Topic-aware video summarization technique for product reviews exploiting the BERTopic and BART models

Yu-Jin Ha[1], Gun-Woo Kim[2*]

[1] *Department of AI Convergence Engineering, Gyeongsang National University, Jinju, Korea*
[2] *School of Computer Science/Department of AI Convergence Engineering, Gyeongsang National University, Jinju, Korea*

Abstract

Recently, there has been a growing trend of consumers seeking product information from video platforms such as YouTube. However, when viewing multiple review videos about the same product, viewers often encounter redundant information, resulting in wasted time. To address these issues, we use BERTopic to eliminate repetitive video content and address the problem of missing subtopics, which has been a limitation of traditional video summarization methods. Subsequently, the topic-aware video contents are summarized using the BART model. The ROUGE metric was used to evaluate the model proposed in this paper, and the experimental results showed improved results compared to previous research.

Keywords

Multi video summarization, Product review summarization, BERTopic, BART

1. Introduction

As the use of mobile internet media has increased since the COVID-19 pandemic, the proportion of people who prefer online video platforms as their main source of news and current affairs information has increased. YouTube is one of the most popular online video platforms, and the percentage of users who use it to get news and current affairs information has increased from before the pandemic. [1], [2] Videos provided by video platforms can be as short as a few seconds and as long as several hours or more. It requires time and effort to watch all these long videos. To reduce this waste of time, video summarization is necessary. A video summary is a short summary of the important content of the entire video or what the user requested. [3]

Video summarization can be based on audio, vision, and textual data (i.e., subtitle) [4], [5]. Most traditional video summarization approaches use only low-level visual images as data. [6], [7], [8], [9], [10] Traditional video summarization lacks the understanding of semantic meaning and relationships in videos. This means that they need to explore not only visual images, but also audio and textual information. Focusing on textual information can outperform traditional video summarization in terms of time and space for specific domains such as education (i.e., lectures, tutorials, and information), knowledge (i.e., news and documentaries), sports, entertainment industry (i.e., movies), medicine, and product reviews.[3]

There are several methods for summarizing textual information in video summaries. Summarization of textual information using video images is possible through image capturing [11]. Audio data is summarized by text-based methods using speech-to-text techniques. Video subtitle summaries are summarized using text summarization algorithms using text data. The resulting summaries exist as short videos or text data. [4], [5]

ISE 2023: 2nd International Workshop on Intelligent Software Engineering, December 4, 2023, Seoul *Corresponding author.

nrt5077@gnu.ac.kr (Y. J. Ha); gunwoo.kim@gnu.ac.kr (G. W. Kim)

0009-0004-8963-1913 (Y. J. Ha); 0000-0001-5643-4797 (G. W. Kim);

The increase in the use of the YouTube platform as a main source of information mentioned earlier is the same for product reviews. Product reviews are a way for buyers to learn about the pros and cons of a product and its features before making a purchase. If a product review is long, it takes a lot of time to read through the content of the entire review. For this reason, there is a need for product review summaries that minimize the length of product reviews while preserving their content. [12], [13] With the increase of product review videos on the YouTube platform, there has been an increase in the number of multiple product review videos for a single product. Multi-video product review summaries provide product reviews from different perspectives for buyers who want to purchase a product and help sellers or companies to improve their quality of products and services [14].

In this paper, we aim to identify the semantic meanings and relationships of product review videos through subtitle summaries and provide product reviews from different perspectives through multi-video product review summaries.

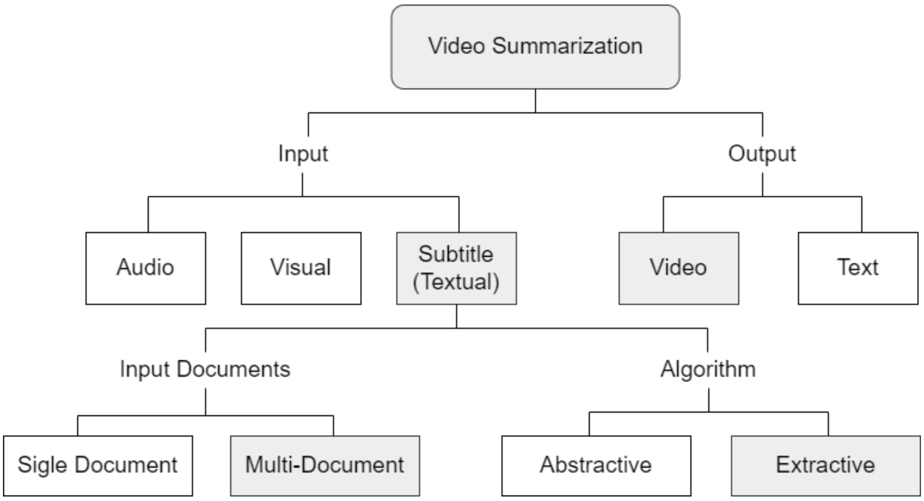


Figure 1: The scope of the video summarization

2. Related Works

Extractive summaries are summaries that attempt to extract the entire summarized document into sentences, phrases, and words from the source document. Generative summaries are summaries in which the summarized document generates a summary using words, sentences, and phrases that are not present in the original document. [15]In this paper, we use extractive summarization to preserve the review video creator's intent by using sentences extracted from the original document to increase the reliability of the review.

Topic modeling is a statistical model that identifies the topics of a set of documents in the field of machine learning and natural language processing. It analyzes large amounts of textual data to help identify the embedded semantic structure of the text and summarize its key content.[16]In this paper, we use topic modeling to identify various embedded topics and subtopics.

Ansamma et al. (2017) [17]maximized relevance and minimized redundancy by representing them as word vectors, and multi-document extractive summarization using Latent Semantic Analysis (LSA) and Non-Negative Matrix Factorization (NMF) as objective functions. Alrumiah et al. (2022) [18]is a single summary of lecture video subtitles using LDA. By generating a keyword list using LDA and extracting sentences from the original document that contain at least one word from the keyword list, they improved the summarization performance in terms of length and quality compared to previous studies. Miller et al. (2019) [19]summarized a single lecture video using BERT. This is the first study in which a large language model (LLM) is used for video summarization, and the LLM model, BERT, is used to embed the input sentences. Then, the n most centroid sentences were selected using k-means to summarize the lecture video subtitles. The performance evaluation was compared with TextRank, but since there was not Golden summary, it was not evaluated using evaluation metrics.

3. Proposed Method

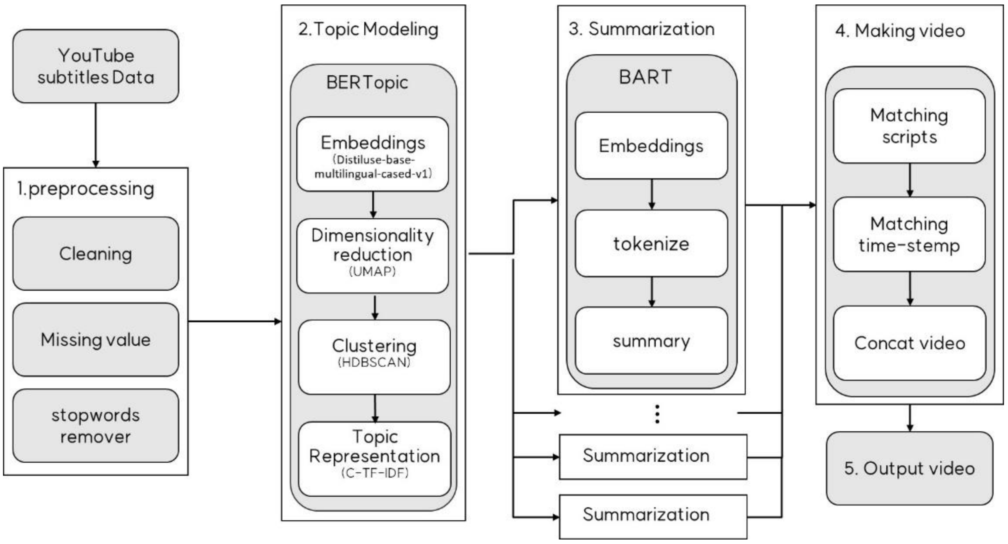


Figure 2: product multi-video summary architecture

3.1. Preprocessing

In the data preprocessing stage, we remove emoticons, onomatopoeia, and onomatopoeia, which are unnecessary information in this summary. For onomatopoeia and onomatopoeia containing more than one consonant and vowel, we removed all but one word. If there were missing values and subtitles other than Korean, the columns were deleted. In addition, we removed stop words, which are information that does not add meaningful information to the sentence. We used the list of Korean stop words provided by NLTK.

3.2. Topic Modeling

BERTopic is a popular BERT-based algorithm for topic modeling. It utilizes Transformer and c-TF-IDF to generate dense clusters while maintaining important words in the topic. The process of BERTopic includes three steps: Document Embedding, Document Clustering, and Topic representation. Each step is designed to be independent and can be customized according to the purpose [20].

In this paper, when comparing topic modeling algorithms, we found that BERTopic shows equally good results in Topic Coherence and Topic Diversity. This means that it generates topics with diversity while maintaining the coherence of the document's topics. In addition, BERTopic has the advantage of requiring relatively little time even as the size of the vocabulary increases. For these reasons, we chose BERTopic for this paper.

3.3. Summarization

BART is a denoising autoencoder based on the Transformer architecture.[21]The corrupted text is input to the encoder and the text representation learned by the encoder is sent to the decoder, which recovers the original undamaged text. BART is a model that can be trained by combining natural language understanding and generation, and performs well on summarization tasks.

3.4. Video Making

This is the process of converting summarized text to video. The process extracts the number of the video where the text is located and the timeline of the video where the text is located. The video is then cut to include the summarized content. The process is iterated over the number of summarized texts, and then the summarized videos are joined together into a single video. Figure 3 is an image that captures the summarized video created through this process.

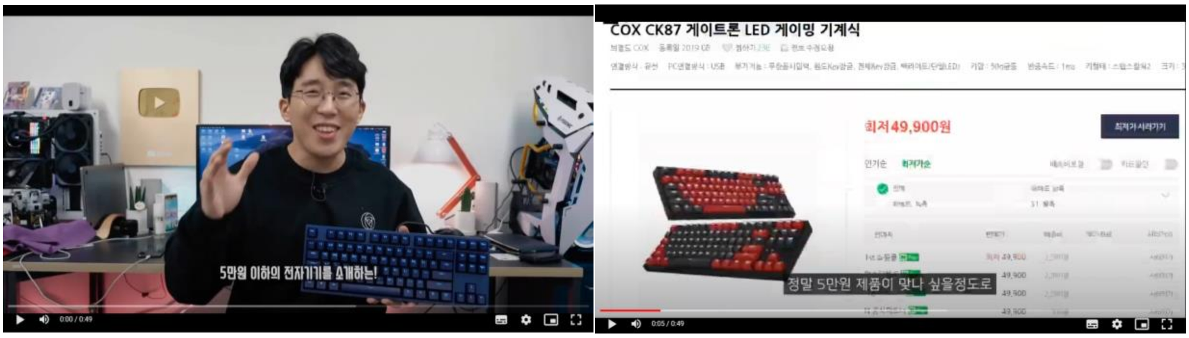
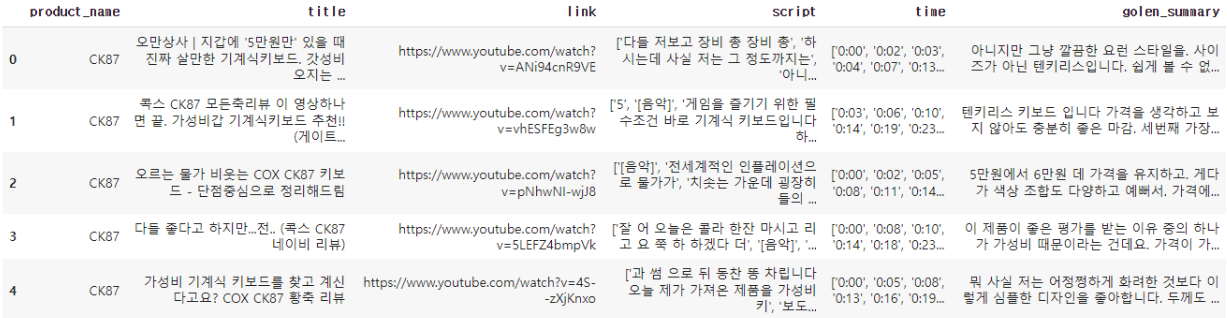

5만원 이하의 전자기기를 소개하는!
0:00 / 0:49
COX CK87 게이트론 LED 게이밍 기계식
최저49,900원
정말 5만원 제품이 맞나 싶을정도로
0:05 / 0:49

Figure 3: Captures of created video

4. Experiments

4.1. Dataset

For Korean data, there is no existing product review data for multi-video summarization, thus we created a dataset in this paper. The dataset is composed of 271 Korean product review videos of 22 different products such as cell phones, game consoles, and robot vacuum cleaners. Shown in Figure 4 below is the shape of the dataset. Each column of the data is in the format of product name, video title, video link, video subtitle, time of video matching with subtitle, and answer data. The answer data was generated randomly.



	product_name	title	link	script	time	golen_summary
0	CK87	오만상사 | 지갑에 '5만원만' 있을 때 진짜 살만한 기계식키보드. 갓성비 오지는 ...	https://www.youtube.com/watch?v=ANi94cnR9VE	['다들 저보고 장비 총 장비 총', '하시는데 사실 저는 그 정도까지는', '아니...	['0:00', '0:02', '0:03', '0:04', '0:07', '0:13...	아니지만 그냥 깔끔한 요런 스타일을. 사이즈가 아닌 텐키리스입니다. 쉽게 볼 수 없...
1	CK87	콕스 CK87 모든축리뷰 이 영상하나면 끝. 가성비갑 기계식키보드 추천!! (게이트...	https://www.youtube.com/watch?v=vhESFEg3w8w	['5', '[음악]', '게임을 즐기기 위한 필수조건 바로 기계식 키보드입니다 하...	['0:03', '0:06', '0:10', '0:14', '0:19', '0:23...	텐키리스 키보드 입니다 가격을 생각하고 보지 않아도 충분히 좋은 마감. 세번째 가장...
2	CK87	오르는 물가 비웃는 COX CK87 키보드 - 단점중심으로 정리해드림	https://www.youtube.com/watch?v=pNhwNI-wjJ8	['[음악]', '전세계적인 인플레이션으로 물가가', '치솟는 가운데 굉장히 들의 ...	['0:00', '0:02', '0:05', '0:08', '0:11', '0:14...	5만원에서 6만원 데 가격을 유지하고. 게다가 색상 조합도 다양하고 예뻐서. 가격에...
3	CK87	다들 좋다고 하지만...전.. (콕스 CK87 네이비 리뷰)	https://www.youtube.com/watch?v=5LEFZ4bmpVk	['잘 어 오늘은 콜라 한잔 마시고 리고 요 쭉 하 하겠다 더', '[음악]', '...	['0:00', '0:08', '0:10', '0:14', '0:18', '0:23...	이 제품이 좋은 평가를 받는 이유 중의 하나가 가성비 때문이라는 건데요. 가격이 가...
4	CK87	가성비 기계식 키보드를 찾고 계신다고요? COX CK87 황축 리뷰	https://www.youtube.com/watch?v=4S--zXjKnxo	['과 썸 으로 뒤 동찬 똥 차립니다 오늘 제가 가져온 제품을 가성비 키', '보도...	['0:00', '0:05', '0:08', '0:13', '0:16', '0:19...	뭐 사실 저는 어정쩡하게 화려한 것보다 이렇게 심플한 디자인을 좋아합니다. 두께도 ...

Figure 4: Structure of Dataset

4.2. Evaluation metric

For evaluation metrics, we used Topic Coherence and Topic Diversity to evaluate Topic modeling, and Rouge for summary evaluation.

4.2.1. Topic Coherence

Topic Coherence evaluates Normalized Pointwise Mutual Information (NPMI) [22], a method proposed by Röder et al. (2015) [23]. The evaluated result has a value between -1 and 1. The closer it is to 1, the more the coherence of the topic.

$$\vec{v}(W') = \left\{ \sum_{w_i \in W'} NPMI(w_i, w_j)^{\gamma} \right\}_{j=1,\ldots,|W|} \tag{1}$$

Let W'be a vector representing the similarity between words in the corpus. $\vec{v}(W')$ computes the similarity of each word in W' to the other words and represents it as a vector.

$$NPMI(w_i, w_j)^{\gamma} = \left(\frac{log \frac{P(w_i, w_j) + \epsilon}{P(w_i) \cdot P(w_j)}}{-\log(P(w_i, w_j) + \epsilon)} \right)^{\gamma} \tag{2}$$

NPMI is a metric that measures the relevance of two words w_i, w_j. $P(w_i, w_j)$ is the probability that two words w_i, w_j co-occurence. $P(w_i)$ is the probability of w_i occurrence and $P(w_j)$ is the probability of w_j occurrence. γ represents an exponent and serves to weight the NPMI values exponentially.

$$\Phi S_i(\vec{u}, \vec{w}) = \frac{\sum_{i=1}^{|W|} u_i \cdot w_i}{||\vec{u}||_2 \cdot ||\vec{w}||_2} \tag{3}$$

i represents the dimension of the vector. $||\vec{u}||_2$ represents the L2 norm of vector $\vec{u}$, $||\vec{w}||_2$ represents the L2 norm of vector $\vec{w}$, and $\Phi S_i(\vec{u}, \vec{w})$ computes the similarity between two vectors $\vec{u}, \vec{w}$ by dividing the dot product of the vectors by the L2 norm of the vectors.

4.2.2. Topic Diversity

Dieng et al. (2020) [24] proposed a metric to measure the diversity of vocabulary or words in the topics of a certain topic model. The metric is the proportion of unique words, excluding duplicate words, that exist in the topics for all topics, and the measure ranges from 0 to 1. 0 represents duplicate topics and 1 represents various topics.

$|\{unique\ words\}|$ is the size of the union of unique words in all topics. topk is the selected number of top topics, and topics is the number of total topics.

$$Topic\ Diversity = \frac{|\{unique\ words\}|}{topk \cdot |topics|} \tag{4}$$

4.2.3. Rouge

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) (Lin, 2004) is a co-occurrence statistical measure of N-grams and is to be defined as follows (5).. $Count_{match}(N - Gram)$ is the number of N-grams that are included in both summarized results and reference summary. $Count(N - Gram)$ is the number of N-grams included in the reference summary. These ROUGE metrics can generate three scores: Recall, Precision, and F-measure [25].

$$\mathrm{ROUGE - N} = \frac{\sum_{S \in Summ_{ref}} \sum_{N-gram \in S} Count_{match}(N - Gram)}{\sum_{S \in Summ_{ref}} \sum_{N-gram \in S} Count(N - Gram)} \tag{5}$$

4.3. BERTopic

We compared the performance difference between embedding models using TC (Topic Coherence) and TD (Topic Diversity), as the performance difference exists depending on the embedding model of BERTopic. As shown in Table 1, the multilingual model of SBERT, distilues-base-multilngual-cased-v1, recorded the best score, thus we compared the topic model with this embedding model.

We calculated the TC and TD of three models, the BERTopic model with distilues-base-multilngual-cased-v1 as the embedding model, the Combined Topic Models (CTM) model, and the LDA model, and found that BERTopic shows the best performance, as shown in Table 2. This shows that the BERTopic model generates various topics and coherent topics in one topic.

Table 1
Comparison of embedding model performance

	TC	TD
Distiluse-base-multilingual-cased-v1	0.7174	0.8152
KR-SBERT-V40K-klueNLI-augSTS	0.5844	0.7
Ko-sbert-multitask	0.7071	0.8014
ko-sroberta-multitask	0.7047	0.8095

We calculated the TC and TD of three models, the BERTopic model with distilues-base-multilngual-cased-v1 as the embedding model, the Combined Topic Models (CTM) model, and the LDA model, and found that BERTopic shows the best performance, as shown in Table 2. This shows that the BERTopic model generates various topics and coherent topics in one topic.

Table 2
Comparison of Topic modeling model performance

	TC	TD
BERTopic	0.7174	0.8152
CTM	0.4978	0.5192
LDA	0.5303	0.752

4.4. Bart

Summarization step uses BART to generate a summary for each topic generated by BERTopic. After sorting the sentences with the highest cosine similarity, we extracted the top n sentences to summarize each topic. In this paper, we set n to 3 arbitrarily. The performance comparison was performed with TextRank [26] and Bert-extractive-summarizer [19]. The results can be seen in Table 3. In precision, TextRank and Bert-extractive-summarizer showed good performance. But, in recall and f1-measure, we can show that the proposed model has the best performance. This means that the proposed model performs the most equally well compared to the other models.

Table 3
ROUGE score of different models

	ROUGE-1			ROUGE-2			ROUGE-L		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Proposed model	0.2488	0.3952	0.2767	0.0254	0.0355	0.0286	0.1758	0.2723	0.1975
TextRank	0.4609	0.0041	0.0081	0.1770	0.0018	0.0034	0.4609	0.0041	0.0081
Bert-extractive-summarizer	0.4385	0.0069	0.0134	0.2782	0.0044	0.0086	0.4385	0.0069	0.0134

5. Conclusion

In this paper, we reduce the duplication of data by grouping the same or similar information into one topic through Topic Modeling. In addition, we compared models such as CTM and LDA, and selected BERTopic with the best topic diversity and topic cohesion to generate coherent and diverse topics. This process reduces wasted time searching for information, provides users with multiple perspectives on the product, and helps eliminate information loss in long videos. We used BART for summarization. In comparison to the previous research such as Text Rank and Bert-extractive-summarizer, ROUGE showed the best performance in recall and F1 measures.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (NRF-2021R1G1A1006381).

References

[1] Parabhoi, Lambodara, et al. "YouTube as a source of information during the Covid-19 pandemic: a content analysis of YouTube videos published during January to March 2020." *BMC Medical Informatics and Decision Making* 21.1 (2021): 1-10.

[2] Khatri, Priyanka, et al. "YouTube as source of information on 2019 novel coronavirus outbreak: a cross sectional study of English and Mandarin content." *Travel medicine and infectious disease* 35 (2020): 101636.

[3] S. Vazarkar and T. Manjusha, "Video to text summarization system using multimodal LDA," Journal of Seybold, vol. 15, no. 9, pp. 3517–3523, 2020

[4] S. Feng, Z. Lei, D. Yi, and S. Z. Li, "Online content-aware video condensation," in IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[5] Y. J. Lee, J. Ghosh, and K. Grauman, "Discovering important people and objects for egocentric video summarization," in IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[6] H. Kang, X. Chen, Y. Matsushita, and X. Tang, "Space-time video montage," in IEEE Conference on Computer Vision and Pattern Recognition, 2006.

[7] Y. Pritch, A. Rav-Acha, and S. Peleg, "Nonchronological video synopsis and indexing," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1971–1984, 2008.

[8] Z. Lu and K. Grauman, "Story-driven summarization for egocentric video," in IEEE Conference on Computer Vision and Pattern Recognition, 2013.

[9] V. B. Aswin, M. Javed, P. Parihar, K. Aswanth, C. R. Druval et al., "NLP-driven ensemble-based automatic subtitle generation and semantic video summarization technique," in Advances in Intelligent Systems & Computing, vol. 1133, Singapore: Springer, pp. 3–13, 2021

[10] Luo, Bo, et al. "Video caption detection and extraction using temporal information." *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*. Vol. 1. IEEE, 2003.

[11] Narwal, Pulkit, Neelam Duhan, and Komal Kumar Bhatia. "A comprehensive survey and mathematical insights towards video summarization." *Journal of Visual Communication and Image Representation* 89 (2022): 103670.

[12] Pawar, Priya, et al. "Online product review summarization." *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*. IEEE, 2017.

[13] Boorugu, Ravali, and G. Ramesh. "A survey on NLP based text summarization for summarizing product reviews." *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2020.

[14] Zhao, Qingjuan, Jianwei Niu, and Xuefeng Liu. "ALS-MRS: Incorporating aspect-level sentiment for abstractive multi-review summarization." *Knowledge-Based Systems* (2022): 109942.

[15] Widyassari, Adhika Pramita, et al. "Review of automatic text summarization techniques & methods." *Journal of King Saud University-Computer and Information Sciences* 34.4 (2022): 1029-1046.

[16] Vayansky, Ike, and Sathish AP Kumar. "A review of topic modeling methods." *Information Systems* 94 (2020): 101582.

[17] John, Ansamma, P. S. Premjith, and M. Wilscy. "Extractive multi-document summarization using population-based multicriteria optimization." *Expert Systems with Applications* 86 (2017): 385-397.

[18] S. S. Alrumiah and A. A. Al-Shargabi, "Educational videos subtitles' summarization using latent dirichlet allocation and length enhancement," Computers, Materials & Continua, vol. 70, no.3, pp. 6205–6221, 2022.

[19] Miller, Derek. "Leveraging BERT for extractive text summarization on lectures." *arXiv preprint arXiv:1906.04165* (2019).

[20] Grootendorst, Maarten. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." *arXiv preprint arXiv:2203.05794* (2022).

[21] Lewis, Mike, et al. "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension." *arXiv preprint arXiv:1910.13461* (2019).

[22] Bouma, Gerlof. "Normalized (pointwise) mutual information in collocation extraction." *Proceedings of GSCL* 30 (2009): 31-40.

[23] Röder, Michael, Andreas Both, and Alexander Hinneburg. "Exploring the space of topic coherence measures." *Proceedings of the eighth ACM international conference on Web search and data mining*. 2015.

[24] Dieng, Adji B., Francisco JR Ruiz, and David M. Blei. "Topic modeling in embedding spaces." *Transactions of the Association for Computational Linguistics* 8 (2020): 439-453.

[25] Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries." *Text summarization branches out*. 2004.

[26] Mihalcea, Rada, and Paul Tarau. "Textrank: Bringing order into text." *Proceedings of the 2004 conference on empirical methods in natural language processing*. 2004