VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science (SCOPE)

ARTIFICIAL INTELLIGENCE PROJECT

REVIEW 1 (C1 SLOT)

SUBTITLE GENERATOR AND TRANSLATOR USING AI

Submitted By

Omprakash 21BCE1950

Submitted To

Dr. Vijayaprabakaran K

Abstract :

Our research introduces an innovative AI solution tailored for digital content creators — a Subtitle Generator and Translator designed to address the growing need for accessible and multilingual content. Utilizing advanced Natural Language Processing (NLP) and machine translation techniques, including transformer architectures and recurrent neural networks, our system autonomously generates precise subtitles and seamlessly translates them into multiple languages.

Highlighted features include real-time processing, adaptability to diverse languages and accents, and an intuitive user interface for straightforward customization. The AI models effectively manage intricate audiovisual contexts, such as background noise and overlapping speech, thereby enhancing accuracy and linguistic fluency in comparison to conventional methods.

As the digital landscape continues to evolve, this AI-driven solution emerges as a transformative tool for content creators, ensuring efficient subtitling and widespread accessibility for diverse global audiences. This abstract underscores the pivotal role of AI in promoting inclusivity and advancing the accessibility of multimedia content in our ever-expanding digital era.

Introduction :

The Subtitle Generator and Translator using AI is an innovative solution tailored for digital content creators, designed to address the growing need for accessible and multilingual content. By leveraging advanced Natural Language Processing (NLP) and machine translation techniques, including transformer architectures and recurrent neural networks, the system autonomously generates precise subtitles and seamlessly translates them into multiple languages.
This AI-driven solution is equipped with features such as real-time processing, adaptability to diverse languages and accents, and an intuitive user interface for straightforward customization, making it a transformative tool for content creators, ensuring efficient subtitling and widespread accessibility for diverse global audiences.

The project's objective is to develop an AI-driven system that can rapidly and accurately generate subtitles in real-time, enable multilingual translation, adapt to diverse content, provide a user-friendly interface, enhance accuracy, improve accessibility, and stay at the forefront of technological advancements in AI.
The working procedure of the system involves user input, API authorization and server processing, response generation and timestamp addition, JSON file creation, language translation, SRT file generation and download option, user-friendly interface and multilingual support, automation and efficiency, enhanced accessibility and dynamic output, and educational and entertaining processing of content.

The project's literature review highlights the use of NLP algorithms, machine learning techniques, and deep learning-based tools for subtitle generation, summarization, and synchronization, demonstrating the effectiveness of AI-based solutions in enhancing video accessibility and indexing, improving language teaching, and bridging communication across different language backgrounds.

Objective :

Efficient Subtitling: Develop an AI-driven system that can rapidly and accurately generate subtitles in real-time, ensuring an efficient subtitling process for various audiovisual content types.

Multilingual Translation: Implement advanced machine translation techniques to enable the automatic translation of subtitles into multiple languages, facilitating broader global accessibility and audience reach.

Adaptability to Diverse Content: Create a versatile system capable of handling diverse audiovisual contexts, including varying speaking rates, background noise, and overlapping speech, ensuring high-quality subtitles in complex scenarios.

User-Friendly Interface: Design an intuitive user interface that allows content creators, filmmakers, and producers to easily customize and fine-tune subtitles according to their preferences, ensuring flexibility and creativity in the subtitling process.

Enhanced Accuracy: Utilize state-of-the-art NLP and machine translation models to achieve a higher level of accuracy in transcribing spoken words and translating them into different languages, surpassing the limitations of traditional subtitling methods.

Accessibility Improvement: Address language barriers and enhance the overall accessibility of digital content by providing automated, accurate, and contextually relevant subtitles, catering to a diverse global audience.

Technological Advancement: Stay at the forefront of technological advancements in AI, incorporating the latest models and techniques to continually improve the performance and capabilities of the Subtitle Generator and Translator.

Model Description:

User Input:

Users provide input, including Google Drive or YouTube links.

API Authorization and Server Processing:

APIs validate the link for authorized access.

The server processes the validated link for content insights.

Response Generation and Timestamp Addition:

The server generates a response based on the content.

Timestamps are added to enhance content accessibility.

JSON File Creation:

The timestamped response is converted into a JSON file.

Language Translation and User Input:

Users choose one of 20 supported languages for translation

SRT File Generation and Download Option:

The model creates SubRip Subtitle (SRT) files.

Users decide whether to download or discard the generated SRT file.

User-Friendly Interface and Multilingual Support:

The model maintains a user-friendly interface throughout.

It supports translation into 20 languages for a diverse user base.

Automation and Efficiency:

Automation enhances efficiency in tasks like timestamp addition and translation.

Enhanced Accessibility and Dynamic Output:

Timestamps and translations contribute to content accessibility.

The output adapts dynamically to user preferences.

Educational and Entertaining:

The model's processing of content adds an educational and entertaining dimension.

Literature Corpus:

1.NLP-Driven Ensemble-Based Automatic Subtitle Generation and Semantic Video Summarization Technique

Problem Statement:

The traditional subtitling process is time-consuming, labor-intensive, and often inaccurate, leading to a lack of accessibility and inclusivity for diverse global audiences. Conventional methods also struggle to handle complex audiovisual contexts, such as background noise and overlapping speech, further reducing accuracy and fluency. Moreover, the need for multilingual content is growing, and traditional subtitling methods are often unable to provide accurate translations, leading to language barriers and reduced accessibility.

Proposed Solution:

The proposed solution is an AI-driven Subtitle Generator and Translator that utilizes advanced NLP and machine translation techniques to autonomously generate precise subtitles and seamlessly translate them into multiple languages. The system is designed to be adaptable to diverse languages and accents, and it can handle intricate audiovisual contexts, thereby enhancing accuracy and linguistic fluency in comparison to conventional methods. The system also features an intuitive user interface for straightforward customization, making it a transformative tool for content creators, ensuring efficient subtitling and widespread accessibility for diverse global audiences.

Approach:

Speech recognition can be used to generate subtitles for videos without them, and natural language processing (NLP) algorithms can be employed for summarization. The proposed model uses NLP-based text summarization algorithms and an Ensemble Technique to summarize videos using subtitles, ensuring a concise and accurate summary.

The use of NLP Algorithms proved to be a very efficient way to form abstracts of videos. The case of no subtitles was by using subtitle generation method to convert speech to text, which turned out to be of great use in normal day to day usage and were successful in attracting more users to the site.

Authors: V. B. Aswin, Mohammed Javed, Parag Parihar, K. Aswanth, C. R. Druval, Anupam Dagar, & C. V. Aravinda

Year of Publication: 2021

Journal: AIDE 2019: Advances in Artificial Intelligence and Data Engineering

2.Subtitle Generation and Video Scene Indexing using Recurrent Neural Networks

Problem Statement:

The accessibility and usability of video content are often hindered by the absence of subtitles and efficient indexing mechanisms. Manual subtitle generation is time-consuming, and existing methods for video scene indexing may lack accuracy and scalability. Consequently, there is a need for automated techniques that can generate subtitles accurately and facilitate efficient video scene indexing, thereby enhancing accessibility and user experience with video content.

Proposed Solution:

We propose a solution that utilizes recurrent neural networks (RNNs) for subtitle generation and video scene indexing. By leveraging RNNs, specifically sequence-to-sequence models, we aim to automate the process of generating subtitles for videos and indexing different scenes accurately. Additionally, we explore the implementation of the Connectionist Temporal Classification (CTC) algorithm for error backpropagation, enhancing the training process and improving the overall performance of our solution.

Approach:

Implementing sequence-to-sequence recurrent neural networks for audio data preprocessing, and network architectures. The use of the CTC (Connectionist Temporal Classification) algorithm for backpropagation of errors through time are explored.

sequence-to-sequence recurrent neural networks has the potential to improve accessibility, simplify subtitle creation, and enhancing the search and indexing capabilities of video files. This approach could positively impact the way users interact with video content and improve their overall experience.

Authors: Sajjan Kiran, Umesh Patil, P Siddarth Shankar, Poonam Ghuli
Year of Publication: 2021
Journal: IEEE

3. An Integrated Model for Text to Text, Image to Text and Audio to Text Linguistic Conversion using Machine Learning Approach

Problem Statement:

Cross-modal linguistic conversion, involving text-to-text, image-to-text, and audio-to-text transformations, poses significant challenges, particularly in multilingual contexts such as India. Existing solutions often lack accuracy, especially when dealing with Indian languages, and may not effectively handle diverse input modalities. Consequently, there is a need for an integrated model capable of accurately converting text, images, and audio into text across various Indian languages to facilitate communication and accessibility.

Proposed Solution:

We propose an integrated model that employs machine learning techniques for text-to-text, image-to-text, and audio-to-text linguistic conversions, with a particular focus on Indian languages. By leveraging machine learning, computer vision, and speech recognition technologies, our solution aims to accurately convert diverse input modalities into textual format, thereby enhancing accessibility and communication across language barriers.

Approach:

The model utilizes machine learning techniques for text-to-text, image-to-text, and audio-to-text conversions, with a focus on Indian languages. The model has been tested on large datasets of various Indian languages and uses machine learning, computer vision, and speech recognition.

the model is effective in accurately translating input data, with potential applications including language learning, accessibility, and cross-language communication. The proposed model aims to bridge communication across different language backgrounds.

Authors: Aman Raj Singh, Diwakar Bhardwaj, Mridul Dixit, Lalit Kumar
Year of Publication: 2023
Journal: ISCON, IEEE

4. Artificial Intelligence Technology Supports the Following Research on the Generation of Subtitles for College English Teaching

Problem Statement:

The integration of artificial intelligence (AI) technology into college English teaching presents opportunities for enhancing learning experiences through automated subtitle generation. However, existing methods may lack efficiency and accuracy in generating subtitles tailored to educational contexts. Therefore, there is a need for advanced AI-powered solutions specifically designed to support college English teaching by generating accurate and contextually relevant subtitles.

Proposed Solution:

We propose the development of an advanced AI-powered subtitle generation model tailored for college English teaching. Our solution leverages cutting-edge AI technology, particularly the self-attention mechanism, to construct a robust subtitle generation model. By focusing on the specific requirements of English language instruction, our solution aims to produce subtitles that accurately capture spoken content, facilitate comprehension, and enhance the learning experience for college students.

Approach:

The self-attention mechanism from artificial intelligence technology is utilized to construct the subtitle generation model tailored for college English teaching. Following model training, its performance is evaluated based on metrics such as accuracy, recall, and F1-Score to assess its effectiveness in generating high-quality subtitles.

The application of the self-attention mechanism in constructing the subtitle generation model has yielded promising results, particularly in the context of enhancing college English teaching. The model demonstrates high levels of accuracy, recall, and F1-Score, indicating its efficacy in generating subtitles that meet the specific requirements of educational settings.

Authors: Guangming Zou, Yingying Qiu
Year of Publication: 2022
Journal: ACAIT (IEEE)

5) Deep-Sync: A novel deep learning-based tool for semantic-aware subtitling synchronization

Problem Statement:

Existing methods for subtitle synchronization often face challenges in accurately aligning subtitles with audio-visual content in real-time. Traditional approaches may lack semantic understanding and struggle to adapt to diverse language nuances and speech variations, leading to inaccurate or delayed subtitles. Additionally, accessibility to content for individuals with hearing impairments or those in noisy environments is limited due to suboptimal synchronization.

Proposed Solution:

We propose Deep-Sync, a novel deep learning-based tool designed specifically for semantic-aware subtitling synchronization. By integrating advanced deep language representation models with real-time voice recognition software, Deep-Sync aims to revolutionize the process of subtitle alignment. This tool seeks to enhance the accuracy and efficiency of subtitle synchronization, ensuring seamless integration with audio-visual content across various languages and speech patterns.

Approach:

The methodology involves integrating a deep language representation model and real-time voice recognition software to create Deep-Sync, an intelligent alignment tool for subtitles and audio-visual content.

Deep-Sync, offers a promising solution to the challenges faced in real-time subtitle generation. By integrating a deep language representation model and real-time voice recognition software, it successfully improves synchronization and accessibility of content.

Authors: Alejandro Martín, Israel González-Carrasco, Victor Rodriguez-Fernandez, Mónica Souto-Rico, David Camacho & Belén Ruiz-Mezcua

Year of Publication: 2021

Journal: S.I. : Data Fusion in the era of Data Science

6) Application of Translation Technology based on AI in Translation Teaching

Problem Statement:

Traditional methods of teaching translation often face challenges in providing students with practical, real-world experience and feedback. Conventional approaches may lack the dynamic adaptability required to keep pace with evolving language nuances and technological advancements. Furthermore, there is a growing need for effective tools and methodologies to enhance the quality and efficiency of translation education.

Proposed Solution:

This paper proposes the application of Translation Technology based on Artificial Intelligence (AI) as a solution to address the shortcomings of traditional translation teaching methods. By integrating AI-driven neural machine translation (NMT) and statistical machine translation (SMT) algorithms into translation pedagogy, students can benefit from immersive learning experiences that closely resemble professional translation tasks. This approach aims to bridge the gap between theoretical knowledge and practical application, empowering students with the skills and tools needed to excel in the rapidly evolving field of translation.

Approach:

Utilized neural machine translation (NMT) and statistical machine translation (SMT) algorithms.

Comparative analysis showed a 97% accuracy rate, surpassing traditional methods, improving student scores, and garnering teacher satisfaction.

AI-based translation technology profoundly impacts language teaching, offering innovative strategies for addressing challenges and opportunities.

Experimental tests in Chinese-English bilingual translation demonstrate the technology's efficacy, with future research aiming for broader applications.

Authors: Yu Yuxiu

Year of Publication: 2024

Journal:Systems and Soft Computing

7) End-to-End Learning of Video-Based Text Generation From Multimodal Inputs

Problem Statement:

Traditional methods for generating text from video inputs often face limitations in effectively leveraging multimodal information and achieving accurate, contextually relevant results. Existing approaches may struggle to integrate diverse visual and textual cues seamlessly, leading to suboptimal performance in tasks such as video captioning, question answering, and audio-visual scene-aware dialog. Furthermore, the lack of end-to-end learning frameworks hinders the holistic understanding and generation of text from complex video data.

Proposed Solution:

This paper proposes an innovative solution for end-to-end learning of video-based text generation from multimodal inputs. By applying a novel differentiable tokenization mechanism and integrating an autoregressive decoder, the proposed framework surpasses existing models in generating open-ended text from videos. This approach aims to address the limitations of traditional methods by providing a unified framework based on transformer networks, capable of effectively processing and synthesizing information from both visual and textual modalities.

Approach:

Applied a differentiable tokenization mechanism for end-to-end training.
Integrated an autoregressive decoder for open-ended text generation, surpassing existing models.
Presents a unified framework for multimodal text generation using transformer networks, outperforming state-of-the-art models.
Demonstrates superior performance in video-based text-generation tasks like captioning, question answering, and audio-visual scene-aware dialog.

Authors: Xudong Lin, Gedas Bertasius, Jue Wang, Shih-Fu Chang, Devi Parikh, Lorenzo Torresani
Year of Publication: 2021
Journal: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 7005-7015

8) A Speech-to-Subtitles corpus

Problem Statement:

Existing speech-to-subtitles corpora for training Neural Machine Translation (NMT) systems often lack comprehensiveness and quality, hindering the development of accurate and robust automated subtitling solutions. The scarcity of task-specific training data poses a significant challenge, limiting the performance and scalability of NMT models in generating subtitles from spoken content. Additionally, there is a need for multilingual corpora that can cater to diverse language pairs and support comprehensive NMT solutions for subtitling across various languages and domains.

Proposed Solution:

This paper proposes the development of a Speech-to-Subtitles corpus, aimed at addressing the deficiencies in existing corpora for automated subtitling using NMT. Leveraging Neural Machine Translation techniques, the proposed corpus

introduces MuST-Cinema, a multilingual speech translation dataset designed to provide high-quality training data for NMT-based subtitling systems. By focusing on task-specific requirements and multilingual support, MuST-Cinema aims to fill the gaps in subtitling corpora, facilitating the development of more accurate and robust automated subtitling solutions across different languages and domains.

Approach:

Leverages NMT for automated subtitling, emphasizing the need for high-quality, task-specific training data.

Introduces MuST-Cinema as a multilingual speech translation corpus for comprehensive NMT solutions.

Introduces MuST-Cinema to address gaps in subtitling corpora for Neural Machine Translation (NMT) solutions.

Proposes an automatic method for annotating other corpora, showcasing the ability to segment full sentences into subtitles.

Authors: Alina Karakanta, Matteo Negri, Marco Turchi

Year of Publication: 2020

Journal: Fondazione Bruno Kessler, Via Sommarive 18, Povo, Trento- Italy, University of Trento, Italy

9) An attention enhanced sentence feature network for subtitle extraction and summarization

Problem Statement:

Existing methods for video subtitle extraction and summarization often struggle to capture and prioritize critical information, leading to suboptimal summary generation and inefficient video retrieval systems. Traditional approaches may overlook important contextual cues and fail to produce concise and informative subtitles, hindering the effectiveness of video content indexing and retrieval. There is a pressing need for advanced techniques that can accurately identify salient sentences from video transcripts and generate coherent summaries to improve the efficiency of video retrieval systems.

Proposed Solution:

This paper proposes an attention-enhanced sentence feature network for subtitle extraction and summarization, aimed at addressing the shortcomings of traditional methods in video content indexing and retrieval. By leveraging Convolutional Neural Networks (CNNs) and Bidirectional Long Short-Term Memory (Bi-LSTM) Networks, the proposed model captures critical information for identifying important sentences in video transcripts. Salient sentence scores obtained through the multiple attention mechanism guide the summary generation process, facilitating the creation of concise and informative summaries for improved video retrieval efficiency.

Approach:

Leverages both CNNs and Bi-LSTM Networks to capture critical information for sentence importance identification.

Salient sentence scores guide the summary generation process, enhancing the efficiency of video retrieval systems.

Introduces a novel multiple attention mechanism for video subtitle summarization using CNNs and Bi-LSTM Networks.
Experimental results showcase the model's superiority over baseline and state-of-the-art methods in terms of ROUGE-1, ROUGE-2, and ROUGE-L scores.

Authors: Chalothon Chootong, Timothy K. Shih, Ankhtuya Ochirbat, Worapot Sommool, Yung-Yu Zhuang
Year of Publication: 2021
Journal: Expert Systems with Applications

10) Direct Speech Translation for Automatic Subtitling

Problem Statement:

Traditional methods for automatic subtitling often lack efficiency and accuracy in translating direct speech, leading to suboptimal subtitle generation for multimedia content. Existing approaches may struggle to capture the nuances of spoken language and produce coherent translations, particularly in educational videos where clear and accurate subtitles are crucial for accessibility and comprehension. There is a pressing need for advanced techniques that can accurately translate direct speech and generate high-quality subtitles to enhance the accessibility and effectiveness of multimedia content.

Proposed Solution:

This paper proposes a novel approach for direct speech translation aimed at improving automatic subtitling in multimedia content, particularly in educational videos. By introducing a groundbreaking multiple attention mechanism and leveraging GloVe embeddings for text data conversion, the proposed model addresses the challenges associated with direct speech translation. The approach combines Convolutional Neural Networks (CNNs) and Bidirectional Long Short-Term Memory (Bi-LSTM) networks to enhance subtitle summarization, ensuring accurate and coherent translations of direct speech in multimedia content.

Approach:

Utilizes GloVe for converting text data into sentence matrices and introduces a groundbreaking multiple attention mechanism.

Experiments demonstrate superior performance over baseline and state-of-the-art models, affirming the effectiveness of the proposed approach.

Introduces a multiple attention mechanism, combining Convolution Neural Networks (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) for subtitle summarization in educational videos.

The novel approach outperforms previous methods in educational video summarization, marking a significant advancement.

Authors: Sara Papi, Marco Gaido, Alina Karakanta, Mauro Cettolo, Matteo Negri, Marco Turchi

Year of Publication: 2023

Journal: Transactions of the Association for Computational Linguistics

11) Integrated Framework for Automatic Bilingual Subtitle Generation for MOOC Lecture Videos.

Problem Statement:

The creation of bilingual subtitles for Massive Open Online Course (MOOC) lecture videos presents significant challenges in terms of time consumption and accuracy. Traditional manual methods for subtitle generation are labor-intensive and time-consuming, leading to delays in content availability and potential errors in translation. Additionally, ensuring consistency and accuracy across bilingual subtitles poses a considerable challenge. There is a need for an integrated framework that can automate the bilingual subtitle generation process for MOOC lecture videos, thereby reducing time expenditure and improving accuracy without compromising quality.

Proposed Solution:

This paper proposes an integrated framework for automatic bilingual subtitle generation for MOOC lecture videos to address the challenges associated with manual subtitle creation. By leveraging Automatic Speech Recognition (ASR), Sentence Boundary Detection (SBD), and Machine Translation (MT) techniques, the proposed framework automates the process of bilingual subtitle generation. This approach aims to streamline the subtitle creation process, reduce time expenditure, and improve accuracy by utilizing automated tools for transcription and translation.

Approach:

Utilizes Automatic Speech Recognition (ASR), Sentence Boundary Detection (SBD), and Machine Translation (MT).

Quantitatively evaluates auto-generated subtitles and manually produced subtitles for accuracy and time expenditure.

Auto-generated subtitles in the original language save 54% of working time with a 54.3% error rate reduction.

Proposed framework shortens bilingual subtitle preparation time by approximately 1/3 without quality decline.

Authors: Xiaoyin Che, Sheng Luo, Haojin Yang, Christoph Meinel

Year of Publication: 2017

Journal: IEEE 17th International Conference on Advanced Learning Technologies (ICALT)

12) Subtitle Generation and Video Scene Indexing using Recurrent Neural Networks.

Problem Statement:

Current methods for video subtitle generation and scene indexing often lack efficiency and accuracy, leading to suboptimal video accessibility and indexing capabilities. Traditional approaches may struggle to capture the contextual nuances of video content and accurately generate subtitles or index scenes. Additionally, there is a lack of comprehensive methodologies that address the challenges in designing and training phase for improved video understanding. There is a pressing need for advanced techniques that can efficiently generate subtitles and index scenes in videos, thereby enhancing video accessibility and indexing capabilities.

Proposed Solution:

This paper proposes a solution for video subtitle generation and scene indexing using Recurrent Neural Networks (RNNs), aiming to address the deficiencies in current methods and improve video accessibility and indexing capabilities. By leveraging sequence-to-sequence RNN architectures, the proposed solution facilitates the generation of accurate subtitles and indexing of video scenes. The solution encompasses various stages, including preprocessing audio data, feature extraction, network architectures, the Connectionist Temporal Classification (CTC) algorithm, and suitable evaluation metrics, to ensure comprehensive coverage and effectiveness.

Approach:

Utilizes sequence-to-sequence recurrent neural networks for video subtitle generation and scene indexing.

Includes preprocessing audio data, extracting features, network architectures, CTC algorithm, and suitable evaluation metrics.

Enhances video accessibility and indexing through sequence-to-sequence models. Addresses challenges in designing and training phase for improved video understanding.

Authors: Sajjan Kiran, Umesh Patil, P Siddarth Shankar, Poonam Ghuli
Year of Publication: 2021
Journal: Third International Conference on Inventive Research in Computing Applications (ICIRCA)

13) Automated Subtitle Generation Using Speech Recognition.

Problem Statement:

Manual subtitle generation for videos is a time-consuming and labor-intensive task, often requiring significant effort to transcribe spoken content accurately. This process becomes even more challenging for large volumes of video content, hindering accessibility for individuals with hearing impairments and limiting the reach of multimedia content. There is a need for automated solutions that can efficiently generate subtitles for videos, reducing the reliance on manual transcription and improving accessibility for a wider audience.

Proposed Solution:

This paper proposes an automated solution for subtitle generation using speech recognition technology, aiming to streamline the process and eliminate manual efforts. The proposed solution follows three main stages: audio extraction, speech recognition, and subtitle generation. By leveraging speech recognition algorithms, the proposed solution converts spoken audio from videos into text, which is then used to generate subtitles automatically. This approach aims to enhance accessibility to video content by providing accurate subtitles without the need for manual transcription.

Approach:

Follows three stages: audio extraction, speech recognition, and subtitle generation.
Focuses on automated subtitle generation using speech recognition to eliminate manual efforts.
Proposes an automated solution for video subtitles.
Emphasizes on the importance of metadata and automatic classification for video understanding.

Authors: Prachi Sharma, Manasi Raj, Pooja Jangam, Sana Bhati, Prof. Neelam Phadnis
Year of Publication: 2019
Journal: SSRG International Journal of Computer Science and Engineering (SSRG - IJCSE) - Volume 6 Issue 4

14) Artificial Intelligence Technology Supports the Following Research on the Generation of Subtitles for College English Teaching.

Problem Statement:

In college English teaching, the integration of subtitles into educational materials can significantly enhance students' learning experiences by providing additional support for comprehension and language acquisition. However, manual subtitle generation for educational content is time-consuming and resource-intensive. Additionally, existing automated methods may not adequately address the specific requirements and nuances of college English teaching materials. There is a need for advanced technologies that can efficiently generate high-quality subtitles tailored to the context and objectives of college English teaching.

Proposed Solution:

This paper proposes the use of Artificial Intelligence (AI) technology to support the generation of subtitles for college English teaching materials, aiming to improve efficiency and effectiveness in language learning. The proposed solution focuses on leveraging the self-attention mechanism to construct an English teaching subtitle generation model. By utilizing AI techniques, the solution aims to automate the process of subtitle generation while ensuring accuracy and relevance to the educational context. This approach facilitates the integration of subtitles into college English teaching materials, enhancing accessibility and comprehension for students.

Approach:

Utilizes the self-attention mechanism for constructing an English teaching subtitle generation model.

Applies the model to sequence and paragraph segmentation in a news dataset.

Self-attention mechanism-based model effectively completes tasks with high accuracy, recall, and FI-Score.

Provides strong support for enhancing English teaching through subtitle generation.

Authors: Guangming Zuo, Yingying Qiu

Year of Publication: 2021

Journal: 5th Asian Conference on Artificial Intelligence Technology (ACAIT)

15. A Novel Approach in the Automatic Generation of Regional Language Subtitles for Videos in English.

Problem Statement:

Despite the widespread availability of videos in English, there is a significant gap in accessibility for individuals who primarily speak regional languages, such as Kannada. Manual generation of subtitles for videos in regional languages is labor-intensive and time-consuming, leading to limited availability of content for non-English speakers. There is a pressing need for automated solutions that can efficiently generate regional language subtitles for videos originally in English, thereby improving accessibility and inclusivity for speakers of regional languages.

Proposed Solution:

This paper proposes a novel approach for the automatic generation of regional language subtitles, focusing specifically on Kannada, for videos originally in English. The proposed solution leverages Natural Language Processing (NLP) and Speech Recognition techniques in computer science to automate the subtitle generation process. By automating the creation of Kannada subtitles, the proposed solution aims to address the accessibility gap and improve the availability of video content for Kannada speakers.

Approach:

Addresses automated creation of Kannada subtitles for videos.
Highlights the use of Natural Language Processing (NLP) and Speech Recognition in computer science for subtitle generation.
Focuses on the significance of subtitles in regional languages.
Emphasizes on the automation of subtitle generation to save time, cost, and labor.

Authors: Bilva Teja R, Nischitha P Chinnari, Prerana Hadadi, Surabhi Shanbhogh, V R Badri Prasad
Year of Publication: 2023
Journal: IEEE 8th International Conference for Convergence in Technology (I2CT)

Reference:

https://ieeexplore.ieee.org/document/9574684 (An Approach for Audio/Text Summary Generation from Webinars/Online Meetings)

https://link.springer.com/chapter/10.1007/978-981-15-3514-7_1 (NLP-Driven Ensemble-Based Automatic Subtitle Generation and Semantic Video Summarization Technique)

https://ieeexplore.ieee.org/document/9731269 (Artificial Intelligence Technology Supports the Following Research on the Generation of Subtitles for College English Teaching)

https://ieeexplore.ieee.org/document/9544837 (Subtitle Generation and Video Scene Indexing using Recurrent Neural Networks)

https://link.springer.com/article/10.1007/s00521-021-05751-y (Deep-Sync: A novel deep learning-based tool for semantic-aware subtitling synchronisation)

https://ieeexplore.ieee.org/abstract/document/8001709 (Integrated Framework for Automatic Bilingual Subtitle Generation for MOOC Lecture Videos)

https://ieeexplore.ieee.org/abstract/document/9544837 (Subtitle Generation and Video Scene Indexing using Recurrent Neural Networks)

https://www.internationaljournalssrg.org/IJCSE/2019/Volume6-Issue4/IJCSE-V6I4P103.pdf (Automated Subtitle Generation Using Speech Recognition)

https://ieeexplore.ieee.org/document/9731269 (Artificial Intelligence Technology Supports the Following Research on the Generation of Subtitles for College English Teaching)

https://ieeexplore.ieee.org/document/10126282 (A Novel Approach in the Automatic Generation of Regional Language Subtitles for Videos in English)

https://www.sciencedirect.com/science/article/pii/S2772941924000012 (Application of Translation Technology based on AI in Translation Teaching)

https://openaccess.thecvf.com/content/CVPR2021/html/Lin_Vx2Text_End-to-End_Learning_of_Video-Based_Text_Generation_From_Multimodal_Inputs_CVPR_2021_paper.html (End-to-End Learning of Video-Based Text Generation From Multimodal Inputs)

https://arxiv.org/abs/2002.10829 (MuST-Cinema: a Speech-to-Subtitles corpus)

https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00607/118115 (Direct Speech Translation for Automatic Subtitling)

https://www.sciencedirect.com/science/article/abs/pii/S0957417421003870 (An attention enhanced sentence feature network for subtitle extraction and summarization)