HiLite Research Paper

By: Leo Milligan, Panav Sheth, Harshit Jain,

Jiahau Liu, Sahil Kuwadia, Yug Jarodiya

Table of Contents

Contents

Problem & Approach

Students are really busy. In today's modern world, many students receive vast amounts of homework including plenty of reading. Many courses throughout college, in all fields, will require many pages of reading every single night. This can be overwhelming to even the most serious of students. To take that stress off students we aim to create a product that can highlight text to allow students to ingest the important parts of texts. To take in the important information when reading so you don't have to highlight and reread, we present HiLite.

HiLite is an auto highlighter that automatically identifies and highlights key phrases and sentences within blocks of text. It uses artificial intelligence to determine which parts of the text are most important based on various factors such as keywords, relevance, and context. Auto highlighters can assist users in quickly summarizing, organizing, and understanding the content of a text. Using HiLite will not only make getting through all your homework quicker but will make you more effective at learning information. This AI ensures that you will not miss any of the important details and will give you the best summary with all the useful information. Using this tool will improve both the effectiveness and the efficiency of studying for students all around the world.

According to 'ABSA toolkit: An open-source tool for aspect-based sentiment analysis' by Zarmeen Nasim and Sajjad Haider, using an AI summarizer and highlighter can save up to 70% of the time it would usually take to read a lengthy document.

> *''… In both domains, the toolkit achieved more than 70% score in … processing tasks including text summarization, entity … API and Aylien API for aspect-based sentiment analysis. The …''*

Using all that saved time, students will have more time to process the information and will spend less time on the nonimportant parts of any given paper. HiLite aims to improve the lives of all students no matter the language they speak or what they are studying, all by saving them time and effort when reading and reviewing important documents.

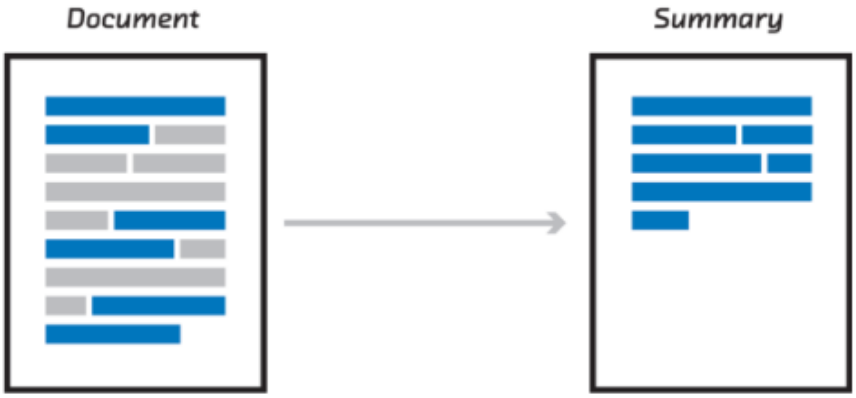
With the amount of information students need to go through, many can suffer from information overload. This can cause mental fatigue and a reduced attention span. Since many professors will assign readings every day before the next class that causes every student to have many reading due every single night which can very easily cause the information to get mixed up and forgotten. Another way that information overload affects students is by reducing their ability to critically evaluate information.

With so much information available, it can be challenging for students to determine which sources are reliable and trustworthy. This can lead to a lack of understanding since students

may simply accept information at face value without questioning its validity. Along with the prior issue there is the topic of burnout within students, especially at the college level. College is stressful enough as is but with the addition of the many homework assignments and exams, students constantly must concentrate on students burn out very quickly and very often.

Using HiLite can assist students with their busy schedules with school allowing more time to focus on hobbies and other activities to help prevent the burnout many may face. With these issues there is a major need for a solution that can save the time of the students while still giving them the knowledge they need to get by. Another major way HiLite can be used in through research and everyday use with articles. Researchers often have a lot of data they need to get through. They receive tons of papers they need to read so HiLite can cut that time down tremendously. Through everyday use people can use it for articles on things they are interested in or to keep up with what may be happening in the world. People have very busy lives so staying informed on current events using HiLite would make staying informed much easier. That is what HiLite is all about

Product Feature



HiLite takes any document you need for a class, or just want to know the information, and turns it into a more digestible and readable document that highlights the important aspects and summarizes the overall paper. Highlighting is an effective tool that can be used when reading through a document or paper to remember the important information to go back to later. 59% of teachers agree that technology should play a role in supporting personalized learning.

Using an AI highlighter in your personalized learning can help you if you were to miss any crucial information that may be important later and make everything easier to read. Using HiLite will also make studying more personalized with how it highlights and summarizes the information. 76% of teachers say that personalized learning is a priority and HiLite will help with that. HiLite can also be used to create study aids, such as flashcards or summaries.

By highlighting key concepts and ideas within a text, students can use HiLite to create concise and comprehensive study aids. This can be particularly helpful for students who are preparing

for exams or writing assignments, as it allows them to quickly review the key points and refresh their memory. Overall, using HiLite can assist with saving time for students by helping them to quickly identify and understand the most important information in academic texts. By using HiLite to organize information, create study aids, and improve reading comprehension, students can improve their academic performance and achieve their academic goals more efficiently.

Relevant Connections

Artificial Intelligence for Automatic Text Summarization

Min-Yuh Day, Chao-Yu Chen

Paper Summery

The purpose of this study is that we proposed an AI approach for automatic text summarization. We developed an AI text summarization system architecture with three models, namely, statistical model, machine learning model, and deep learning model as well as evaluating the performance of three models.

Relation

By leveraging advanced natural language processing algorithms, machine learning models, and deep learning architectures, we have an insight of how AI-driven text summarization has the potential to significantly improve the efficiency and accuracy of information consumption. The application of AI in text summarization can lead to better comprehension and knowledge retention, as well as increased productivity across a variety of industries and fields, such as academia, journalism, business, and personal information management.

Cycle-SUM: Cycle-Consistent Adversarial LSTM Networks for Unsupervised Video Summarization

National University of Singapore： Li Yuan, Francis EH Tay, Ping Li, Li Zhou, Jiashi Feng

Paper Summery

This paper presents a novel unsupervised video summarization model that requires no manual annotation. The proposed model termed Cycle-SUM adopts a new cycle consistent adversarial LSTM architecture that can effectively maximize the information preserving and compactness of the summary video. It consists of a frame selector and a cycle-consistent learning-based evaluator. The selector is a bi-direction LSTM network that learns video representations that embed the long-range relationships among video frames. The evaluator defines learnable information preserving metric between original video and summary video and “supervises” the selector to identify the most informative frames to form the summary video. In particular, the evaluator is composed of two generative adversarial networks (GANs), in which the forward GAN is learned to reconstruct original video from summary video while the backward GAN learns to invert the processing. The consistency between the output of such cycle learning is adopted as the information preserving metric for video summarization.

Relation

This paper utilizes a method cycle-consistent adversarial training and LSTM networks to create coherent

and representative video summaries without relying on supervised data. Similarly, in the context of AI text summarization, incorporating cycle-consistency and adversarial training techniques could potentially improve the quality of generated summaries by encouraging the model to focus on the most relevant and informative portions of the text. This approach, combined with the inherent capabilities of LSTM networks, could lead to more accurate and contextually aware AI-generated text summaries.

Text Summarization Approaches Using Machine Learning & LSTM

Shobhit Institute of Engineering & Technology, Meerut, India: Neeraj Kumar Sirohi, Dr. Mamta Bansal, Dr.S.N. Rajan

Paper Summery

This article mainly focuses on different ATS (Automatic text summarization) techniques that have been instigated in the present are argue. The paper begins with a concise introduction of automatic text summarization, then closely discussed the innovative developments in extractive and abstractive text summarization methods, and then transfers to literature survey, and it finally sum-up with the proposed techniques using LSTM with encoder Decoder for abstractive text summarization are discussed along with some future work directions.

Relation

This paper offers an in-depth exploration of various machine learning techniques and the application of LSTM networks in text summarization tasks. By analyzing the methodologies, performance, and outcomes discussed in this paper, researchers can gain a comprehensive understanding of the current state of AI summarization and the benefits of utilizing LSTM networks. Furthermore, this paper can inspire new ideas, approaches, and optimizations for future AI summarization research, fostering the development of more accurate, efficient, and contextually aware AI-driven summarization tools.

Potential Flaws & Ethical Risks

As we are designing HiLite There are few ethical risks we need to notice when using AI summarizers continue to gain prominence in information processing. It would be crucial to consider the potential flaws and ethical risks associated with their use. There are some concerns that can impact the quality and reliability of generated summaries, as well as raise questions about data privacy and the potential perpetuation of existing biases.

Aspect

Biases and inaccuracies

AI summarizers rely on training data, which can contain biases and inaccuracies stemming from the original sources. Consequently, these biases may be reflected in the summaries generated by the AI, potentially leading to the propagation of misleading or harmful information. Since many data consists of a variety of information, it would be crucial to choose which training data we need to use.

Loss of context and nuance

Using AI summarizers may struggle to capture the subtleties and complexities of certain topics, resulting possibly in oversimplified or incomplete summaries that misrepresent the original content and it often result to lack of lead to misunderstandings or misinterpretations by users.

Overreliance on AI-generated summaries

As individuals increasingly depend on AI summarizers to process information, there is a risk that deep reading and critical thinking skills may decline. This also applies to writing skills when using a AI writer to write an article or paper. This reliance on summarized content could result in a superficial understanding of complex issues and hinder the development of informed opinions.

Data privacy concerns

As AI summarizers often require access to large amounts of textual data, this raises concerns about data privacy and security. When building out product, we must ensure the protection of sensitive information, particularly in business or legal contexts, it is crucial to maintain trust and compliance with data protection regulations.

Intellectual property and attribution

The use of AI-generated summaries may blur the lines between original content and AI-generated content, creating challenges in attributing ideas and recognizing intellectual property rights. This raises ethical questions about the ownership and appropriate use of summarized content.

Conclusion

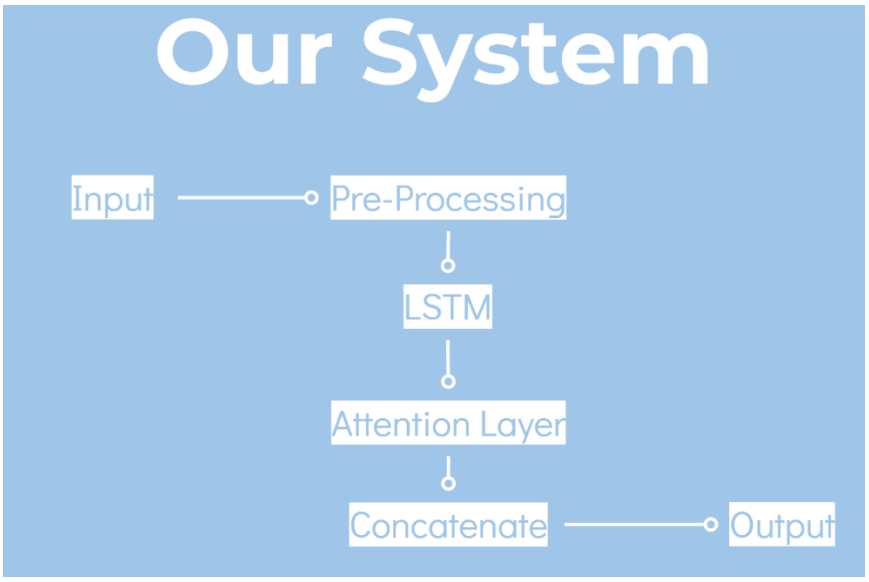
To address these potential flaws and ethical risks, it is essential for us to develop robust and transparent AI summarization algorithms, continuously refining their performance and addressing biases. Moreover, we should foster collaboration between AI developers, users, and regulatory bodies will be crucial in establishing guidelines and best practices for the responsible use of AI summarizers, ultimately ensuring that the technology remains a valuable tool for information processing while minimizing its potential negative impacts.

Solution

LSTM Introduction

As the demand for efficient and accurate information processing continues to grow, advancements in artificial intelligence have paved the way for innovative solutions, such as AI summarizers. One particularly promising approach involves the use of Long Short-Term Memory (LSTM) networks, a type of recurrent neural network architecture specifically designed to handle sequential data and learn long-term dependencies. In this context, the integration of LSTM networks into AI summarizers has the potential to significantly enhance their performance, leading to more contextually aware and coherent summaries. This introduction will explore the benefits of incorporating LSTM networks into AI summarizers and the implications of this cutting-edge technology for the future of information management and consumption.

Our System



To create a text summarization solution using LSTMs, We implement an Encoder-Decoder architecture with attention mechanism. This approach is commonly used in sequence-to-sequence learning tasks, such as machine translation and abstractive summarization.

Data Pre-Processing:

Preprocess your text data (e.g., tokenization, lowercasing, and removing special characters).

Split your dataset into training and validation (and optionally, test) sets.

Convert the text data into sequences of integer values, using a pre-defined vocabulary and word-to-integer mapping. Pad the sequences to have the same length.

LSTM:

Create an Encoder using LSTM layers. The Encoder processes the input text and generates a context vector that captures the information from the input sequence.

Create a Decoder, also using LSTM layers. The Decoder takes the context vector generated by the Encoder and produces the output summary.

Add an Attention mechanism between the Encoder and Decoder. Attention helps the model to focus on different parts of the input sequence when generating the output summary, improving the model's performance.

Attention Layer:

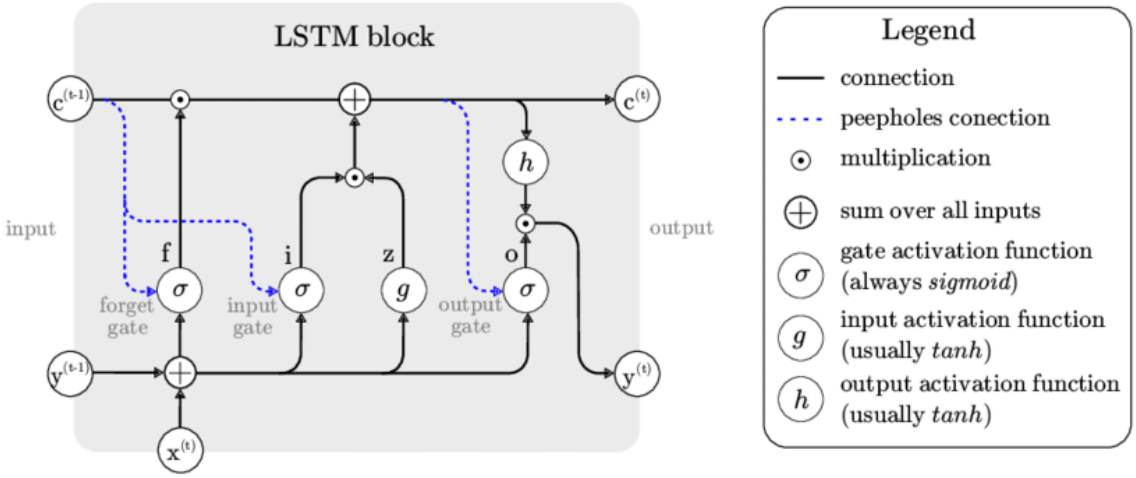
Compile the model using an optimizer (e.g., Adam), a loss function (e.g., categorical cross-entropy), and evaluation metrics (e.g., accuracy).

Train the model on your training set, using the validation set to monitor its performance and prevent overfitting.

Concatenate:

Implement a method for generating summaries from the trained model, such as the beam search or greedy search algorithms.

Use this method to generate summaries for new, unseen text data.



LSTM Component

Memory Cell: The memory cell is the central component of an LSTM cell, responsible for storing the long-term information. It can maintain its state over time, which helps in retaining information for longer durations.

Input Gate: The input gate determines how much of the new input data should be incorporated into the memory cell. It uses a sigmoid activation function, which produces values between 0 and 1, to weigh the importance of new input data. A value close to 0 means the input data is not important, while a value close to 1 indicates that the input data is significant and should be stored in the memory cell.

Forget Gate: The forget gate controls how much of the previous memory cell state should be retained or discarded. It also uses a sigmoid activation function, with values closer to 0 indicating that the past information should be forgotten, and values closer to 1 indicating that the information should be retained.

Output Gate: The output gate determines how much of the memory cell's state should be passed on to the next hidden state. It uses a sigmoid activation function to weigh the importance of the memory cell's

information and then applies an Adam optimizer to the memory cell state, scaling the output between -1 and 1.

LSTM Training

During training, the LSTM learns the optimal weights for each of these gates, which allows it to determine which information is important and should be retained or discarded. This mechanism helps LSTMs to capture long-range dependencies in data effectively and makes them suitable for various applications, such as text generation, sentiment analysis, and time series forecasting.

Despite their effectiveness, LSTMs can be computationally expensive, especially for long sequences. This limitation led to the development of more advanced architectures like the Transformer, which relies on self-attention mechanisms to efficiently process and model long-range dependencies in data. As a result, Transformer-based models have become state-of-the-art in many natural language processing tasks, but LSTMs still have their place in certain applications and continue to be researched and improved upon.

LSTM Conclusion

In conclusion, the incorporation of LSTM (Long Short-Term Memory) networks in AI summarizers presents numerous benefits, enabling the development of more sophisticated and accurate summarization tools. LSTM's ability to learn long-term dependencies and capture the underlying structure of complex textual data allows for a more nuanced understanding of context and meaning. This, in turn, leads to higher-quality summaries that retain the essence of the original content. By leveraging the power of LSTM networks, AI summarizers can be further enhanced to provide even more value to users across diverse industries and contexts. Ultimately, the marriage of LSTM technology and AI summarization promises to contribute significantly to the advancement of information processing, further revolutionizing the way we consume and share knowledge.

Utopias and Dystopias

An interesting topic to cover is the possibility of utilizing HiLite, or a similar product, in the context of both a future utopia and a dystopia.

In terms of a dystopia: Any state controlled by a strict government would have the potential for regular, in-depth surveillance of its citizens. An AI auto-summarizer technology here could be vital in two main different ways. First, the tool could easily be used to make this surveillance of citizens much more efficient. Any piece of writing - whether online or in letter form – that is considered to be suspicious, can be scrutinized quickly by a summarizer. If a long article is fed into the tool, it could easily summarize main concepts, and it would make detection of “illegal” texts very easy. An AI summarize could even be combined with some kind of auto-flagger that would be able to quickly check through any sent email or letter, since instead of reading the whole thing, the flagger could just look through summarized text.

Another way that this product might be used in a dystopia is if the government trains it on a dataset of people’s personal information. After this training has been done, the tool would be capable of quickly summarizing a citizen’s private data for the government and make mass-tracking much easier.

However, if HiLite were to be used for utopian purposes, things would look much different. This scenario is closest to how we envisioned HiLite's ideal usage – to help people gain knowledge quickly and efficiently. An AI-summarizer could be used to get an overview of a topic, identify areas of interest, and focus the user's attention on the most important information. This scenario highlights how AI summarizers/highlighters can be used to enhance or improve educational outcomes in a society.

Conclusion

In conclusion, the utilization of AI summarizers in HiLite has revolutionized the way we process and digest vast amounts of information, offering time-saving and efficient solutions for a variety of industries. HiLite has the potential to greatly impact not only individual productivity, but also collective knowledge sharing and accessibility. However, as we continue to reap the benefits of AI-powered summarization, it is essential to remain vigilant about potential biases and inaccuracies, and to continually refine these technologies to ensure their ethical and responsible use. By striking a balance between leveraging the capabilities of AI summarizers and addressing potential concerns, we can harness their full potential and create a brighter future for information consumption and dissemination.

Certainly, the rise of AI summarizers has fundamentally transformed the landscape of information management and communication. By employing LSTM, these tools are capable of condensing lengthy documents and articles into concise, easily digestible summaries. This enables users to quickly grasp key points and ideas without having to comb through excessive amounts of text.

The application of HiLite extends across various fields, we truly believe in fields such as academia, journalism, business, and even personal use. For instance, researchers can utilize these tools to stay updated on the latest findings in their fields without having to read entire research papers. Journalists, on the other hand, can utilize AI summarizers to extract essential information from multiple sources to create comprehensive news stories. In the business world, these tools can be employed to condense lengthy reports and emails, enabling decision-makers to focus on critical information.

The democratization of knowledge access is another significant advantage of HiLite. By breaking down barriers to information consumption, they allow individuals with diverse backgrounds and reading abilities to engage with complex ideas and stay informed. This can potentially result in a more knowledgeable and empowered society, fostering better decision-making and collaboration.

However, alongside the many benefits of HiLite come several challenges and concerns. For one, the potential for biases and inaccuracies in the summaries generated by these tools must be acknowledged. It is crucial to recognize that AI algorithms are not immune to the biases inherent in the data they are trained on. Consequently, it is essential to ensure that the training data is diverse and representative, and that the algorithms are continuously refined and monitored to mitigate these risks.

Furthermore, HiLite may inadvertently contribute to the decline of deep reading and critical thinking, as individuals become increasingly reliant on summarized content. To counteract this possibility, it is important to strike a balance between using AI-generated summaries for quick information gathering and engaging with original sources for a more in-depth understanding.

In the grand scheme of things, HiLite has the potential to significantly impact the way we process and share information. By addressing the challenges and concerns associated with their use, we can harness their power to create a more informed, efficient, and connected world.

References:

M. -Y. Day and C. -Y. Chen, "Artificial Intelligence for Automatic Text Summarization," 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, USA, 2018, pp. 478-484, doi: 10.1109/IRI.2018.00076.

L. Yuan, F. E. Tay, P. Li, L. Zhou, and J. Feng, “Cycle-SUM: Cycle-Consistent Adversarial LSTM Networks for Unsupervised Video Summarization”, AAAI, vol. 33, no. 01, pp. 9143-9150, Jul. 2019.

Sirohi, Neeraj Kumar, Mamta Bansal, and S. N. Rajan. "Text summarization approaches using machine learning & LSTM." Revista Geintec-Gestao Inovacao E Tecnologias 11.4 (2021): 5010-5026.

Nasim, Zarmeen, and Sajjad Haider. "ABSA toolkit: An open source tool for aspect based sentiment analysis." *International Journal on Artificial Intelligence Tools* 26.06 (2017): 1750023.