Rhetorical Figuration as a Metric in Text Summarization

Mohammed Alliheedi[1] and Chrysanne Di Marco[1]

Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
malliheedi@uwaterloo.ca
cdimarco@uwaterloo.ca

Abstract. We show that surface-level markers of pragmatic intent can be used to recognize the important sentences in text and can thereby improve the performance of text summarization systems. In particular, we focus on using automated detection of rhetorical figures—characteristic syntactic patterns of persuasive language—to provide information for an additional metric to enhance the performance of the MEAD summarizer.

1 Introduction

Extractive summarization attempts to find the most "important" sentences in a document or document set for inclusion in a summary. One cue to importance which has hitherto received little attention is the writer's or speaker's use of rhetorical devices, especially those that involve repetition, to deliberately draw attention to important points. Over the past 20 years, research in automated text summarization has grown significantly in the field of Natural Language Processing. However, because the information available on the Web is ever-expanding, reading the sheer volume of information is a significant challenge. Although many automated text summarization systems have been proposed in the past twenty years, most of these systems have relied only on statistical approaches without incorporating much use of detailed linguistic knowledge. Our hypothesis is that rhetorical figuration, which involves the persuasive presentation of information in a text, generally at the sentence level, can provide such linguistic detail and can be detected computationally relatively easily. In particular, this research investigates the role of rhetorical figuration in determining the most important information to preserve in text summarization. Our experiments show that extractive summarization at the sentence level can be improved using the inclusion of metrics based on rhetorical figuration.

2 Related work

Researchers have proposed various approaches in the area of text summarization. Luhn states that the frequency of certain words in the document reflects its relevance [15]. Subsequent researchers (e.g., [10], [14], [12], and [24]) have proposed various frequency measures. The position of sentences is another feature that has been proposed in different systems [2], [3], [14], and [17]. Lexical chains are also used in several summarization systems [1], [23], [20], and [26] as a representation of the text to produce a summary. Several researchers in text summarization used Machine Learning methods such as Naive Bayes in [14], Neural Networks in [25], and Hidden Markov Models in [7]. Maximal marginal relevance (MMR) is proposed in [8] which combines measurements for both query relevance and the novelty of information for a specific topic. Different systems have been proposed as well, such as SUMMON [18], MEAD [22], multi-lingual summarization system [11], Hub/Authority Framework [27], and a progressive summarization system [6].

3 MEAD

MEAD is an extractive multi-document summarizer that was developed by [21]. We built our approach on top of the MEAD summarizer [21] for the following reasons. Most significantly, MEAD is open source, can be downloaded online, and is capable of summarizing multiple natural language texts. MEAD is also domain-independent so different types of corpora can be used. In addition, MEAD implements various summarization algorithms such as position-based, term frequency, inverse document frequency, and others[21]. Moreover, MEAD has the capability to modify, add, include, and exclude new algorithms (features).

3.1 MEAD Components

MEAD uses centroid-based summarization to identify important words in a cluster of documents based on the cosine overlap between the words and the vector of that cluster. In other words, a centroid represents a group of words that are statistically significant in the cluster [21].

Various features are used to score a sentence to determine whether it should be retained in the summary: centroid value, positional value, and length feature.

The centroid value C_i is the sum of the centroid values of all words $\mathrm{C}_{w,i}$ in a sentence S_i. For instance, assume that we have the centroid values of words (Obama = 44.68; Hillary = 32.50; Clinton = 29.40; morning = 12.25), and the sentence "President Obama meets with Secretary of State Hillary Rodham Clinton in the morning." In this case the total score of this sentence is 118.83 based on the following formula:

$$C_i = \sum_{w} C_{w,i} \tag{1}$$

The positional value for each sentence in a document is calculated as follows:

$$P_i = \frac{(n - i + 1)}{n} * C_{max} \quad (2)$$

While n, i are the total number of sentences in a cluster of documents and the order number of a sentence in the cluster respectively, C_{max} is the highest centroid value in the cluster. Note that the position value of the first sentence in the cluster is assigned C_{max} since the first sentence tends to possess more important information than other sentences.

The length feature is a cut-off feature that means every sentence with a length shorter than the threshold will receive a score of zero regardless of other feature scores. Thus, if a sentence has a length greater than the default threshold, which is nine words, it will receive a score that is the combination of other feature scores.

As no learning algorithm has been incorporated to predict the weight for each feature, we assign an equal weight for all features equal to one.

$\mathrm{Score}(\mathrm{S}_i) = \mathrm{W}_c\ \mathrm{C}_i + \mathrm{W}_p\ \mathrm{P}_i$; if $\mathrm{Length}(\mathrm{S}_i) > 9$

$\mathrm{Score}(\mathrm{S}_i) = 0$; if $\mathrm{Length}(\mathrm{S}_i) < 9$

While the input is the cluster of d documents with n sentences, the output is the number of sentences from the cluster with the highest score multiplied by compression rate r.

4 What is Rhetorical Figuration?

Rhetoric is the art of persuasive discourse, i.e., using language for the purpose of persuading, motivating, or changing the behaviour and attitudes of an audience. Corbett defines rhetorical figuration as "an artful deviation" from normal usage [9]. Rhetorical figures can be categorized into several types: schemes, tropes, and colours. Schemes are defined as "deviations from the ordinary arrangement of words" [4] and are generally characteristic syntactic patterns, such as repetition of the same word in successive sentences. We focus on figures of repetition because these are among the most commonly used and most effective rhetorically. Since figures of repetition often occur at the surface level, they can be detected and classified computationally with relative ease. Although there are over 300 rhetorical figures, we chose to focus on the following four figures of repetition in our experiments since they are among the most common and JANTOR demonstrated strong performance in detecting these figures:

Antimetabole

A repetition of words in adjacent sentences or clauses but in reverse grammatical order[4]. For instance, *But I am here tonight and I am your candidate because the most important work of my life is to complete the mission we started in 1980. How do we complete it?* [5].

Isocolon
A series of similar structured elements that have the same length, which is a kind of parallelism [4]. For example, in the sentence, *The ones who raise the family, pay the taxes, meet the mortgage* [5], the phrases *raise the family, pay the taxes*, and *meet the mortgage* have a parallel structure.

Epanalepsis
The repetition at the end of a line, phrase, or clause of a word or words that occurred at the beginning of the same line, phrase, or clause [4]. For example, in the sentence, *I am a man who sees life in terms of missions - missions defined and missions completed* [5], the word *missions* appears three times.

Polyptoton
The repetition of a word, but in a different form, using a cognate of a given word in close proximity [4]. For example, *I may not be the most eloquent, but I learned early that eloquence won't draw oil from the ground* [5], the words *eloquent* and *eloquence* are derived from the same root, *loqui*[1], but appear in different forms.

5 Methodology

5.1 Basis of the Approach

Our proposed multi-document summarization system is divided into two components: (1) an annotator of rhetorical figures; and (2) the basic multi-document summarizer itself. For the first component, we make use of JANTOR [13], a computational annotation tool for detecting and classifying rhetorical figures. We called our system JANTOR-MEAD.

5.2 Rhetorical Figure Value

The rhetorical figure value RF_j is the new feature added to our system JANTOR-MEAD along with the basic MEAD features above. RF_j is the sum of occurrences of a specific rhetorical figure in a document divided by the total number of all figures in the document. It is calculated as follows:

$$RF_j = \frac{n}{N} \tag{3}$$

where j is a type of rhetorical figure such as polyptoton or isocolon, and n is the total number of i occurrences in a document. N is the total number of all figures, which includes antimetabole, epanalepsis, isocolon, and polyptoton, that occurred in the document. For example, suppose we have a document containing 149 rhetorical figures (12 antimetabole; 2 epanalepsis; 100 isocolon; 35 polyptoton) and we would like to have the value of each figure. We apply the rhetorical figure value as follows:

$\mathrm{RF}_{antimetabole} = 12/149 = 0.081$

[1] http://www.merriam-webster.com/dictionary/eloquent

$\mathrm{RF}_{epanalepsis} = 2/149 = 0.013$
$\mathrm{RF}_{isocolon} = 100/149 = 0.671$
$\mathrm{RF}_{polyptoton} = 35/149 = 0.235$

Since there are four rhtorical figures have been considered in this study, the score of each sentence will include the total value of all rhtorical figures that occurred in that sentence as follows:

$$RF_i = \sum_{j=0}^{4} RF_j \tag{4}$$

Thus, the sentence score equation is modified to include rhetorical figure feature as follows:

$\mathrm{Score}(\mathrm{S}_i) = \mathrm{W}_c\ \mathrm{C}_i + \mathrm{W}_p\ \mathrm{P}_i + \mathrm{W}_{rf}\ \mathrm{RF}_i$

6 Data Set

We created a set of clusters for several U.S. presidents using presidential speeches from The American Presidency Project[1]. In our analysis we used four collections for the following presidents: Barack Obama, George W. Bush, Bill Clinton, and George Bush. Each collection contains three different types of speeches: State of the Union, campaign, and inaugural speeches, i.e., we have a cluster for each of the three types of speeches. Each cluster contains from 6 to 17 speeches, and these speeches contain from 90 to 700 sentences. These tested clusters were used for the evaluation process such that each cluster is summarized with ratio of 10% of actual texts.We choose presidential speeches because they tend to be rhetorical and persuasive texts. JANTOR has been tested previously by [13] using presidential speeches as data set and JANTOR was able to detect successfully many instances of different rhetorical figures include antimetabole, epanalepsis, polyptoton and isocolon.

7 Experiments with Rhetorical Figures

The following sections show the results of our experiments done for various rhetorical figures. Our tests involved four types of rhetorical figures: antimetabole, epanalepsis, isocolon, and polyptoton. JANTOR was used to detect these figures in all clusters of presidential speeches. However, the quality of performance varied for the different figures.

[1] http://www.presidency.ucsb.edu

7.1 Antimetabole

Although antimetabole is defined as the repetition of words in adjacent clauses or phrases in reverse grammatical order, this definition does not specify whether or not different word forms and word types should be considered[13]. Gawryjolek [13] addressed this issue by considering all types of words, such as determiners and prepositions, and by looking for only the repetition of the exact words. Although it is obvious that including prepositions and determiners will definitely produce more antimetaboles, which are not salient in a text, it is important to include them in order not to decrease the recall value. Therefore, JANTOR identifies correct antimetaboles, but it may also identifies antimetaboles that are not rhetorically genuine.

7.2 Epanalepsis

The simple definition of epanalepsis—repetition of the same word or phrase in the beginning and the end of a sentence, clause, or line—allowed JANTOR to correctly detect most of the epanalepsis instances in our corpus.

7.3 Isocolon

Isocolon requires a set of similarly structured phrases or sentences that have the same length. JANTOR was able to successfully detect isocolon instances in our corpus. However, JANTOR sometimes encountered the same instance repeated and counted it as a new isocolon occurrence. This happens because JANTOR is not able to distinguish between a sentence and the same sentence with punctuation such as dash and semicolon. For example, "for all of our economic dominance" and "for all of our economic dominance -". In this example, JANTOR detects the same instance of isocolon twice, and is not able to identify them as one instance.

7.4 Polyptoton

JANTOR was able to correctly detect several instances of polyptoton in our test corpus. From our observation of the detection of polyptoton, however, we found that JANTOR has one problem with detecting this figure: it fails to detect some words that are in different stem forms.

8 Evaluation

Evaluation of summarization requires human judgment in several quality metrics such as coherence, grammaticality, and content. Although the quality metrics are significant for the evaluation of summarization systems, these metrics would require an enormous amount of human effort which is very expensive and difficult to manage [16]. Thus, over the past years, many researchers in the area of text summarization have attempted to develop an automatic evaluation metric for summarization tasks.

8.1 ROUGE

We used the intrinsic evaluation system ROUGE [16] to evaluate our summarization system. ROUGE metrics include ROUGE-N, ROUGE-L, ROUGE-S, ROUGE-SU. ROUGE-N is an n-gram recall measure that calculates the number of n-grams in common between the human summaries and system summaries. ROUGE-L is a measure based on Longest Common Subsequence (LCS) between reference summaries and model summaries. It uses the concept that the longer the subsequence of words that are in common between summaries, the greater the similarity will be. ROUGE-S is a skip-bigram co-occurrence statistical measure that counts any pair of words in their sentence order in a model summary that overlaps with a reference summary. ROUGE-S allows arbitrary gaps between bigrams. ROUGE-SU is an extension of ROUGE-S with a unigram as a counting unit.

9 Results

As we discussed in the data set section, the presidential speeches are annotated by the JANTOR tool for different rhetorical figures that include antimetabole, epanalepsis, isocolon, and polyptoton. We have two types of summaries for our data set: MEAD summaries, which are the baseline summaries, and our rhetorically based summaries. Two human assessors provided manual summaries for all data sets with a ratio of 10% of actual texts. These assessors are two graduate students from the English Language and Literature Department at the University of Waterloo. One is a PhD student and the other is a Master's student. The task of assessors was to review the presidential speeches and identify the most important sentences in each speech and manually extract 10% of the actual text. The assessors were able to create summaries for all presidential speeches in the data set. Thus, we have a total of 33 human-based summaries. Both summaries produced by the MEAD baseline and JANTOR-MEAD are compared with the human-based summaries, which are the gold standard, using the evaluation system ROUGE. We denote JMEAD in the tables for JANTOR-MEAD and MEAD for the baseline summarizer.

ROUGE	**MEAD**	**JMEAD**
R-1	0.690	0.715
R-L	0.679	0.702
R-S4	0.483	0.515
R-SU4	0.483	0.515

Table 1. Results of ROUGE metrics for both MEAD and JANTOR-MEAD.

Table 1 shows the results of ROUGE metrics for both the baseline, which is the MEAD summarization system, and our rhetorically enhanced JANTOR-MEAD summarization system using all rhetorical figures includes antimetabole,

ROUGE	MEAD	JMEAD
R-1	0.714	0.743
R-L	0.704	0.732
R-S4	0.521	0.560
R-SU4	0.521	0.560

Table 2. Results of ROUGE metrics for both MEAD and JANTOR-MEAD with Porter Stemmer.

epanalepsis, isocolon, and polyptoton. Table 1 demonstrates that JANTOR-MEAD summarization system slightly outperforms the MEAD system in every ROUGE metric. The correlation in R-SU4 and R-S4 between human-based summaries and JANTOR-MEAD is higher than MEAD which indicates that rhetorical figures tend to preserve salient information that can contribute to improve summaries.Table 2 show how the results changed when the Porter stemmer algorithm applied in ROUGE metrics. Once Porter stemmer is incorporated, it penalizes the polyptoton score and reduce it to minimal since polyptoton concerns the occurrences of words in different forms or part of speech and the stemmer counts words with same root as same words. Table 2, however, shows that JANTOR-MEAD still slightly outperforms MEAD.

In summary, we found that our system using all figures include antimetabole, epanalepsis, isocolon, and polytoton provides better results than MEAD in every ROUGE method. Surprisingly, Porter stemmer did imporve the performance of JANTOR-MEAD in Table 2. Thus, JANTOR-MEAD is better than MEAD in every ROUGE evalutation metric. Since only two human assessors were involved in this experiment (due to funding constraints) and there was no consideration of inter-agreement between assessors, we hypothesize that if there were more assessors participating and each assessor evaluates all documents in the data set, the results of JANTOR-MEAD would improve significantly.

10 Conclusion

We propose the use of rhetorical devices in the form of figuration patterns as additional metrics for improving the performance of text summarization. This hypothesis is based on the observation that rhetoric is persuasive discourse and so may be expected to highlight significant sentences in a text that should be preserved in a summary. Our experiment showed that rhetorical figures can provide additional information that in some cases allowed our summarizer to outperform the MEAD system. These results are promising and indicate that rhetoric, which provides additional linguistic information, may be a useful tool in natural language processing systems.

11 Future Work

Further investigation is planned to expand the work of JANTOR-MEAD and include new rhetorical figures such as Anaphora and Ploche as well as increasing the number of assessors to provide more accurate concrete gold standard summaries. We also plan to test our approach on other corpora. It is also feasible to test other corpora such as Document Understanding Conference (DUC) data sets since they are standard data sets and have been widely used by many summarization systems.

12 Acknowledgements

We would like to acknowledge and thank the first author's sponsor Al Baha University, Albaha, Saudi Arabia for funding this research. We would like also to thank Graeme Hirst for his valuable comments on earlier versions of the manuscript.

References

[1] Regina Barzilay and Michael Elhadad. 1997. Using lexical chains for text summarization. *In Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization*, 10–17.

[2] P. B. Baxendale. 1958. Machine-made index for technical literature: An experiment. *IBM Journal of Research and Development*, 2(4): 354–361.

[3] Ronald Brandow and Karl Mitze and Lisa F. Rau 1995. Automatic condensation of electronic publications by sentence selection. *Information Processing and Management*, 31(5), 675–685.

[4] Gideon Burton. 2007. *Silva rhetoricae: The forest of rhetoric.*

[5] George Bush. 1988. Address Accepting the Presidential Nomination at the Republican National Convention. *Online by Gerhard Peters and John T. Woolley, The American Presidency Project* New Orleans, Louisiana.

[6] Praveen Bysani. 2010. Detecting novelty in the context of progressive summarization. *Proceedings of the NAACL HLT 2010 Student Research Workshop*, 13–18. Los Angeles, California.

[7] John Conroy, and Dianne P. O'Leary. 2001. Text summarization via Hidden Markov Models. *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 406–407. New Orleans, Louisiana.

[8] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 335–336. Los Angeles, California.

[9] Edward P. J. Corbett. 1990. *Classical rhetoric for the modern student, 3rd edition.* Oxford University Press, New York, USA.

[10] H. P. Edmundson. 1969. New methods in automatic extracting. *Journal of the ACM*, 16(2), 264–285. New York, NY, USA.

[11] David Kirk Evans, Kathleen McKeown, and Judith L. Klavans. 2005. Similarity-based multilingual multidocument summarization. *IEEE Transactions on Information Theory*, 49.

[12] Eduard Hovy, and Chin-Yew Lin. 1999. Automated text summarization in SUMMARIST. *Advances in automatic text summarization*, 81–94. MIT Press, Cambridge, MA, USA.

[13] Jakub J. Gawryjolek. 2009. Automated annotation and visualization of rhetorical figures. Master's Thesis, University of Waterloo.

[14] Julian Kupiec, Jan Pedersen, and Francine Chen. 1995. A trainable document summarizer. *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 68–73. Seattle, Washington.

[15] H. P. Luhn. 1958. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2): 159–165.

[16] Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. *Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*, 16.

[17] Chin-Yew Lin, and Eduard Hovy. 1997. Identifying topics by position. *Proceedings of the Fifth Conference on Applied Natural Language Processing*, 283–290. Washington, DC, USA.

[18] Kathleen McKeown, and Dragomir R. Radev. 1995. Generating summaries of multiple news articles. *Proceedings SIGIR '95 Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, 74–82. Seattle, Washington, United States.

[19] Edward F. McQuarrie and David Glen Mick. 1996. Figures of rhetoric in advertising language. *Journal of Consumer Research*, 22(4): 424–438.

[20] Thade Nahnsen, Ozlem Uzuner, and Boris Katz. 2005. Lexical chains and sliding locality windows in content-based text similarity detection. *Companion Volume to the Proceedings of Conference Including Posters/Demos and Tutorial Abstracts*, 150–154.

[21] Dragomir Radev, Timothy Allison, Sasha Blair-Goldensohn, John Blitzer, Arda Celebi, Stanko Dimitrov, Elliott Drabek, Ali Hakim, Wai Lam, and Danyu Liu. 2004. MEAD-a platform for multidocument multilingual text summarization. *Proceedings of the International Conference on Language Resources and Evaluation, Vol. 2004* , 1–4.

[22] Dragomir R. Radev, Hongyan Jing, and Malgorzata Budzikowska. 2000. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. *Proceedings of the 2000 NAACL-ANLP Workshop on Automatic summarization - Volume 4*. Seattle, Washington.

[23] H. Gregory Silber, and Kathleen F. McCoy. 2002. Efficiently computed lexical chains as an intermediate representation for automatic text summarization. *Computational Linguistics - Summarization archive Volume 28 Issue 4*, 487–496. MIT Press Cambridge, MA, USA.

[24] Tomek Strzalkowski, Jin Wang, and Bowden Wise. 1998. A robust practical text summarization. *Proceedings of the AAAI Symposium on Intelligent Text Summarization*, 26–33.

[25] Krysta Svore, Lucy Vanderwende, and Chris Burges. 2007. Enhancing single-document summarization by combining RankNet and third-party sources. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 448–457.

[26] Rakesh Verma, Ping Chen, and Wei Lu. 2007. A semantic free-text summarization using ontology knowledge. *Proceedings of the Document Understanding Conference.*

[27] Junlin Zhang, Le Sun, and Quan Zhou. 2005. A cue-based hub-authority approach for multi-document text summarization. *Natural Language Processing and Knowledge Engineering. Proceedings of IEEE International Conference*, 642-645.