp. 1–5.
- [6] A. Zamuda, J. D. H. Sosa, Success history applied to expert system for underwater glider path planning using differential evolution, *Expert Systems with Applications* 119 (2019) 155–170.
- [7] A. Zamuda, J. Brest, Population Reduction Differential Evolution with Multiple Mutation Strategies in Real World Industry Challenges, in: L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, J. Zurada (Eds.), *Swarm and Evolutionary Computation, Lecture Notes in Computer Science*, Springer, 2012, pp. 154–161.
- [8] L. Padro, E. Stanilovsky, Freeling 3.0: Towards wider multilinguality, in: Proceedings of the Language Resources and Evaluation Conference (LREC 2012), ELRA, Istanbul, Turkey, pp. 2473–2479.
- [9] M. Allahyari, S. A. Pouriye, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, K. Kochut, Text summarization techniques: A brief survey, *CoRR* abs/1707.02268 (2017).



- [10] W. Dressler, Current trends in textlinguistics, Research in text theory, W. de Gruyter, 1978.
- [11] C. Leopold, A. Brückner, S. Dutke, Summarizing as a strategy for science text comprehension: Text-based versus content-based processing, Discourse Processes 0 (2019) 1–20.
- [12] K. S. Jones, Automatic summarising: factors and directions, CoRR cmp-lg/9805011 (1998).
- [13] I. Mani, M. T. Maybury (Eds.), Advances in Automatic Text Summarization, MIT Press, 1999.
- [14] R. Ferreira, L. de Souza Cabral, F. Freitas, R. D. Lins, G. de França Silva, S. J. Simske, L. Favaro, A multi-document summarization system based on statistics and linguistic treatment, Expert Systems with Applications 41 (2014) 5780 – 5787.
- [15] W. Yin, Y. Pei, Optimizing sentence modeling and selection for document summarization, in: Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI’15, AAAI Press, 2015, pp. 1383–1389.
- [16] X. Zhang, M. Lapata, F. Wei, M. Zhou, Neural latent extractive document summarization, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 779–784.
- [17] M. Zhong, P. Liu, D. Wang, X. Qiu, X. Huang, Searching for effective neural extractive summarization: What works and what’s next, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 1049–1058.
- [18] R. M. Alguliev, R. M. Aliguliyev, C. A. Mehdiyev, Sentence selection for generic document summarization using an adaptive differential evolution algorithm, Swarm and Evolutionary Computation 1 (2011) 213–222.
- [19] R. M. Alguliev, R. M. Aliguliyev, N. R. Isazade, CDDS: Constraint-driven document summarization models, Expert Systems with Applications 40 (2013) 458–465.

- [20] R. M. Alguliev, R. M. Aliguliyev, M. S. Hajirahimova, GenDocSum+MCLR: Generic document summarization based on maximum coverage and less redundancy, *Expert Systems with Applications* 39 (2012) 12460 – 12473.
- [21] R. M. Alguliev, R. M. Aliguliyev, N. R. Isazade, DESAMC+DocSum: Differential evolution with self-adaptive mutation and crossover parameters for multi-document summarization, *Knowledge-Based Systems* 36 (2012) 21–38.
- [22] R. M. Alguliev, R. M. Aliguliyev, N. R. Isazade, Multiple documents summarization based on evolutionary optimization algorithm, *Expert Systems with Applications* (2012). DOI 10.1016/j.eswa.2012.09.014.
- [23] R. M. Alguliev, R. M. Aliguliyev, N. R. Isazade, Formulation of document summarization as a 0-1 nonlinear programming problem, *Computers & Industrial Engineering* (2012). DOI 10.1016/j.cie.2012.09.005.
- [24] A. Zamuda, C. Zarges, G. Stiglic, G. Hrovat, Stability selection using a genetic algorithm and logistic linear regression on healthcare records, in: *Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO 2017)*, pp. 143–144.
- [25] G. Zhang, H. Rong, F. Neri, M. J. Pérez-Jiménez, An optimization spiking neural p system for approximately solving combinatorial optimization problems, *International Journal of Neural Systems* 24 (2014) 1440006.
- [26] V. Roostapour, A. Neumann, F. Neumann, On the performance of baseline evolutionary algorithms on the dynamic knapsack problem, in: *International Conference on Parallel Problem Solving from Nature*, Springer, pp. 158–169.
- [27] A. Zamuda, G. Hrovat, E. Lloret, M. Nicolau, C. Zarges, Examples implementing black-box discrete optimization benchmarking survey for BB-DOB@GECCO and BB-DOB@PPSN, in: *Black Box Discrete Optimization Benchmarking (BB-DOB) Workshop at 15th International Conference on Parallel Problem Solving from Nature (PPSN 2018)*, September 8-12, 2018, Coimbra, Portugal, p. 1.

- [28] K. M. Hermann, T. Kočiský, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, P. Blunsom, Teaching machines to read and comprehend, in: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, NIPS'15, MIT Press, Cambridge, MA, USA, 2015, pp. 1693–1701.
- [29] C.-Y. Lin, ROUGE: A package for automatic evaluation of summaries, in: Proceedings of the ACL-04 Workshop on Text Summarization Branches Out, Barcelona, Spain, pp. 74–81.
- [30] G. Giannakopoulos, V. Karkaletsis, G. Vouros, P. Stamatopoulos, Summarization system evaluation revisited: N-gram graphs, *ACM Trans. Speech Lang. Process.* 5 (2008) 5:1–5:39.
- [31] L. A. Cabrera-Diego, J. Torres-Moreno, Summtriver: A new trivergent model to evaluate summaries automatically without human references, *Data Knowl. Eng.* 113 (2018) 184–197.
- [32] E. Lloret, L. Plaza, A. Aker, The challenging task of summary evaluation: an overview, *Language Resources and Evaluation* 52 (2018) 101–148.
- [33] R. Storn, K. Price, Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, *Journal of Global Optimization* 11 (1997) 341–359.
- [34] J. Holland, *Adaptation In Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, 1975.
- [35] A. E. Eiben, J. E. Smith, *Introduction to Evolutionary Computing (Natural Computing Series)*, Springer, 2003.
- [36] A. Zamuda, M. Nicolau, C. Zarges, A black-box discrete optimization benchmarking (BB-DOB) pipeline survey: taxonomy, evaluation, and ranking, in: Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO 2018), pp. 1777–1782.
- [37] S. Das, S. S. Mullick, P. Suganthan, Recent advances in differential evolution – An updated survey, *Swarm and Evolutionary Computation* 27 (2016) 1–30.

- [38] A. P. Piotrowski, Review of differential evolution population size, *Swarm and Evolutionary Computation* 32 (2017) 1–24.
- [39] A. P. Piotrowski, J. J. Napiorkowski, Some metaheuristics should be simplified, *Information Sciences* 427 (2018) 32–62.
- [40] R. D. Al-Dabbagh, F. Neri, N. Idris, M. S. Baba, Algorithmic design issues in adaptive differential evolution schemes: Review and taxonomy, *Swarm and Evolutionary Computation* 43 (2018) 284–311.
- [41] A. P. Piotrowski, J. J. Napiorkowski, Step-by-step improvement of JADE and SHADE-based algorithms: Success or failure?, *Swarm and Evolutionary Computation* 43 (2018) 88–108.
- [42] M. Weber, F. Neri, V. Tirronen, A Study on Scale Factor in Distributed Differential Evolution, *Information Sciences* 181 (2011).
- [43] F. Neri, G. Iacca, E. Mininno, Disturbed exploitation compact differential evolution for limited memory optimization problems, *Information Sciences* 181 (2011) 2469–2487.
- [44] M. Weber, F. Neri, V. Tirronen, A study on scale factor/crossover interaction in distributed differential evolution, *Artificial Intelligence Review* 39 (2013) 195–224.
- [45] A. Zamuda, J. Brest, Self-adaptive control parameters’ randomization frequency and propagations in differential evolution, *Swarm and Evolutionary Computation* 25 (2015) 72–99.
- [46] A. Viktorin, R. Senkerik, M. Pluhacek, A. Zamuda, Steady success clusters in Differential Evolution, in: 2016 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, pp. 1–8.
- [47] R. Tanabe, A. S. Fukunaga, How Far Are We From an Optimal, Adaptive DE?, in: 14th International Conference on Parallel Problem Solving from Nature (PPSN XIV), IEEE, p. accepted.
- [48] K. R. Opara, J. Arabas, Differential Evolution: A survey of theoretical analyses, *Swarm and Evolutionary Computation* 44 (2019) 546–558.

- [49] J. Brest, S. Greiner, B. Bošković, M. Mernik, V. Žumer, Self-Adapting Control Parameters in Differential Evolution: A Comparative Study on Numerical Benchmark Problems, *IEEE Transactions on Evolutionary Computation* 10 (2006) 646–657.
- [50] E. Mezura-Montes, B. C. Lopez-Ramirez, Comparing bio-inspired algorithms in constrained optimization problems, *The 2007 IEEE Congress on Evolutionary Computation (25-28 Sept. 2007)* 662–669.
- [51] X. Yao, Y. Liu, G. Lin, Evolutionary Programming Made Faster, *IEEE Transactions on Evolutionary Computation* 3 (1999) 82–102.
- [52] Y. Liu, X. Yao, Q. Zhao, T. Higuchi, Scaling Up Fast Evolutionary Programming with Cooperative Coevolution, in: *Proceedings of the 2001 Congress on Evolutionary Computation CEC 2001*, IEEE Press, 2001, pp. 1101–1108.
- [53] J. Brest, P. Korošec, J. Šilc, A. Zamuda, B. Bošković, M. S. Maučec, Differential evolution and differential ant-stigmergy on dynamic optimisation problems, *International Journal of Systems Science* 44 (2013) 663–679.
- [54] A. Viktorin, R. Senkerik, M. Pluhacek, T. Kadavy, A. Zamuda, Distance Based Parameter Adaptation for Success-History based Differential Evolution, *Swarm and Evolutionary Computation* 50 (2019) 100462.
- [55] K. V. Price, R. M. Storn, J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*, Natural Computing Series, Springer-Verlag, Berlin, Germany, 2005.
- [56] S. Das, P. N. Suganthan, Differential Evolution: A Survey of the State-of-the-art, *IEEE Transactions on Evolutionary Computation* 15 (2011) 4–31.
- [57] R. Joshi, A. Sanderson, Minimal representation multisensor fusion using differential evolution, *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 29 (1999) 1083–4427.
- [58] B. Bošković, J. Brest, A. Zamuda, S. Greiner, V. Žumer, History Mechanism Supported Differential Evolution for Chess Evaluation Function Tuning, *Soft Computing – A Fusion of Foundations, Methodologies and Applications* 15 (2011) 667–682.

- [59] A. Zamuda, J. D. H. Sosa, L. Adler, Constrained Differential Evolution Optimization for Underwater Glider Path Planning in Sub-mesoscale Eddy Sampling, *Applied Soft Computing* 42 (2016) 93–118.
- [60] D. Zaharie, Influence of crossover on the behavior of Differential Evolution Algorithms, *Applied Soft Computing* 9 (2009) 1126–1138.
- [61] Z. Michalewicz, M. Schoenauer, Evolutionary Algorithms for Constrained Parameter Optimization Problems, *Evolutionary Computation* 4 (1996) 1–32.
- [62] Z. Michalewicz, D. B. Fogel, *How to Solve It: Modern Heuristics*, Springer, Berlin, 2000.
- [63] C. A. Coello Coello, Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art, *Computer Methods in Applied Mechanics and Engineering* 191 (2002) 1245–1287.
- [64] E. Mezura-Montes, C. A. C. Coello, Constraint-handling in nature-inspired numerical optimization: past, present and future, *Swarm and Evolutionary Computation* 1 (2011) 173–194.
- [65] J. Brest, Constrained Real-Parameter Optimization with  $\epsilon$ -Self-Adaptive Differential Evolution, in: *Constraint-Handling in Evolutionary Optimization*, Springer, 2009, pp. 73–93.
- [66] J. Brest, V. Žumer, M. S. Maučec, Self-adaptive Differential Evolution Algorithm in Constrained Real-Parameter Optimization, in: *The 2006 IEEE Congress on Evolutionary Computation CEC 2006*, IEEE Press, 2006, pp. 919–926.
- [67] T. Takahama, S. Sakai, N. Iwane, Solving Nonlinear Constrained Optimization Problems by the  $\epsilon$  Constrained Differential Evolution, *IEEE International Conference on Systems, Man and Cybernetics 2006 (SMC 2006)* 3 (2006) 2322–2327.
- [68] S. Spolaor, M. Gribaudo, M. Iacono, T. Kadavy, Z. K. Oplatkova, G. Mauri, S. Pillana, R. Senkerik, N. Stojanovic, E. Turunen, A. Viktorin, S. Vitabile, A. Zamuda, M. S. Nobile, Towards human cell simulation, in: *Lecture Notes In Computer Science*, volume 11400 of *High-*

*Performance Modelling and Simulation for Big Data Applications: Selected Results of the COST Action IC1406 cHiPSet (Kolodziej, Joanna, Gonzalez-Velez, Horacio (Eds.)), pp. 221–249.*

- [69] A. Biasizzo, F. Novak, P. Korošec, A multi-alphabet arithmetic coding hardware implementation for small fpga devices, *Journal of Electrical Engineering* 64 (2013) 44–49.
- [70] D. Padua, *Encyclopedia of Parallel Computing*, Springer Publishing Company, Incorporated, 2011.
- [71] C. Kessler, U. Dastgeer, S. Thibault, R. Namyst, A. Richards, U. Dolinsky, S. Benkner, J. L. Träff, S. Pllana, Programmability and performance portability aspects of heterogeneous multi-/manycore systems, in: *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2012, IEEE, pp. 1403–1408.
- [72] S. Benkner, S. Pllana, J. L. Traff, P. Tsigas, U. Dolinsky, C. Augonnet, B. Bachmayer, C. Kessler, D. Moloney, V. Osipov, Peppher: Efficient and productive usage of hybrid computing systems, *IEEE Micro* 31 (2011) 28–41.
- [73] S. Memeti, S. Pllana, Hstream: A directive-based language extension for heterogeneous stream computing, in: *2018 IEEE International Conference on Computational Science and Engineering (CSE)*, pp. 138–145.
- [74] M. S. Nobile, P. Cazzaniga, A. Tangherloni, D. Besozzi, Graphics processing units in bioinformatics, computational biology and systems biology, *Brief. Bioinform.* 18 (2017) 870–885.
- [75] S. Haug, M. Hostettler, F. Sciacca, M. Weber, The atlas arc backend to hpc, *Journal of Physics: Conference Series* 664 (2015) 062057.
- [76] S. Memeti, L. Li, S. Pllana, J. Kolodziej, C. Kessler, Benchmarking OpenCL, OpenACC, OpenMP, and CUDA: Programming Productivity, Performance, and Energy Consumption, in: *Proceedings of the 2017 Workshop on Adaptive Resource Management and Scheduling for Cloud Computing, ARMS-CC '17*, ACM, New York, NY, USA, 2017, pp. 1–6.
- [77] S. Wienke, P. Springer, C. Terboven, D. an Mey, OpenACC: First Experiences with Real-world Applications, in: *Proceedings of the 18th*

International Conference on Parallel Processing, Euro-Par'12, Springer-Verlag, Berlin, Heidelberg, 2012, pp. 859–870.

- [78] J. E. Stone, D. Gohara, G. Shi, OpenCL: A parallel programming standard for heterogeneous computing systems, *Computing in science & engineering* 12 (2010) 66–73.
- [79] OpenMP, OpenMP 4.0 Specifications, <http://www.openmp.org/specifications/>, 2013. Accessed: 2019-02-28.
- [80] NVIDIA, CUDA C Programming Guide, <http://docs.nvidia.com/cuda/cuda-c-programming-guide/>, 2016. Accessed: 2019-02-28.
- [81] W. Gropp, E. Lusk, A. Skjellum, Using MPI: portable parallel programming with the message-passing interface, volume 1, MIT press, 1999.
- [82] Apache, Apache Hadoop project, <https://hadoop.apache.org/>, 2019. Accessed: 2019-02-28.
- [83] Apache, Apache Spark project, <https://spark.apache.org/>, 2019. Accessed: 2019-02-28.
- [84] A. Kos, S. Tomažič, J. Salom, N. Trifunovic, M. Valero, V. Milutinovic, New benchmarking methodology and programming model for big data processing, *International Journal of Distributed Sensor Networks* 11 (2015) 271752.
- [85] V. Milutinovic, A. Hurson, *Dataflow Processing*, volume 1st edition, Academic Press, 2015.
- [86] E. Hovy, Text summarization, in: *The Oxford Handbook of Computational Linguistics* 2nd edition, Oxford University Press, 2003, pp. 583–598.
- [87] T. A. Mogensen, *Introduction to Compiler Design*, Springer Publishing Company, Incorporated, 1st edition, 2011.
- [88] D. McCarthy, R. Navigli, Word sense disambiguation: An overview, *Proceedings of the 4th International Workshop on Semantic Evaluations* (2007) 7–12.



- [89] G. Pampara, A. P. Engelbrecht, N. Franken, Binary differential evolution, in: 2006 IEEE International Conference on Evolutionary Computation, IEEE, pp. 1873–1879.
- [90] E. Lloret, M. Palomar, COMPENDIUM: a text summarisation tool for generating summaries of multiple purposes, domains, and genres, *Natural Language Engineering* 19 (2013) 147–186.