



Calculating the Upper Bounds for Portuguese Automatic Text Summarization Using Genetic Algorithm

Jonathan Rojas-Simón^(✉), Yulia Ledeneva^(✉),
and René Arnulfo García-Hernández^(✉)

Autonomous University of the State of Mexico, Instituto Literario no. 100,
50000 Toluca, State of Mexico, Mexico
ids_jonathan_rojas@hotmail.com, yledeneva@yahoo.com,
rearnulfo@hotmail.com

Abstract. Over the last years, Automatic Text Summarization (ATS) has been considered as one of the main tasks in Natural Language Processing (NLP) that generates summaries in several languages (e.g., English, Portuguese, Spanish, etc.). One of the most significant advances in ATS is developed for Portuguese reflected with the proposals of various state-of-art methods. It is essential to know the performance of different state-of-the-art methods with respect to the upper bounds (*Topline*), lower bounds (*Baseline-random*), and other heuristics (*Baseline-first*). In recent works, the significance and upper bounds for Single-Document Summarization (SDS) and Multi-Document Summarization (MDS) using corpora from Document Understanding Conferences (DUC) were calculated. In this paper, a calculus of upper bounds for SDS in Portuguese using Genetic Algorithms (GA) is performed. Moreover, we present a comparison of some state-of-the-art methods with respect to the upper bounds, lower bounds, and heuristics to determinate their level of significance.

Keywords: *Topline* · Single-document summarization · Portuguese
Genetic algorithms · State-of-the-art methods

1 Introduction

Automatic Text Summarization (ATS) has been considered one of the most critical tasks in Natural Language Processing (NLP) that continues to be open. In the last three decades, a great variety of advances has been presented over the Document Understanding Conferences (DUC) and Text Analysis Conferences (TAC) workshops,¹ organized by the National Institute of Standards and Technology (NIST). These workshops have been focused in generate summaries in English. However, other organizations have been reported several advances in the state-of-the-art. One of the primary organization of this area is the Interinstitutional Center for Computational

¹ DUC website: <https://www-nlpir.nist.gov/projects/duc/>, TAC website: <https://tac.nist.gov/>.

Linguistics² (NILC, abbreviation in Portuguese) has been performed several advances and resources for NLP in Portuguese.

Since 1993, the researchers of NILC and others have been performed several applications for Portuguese ATS. Some of them include the use of supervised and unsupervised machine learning methods [1, 2], discursive knowledge models [3–5], identification of “gist sentence” from the source documents to generate extractive summaries [6], Text Simplification (TS) [7], complex networks and graph-based methods to text analysis [8–11]. On the other hand, some ATS systems have been proposed to generate extractive summaries through optimization-based methods [12–15]. In the most of these works have been presented ATS systems for Single-Document Summarization (SDS) and Multi-Document Summarization (MDS) using corpora like TeMario and CSTNews respectively [16–18].

In [19, 20] has been mentioned that the primary challenge of ATS is to generate extractive summaries of better similarity in comparison to the summaries created by humans (gold-standard summaries). However, for several domains, the gold-standard summaries are made by substituting some terms (words and phrases) from the source documents. This case is happened with DUC, TeMario and CSTNews corpus [16, 18, 21]. Consequently, the level of maximum similarity will be less than 100%, and therefore the upper bounds will be lower for any method. To determine the maximum similarity to the gold-standard summaries involves the search and evaluation of several numbers of possible sentence combinations from a document to generate the best extractive summary.

Currently, some heuristics have been used to compare the performance of several state-of-the-art methods to know their level of advance. These heuristics are known as *Baseline-first* and *Baseline-random* that reflects the standard and lower bounds respectively [19]. On the other hand, the use of *Topline* heuristic has been introduced in recent works, with the purpose of reflecting the upper bounds [22]. These one have been used to calculate the significance of SDS and MDS tasks [19, 20]. However, for Portuguese SDS has not performed a significant analysis to compare the best state-of-the-art methods due to that *Topline* was unknown.

The use of optimization-based methods for SDS and MDS have been represented a viable solution to generate extractive summaries of superior performance. These ones include the use of Genetic Algorithms [13]. Therefore, the use of optimization-based methods represents a viable solution to obtain extractive summaries closest to the human-written summaries. In this paper, a GA is used with some adjustment of parameters of [19] to get the sentence combinations of best similarity to the sentences selected by humans in Portuguese, using some evaluation measures of the ROUGE system. Moreover, different lengths of summaries and sentence segmentations as constraints were considered to calculate the upper bounds.

The rest of the paper is organized as follows: Sect. 2 presents some related works and previous works that have used techniques based on exhaustive searches and optimized based techniques to determine the best sentence combinations to calculate the significance for SDS and MDS methods. Section 3 describes the structure and

² <http://www.nilc.icmc.usp.br/nilc/index.php>.

development of proposed GA. Section 4 shows the experimental configuration of GA to determine the *Topline* for TeMario corpus. Moreover, a significant analysis to identify the best state-of-the-art methods with the use of *Baseline-first*, *Baseline-random*, and *Topline* heuristics. Finally, Sect. 5 describes the conclusions and future work.

2 Related Works

Over of the last two decades, many problems have been treated around the ATS (e.g., automatic evaluation of summaries [23, 24], sentence boundary detection, language analysis, etc.). However, few studies have been performed to determine the best extractive summaries. Some related works use techniques based on exhaustive searches to represent the summaries made by humans [25, 26].

In the work of Ceylan [25], an exhaustive search-based method was presented to obtain the best combinations of sentences. Unlike of Lin and Hovy [26], this method employs a probability density function to reduce the number of all possible combinations of sentences in different domains (literary, scientific, journalistic and legal). Each sequence is evaluated with some metrics of the ROUGE system. A similar approach has been performed in [27]. Nevertheless, the main problem of this method involves the partial processing of several source document subsets to reduce their handling [19, 20]. Therefore, the use of this strategy can generate biased results.

In the work of [28], nine heuristic methods to reduce and assign scores to the sentence combinations for SDS and MDS have been presented. First, the redundant sentences are removed. Subsequently, the remaining sentences are introduced into eight methods to assign them a score according to the gold-standard summaries, with the purpose of eliminating the low scoring sentences. However, the use of several heuristics to determine the best combinations of sentences in different domains and different entries allows the increase of computational cost to find the best sentence combinations. Furthermore, for SDS only a single gold-standard summary was used. In the case of MDS, only 533 documents of 567 on DUC02 were used, generating more biased results.

Finally, a calculus of significance and upper bounds for SDS and MDS using GAs were presented in [19, 20]. Using three different heuristics (*Baseline-random*, *Baseline-first*, and *Topline*) that represent the lower, standard and upper bounds it has been calculated the percentage of advance of several state-of-the-art methods for SDS and MDS, using DUC01 and DUC02 as test datasets. Unlike the previous works, all sentences were considered as candidates to construct the best extractive summaries and calculate the upper bounds using GAs. In this paper, we propose the calculus of upper bounds in Portuguese using GAs to find the best combinations of sentences that can be generated from the single-document summaries of TeMario corpus and rank the best SDS methods for Portuguese.

3 Calculating Upper Bounds

To calculate the upper bounds for SDS in Portuguese, we propose the use of typical steps and procedures of basic GA described in [29], to evaluate several combinations of sentences in an optimized search space. In this section, the main stages and descriptions of the proposed GA to calculate the upper bounds are shown.

Solution Representation. In previous works [19, 20], the solution is presented using a coding of individuals considering the order of sentences that can appear in the extractive summary. Therefore, each individual X_i (a candidate of best extractive summary) is represented in a vector of n positions $[P_1, P_2, \dots, P_n]$, where each position includes a sentence $\{S_1, S_2, \dots, S_n\}$ of original document D . For each coding to be considered like an extractive summary, the first sentences are considered according to a limit of words.

Fitness Function. The evaluation of individuals is an essential stage of GA where each candidate summary X_i is evaluated according to the F-measure score from ROUGE system metrics [30]. The maximum F-measure score of summary X_k obtained from g generations determine the best combination of sentences found by GA. This maximization is shown in Eq. (1), where n is the length of n-grams for evaluation of candidate summaries.

$$\text{Max}(F(X_k(g))) = \frac{\sum_{S \in S_{ref}} \sum_{gram_n \in S} \text{Count}_{match}(gram_n)}{\sum_{S \in S_{ref}} \sum_{gram_n \in S} \text{Count}(gram_n)}, g = \{0, \dots, G\} \quad (1)$$

In this case, we have focused in optimize through ROUGE-1 and ROUGE-2 metrics (evaluation based on bag-of-words and bigrams respectively) due to that these metrics have been obtained the maximum correlations with respect to human judgments [30]. F is the F-measure score of the ROUGE system, and $\text{Count}_{match}(gram_n)$ is the number of co-occurrent n-grams between a candidate summary X_i and gold-standard summary. If the candidate summary $X_k(g)$ has the highest co-occurrence of n-grams from all populations $X_i(g)$, then it will have the best combination of sentences due to that it has the most substantial of retrieved n-grams.

Initialization of Individuals. To initialize the population of individuals (when $g = 0$) must be generated with codifications of random real numbers for signature each sentence of source document $D = \{S_1, S_2, \dots, S_n\}$ in each position P_i of $[P_1, P_2, \dots, P_n]$. Therefore, the first generation of individuals will be according to Eq. (2), where a_s represents a real integer number $\{1, 2, \dots, n\}$ that corresponds to the number of the selected sentence in document D , $c = 1, 2, \dots, N_{pop}$, $s = 1, 2, \dots, n$, n is the number of the n -th sentence from the source document.

$$X_c(0) = [X_{c,1}(0), X_{c,2}(0), \dots, X_{c,n}(0)], X_{c,s} = a_s \quad (2)$$

Therefore, each sentence has the same probability of being included as part of an extractive summary according to a number W of requested words (see Eq. (3)).

$$\sum_{S_i \in \text{Summary}} l_i \leq W \quad (3)$$

where l_i is the length of the sentence S_i (measured in words) and W is the maximum number of words allowed to generate an extractive summary. In this case, we considered the use of several numbers of words per document as a constraint, due to that the lengths of each document of TeMario (gold-standard summaries and source documents) are made up of different compression rates.

Selection. In the selection stage, we propose the use of two selection operators to obtain the best subsets of individuals for each population of individuals. The first one consists in selecting a small subset of individuals through the elitism operator, which has the feature to choose minimal subgroups of individuals of best aptitude from generation g to pass the next generation ($g + 1$). To select the remaining individuals from each generation, we propose several select of individuals from the tournament selection operator. This operator generates several subsets of N_{Tor} randomly picked individuals to retrieve the individual with the best fitness value, as shown in Eq. (4), where $X_b(g)$ is the individual with the best fitness value and F is the F-measure score of ROUGE metric.

$$X_b(g) = \operatorname{argmax}(F(X_1(g)), F(X_2(g)), \dots, F(X_{N_{Tor}}(g))) \quad (4)$$

To integrate the selection stage, we propose to use the elitism operator to choose the best individuals of each population, using a percentage of them. Finally, the remaining individuals are obtained from the tournament selection operator, using samples of two randomly obtained individuals.

Crossover. For the crossing of individuals, we use the cycle crossover algorithm (CX) to interchange a subset of genes according to a start point (initial gene). For the CX operator to be started, it is necessary considering a crossover probability P to determine the subset of individuals who will perform the genetic exchange. Therefore, if b_{rand} (a random number) is between 0 and P , then the operator must select a starting point to perform the genetic exchange of parents $X_{p1}(g)$ and $X_{p2}(g)$ to generate an offspring $Y_i(g)$, otherwise, the first parent ($X_{p1}(g)$) will be $Y_i(g)$. To produce the second offspring, the roles of $X_{p1}(g)$ and $X_{p2}(g)$ are exchanged.

Mutation. For the mutation stage, we propose taking a set of individuals $Y_i(g)$ to generate individuals $Z_i(g)$ modifying some genes of each population of individuals. To the mutation of individuals, we used the insertion mutation operator to select a pair of genes of the individual $Y_{i,t}(g)$ and $Y_{i,r}(g)$ randomly to insert the gene $Y_{i,t}(g)$ in the gene $Y_{i,r}(g)$, as shown in Eq. (5), where r is the variable that relates the gene to be inserted, the variable t represents the target gene to be inserted, which are an element of subset $s = \{1, 2, \dots, n\}$, and n is the number of the sentence S_i from the source document D .

$$Z_{i,s}(g) = \begin{cases} Y_{i,t}(g) = Y_{i,r}(g), Y_{i,t\pm 1}(g) = Y_{i,t}(g), \dots, Y_{i,r}(g) = Y_{i,r\pm 1}(g); & \text{if } 0 < rand \leq P \\ Y_{i,s}(g) & \text{otherwise} \end{cases} \quad (5)$$

Therefore, if *rand* (a random number) is between 0 and *P*, then the mutation of individuals is performed by insertion operator, otherwise, the individual is not modified.

Replacement of Individuals. Taking as reference the previous works [19, 20], the replacement of individuals step, we propose to integrate the set of individuals generated by elitist selection ($E(g + 1)$) and the set of individuals $Z_i(g)$ from the mutation stage, to integrate the population of the next generation ($X_i(g + 1) = X_i(g + 1) \cup Z_i(g)$).

Termination Criterion. The termination criterion used to halt the GA iterations is determined by the number of *G* generations established as a constraint of stop. In the experimentation stage, 50 generations were used for each document of TeMario due to that was the best parameter found.

4 Experimental Results

In this section, we describe TeMario corpus and the experiments performed to generate the best extractive summaries. Moreover, the performance of some state-of-the-art methods and heuristics are presented to determine which methods of the state-of-the-art are more significant.

4.1 TeMario Corpus

TeMario (derived from “TExtos com suMÁRIOS”) is a corpus of 100 newspaper articles written in Brazilian Portuguese. 60 documents were written by the online Brazilian newspaper *Folha de São Paulo*, and 40 documents were written by the newspaper *Journal do Brasil* (Brazilian Newspaper) [16]. The TeMario documents are distributed into equitably five sections (Special, World, Opinion, International, and Politics).³ Moreover, a gold-standard summary was generated for each document of TeMario by a professional writer. Table 1 shows some features of TeMario corpus.

Unlike DUC datasets, TeMario was not created with specific constraints to indicate the comparison of the performance of the state-of-the-art methods. One of the main problems is derived from the lack of explicit identification of sentences or phrases to generate summaries because it was not determined the sentence labeling. Due to this, we present the segmentation of sentences in three different cases.⁴ The first segmentation consists in divide the documents by paragraph, the second segmentation includes in split the source documents into several sentences manually (Tagged), and finally, the third division consists in divide the documents into sentences through an automatic sentence boundary detection tool (SEnTER) developed by the same author of TeMario.⁵ Table 2 shows the number of sentences of each segmentation.

³ <https://www.linguateca.pt/Repositorio/TeMario/>.

⁴ Each segmentation can be downloaded from <https://gitlab.com/JohnRojas/Corpus-TeMario>.

⁵ http://conteudo.icmc.usp.br/pessoas/taspardo/SEnTER_Por.zip.

Table 1. TeMario corpus description [16].

Journal	Section	Number of documents	Number of words	Mean of words per document
<i>Folha de São Paulo</i>	Special	20	12340	617
	World	20	13739	686
	Opinion	20	10438	521
<i>Journal do Brazil</i>	International	20	12098	604
	Politics	20	12797	439
	Total	100	61412	
	Mean		12282	613

Table 2. Number of sentences obtained from TeMario corpus using different segmentations.

	Paragraph	Tagged	SENDER
Total	1275	2896	2899
Mean	12.75	28.96	28.99
Std	6.20	9.08	9.22

According to the number of sentences obtained from different segmentations (see Table 2), the use of varying segmentations of terms generates different sequences to construct extractive summaries, and therefore the performance of SDS methods can be affected. The division by paragraphs presents fewer sentences to combine. Tagged and SENTER segmentations presents a similar number of sentences. However, these segmentations capture different sequences of terms due to that Std indicator is different between segmentations.

4.2 Parameters of GA

With respect to the GA, different parameters were carried out; however, the best parameters performed are presented in Table 3. Unlike the previous works [19, 20], the *Topline* was calculated considering different segmentations of sentences described above. Moreover, the gold-standard summaries were written with different lengths and therefore were not possible to determine the upper bounds. Considering this constraint, the *Topline* was calculated for different lengths. These are: 1. Summaries with a compression rate of 30% (parameter proposed in [1, 10]). 2. Summaries with 100 words. 3. Summaries according to the length of words to the gold-standard summaries.

4.3 State-of-the-Art Methods and Heuristics

In this paper, we determine the level of advance with respect to other heuristics (*Baseline-first*, *Baseline-random*, and *Topline*). The methods and heuristics taken into consideration for this comparison are the following:

Table 3. GA parameters to calculate the upper bounds of TeMario corpus.

N_{pop}	Selection				Crossover		Mutation	
200	Operator	e	Operator	N_{Tor}	Operator	P	Operator	P
	Elitism	1%	Tournament	2	CX	85%	Insertion	0.012%

Baseline-first: This heuristic uses the first sentences from the source text to generate an extractive summary, according to a length of words. The performance of this heuristic has been generated good results in SDS and MDS [19]. However, this heuristic must be overcome by state-of-the-art methods.

Baseline-random: This heuristic consists in selecting a random number of sentences according to a length of words to generate an extractive summary. This heuristic allows us to determine how significant is the performance of the state-of-the-art methods [29].

Topline: It is an heuristic that allows to obtain the upper bounds for SDS and MDS that any state-of-the-art method can achieve, due to the lack of concordance between evaluators [22].

GA-Summarization: The method presented in [13] uses a GA to generate extractive independent-language summaries. This method evaluates the quality of each candidate summary considering three features: 1. Frequency of terms in sentence. 2. Frequency of terms in summary. 3. Importance of sentences according to the position from the source document.

GistSumm: The method presented in [6] uses a gist-sentence approach to generate extractive summaries. First, the identification of the “gist-sentence” is performed through simple statistical measures. Then, the gist sentence is used as a guideline to identify and select other sentences to integrate the extractive summary. This method can generate extractive summaries in three different forms: 1. Intrasentential Summarization (*GistSumm-1*). 2. Query-based summarization (*GistSumm-2*). 3. Average keywords ranking (*GistSumm-3*).

Shvoong: It is an online tool founded by Avi Shaked and Avner Avrahami to generate extractive summaries in 21 different languages.⁶ Some of them include the English, French, German, Portuguese, and others.

Open Text Summarizer (OTS): It is an open-source application to generate multilingual extractive summaries that can be downloaded online.⁷ This tool allows constructing extractive summaries based on the detection of the main ideas from the source document, considering the reduction of redundant information.

To compare the performance of heuristics and the state-of-the-art methods previously described, the evaluation based on the statistical co-occurrence of bag-of-words and bigrams (ROUGE-1 and ROUGE-2) from the ROUGE system was performed [30]. ROUGE-1 and ROUGE-2 use the ROUGE-N evaluation method, based on the

⁶ <http://www.shvoong.com/summarizer/>. (URL viewed May 7th, 2017).

⁷ <https://github.com/neopunisher/Open-Text-Summarizer/> (URL viewed February 10th, 2018).

statistical co-occurrence of terms included between a candidate summary and the gold-standard summaries (see Eq. (6)).

$$ROUGE - N = \frac{\sum_{S \in \text{Summ}_{ref}} \sum_{gram_n \in S} \text{Count}_{match}(gram_n)}{\sum_{S \in \text{Summ}_{ref}} \sum_{gram_n \in S} \text{Count}(gram_n)} \quad (6)$$

Tables 4, 5 and 6 show the average results of ROUGE-1 and ROUGE-2 (R1 and R2) scores of *Baseline-first*, *Baseline-random*, and *Topline* heuristics considering different segmentations of TeMario. As we can see, the performance of *Baseline-random* and *Topline* was affected, due to that the selection of sentences was obtained from different criteria (paragraph, automatic and manual form). On the other hand, the performance of *Baseline-first* was not affected significantly, due to this heuristic only uses the length of words to construct extractive summaries. However, each segmentation of sentences generated a higher number of words (some words were split), and therefore, the evaluation step generates different results (but it is not significant). Moreover, the use of different compression rates (100 words, 30% of the source text and according to the length of gold-standard summaries) affects the performance of all heuristics, due to the gold-standard summaries has different lengths of words and therefore must be evaluated with varying rates of compression.

To compare the state-of-the-art methods and heuristics previously described, we generated summaries according to human segmentation (Tagged) with a compression rate of 30% from the source documents (see Table 7). Table 7 shows the performance of GA-Summarization method (48.791) is better than other state-of-the-art methods in ROUGE-1. However, the performance of GistSumm-2 method (18.375) is better than other state-of-the-art methods in ROUGE-2. On the other hand, the performance of *Baseline-first* outperforms all state-of-the-art methods in ROUGE-1 (48.986) and ROUGE-2 (18.948). Furthermore, some methods have been obtained worse performance than *Baseline-random* heuristic (Gist-Summ-3 and *GistSumm-1*).

To unify the performance of the state-of-the-art methods in ROUGE-1 and ROUGE-2, the Eq. (7) was used to rank the best ones according to the position of each method (See Table 7).

$$\text{Ran}(\text{method}) = \sum_{r=1}^6 \frac{(6-r+1)R_r}{6} \quad (7)$$

where R_r refers the number of times the method occurs in the r -th rank. The number 6 represents the total number of methods involved in this comparison.

Table 8 shows the result rank of each state-of-the-art method. As we can see, the performance of *GA-Summarization* (1.833) and *GistSumm-2* (1.833) methods show the best positions in the method rankings. However, *GA-Summarization* performs some independent-language features, while the performance of *GistSumm-2* depends on some language features. On the other hand, the methods *Shvoong*, *OTS*, *GistSumm-3* and *GistSumm-1* present the same positions across ROUGE metrics.

Table 4. Results of heuristics considering the segmentation by paragraph.

	100 words		30% of source text		Gold-standard summary	
	R1	R2	R1	R2	R1	R2
<i>Topline</i>	54.653	27.808	58.253	28.693	58.848	28.647
<i>Baseline-first</i>	46.448	19.881	49.001	18.952	49.514	18.883
<i>Baseline-random</i>	37.605	10.497	46.242	15.540	47.952	16.669

Table 5. Results of heuristics considering the tagged segmentation.

	100 words		30% of source text		Gold-standard summary	
	R1	R2	R1	R2	R1	R2
<i>Topline</i>	59.986	32.401	62.342	32.558	63.223	32.870
<i>Baseline-first</i>	46.452	19.881	48.986	18.948	49.493	18.866
<i>Baseline-random</i>	38.324	10.731	45.721	14.616	48.527	16.515

Table 6. Results of heuristics considering the segmentation of SENTER.

	100 words		30% of source text		Gold-standard summary	
	R1	R2	R1	R2	R1	R2
<i>Topline</i>	59.765	32.437	62.282	32.598	63.155	32.779
<i>Baseline-first</i>	46.448	19.881	49.005	18.953	49.515	18.883
<i>Baseline-random</i>	38.009	10.507	45.743	15.126	47.657	16.227

Table 7. Results of the state-of-the-art methods and heuristics considering a tagged segmentation of sentences with a compression rate of 30%.

	ROUGE-1	ROUGE-2
<i>Topline</i>	62.342	32.558
<i>Baseline-first</i>	48.986	18.948
GA-Summarization	48.791 (1)	18.375 (2)
GistSumm-2	48.552 (2)	18.862 (1)
Shvoong	47.819 (3)	17.923 (3)
OTS	47.199 (4)	17.401 (4)
<i>Baseline-random</i>	45.721	14.616
GistSumm-3	45.021 (5)	15.651 (5)
GistSumm-1	35.864 (6)	11.563 (6)

Table 8. Ranking of the state-of-the-art methods.

Method	R_r						Resultant rank
	1	2	3	4	5	6	
GA-Summarization	1	1	0	0	0	0	1.833
GistSumm – 2	1	1	0	0	0	0	1.833
Shvoong	0	0	2	0	0	0	1.333
OTS	0	0	0	2	0	0	1.000
GistSumm – 3	0	0	0	0	2	0	0.666
GistSumm – 1	0	0	0	0	0	2	0.333

5 Conclusions and Future Work

In several works have been presented several ATS methods for SDS and MDS tasks to generate extractive summaries in Portuguese. However, the calculus of upper bounds was unknown. In this paper, a calculus of upper bounds for SDS in Portuguese was presented. Furthermore, it was possible to generate a general ranking of the state-of-the-art methods according to their position.

In the process of calculating the upper bounds, it was necessary the use of different segmentation of sentences to obtain the best extractive summaries in Portuguese, due to that TeMario has not a specific delimitation of items to generate extractive summaries. Nevertheless, in this work, we proposed the use of three different segmentation of sequences (Paragraph, Tagged and Automatic Segmentation of sentences) to generate extractive summaries in TeMario.

The length of gold-standard summaries affects the performance of lower bounds and upper bounds (*Topline* and *Baseline-random* respectively) (see Tables 4, 5 and 6), due to that these summaries were not written with a specific compression rate. The use of different segmentations of sentences with different compression rates affects the performance of all state-of-the-art methods and heuristics, therefore, it is necessary consider these constraints to generate and evaluate summaries.

The performance of *Baseline-first* was not affected significantly by the segmentation of sentences, because this heuristic employs the number of the first words to generate an extractive summary. Moreover, the performance of this heuristic it was better with respect to all state-of-the-art methods (see Table 7).

In Table 7 it is observed that *Baseline-first* heuristic outperforms all state-of-the-art methods involved in this comparison, therefore to generate summaries with better performance we propose the use of other methods (or combinations of them) to generate summaries to outperform this heuristic. Finally, we propose the generation and evaluation of summaries in TeMario considering the constraints mentioned above to generate a comparison with respect the upper bounds and lower bounds.

Acknowledgements. Work done under partial support of Mexican Government CONACyT Thematic Network program (Language Technologies Thematic Network project 295022). We also thank UAEMex for their support.

References

1. Pardo, T.A.S., Rino, L.H.M., Nunes, M.G.V.: NeuralSumm: Uma Abordagem Conexcionista para a Sumarização Automática de Textos. An. do IV Encontro Nac. Inteligência Artif., no. 1 (2003)
2. Orrú, T., Rosa, J.L.G., de Andrade Netto, M.L.: SABio: an automatic portuguese text summarizer through artificial neural networks in a more biologically plausible model. In: Vieira, R., et al. (eds.) PROPOR 2006. LNCS (LNAI), vol. 3960, pp. 11–20. Springer, Heidelberg (2006). https://doi.org/10.1007/11751984_2
3. Pardo, T.A.S., Rino, L.H.M.: DMSumm: review and assessment. In: Ranchhod, E., Mamede, N.J. (eds.) PorTAL 2002. LNCS (LNAI), vol. 2389, pp. 263–273. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45433-0_36
4. Cardoso, P.C.F.: Exploração de métodos de sumarização automática multidocumento com base em conhecimento semântico- discursivo. USP (2014)
5. Nunes, M.D.G.V., Aluisio, S.M., Pardo, T.A.S.: Um panorama do Núcleo Interinstitucional de Linguística Computacional às vésperas de sua maioridade. Linguamática **2**(2), 13–27 (2010)
6. Pardo, T.A.S., Rino, L.H.M., Nunes, M.D.G.V.: GistSumm: a summarization tool based on a new extractive method. In: Mamede, N.J., Trancoso, I., Baptista, J., das Graças Volpe Nunes, M. (eds.) PROPOR 2003. LNCS (LNAI), vol. 2721, pp. 210–218. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-45011-4_34
7. Margarido, P.R., et al.: Automatic summarization for text simplification. In: Companion Proceedings of the XIV Brazilian Symposium on Multimedia and the Web, pp. 310–315 (2008)
8. Pardo, T.A.S., Antiquiera, L., Nunes, M.D.G.V., Oliveira, O.N., Costa, L.D.F.: Using complex networks for language processing: the case of summary evaluation. In: International Conference on Communications, Circuits and Systems, pp. 2678–2682 (2006)
9. Antiquiera, L.: Desenvolvimento de técnicas baseadas em redes complexas para sumarização extrativa de textos. USP – São Carlos (2007)
10. Amancio, D.R., Nunes, M.G., Oliveira, O.N., Costa, L.D.F.: Extractive summarization using complex networks and syntactic dependency. Physica A: Stat. Mech. Appl. **391**(4), 1855–1864 (2012)
11. Mihalcea, R., Tarau, P.: A language independent algorithm for single and multiple document summarization. Department of Computer Science and Engineering, vol. 5, pp. 19–24 (2005)
12. Leite, D., Rino, L.: A genetic fuzzy automatic text summarizer. In: CSBC 2009. Inf. UFRGS, Brazil, vol. 2007, pp. 779–788 (2009)
13. Matías, G.A.: Generación Automática de Resúmenes Independientes del Lenguaje. Universidad Autónoma del Estado de México (2016)
14. Oliveira, M.A.D., Guelpeli, M.V.: BLMSumm – Métodos de Busca Local e Metaheurísticas na Sumarização de Textos. In: Proceedings of ENIA - VIII Encontro Nac. Inteligência Artif., vol. 1, no. 1, pp. 287–298 (2011)
15. Oliveira, M.A., Guelpeli, M.V.C.: The performance of BLMSumm: distinct languages with antagonistic domains and varied compressions. In: Information Science and Technology, ICIST 2012, pp. 609–614 (2012)
16. Pardo, T., Rino, L.: TeMário: Um Corpus para Sumarização Automática de Textos. NILC - ICMC-USP, São Carlos (2003)
17. Maziero, E.G., Volpe, G.: TeMário 2006 : Estendendo o Córpus TeMário (2007)

18. Aleixo, P., Pardo, T.A.S.: CSTNews: um Córpus de Textos Jornalísticos Anotados segundo a Teoria Discursiva Multidocumento CST (cross-document structure theory), *Structure*, pp. 1–12 (2008)
19. Rojas-Simón, J., Ledeneva, Y., García-Hernández, R.A.: Calculating the significance of automatic extractive text summarization using a genetic algorithm. *J. of Intell. Fuzzy Syst.* **35**(1), 293–304 (2018)
20. Rojas Simón, J., Ledeneva, Y., García Hernández, R.A.: Calculating the upper bounds for multi-document summarization using genetic algorithms. *Comput. y Sist.* **22**(1), 11–26 (2018)
21. Verma, R., Lee, D.: Extractive summarization: limits, compression, generalized model and heuristics, p. 19 (2017)
22. Sidorov, G.: *Non-linear construction of n-grams in computational linguistics*, 1st edn. Sociedad Mexicana de Inteligencia Artificial, México (2013)
23. Louis, A., Nenkova, A.: Automatically evaluating content selection in summarization without human models. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, no. August, pp. 306–314 (2009)
24. Torres-Moreno, J.M., Saggion, H., Cunha, I.D., SanJuan, E., Velázquez-Morales, P.: Summary evaluation with and without references. *Polibits Res. J. Comput. Sci. Comput. Eng. Appl.* **42**, 13–20 (2010)
25. Ceylan, H., Mihalcea, R., Özertem, U., Lloret, E., Palomar, M.: Quantifying the limits and success of extractive summarization systems across domains. In: *Human Language Technologies*, no. June, pp. 903–911 (2010)
26. Lin, C.-Y., Hovy, E.: The potential and limitations of automatic sentence extraction for summarization. In: *Proceedings of the HLT-NAACL 2003 on Text Summarization Workshop*, vol. 5, pp. 73–80 (2003)
27. Hong, K., Marcus, M., Nenkova, A.: System combination for multi-document summarization, pp. 107–117, September 2015
28. Wang, W.M., Li, Z., Wang, J.W., Zheng, Z.H.: How far we can go with extractive text summarization? Heuristic methods to obtain near upper bounds. *Expert Syst. Appl.* **90**, 439–463 (2017)
29. Ledeneva, Y., García-Hernández, R.A.: *Generación automática de resúmenes Retos, propuestas y experimentos* (2017)
30. Lin, C.: Rouge: a package for automatic evaluation of summaries. In: *Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*, no. 1, pp. 25–26 (2004)