# Exploring sentiments in summarization: SentiTextRank, an Emotional Variant of TextRank

Md. Murad Hossain[1,*,†], Luca Anselma[1,†] and Alessandro Mazzei[1,†]

[1] Department of Computer Science, University of Turin, Corso Svizzera 185, 10149 Torino, Italy

## Abstract

**English.** A summary that aims at preserving the emotions of the original text can be interesting in certain application scenarios, such as in the generation of metareviews, both in academic and commercial domains. TextRank is a well-studied algorithm for automatic extractive summarization. This work introduces SentiTextRank, an emotional variant of TextRank, to enhance the extractive technique for both single-document and multi-document summarization. SentiTextRank incorporates emotions into the summarization process by classifying sentences into the eight emotional categories used in SenticNet. The preliminary evaluation of SentiTextRank yields encouraging results. In particular, our method generates informative summaries composed of sentences that preserve the emotional content of the original document.

**Italian.** Un riassunto che miri a preservare le emozioni del testo originale può essere interessante in alcuni scenari applicativi, come ad esempio nella generazione di meta-recensioni sian nel dominio accademico che in quello commerciale. TextRank è un algoritmo per il riassunto automatico estrattivo molto studiato. Questo lavoro introduce SentiTextRank, una variante emozionale di TextRank, per potenziare la tecnica estrattiva sia per il riassunto di singoli documenti che per il riassunto di documenti multipli: SentiTextRank integra le emozioni nel processo di sintesi, classificando le frasi nelle otto categorie emotive utilizzate in SenticNet. La valutazione preliminare di SentiTextRank produce dei risultati incoraggianti. In particolare, il nostro metodo produce dei riassunti informativi formati da frasi che rispettano il contenuto emozionale del documento originale.

## Keywords

Extractive summarization, SentiTextRank, emotional variant, single and multi-document Summary, emotional content.

## 1. Introduction

Summarization is the process of reducing a larger body of information into a concise and coherent summary that captures the essential points and main ideas. Extractive summarization involves selecting and combining sentences or phrases directly from the source text to form the summary [1] and plays an important role in condensing news articles into concise summaries, allowing readers to quickly grasp the key information. Traditional extractive methods primarily rely on lexical word distance to select important sentences for summarization. In many cases the emotional aspects found in the documents are not considered in summarization, and this can affect how readers engage with and understand the information.

Sentiment analysis is the extraction of subjective information from text, encompassing emotions and opinions, and the classification based on the expressed emotions, such as happiness, sadness, anger, fear, or surprise, to capture the overall emotional sentiment [2] and [3]. Sentiment analysis had a huge impact on many applications of NLP in the last years, but there is still space for understanding the details of its implementations [4].

The current study employs SenticNet [3], a multidisciplinary approach to opinion mining that lies at the intersection of affective and common-sense computing. This approach integrates elements from semiotics, psychology, linguistics, and machine learning. Unlike statistical sentiment analysis, Sentic computing focuses on preserving the semantic representation of natural language concepts and sentence structure. The foundation of SenticNet is the Hourglass of Emotions, an emotion categorization model designed to accurately express the affective information present in natural language text.

To the best of our knowledge, sentiment generation is an understudied argument in the field of automatic summarization. Despite the advancements in text summarization techniques, there is a gap in research when it comes to considering emotions in the process. However, we believe that there are a number of applications in which the emotions in a summary should correspond to the emotions of the original document(s). For instance, this is the case of meta-reviews in conference management or the summarization of product reviews in the case of e-commerce. For these application domains, it is

*CLiC-it 2023: 9th Italian Conference on Computational Linguistics, Nov 30 — Dec 02, 2023, Venice, Italy*

* Corresponding author.
† These authors contributed equally.
mdmurad.hossain@unito.it (Md. M. Hossain); luca.anselma@unito.it (L. Anselma); alessandro.mazzei@unito.it (A. Mazzei)
https://shorturl.at/bEGZ0 (Md. M. Hossain); http://www.di.unito.it/~anselma/ (L. Anselma); http://www.di.unito.it/~mazzei/ (A. Mazzei)
0000-0002-8224-3246 (Md. M. Hossain); 0000-0003-2292-6480 (L. Anselma); 0000-0003-3072-0108 (A. Mazzei)





CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

important to consider emotions to create summaries that truly reflect the essence of the source texts.

In this paper, by incorporating sentiment scoring between sentences, we generate summaries that capture the emotional tone and impact of the original text. We believe that exploring this aspect further would lead to more comprehensive and effective text summarization methods.

This paper has two main goals. First, we define a new algorithm called *SentiTextRank*, which is an *emotional* variant of TextRank [5]. Second, we provide an initial evaluation of SentiTextRank by considering two automatic metrics based on content distance.

Note that modern LLMs showed some abilities in summarizing texts by using a specific style, with some limitations in producing a summary that is truly extractive. Moreover, LLMs showed also a big impact from the point of view of the required computational resources. We believe that the work presented in this paper, that requires just few hours to conduct all the experiments, can be seen as a cheap (in many senses) alternative to the use of modern expensive (in many senses) LLMs[1].

The paper is structured as follows. In Section 2, we define the new SentiTextRank algorithm, in Section 3 we report the result of a first experimental evaluation of the SentiTextRank algorithm and the Section 4 ends the paper pointing out to work in progress.

## 2. SentiTextRank: a variant of TextRank accounting for emotions

TextRank is a popular algorithm for extractive summarization which constructs a graph of sentences or words from a text and assigns scores to each node based on their importance in the graph structure. Finally, it ranks the nodes and selects the top-ranked sentences or words as the summary [5]. The TextRank algorithm is based on the PageRank algorithm, where the sentences of the documents play the role of web pages, and a similarity score plays the role of hyperlink connectivity. Our approach enhances traditional TextRank by incorporating emotions. In particular, we categorize sentences of the original source(s) on the basis of emotions using SenticNet [6]. On the basis of this classification, we obtain a number of distinct emotion sets of sentences. The main idea is to build one single final summary by merging in a selective way the results of TextRank on each one of these emotions sets.

So, the proposed *SentiTextRank* algorithm generates extractive summaries with emphasis on emotion categories through the following steps:

[1]We thank an anonymous reviewer for pointing out this point.

**SentiTextRank: Input**=Source, **Output**=$Sum_F$

1. Set compression ratio parameter $C$ between the source(s) and the final summary $Sum_F$.
2. Classify sentences of the source(s) into different SenticNet emotion categories $CAT_{em}$ with $em \in$ {joy, admiration, surprise, fear, disgust, anger, sadness, interest}.
3. Generate a summary $Sum_{em}$ for each emotion category $CAT_{em}$ by using TextRank.
4. Build $Sum_F$ by picking a number of sentences proportional to $C$ from each $Sum_{em}$ maintaining the original sentence order of the source document.

## 3. Experimental Result and Discussion

In this section, we present the experimental results of single-document summarization using two datasets, the CNN/Daily Mail dataset (CNN) and the DUC2001 single document dataset (D01), as well as the results of multi-document summarization using two datasets, the DUC2001 multi-document dataset (MD01) and the DUC2004 multi-document dataset (MD04).

The DUC 2001 single document and DUC 2001 Multi-document datasets were collected from the website[2] and consist of news datasets. For our experiments with single documents, we utilized a sample data set of 54 documents. The DUC 2004 multi-document summarization dataset[3] includes 50 items with multiple files and four reference files per item, from which we utilized the first reference for each item. Additionally, we used the CNN/Daily mail dataset[4], where we considered the “highlights” column as the reference summary. Our experiments were conducted on the first 100 rows of text in the CNN/Daily mail dataset. Since the datasets provide just abstractive gold summaries, in order to provide a fair comparison we have converted the abstractive summaries into extractive summaries. This procedure has been proposed in [7, 8]. An extractive reference summary should yield the highest Rouge score when compared to the gold abstractive summary. As finding the globally optimal subset of sentences that maximize the Rouge score is computationally intractable, we adopt a greedy approach: we iteratively add one sentence at a time to the summary, ensuring that the Rouge score of the current set of selected sentences is maximized in relation to the entire gold summary. We repeat this process until there are no more candidate sentences that could enhance the Rouge score when added

[2]https://duc.nist.gov/data.html
[3]https://rb.gy/gp1gbt
[4]https://rb.gy/v4u2g

| | |
|---|---|
| **Original Text** | Ever noticed how plane seats appear to be getting smaller and smaller? With increasing numbers of people taking to the skies, some experts are questioning if having such packed out planes is putting passengers at risk. |
| **Gold Abstractive Summary** | Experts question if packed out planes are putting passengers at risk. U.S consumer advisory group says minimum space must be stipulated. |
| **Reference Extractive Summary** | Ever noticed how plane seats appear to be getting smaller and smaller?. This week, a U.S consumer advisory group set up by the Department of Transportation said at a public hearing that while the government is happy to set standards for animals flying on planes, it doesn't stipulate a minimum amount of space for humans. |
| **Lead** | Ever noticed how plane seats appear to be getting smaller and smaller? With increasing numbers of people taking to the skies, some experts are questioning if having such packed out planes is putting passengers at risk. |
| **TR** | They say that the shrinking space on aeroplanes is not only uncomfortable - it's putting our health and safety in danger. This week, a U.S consumer advisory group set up by the Department of Transportation said at a public hearing that while the government is happy to set standards for animals flying on planes, it doesn't stipulate a minimum amount of space for humans. |
| **STR** | They say that the shrinking space on aeroplanes is not only uncomfortable - it's putting our health and safety in danger. 'It is time that the DOT and FAA take a stand for humane treatment of passengers. |

**Table 1**
An excerpt from the Original Text from the CNN dataset, the existing reference summary (Gold Abstractive Summary), the generated reference summary (Reference Extractive Summary), the lead baseline (Lead), the summary generated by TextRank (TR), and the summary generated by SentiTextRank (STR).

to the current summary set. The subset of sentences that we have at this point is then considered the extractive reference summary for the evaluation. In Table 1 we report an example of summaries generated with the different methods.

Table 1 reports an excerpt from the CNN dataset (Original Text) and the corresponding reference summary (Gold Abstractive Summary). Moreover, Table 1 contains the corresponding generated reference extractive summary (Reference Extractive Summary), the prefix baseline (Lead), the text generated with the TextRank baseline (TR) and, finally, the SentiTextRank generated summary (STR).

| Dataset | Algorithm | RL-F1 | BERT-F1 |
|---|---|---|---|
| D01 | Lead | **0.600** | **0.729** |
| | TR | *0.382* | *0.649* |
| | STR | 0.366 | 0.605 |
| CNN | Lead | **0.711** | **0.794** |
| | TR | 0.345 | *0.642* |
| | STR | *0.372* | 0.608 |
| MD01 | Lead | **0.802** | **0.851** |
| | TR | 0.061 | *0.560* |
| | STR | *0.163* | 0.505 |
| MD04 | Lead | **0.511** | **0.683** |
| | TR | 0.123 | *0.575* |
| | STR | *0.227* | 0.542 |

**Table 2**
The results of summarization experiments. Lead = Lead Baseline, TR = TextRank, STR = SentiTextRank.

Table 2 presents the experimental results of different summarization methods, namely the Baseline (Lead), TextRank (TR), and our proposed method, SentiTextRank (STR) evaluated on single-document datasets DUC-2001 (D01) and CNN, and the multi-document datasets DUC-2001 (MD01) and DUC-2004 (MD04). As a baseline, we selected the leading sentences from the original documents based on the compression ratio. We evaluated the summaries using two measures: Rouge-L F1 (RL-F1) and BERT F1. ROUGE (Recall Oriented Understudy for Gisting Evaluation) is frequently used to assess how well summarization techniques perform. Rouge-L computes ROUGE for the longest sequence of n-grams [9]. BERT F1 score is a metric commonly used in text classification tasks. It measures the token-level similarity between the generated summary and the reference summary, considering both precision and recall [10].

The results consistently indicate that the Lead method outperforms the other methods across the datasets, showcasing its superiority in generating high-quality summaries. Specifically, Lead achieves the highest scores in Rouge-L F1 and BERT-F1 for D01, CNN, MD01, and MD04. The TR and STR methods exhibit moderate performance in specific evaluation metrics.

Note that the better performance of the Lead method can be attributed to the fact that all the experiments were conducted using news datasets; indeed this result is consistent with the results reported in the literature, where a baseline composed of the leading sentences frequently outperforms extractive and abstractive models on news datasets [11]. However, we think that the comparison between original TR and STR shows encouraging results.

Indeed, the fact that using emotions does not degrade the performance with regard to TextRank shows that we can produce a summary that represents the content as well as the emotions of the source documents. In order to experimentally prove this intuition, we need to formalize an *emotional distance* between summaries and source documents. We plan to develop this point in the future by both (1) using LLM and (2) considering human evaluation.

Further research is necessary to evaluate the performance of our proposed STR method on another domain dataset to provide a comprehensive understanding of its effectiveness.

## 4. Conclusion and Future Work

This paper introduces the SentiTextRank algorithm, which integrates emotions into the extractive summarization process to create more informative and emotionally rich summaries. The experimental results are encouraging with respect the effectiveness of SentiTextRank in capturing factual information.

The ongoing work on SentiTextRank is following different directions.

First, we want to design a new version of the algorithm that will not be based on the classification of a sentence in one single prevalent emotion. The idea that we want to develop is to define one single measure that combines both *content* and *emotion* similarities. By using this combined measure, we can apply the original TextRank algorithm on the entire set of sentences from the source(s) and obtain one single ranking structure accounting for both content and emotion.

Second, we want to conduct more extensive experiments also on datasets from different domains. In particular, we are considering medical applications since the affective component of medical information can represent a relevant biopsychosocial feature [12].

Third, we are aware that automatic metrics not always measure a real *quality* of the summarized text with respect to human judgment [13, 14]. So, we plan in future to conduct human-based evaluation too.

## References


[1] A. Nenkova, K. McKeown, et al., Automatic summarization, Foundations and Trends® in Information Retrieval 5 (2011) 103–233.

[2] M. Wankhade, A. C. S. Rao, C. Kulkarni, A survey on sentiment analysis methods, applications, and challenges, Artificial Intelligence Review 55 (2022) 5731–5780.

[3] E. Cambria, D. Das, S. Bandyopadhyay, A. Feraco, Affective computing and sentiment analysis, A practical guide to sentiment analysis (2017) 1–10.

[4] K. Kenyon-Dean, E. Ahmed, S. Fujimoto, J. Georges-Filteau, C. Glasz, B. Kaur, A. Lalande, S. Bhanderi, R. Belfer, N. Kanagasabai, et al., Sentiment analysis: It's complicated!, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2018, pp. 1886–1895.

[5] R. Mihalcea, P. Tarau, Textrank: Bringing order into text, in: Proceedings of the 2004 conference on empirical methods in natural language processing, 2004, pp. 404–411.

[6] E. Cambria, Y. Li, F. Z. Xing, S. Poria, K. Kwok, Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis, in: Proceedings of the 29th ACM international conference on information & knowledge management, 2020, pp. 105–114.

[7] R. Nallapati, F. Zhai, B. Zhou, Summarunner: A recurrent neural network based sequence model for extractive summarization of documents, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI'17, AAAI Press, 2017, p. 3075–3081.

[8] M. Isonuma, T. Fujino, J. Mori, Y. Matsuo, I. Sakata, Extractive summarization using multi-task learning with document classification, in: Proceedings of the 2017 Conference on empirical methods in natural language processing, 2017, pp. 2101–2110.

[9] C.-Y. Lin, Rouge: A package for automatic evaluation of summaries, in: Text summarization branches out, 2004, pp. 74–81.

[10] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[11] Y. Liu, M. Lapata, Text summarization with pretrained encoders, arXiv preprint arXiv:1908.08345 (2019).

[12] D. Caldo, S. Bologna, L. Conte, M. S. Amin, L. Anselma, V. Basile, M. M. Hossain, A. Mazzei, P. Heritier, R. Ferracini, et al., Machine learning algorithms distinguish discrete digital emotional fingerprints for web pages related to back pain, Scientific Reports 13 (2023) 4654.

[13] J. Novikova, O. Dušek, A. Cercas Curry, V. Rieser, Why we need new evaluation metrics for nlg, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 2241–2252. doi:10.18653/v1/D17-1238.

[14] F. Moramarco, A. Papadopoulos Korfiatis, M. Perera, D. Juric, J. Flann, E. Reiter, A. Belz, A. Savkov, Human evaluation and correlation with automatic metrics in consultation note generation, in: Proceedings of the 60th Annual Meeting of the Asso-

ciation for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 5739–5754. URL: https://aclanthology.org/2022.acl-long.394. doi:10.18653/v1/2022.acl-long.394.