A Study on Methods for Summarizing Information from YouTube Channels

Hyoung-Sik Park

ABSTRACT

This study explored the possibility of summarizing information from YouTube channels. To achieve this, the research compared and analyzed speech-to-text (STT) data from top-viewed videos and comment data from recent videos. The keyword network analysis revealed that comment data could represent the main content of the videos, thereby validating the feasibility of information summarization using only comments.
To enhance the reliability of the study, text similarity analysis and statistical verification were conducted. The results of this research are expected to contribute to the utilization of influencer marketing by businesses and to help individual subscribers quickly grasp the main content of channels. Furthermore, it is anticipated that this study will make a practical contribution to the development of efficient processing and analysis methodologies for digital media data.

☞ keyword : YouTube Channels, Information Summarization, STT (Speech-to-Text), Keyword Network, Similarity Analysis, Text Analysis, Comment Analysis, Influencer Marketing

1. Introduction

Digital media, particularly social media platforms like YouTube, have established themselves as a crucial means of information delivery and marketing tools in modern society. The diverse content on YouTube channels significantly influences the opinion formation, consumption patterns, and cultural trends of millions of users. Especially with the growing interest in healthcare and health-related content, such information plays a vital role in users' health-related decision-making (Mohamed & Shoufan, 2024). However, the vast amount of content generated on YouTube channels is reducing the efficiency of information delivery. While the need for effective video summarization methods has emerged, existing research has mainly focused on summarizing individual videos or converting audio information to text, with limited research on methods to summarize information across entire channels.

The purpose of this study is to present and evaluate the validity of a method for comprehensively summarizing information from large amounts of video and comment data generated on YouTube channels. Specifically, for YouTube channels providing health and dietary supplement-related content in Korean, this study explores the possibility of summarization using only comments by comparing and analyzing speech-to-text (STT) data from top-viewed videos with recent video comments.

This research moves beyond fragmentary video summarization to present a new methodology for understanding the characteristics of entire channels, which can help businesses analyze YouTube content more quickly and develop influencer marketing strategies. Additionally, it provides ways for individual users to effectively grasp the core content of subscribed channels and offers foundational data that can be used in developing influencer marketing and personalized content recommendation systems. This study is expected to make a practical contribution to the development of efficient processing and analysis methodologies for digital media data.

2. Theoretical Background

2.1 Previous Research

Existing YouTube summarization research has primarily followed three approaches: summarizing videos themselves, summarizing captions uploaded by content creators, and extracting audio information to convert it to text before applying summarization techniques. The video summarization method typically involves condensing videos longer than an hour into approximately 10-minute versions. However, this approach has limitations in terms of information consistency and efficiency from a channel perspective, where multiple videos with similar content need to be quickly and effectively summarized, as it provides summaries through different views of video and text.

Previous research on text summarization of YouTube videos has utilized two methods commonly used in literature information and text sentence summarization: Extractive Summarization and Abstractive Summarization. Extractive summarization identifies and extracts important sentences or phrases from the original text, while abstractive summarization understands the meaning of the original text and generates new sentences to create summaries. While earlier studies primarily used extractive summarization methods, recent research has adopted abstractive summarization approaches using artificial neural network-based models, particularly Transformer models, to generate more natural and human-readable summaries. Recently, information summarization using Large Language Models (LLMs) based on Transformer architecture has seen significant growth.

In video summarization research, Kaiyang Zhou et al. (2018) proposed a reinforcement learning-based unsupervised video summarization method which considers the diversity and representativeness of video frames. Their study used a network combining CNN and LSTM to select video frames and evaluated summary quality through various reward functions. Experimental results showed superior performance compared to other unsupervised learning methods.

Research on summarizing video information through caption or audio text conversion includes the work of Lavish Mangal et al. (2022), who developed a YouTube Transcript Summarizer that extracts video captions and generates summaries using abstractive summarization. Their research uses the Hugging Face Transformer model for caption summarization and provides results through a REST API. Additionally, Sulochana Devi et al. (2023) proposed a model applying abstractive summarization to extracted YouTube video captions. Their study presents methods for converting video audio to text and summarizing the converted text to effectively provide users with necessary information. They also explain the two main approaches to text summarization and demonstrate situations where abstractive summarization may be more useful than extractive summarization.

KIM, JI YEON (2023) proposed a new abstractive summarization model emphasizing the importance of summarization without information loss. This enables searching and identifying appropriate information from vast amounts of text data in the big data era. In similar research on literature information summarization, Lee Jong-won, Kim Tae-hyeon, Shin Dong-gu, and Cho Woo-seung (2022) proposed a Korean information summarization system for summarizing national R&D project information. This system developed a preprocessor analyzing Korean linguistic characteristics and a natural language processing-based project information summarization model.

However, Transformer models commonly used in abstractive summarization have fundamental limitations in input token count, or input sentence size, making them impractical for summarizing large amounts of information without additional summarization methodologies. To summarize videos longer than an hour, methods must be devised to divide the video into 10-minute segments, extract and convert audio information to text, create summaries within the token size limits of

Transformer models, and merge these results while preserving the original content.

This study focuses on developing ways for users to quickly grasp necessary information, considering both the advantages and limitations of existing research methods, to address the problem of extensive information volume and time-consuming comprehension in YouTube channels. The aim is to help users understand main content without watching videos while improving information consistency and transmission efficiency.

2.2 Theoretical Background

In this paper, we aim to verify hypotheses about information summarization methods based on key theoretical foundations including information summarization, keyword network analysis, and document similarity analysis. Therefore, we will provide an overview of these major theoretical frameworks..

2.2.1 Information Summarization Methodology

The theoretical background of information summarization can be broadly categorized into two methods: Extractive Summarization and Abstractive Summarization.

Extractive summarization is an approach that identifies and extracts important sentences or phrases from the original text to generate a summary. It utilizes statistical techniques and Natural Language Processing (NLP) technology. A representative method is TF-IDF (Term Frequency-Inverse Document Frequency), which is used to extract important keywords by evaluating the significance of specific terms within a document.

Abstractive summarization creates summaries by deeply understanding the meaning of the original text and generating new sentences. This method incorporates Natural Language Generation (NLG) technology and can better convey meaning while reducing redundancy. Unlike extractive summarization, which simply takes important sentences or phrases directly from the original text, abstractive summarization generates summaries using new expressions based on the core content of the original text. This allows for not only information compression but also provides readers with clearer and more coherent information through restructuring.

The Transformer model based on deep learning is a prime example. Transformer models efficiently learn relationships between words within sentences using the Self-Attention Mechanism. Compared to traditional sequential models like RNN (Recurrent Neural Network) or LSTM (Long Short-Term Memory), it offers advantages in parallel processing capabilities and better understanding of long contexts.

Notably, models that emerged from the development of Transformer architecture, such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), have shown exceptional performance in various natural language processing tasks. BERT's ability to understand context bidirectionally makes it advantageous for deep text comprehension, while GPT excels at sentence generation.

Other examples of abstractive summarization models include PEGASUS and T5 (Text-To-Text Transfer Transformer). PEGASUS maximizes summarization performance by evaluating the importance of each sentence and learning through masking important sentences for generation. The T5 model redefines all natural language processing problems as text-input and text-output formats, enabling various language tasks to be processed within a unified framework.

These advanced technologies contribute to improving the accuracy of abstractive summarization and effectively summarizing various forms of text data. Therefore, abstractive summarization plays a crucial role in efficiently extracting and conveying useful information from text data.

2.2.2 Keyword and Network Analysis Techniques

Two-Mode Keyword Network Analysis is a methodology that uses two types of nodes to analyze relationships between keywords. It is particularly useful for understanding deeper network structures by simultaneously analyzing relationships between keywords and documents.
The two-mode network can be represented by the following matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \quad (1)$$

In Equation (1), a_ij represents the relationship between document i and keyword j. Rows represent documents, and columns represent keywords. Based on this matrix, the co-occurrence frequency between keywords is calculated and visualized to identify major hub keywords.
Two-mode keyword network analysis is particularly useful for identifying important topics and trends in large volumes of data. For example, Nam Hye-ju and Hong Dae-seok (2023) utilized keyword network analysis to explore research trends related to sports injuries.

2.2.3 Document Similarity Analysis

Document similarity is a methodology for measuring the similarity between two documents, primarily using techniques such as cosine similarity, Pearson correlation coefficient, and k-NN.

● Cosine Similarity
Cosine similarity is a method that measures similarity by calculating the cosine angle between two vectors. When there are two vectors A and B, the cosine similarity is calculated as follows:

Cosine Similarity(A, B) = cos(θ) = (A · B) / (||A|| ||B||)

Here, A · B is the dot product of two vectors, and ||A|| and ||B|| are the magnitudes of vectors A and B, respectively.

● Pearson Correlation Coefficient
The Pearson correlation coefficient is a method for measuring the linear correlation between two variables. The Pearson correlation coefficient between variables X and Y is calculated as follows:.

r=∑(Xi−X¯)2∑(Yi−Y¯)2∑(Xi−X¯)(Yi−Y¯)
Here, X¯\bar{X}X¯ and Y¯\bar{Y}Y¯ are the means of variables XXX and YYY, respectively.

● k-NN (k-Nearest Neighbors)
The k-NN algorithm is used to classify or predict similar data based on distance measurements between data points. This algorithm refers to the k nearest neighbors to determine the class of a new data point. Distance is primarily calculated using Euclidean Distance:.

d(A,B)=∑i=1n(Ai−Bi)2

Here, d(A,B)d(A, B)d(A,B) is the Euclidean distance between vectors AAA and BBB.

These theories contribute to efficiently summarizing and utilizing the vast amount of data provided by digital media platforms like YouTube. These studies provide deeper insights by integrating text and video data, supporting information-based decision-making.

3. Research Methods

3.1 Research Hypothesis

This study hypothesizes that analyzing only comments can identify a channel's main

characteristics and topics without watching all videos or converting them to text for analysis. To verify this hypothesis, we analyze comment data from top-viewed videos over the past year and compare it with keyword network analysis of speech-to-text (STT) converted audio data from the same videos. By comparing the similarity between keyword networks from these two datasets (comments and video text), we assess whether comment analysis alone can sufficiently summarize a YouTube channel's key information. Furthermore, we aim to validate whether this method can be utilized as an effective information summarization approach for YouTube channels.

3.2 Data Collection and Preprocessing

3.2.1 Data Collection Scope and Targets

Five YouTube channels providing content related to nutritional supplements and health supplements were selected as data collection targets. The channel selection criteria included channels with varying subscriber scales (over 1 million, over 100,000, over 10,000, and several thousand subscribers). The collected data includes videos posted during the past year on each channel and their associated comment data. Specifically, we collected channel information, video information, view counts, like counts, and comment content, with data analysis focusing on videos with the highest view counts. In particular, we analyzed the 3-5 most-viewed videos from each channel by converting their audio data to text using STT (Speech-to-Text) tools, along with their comment data. These top-viewed videos were assumed to be representative content for their respective channels, and we determined that the topics and content of these videos well reflect the overall character of the channel.

3.2.2 Data Preprocessing Process

The collected data underwent preprocessing to analyze text data characteristics. First, comment data and video text data were preprocessed separately. Comment data was collected along with video view counts, and de-identified IDs (A, B, C, D, E) were assigned to each channel for analysis. A Korean stopwords list was defined to remove unnecessary words and greetings, and a keyword dictionary related to nutritional supplements and health supplements was created for key keyword extraction.

Video text data was preprocessed in the same way as comment data after converting audio data to text using STT tools. The Korean morphological analyzer (Mecab) was used to separate text into morpheme units and remove stopwords. Based on this, key keywords were extracted from the preprocessed data and visualized through word clouds. This process provided foundational data for identifying key keywords for each channel..

3.3 Analysis Methods

3.3.1 Keyword Network Analysis Method

Keyword network analysis is the process of extracting important keywords from collected comment data and video text data and visualizing their relationships. First, using the TF-IDF (Term Frequency-Inverse Document Frequency) technique, important keywords are extracted from each text based on preprocessed text data. TF-IDF is a statistical method that evaluates how important a word is within a document, assigning higher weights to words that appear frequently in a document but less frequently in other documents. Networks are constructed based on co-occurrence frequency between extracted keywords to identify how often certain keywords appear together.

By visualizing this network, relationships between major keywords and important keywords can be intuitively understood. This allows for quick comprehension of each channel's topics and characteristics, and identification of hub keywords

(major keywords) that play important roles within the network. Keyword network analysis is useful for information compression and summarization, effectively summarizing core topics and content from vast amounts of channel information. It is particularly suitable for simultaneously summarizing and visualizing extensive video content and comment data from YouTube channels.

3.3.2 Text Similarity Analysis Method

Text similarity analysis measures the similarity between comment data and video text data generated through keyword network analysis. This method verifies whether comment analysis alone can effectively summarize the main content of videos. For text similarity analysis, cosine similarity and Pearson correlation coefficient are used. Cosine similarity measures similarity by vectorizing two texts and calculating the angle between vectors, while the Pearson correlation coefficient is a statistical indicator measuring the strength of linear relationships between two data sets. Both methods numerically express the similarity between comment and video text keyword networks, enabling evaluation of how well comment data represents video content. Higher similarity indicates that comment analysis alone can sufficiently grasp the video's topic and content, thereby proving whether the information summarization method using comment analysis can effectively summarize the main content of YouTube channels.

4. Analysis Results

4.1 Analysis Results by YouTube Channel

For each channel, keyword networks were generated using comment data and video text data, and important keywords and their relationships were visually analyzed. The 2-mode network shows relationships between keywords extracted from each data source. The main elements of network analysis are as follows::

- Nodes: Keywords extracted from comments or video text
- Links: Represent relationships between keywords, with relationship strength expressed through color and thickness
- Colors and Links: Nodes represent keywords, yellow links represent comments, green links represent video text, and blue links represent keyword relationships between the two sources.

4.1.1 Channel A

As shown in Figure 1, keywords such as 'vitamin', 'calcium', 'skin', and 'blood pressure' frequently appear in both comments and video text, indicating their importance as key topics. Thick links represent strong relationships between these keywords, and keyword clustering suggests that related topics form groups. Blue links demonstrate high correspondence between the two sources.

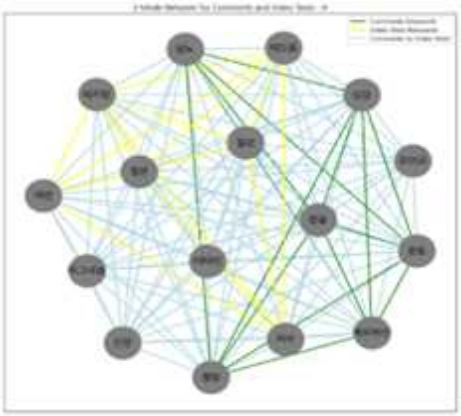


Figure 1: Keyword Network Analysis Results for Channel A

4.1.2 Channel B

As shown in Figure 2, keywords such as 'muscle', 'exercise', 'skin', and 'vitamin' are centrally positioned in the network and are treated as important topics in both sources. 'Muscle' and 'exercise' are frequently mentioned together, and there are many blue links indicating high correspondence between comments and video text.

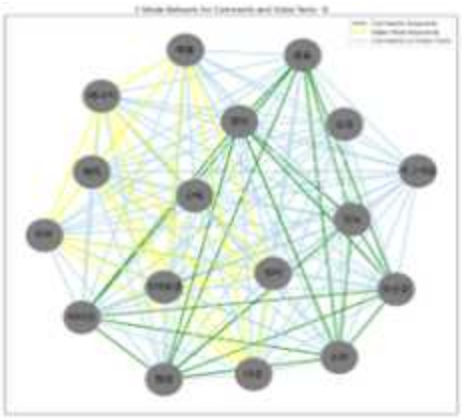


Figure 2: Keyword Network Analysis Results for Channel B

4.1.3 Channel C

As shown in Figure 3, keywords such as 'melatonin', 'magnesium', and 'energy' emerge as major discussion topics, with strong relationships appearing between them. The keyword clusters indicate that specific topics are being discussed in groups, and high similarity can be observed between comments and video text..

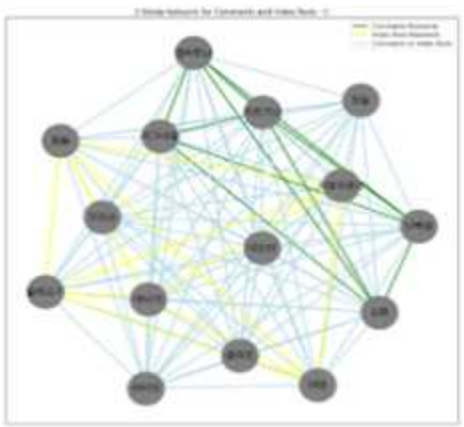


Figure 3: Keyword Network Analysis Results for Channel C

4.1.4 Channel D

As shown in Figure 4, keywords such as 'skin', 'exercise', and 'vitamin' are centrally positioned, with thick links between them indicating strong relationships. Keyword clusters have formed around terms like 'cholesterol', and there is a high degree of information correspondence between comments and video text..

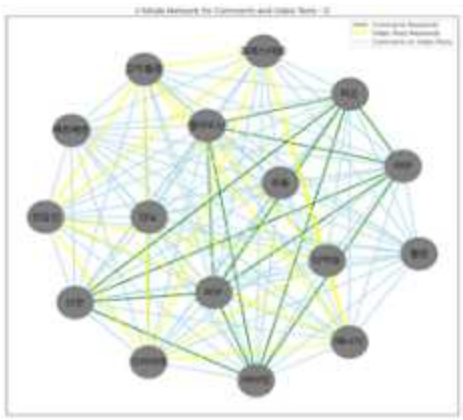


Figure 4: Keyword Network Analysis Results for Channel D

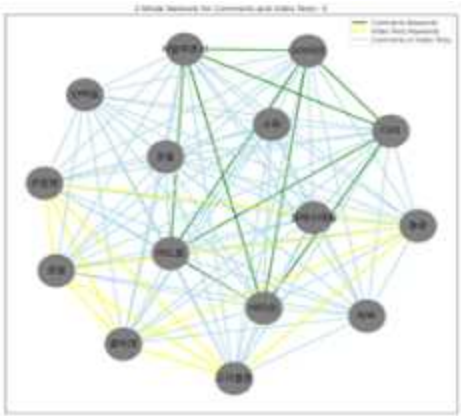
4.1.5 Channel E

As shown in Figure 5, keywords such as 'vitamin', 'exercise', and 'diet' emerge as important topics, with strong relationships appearing between them. Keyword clusters can be identified around terms like 'diet' and 'hyaluronic acid', and the abundance of blue links between the two sources suggests high similarity..

Figure 5: Keyword Network Analysis Results for Channel E

4.2 Text Similarity Analysis

The similarity between comment and video text keyword networks was measured using cosine similarity and Pearson correlation coefficient.



Channel B showed the highest similarity with a cosine similarity of 0.699, while Channel E showed the lowest at 0.424. This indicates that comment data alone can be sufficient to identify a channel's main topics.

채널구분	Cosine Similarity	Pearson Correlation	k-NN Similarity
A	0.676180	0.663575	0.521086
B	0.699733	0.688897	0.545671
C	0.614504	0.595207	0.537723
D	0.546431	0.526614	0.491970
E	0.424739	0.392427	0.472317

Table 1.: Similarity Measurement Results by Channel

4.3 Statistical Verification

The t-test results showed a significant difference with a p-Value of 0.0456. This suggests that comment data could potentially substitute for video text data, particularly in channels with high similarity scores, where comment data alone can effectively identify channel characteristics and topics..

5. Conclusion and Future Research Directions

5.1 Research Summary and Implications

This study explored methods for summarizing YouTube channel information by comparing speech-to-text (STT) data from top-viewed videos with comment data. Particular emphasis was placed on the possibility of information summarization using comments. The research findings demonstrated that keyword networks extracted from comments showed high similarity with STT data, confirming that comments can effectively summarize the main content of YouTube channels. This method can be rapidly utilized by businesses in influencer marketing and help individual subscribers quickly grasp the core content of channels.

This research has proven that comment data serves as a valid information summarization tool. Furthermore, it suggests that keyword network and text similarity analyses can be applied to various topics and channels. Future research will require multimodal data analysis integrating video, audio, and text, which could lead to the development of more sophisticated content summarization methods.

5.2 Research Limitations and Future Research Directions

This study analyzed data limited to certain YouTube channels, and research including channels with diverse topics is needed. Additionally, the heavy reliance on keyword networks and text analysis necessitates the introduction of multimodal data analysis. Future research needs to develop more comprehensive methods for analyzing and summarizing YouTube content by integrating video, audio, and text data. This can be utilized across various fields including digital marketing, education, and information delivery..

Reference

Kim, J. Y. (2023). "Enhancing Abstract Summarization through Triplet loss." Master's Thesis, Graduate School of Computer and Information Technology, Korea University, Seoul.

Lee, J. W., Kim, T. H., Shin, D. G., & Cho, W. S. (2022). Korean Information Summarization System for Summarizing National R&D Project Information. Journal of the Korea Institute of Information and Communication Engineering, Jeju.

Jung, J. I. (2016). "Improved Tag Search System through Keyword Extraction and Similarity Evaluation." Master's Thesis, Soongsil University, Seoul.

Baek, S. H., & Kim, E. J. (2021). Research Trends of Korean Elderly Exercise through

Keyword Network Analysis (1995-2019). Journal of the Korea Convergence Society, 19(1).

Nam, H. J., & Hong, D. S. (2023). A Study on Keyword Network Analysis for Exploring Research Trends Related to Sports Injuries. Korean Journal of Sports Science, 32(6), 755-766. https://doi.org/10.35159/kjss.2023.12.32.6.755

Kim, G. Y. (2019). Analysis of Tourist Destination Network Characteristics through Two-Mode Network Analysis - Focusing on Japanese Tourists Visiting Korea. Journal of Tourism Research, 31(4), 23-46. https://doi.org/10.21581/jts.2019.11.31.4.23

Yoo, J. Y., & Kim, M. K. (2019). Convergence of Korean Traditional Dance and K-pop Dance: Analysis of YouTube Comments on BTS's 'IDOL' at 2018 MMA. Journal of Korea Entertainment Industry Association, 13(8), 189-198. https://doi.org/10.21184/jkeia.2019.12.13.8.189

Lavish Mangal, Dhruv Aggarwal, Gulshan Kumar, Jai Kumar, Meenakshi Aggarwal. (2022). YouTube Transcript Summarizer. 2022 International Mobile and Embedded Technology Conference (MECON)

J. H. Shin, E. H. Kim, and M. J. Lim, “Improving the effectiveness of document extraction summary based on the amount of sentence information,” Smart media journal, Vol. 11, No. 3, pp. 31-38, Apr. 2022.

Mohamed, Fatma, and Abdulhadi Shoufan. (2024). Users’ experience with health-related content on YouTube: an exploratory study. BMC Public Health, 24:86.

Sulochana Devi, Rahul Nadar, Tejas Nichat, and Alfredprem Lucas. (2023). Abstractive Summarizer for YouTube Videos. In Proceedings of the International Conference on Applied Mathematics, Intelligent Systems and Computing Applications (ICAMIDA), vol. 105, pp. 431-438.

Kaiyang Zhou, Yu Qiao, and Tao Xiang. (2018). Deep Reinforcement Learning for Unsupervised Video Summarization with Diversity-Representativeness Reward. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1).

◐ ABOUT THE AUTHOR ◑

Hyoung-sik Park

EDUCATION

- Master of Engineering in AI Big Data
- Department of Advanced AI, Graduate School of AI
- aSSIST (Seoul School of Integrated Sciences & Technologies)

RESEARCH INTERESTS

- Artificial Intelligence
- Natural Language Processing
- Large Language Models (LLMs)
- And related fields

CONTACT E-mail: saintphs@naver.com