Education and Information Technologies
https://doi.org/10.1007/s10639-023-11855-w

Abstractive video lecture summarization: applications and future prospects

Irene Benedetto[1] · Moreno La Quatra[1] · Luca Cagliero[1] · Lorenzo Canale[1] · Laura Farinetti[1]

Received: 22 October 2022 / Accepted: 25 April 2023

Abstract
Modern educational technology systems allow learners to access large amounts of learning materials such as educational videos, learning notes, and teaching books. Automated summarization techniques simplify the access and exploration of complex data collections by producing synthetic versions of the original content. This paper addresses the problem of video lecture summarization by means of abstractive techniques. To enhance the accessibility of the video lecture content in challenging contexts or while coping with learners with special needs it produces a synthetic textual summary condensing the key concepts mentioned in the lecture's speech. Unlike prior works based on extractive methods, the proposed method can produce more readable and actionable summaries, not necessarily composed of existing portions of speech content. To compensate the lack of annotated data, it also opportunistically reuses the pretrained models available for meeting summarization. The experimental results achieved on a benchmark dataset show that the proposed method generates more fluent and actionable summaries than prior approaches simply relying on content extraction. Finally, we also envision further applications of summarization techniques to learning content. The future prospects of use of summarization techniques in education have shown to go well beyond video summarization.

✉ Irene Benedetto
irene.benedetto@polito.it

Moreno La Quatra
moreno.laquatra@polito.it

Luca Cagliero
luca.cagliero@polito.it

Lorenzo Canale
lorenzo.canale@polito.it

Laura Farinetti
laura.farinetti@polito.it

[1] Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy

Published online: 16 June 2023

Keyword Learning analytics, Summarization, Blended learning, Educational video lectures

1 Introduction

Learning Analytics (LA) techniques entail measuring, collecting, analyzing, and reporting data about learners and their contexts. The goal is to understand and optimize learning activities and environments (Romero and Ventura, 2020). Since its official launch in 2011, hundreds of LA frameworks have been designed, implemented, and tested (Khalil et al, 2022). Among them, several studies address the analysis of video lectures in higher education.

The use of Online Video Lectures (OVLs) in education has become established with the advent of Massive Open Online Courses (MOOCs) and similar digital educational platforms. Learners nowadays accept OVLs to a high degree (Pedrotti and Nistor, 2014). However, the learning experience can be sub-optimal when (1) Learners have to hold the attention on the same video lecture for a long time, or (2) Content accessibility is limited because, for instance, poor Internet performance, linguistic barriers, or presence of visually-impaired learners.

Automated summarization techniques have recently attracted the attention of the Learning Analytics community. They consist of Machine Learning or Information Retrieval techniques whose main purpose is to automatically generate concise summaries of large data collections. The output summary is expected to be on-topic, highly informative and minimally redundant (El-Kassas et al, 2021). Producing summaries of video lectures not only improves content accessibility but also provide learners' with content outlines that are beneficial for learning performance and engagement (Pi et al, 2022).

Existing video lecture summarization methodologies, e.g., (Lv et al, 2021; Lee et al, 2021; Cagliero et al, 2019)) are *extractive*, i.e., they compose a summary by shortlisting part of the existing content such as the most relevant video or audio speech segments. Here we present an application of text summarization techniques to the audio recording of the lecture's speech.

Extractive text-based approaches have three main drawbacks:(1) The *fluency* and *readability* of the summaries are limited because, in most cases, the order of appearance and the relations between the shortlisted pieces of text does not meet any specific constraint. (2) *Speech content* is inherently *repetitive* and *less organized* thus making the summarization process much more complex. (3) Applying text summarization techniques to the audio speech entails adopting a *speech transcription* tool *first*. This step could accidentally add *noise* or *errors* to the source data.

This paper explores the applicability of *abstractive* summarization techniques to university-level video lectures. The goal is to generate a summary consisting of an abstract of the most salient portions of the video speech transcript. Abstractive techniques, such as Lewis et al (2019); Zhang et al (2019a), rely on supervised Deep Learning architecture capable of generating new pieces of text reflecting the key concepts in the source data. They overcome issues (1) and (2) because they are more likely to produce fluent and organized summaries. As a drawback, abstractive methods require domain-specific training data consisting of on-topic video lectures annotated with the human generated summaries. However, due to privacy reasons, the availability of open-source video datasets is rather limited. To address the above-mentioned issue, we explore the use of a transfer learning approach, which entails reusing summarization models pretrained on video meeting recordings. The results achieved on benchmark datasets (Carletta et al, 2005; Janin et al, 2003) show that combining extractive and abstractive approaches yield qualitative and quantitative performance improvements compared to extractive-only methodologies.

Finally, we also give a look to the future. We examine the role of summarization in data-driven educational technology systems. A review of existing summarization frameworks highlights various ways to exploit summarized knowledge in education (not only for video summarization).

The main paper contributions are summarized below.

- **Abstractive summarization**. It proposes to summarize video lectures using abstractive summarization methods. Unlike prior works based on extractive methods, abstractive models can produce more readable and actionable summaries not necessarily composed of existing portions of speech content. To the best of our knowledge, this paper is the first attempt to use abstractive summarizers in a video lecture summarization framework.
- **Transfer learning**. It opportunistically reuses the pretrained models available for meeting summarization to compensate the lack of annotated data.
- **Empirical validation**. It presents the summarization pipeline and validates the proposed method on benchmark data. The results confirm the applicability of the proposed approach to real video lectures.
- **Prospects of summarization in education**. It examines the state-of-the-art related to summarization in the learning analytics fields and discusses the prospects of extension of existing approaches.

The rest of the paper is organized as follow: Section 2 presents the video lecture summarization methodology based on abstractive techniques and transfer learning and reports the main results achieved on benchmark datasets. Section 3 reviews the state of the art of summarization techniques applied to learning data and Section 4 gives insights into the prospects of use of summarization techniques in education. Finally, Section 5 draws conclusions and discusses the future research agenda.

2 Abstractive video lecture summarization

Given a video lecture, we propose a methodology to produce an *abstractive summary* of its audio speech transcription. Unlike extractive summarization methods, e.g., Lv et al (2021); Yoo et al (2021), which shortlist existing portions of the audio speech content, abstractive summarization focuses on producing new pieces of text reflecting the most relevant speech content.

Abstractive summarization of learning content is particularly challenging because entails training ad hoc deep learning models on a large human-annotated datasets. Due to the high variety of learning contents and environments and the concerns on data privacy and intellectual property, currently there is a *lack of pre-trained summarization models* tailored to learning scenarios.

We propose to overcome the aforesaid issue by leveraging *transfer learning*, i.e., we explore the portability of pre-trained video meeting summarization models to abtractive video lectures.

2.1 Dataset

We manually revised and enriched, with the help of a domain expert, a selection of open-source video lectures available in the MIT OpenCourseWare repository[1]. The original repository consists of a set of educational video courses. Each course comprises a set of video lectures, whereas each lecture is accompanied by a transcript and a description of the video content. Our extended version, hereafter denoted as *EduSum*, enriches the university-level video lectures with the corresponding (human-generated) summary. Summaries consist of refined (and more concise) versions of the original video lecture descriptions.

The courses shortlisted in the EduSum benchmark have the following characteristics:

- Each video lecture in the course includes the speech transcription, which will be used as input for the text summarization process.
- The speech transcriptions are enhanced to include punctuation marks for better content understanding.
- Each video lecture includes a refined version of the textual description, which will be hereafter used as reference abstractive summary of the lecture.
- Each course covers a different topic.

The average number of words in the transcriptions is 10170 (with a standard deviation of 3688 words), whereas the average summary length is 100 words (with a standard deviation of 60 words). Table 1 outlines the main features of EduSum dataset.

Comparison with existing datasets in the educational domain Table 2 compares EduSum with other existing benchmarks. The majority of freely accessible datasets are primarily focused on the Information and Communication Technology domain (Fujii et al, 2008; Miller, 2019; Abhilash et al, 2021; Lv et al, 2021), whereas EduSum covers a wide range of different topics according to the International Standard Classification of Education (ISCED)[2], ranging from *Humanities and Arts* to *Blockchain and Money*. Similar to EduSum, Lv et al (2021) collects lectures from VideoLectures.NET for summarization purposes. Unlike EduSum, they focus on the task of extractive summarization, where the system identifies important snippets from the lecture, which

1 https://ocw.mit.edu/about/ Latest access: August 2022

2 http://uis.unesco.org/en/topic/international-standard-classification-education-isced Latest access: May 2022

Table 1 EduSum: detailed characteristics

Course ID	ISCED level	ISCED-F 2013 broad education field	Num. of lectures
STS-081	6	Social sciences, journalism and information	12
21L-011	6	Humanities and Arts	20
5-111SC	6	Natural sciences, mathematics and statistics (Physical science - Chemistry)	35
6-006	6	Information and Communication Technologies	24
6-S897	7	Information and Communication Technologies	23
7-91J	6-7	Information and Communication Technologies	20
15-S12	7	Social sciences, journalism and information (Economics)	22

Table 2 Comparison between EduSum and existing educational video datasets

Paper	Content description	Summary	ICSED category	Dimension
Fujii et al (2008)	Corpus of Japanese Lecture Contents	N.A.	ICT	40
Miller (2019)	Lectures from Udacity	N.A.	ICT	2
Abhilash et al (2021)	National Programme on Technology Enhanced Learning	N.A.	ICT	4
Lv et al (2021)	Lectures from VideoLectures.NET	Extractive (from slides)	Mainly ICT	9616
EduSum	Lectures from MIT courses	Abstractive (from video descriptions)	Stratified over the following categories: (1) Social sciences, journalism and information. (2) Humanities and Arts. (3) Natural sciences, mathematics and statistics (Physical science - Chemistry). (4) Information and Communication Technologies	156

are then concatenated to compose the summary. Hence, the reference summaries are generated by selecting the portions of the speech transcripts that best match the slides content. Conversely, EduSum is designed for abstractive summarization. Hence, it relies on human-written descriptions of the video content, which typically contain a higher level of abstraction, and are not tied to specific portions of the audio transcript or presentation slides.

2.2 Summarization pipeline

To address the abstractive video lecture summarization task we adopt the pipeline depicted in Fig. 1. Firstly, we generate the audio speech transcription from the raw educational video content. Secondly, we apply a cleaning process to fix errors and reconstruct punctuation marks (whether they are not automatically reconstructed by the transcription process). Thirdly, we adopt a sentence-level filter, based on extractive summarization techniques, to remove the content that unlikely conveys relevant information. Finally, an abstractive summarization model, trained on video meeting summarization data, is executed on top of the extracted sentences to generate the abstractive summary of the video lecture.

A more detailed description of each step follows.

1. **Speech-to-text transcription**: We first apply automatic speech recognition to transcribe the audio of the video lectures. The goal is to produce a textual version of the lecturers' speech. This step relies on dedicated services (e.g., the Google Cloud Platform APIs[3]).
2. **Punctuation restoration and tokenization**: This step aims at fixing errors and incomplete punctuation of the transcribed text. It also entails splitting the textual content into sentences. This procedure is instrumental for the extractive summarization step, which requires tokenization of the source text to extract the salient portions of the original content. To restore punctuation marks we rely on the following state-of-the-art libraries:

   - The *Punctuator* (Tilk and Alumäe, 2016)
   - The *FastPunct* library[4].
   - Transformer-based model, first proposed by Wang et al (2018) and then extended by Alam et al (2020).

3. **Extractive summarization**. This step applies a text summarization technique to shortlist the most relevant sentences in each transcription. Notice that content extraction is just a preliminary step (i.e., the extracted text is *not* the output of the pipeline). The aim is twofold: (1) Remove the redundant or less relevant content present in the audio transcription (2) Limit the transcription length by condensing the key information into a few sentences thus enabling the subsequent abstractive summarization step. To this aim, we apply the following well-established algorithms:

[3] https://cloud.google.com/speech-to-text Latest access: May 2022

[4] https://pypi.org/project/fastpunct/ Latest access: May 2022

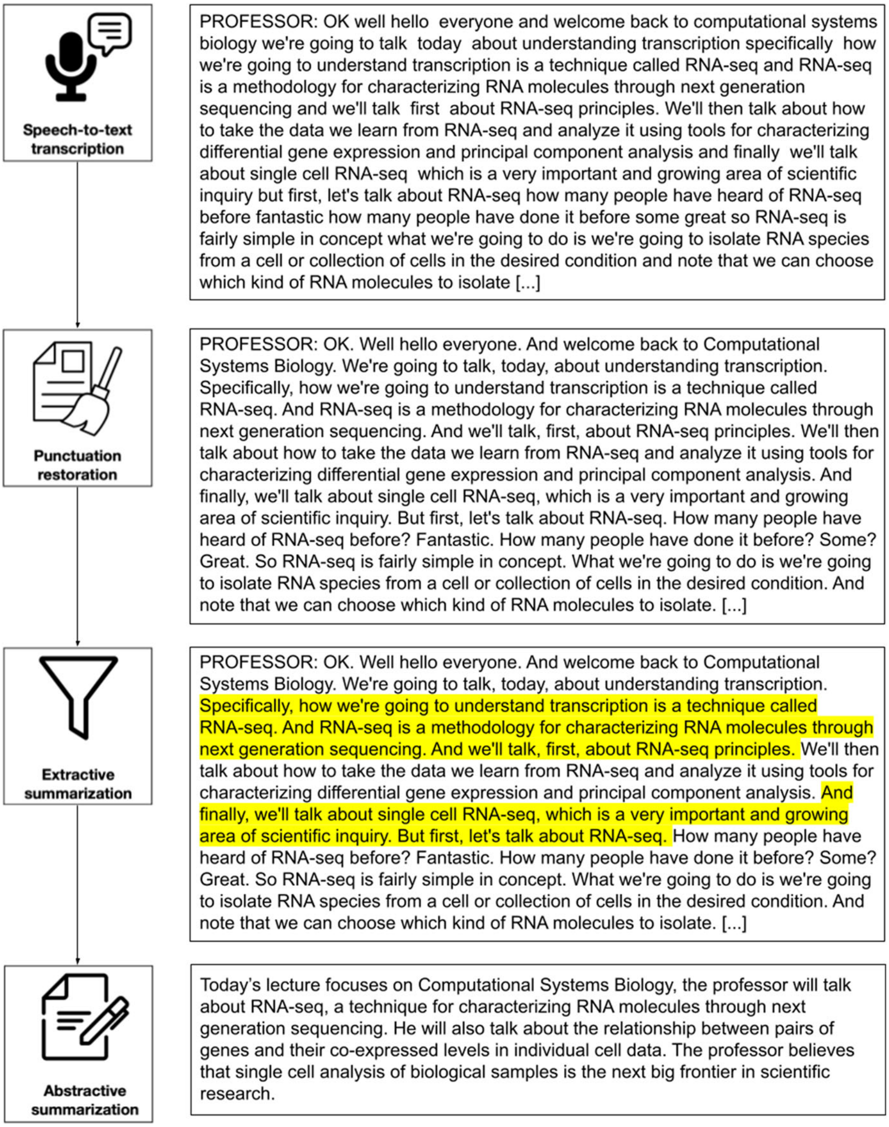

Fig. 1 Sketch of the summarization pipeline

- *LSARank*: It relies on Latent Semantic Analysis to select the sentences that best represent the most salient topics.
- *TextRank* (Mihalcea and Tarau, 2004)[5]: It first models the tokenized sentences as nodes in a graph and then applies a popular graph ranking strategy (Page et al, 1999) to evaluate the sentence authoritativeness in the analyzed text.

[5] https://pypi.org/project/pytextrank/ Latest access: June 2022

- *Cluster-BERT* (Miller, 2019)[6]: it leverages the semantic representations obtained by the popular BERT encoder (Devlin et al, 2019) to cluster input sentences using their semantic similarity. Next, it picks the cluster representatives to compose the extractive summary.

4. **Abstractive summarization**. It produces a video lecture summary consisting of a rephrase of the previously extracted content. The purpose is to generate a fluent description of the video lecture content. We leverage the state-of-the-art neural network-based summarization model, namely BART (Lewis et al, 2019), pre-trained on the SAMSum dataset (Gliwa et al, 2019), which contains spoken conversation transcripts. The pre-training step is aimed at learning the key properties of spoken conversation transcripts. However, the model requires an ad hoc fine-tuning stage to further specialize the summarizer on the main lexical features. To overcome the lack of domain-specific data, we fine-tune the model on a benchmark video meeting summarization dataset, i.e., the AMI (Carletta et al, 2005) meeting corpus. The dataset contains transcriptions of business meetings as well as their corresponding summaries. Although the application domain is different from those observed in educational video lectures, they share high-level lexical features (e.g., they both contain transcriptions including one or more speakers and formal speaking style). To run BART (Lewis et al, 2019) we rely on the open-source implementation of available on the `transformers` library[7]. Specifically, we use the *bart-large-cnn-samsum* checkpoint, which corresponds to the largest configuration of BART (400 millions parameters) trained on a mix of news articles (Hermann et al, 2015) and dialogue datasets (Gliwa et al, 2019).

2.3 Qualitative summary examples

Table 3 reports a qualitative comparison between the automatically and human-generated summaries. The summary generated by our pipeline captures the most salient topics covered by the video lecture and expresses them concisely. The summary is easy to read, even though it style is formal. The last summary sentence helps learners to better understand the importance of the underlying topic. Notice that the summary is a little bit more concise than the description and ignores secondary aspects like the usability of principal component analysis in RNA-seq data analytics.

2.4 Quantitative summary evaluation

We adopt established metrics to evaluate the quality of the summarization process. Specifically, we compute the syntactic and semantic similarities between the human-generated summary and the automatically generated ones. Intuitively, the more similar the expected output and the automatically generated summaries the better the summarization process (Chandrasekaran and Mago, 2021).

[6] https://pypi.org/project/sentence-transformers/0.3.0/ Latest access: June 2022

[7] https://huggingface.co/docs/transformers/ Latest access: September 2022

Table 3 Examples of summaries: qualitative comparison

Summary source	Content
Handwritten description (ground truth)	This lecture is about RNA-seq (RNA sequencing), a method of characterizing RNA molecules (isoforms) through next-generation sequencing. He begins with the principles of RNA-seq, and then moves on to how to analyze the data generated by RNA-seq. The professor explains how it is possible characterize expressed genes via principal component analysis. He ends the lecture talking about the benefits and challenges of working with single cells for RNA-seq.
EduSum (our)	Today's lecture focuses on Computational Systems Biology, the professor will talk about RNA-seq, a technique for characterizing RNA molecules through next generation sequencing. He will also talk about the relationship between pairs of genes and their co-expressed levels in individual cell data. The professor believes that single cell analysis of biological samples is the next big frontier in scientific research.
Handwritten description (ground truth)	During this class the professor continues the discussion of Charlie Chaplin, comparing his films to those of Buster Keaton. The Films discussed during the class are: Keaton's Cops, Chaplin's The Gold Rush, City Lights, and particularly Modern Times. The class is focused on analyzing the context, the sound, the structure and the complexity of Modern Times.
EduSum (our)	Modern Times is a Chaplin film. It invokes the tradition of silent films and alludes to earlier Chaplin films. The soundtrack has a quality in which particular themes recur when certain characters appear on screen, and you begin to associate certain melodies with certain characters. Modern Times manages to dramatize that the principle of repetition never ends, that Charlie is on a kind of treadmill, and that talk and talking films are associated with what the film identifies as evil or dangerous. The Last Laugh is Murnau's silent film from Germany.
Handwritten description (ground truth)	The lecture introduces the concept of pH and we measure the pH of various common solutions. The topic covered by the professor are the following: definitions and relationships between pKw, pH, and pOH, strengths of acids and bases and equilibrium acid-base problems (weak acids and weak bases).
EduSum (our)	There are five types of acid-base problems: weak acid in water, weak basin water, strong acids in water and strong bases in water. Water is an important solvent because it acts as an acid and a base. The relationship between pKw, pH, and pOH, the strengths of acids and bases, and the equilibrium Acid-Base problems are discussed today. Next week, the class will do strong acids and strong bases, and then they will have all five kinds of equations.

Notice that the EduSum dataset already includes the speech transcriptions (see Table 1). Hence, part of the preprocessing stages are not necessary. However, the presented pipeline is general enough to handle arbitrary video lectures, including those in which speech transcriptions are missing or of low quality.

Syntactic scores ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is the most established summary evaluation metric (Lin, 2004). It computes the overlap between human- and model-generated summaries by counting the number of overlapping textual units. The ROUGE metrics considered in this study are defined as follows:

- ROUGE-N: let R_S be the set of reference summaries, and let S be the model output.

$$\text{ROUGE-N} = \frac{\sum_{S \in R_S} \sum_{gram_n \in S} count_{\text{match}}(gram_n)}{\sum_{S \in R_S} \sum_{gram_n \in S} count(gram_n)} \tag{1}$$

  where n represents the length of the n-$gram$ and $count_{\text{match}}(gram_n)$ is the number of times in which the n-gram occurs in both reference summaries and model output.
- ROUGE-L: a variation on the ROUGE-N metric, which evaluates the Longest Common Subsequence (LCS) by computing the length of the longest sequence of words that are common to the model output S and the reference summaries R_S.

Semantic score BERTScore Zhang et al (2019b) is an established evaluation metric based on contextual embedding (Devlin et al, 2019), which is commonly used to capture semantic text similarities. It is computed as follows:

1. Firstly, BERT generates the contextual embeddings of both reference and candidate summaries, represented by a sequence of vectors $\langle \mathbf{x}_1, ..., \mathbf{x}_k \rangle$ and $\langle \hat{\mathbf{x}}_1, ..., \hat{\mathbf{x}}_l \rangle$.
2. Secondly, it computes the token-level similarity using a pre-normalized cosine similarity $\mathbf{x}_i^T \hat{\mathbf{x}}_j$.
3. Thirdly, it computes the recall, precision, and F1-measure scores. The recall is defined as the number of tokens in $\mathbf{x}$ that match with tokens in $\hat{\mathbf{x}}$. To compute the precision it looks for the matching of each token in $\hat{\mathbf{x}}$ to a token in $\mathbf{x}$. Finally, the F1-measure is computed as the harmonic mean between precision and recall. The formulae for each metric are defined as follows:

$$R_{BERT} = \frac{1}{\|\mathbf{x}\|} \sum_{x_i \in \mathbf{x}} \max_{\hat{x}_j \in \hat{x}} \mathbf{x}_i^T \hat{\mathbf{x}}_j \tag{2}$$

$$P_{BERT} = \frac{1}{\|\hat{\mathbf{x}}\|} \sum_{\hat{x}_i \in \hat{\mathbf{x}}} \max_{x_j \in \hat{x}} \mathbf{x}_i^T \hat{\mathbf{x}}_j \tag{3}$$

$$F_{BERT} = 2 \frac{P_{BERT} \cdot R_{BERT}}{P_{BERT} + R_{BERT}} \tag{4}$$

4. Finally, the scores are normalized between between 0 and 1.

2.4.1 Baseline models

We compare the summaries generated by the proposed summarization pipeline with those produced by the following baseline methods:

- HMNet (Zhu et al, 2020): to the best of our knowledge, it is the most recent abstractive summarization model tailored to speech transcriptions.

Table 4 Performance comparisons in terms of ROUGE and BERT F1-Scores

Approach	Extractive	Abstractive	ROUGE-1	ROUGE-2	ROUGE-L	BERTScore
Baseline	LEAD	-	0.280*	0.053*	0.160	0.845
Baseline	-	HMNet	0.188*	0.026*	0.17	0.806*
EduSum	BERT	BART-L	0.293	0.050	0.169	0.843
EduSum	LSA	BART-L	0.275	0.042	0.158	0.838
EduSum	TextRank	BART-L	0.298	**0.061**	**0.174**	**0.846**

The best performance results are written in boldface. The starred performance worsening is statistically significant (p-value: 0.05)

- LEAD: A popular extractive summarization method, which consists in shortlisting the very first sentences of the speech transcription. Since at the beginning of the lecture the teacher likely overviews the main topics covered in the lecture, the first sentences are deemed as a valid candidate summary in the learning analytics context.

2.4.2 Hardware and execution times

All the experiments were run on a single NVidia© Tesla© V100 GPU equipped with 32 GB memory.

The inference time ranges between few seconds and few tens of seconds for all the tested methods.

2.5 Quantitative summary evaluation

Table 4 compares the results achieved by the proposed summarization pipeline and the tested abstractive and extractive competitors on the EduSum dataset (see Section 2.1) using both syntactic and semantic evaluation metrics (see Section 2.4). The top-2 rows report the results achieved by the baseline methods, i.e., LEAD (1st row), and HMNet (2nd row). The subsequent rows respectively report the performance of different variants of the proposed summarization pipeline using various extractive summarizer and a fine-tuned abstractive BART model.

The proposed summarization pipeline performs best against HMNet and LEAD in terms of both semantic and syntactic metrics. The TextRank summarizer turns out to be the most effective extractive summarization module. Its performance is superior to that of LEAD in terms of Rouge-2 whereas they are comparable in terms of BERTScore. The reason is that at the beginning of the video lecture teachers likely mention most of the salient concepts covered during the lesson. However, they omit relevant details which can be retained by applying an abstractive summarization stage applied on top of the entire speech transcript.

We have also analyzed the impact of the punctuation restoration step on the quality of the automatically generated summary. Specifically, in Table 5 we report the results achieved by our best performing summarization pipeline relying on transformer-based

Table 5 Effect of punctuation restoration

Approach	Punctuation	BERT-Score	ROUGE-1	ROUGE-2	ROUGE-L
EduSum	No restoration	0.801	0.091	0.023	0.060
	Automatic	0.811	0.095	0.025	0.062
	Original	0.812	0.096	0.026	0.062

Extractive method: TextRank (Mihalcea and Tarau, 2004)

restoration (namely, *Automatic*), without punctuation restoration (namely, *No punctuation*), and by restoring the original punctuation (upper-bound performance limit, namely *Original*). The results show that automatic punctuation restoration is beneficial (*Automatic* is better than on *No punctuation*), despite it is not as reliable as using the original punctuation marks (*Automatic* is slightly worse than *Original*).

3 Summarization in education: an overview

Table 6 enumerates the prior works on summarization of learning data. According to their learning goal, existing applications can be categorized as follows:

- *Content Curation* (CC): Support teachers in creating and updating new learning materials based on the revision and extension of existing contents (e.g., (Lv et al, 2021; Lee et al, 2021)).
- *Decision Making* (DM): Support teachers in assessing the level of knowledge of the learners and planning future activities (e.g., (Miller, 2019; Cagliero et al, 2019)).
- *Personalization* (P) and *Accessibility* (A): Support learners with special needs by providing them with one-to-one support and domain-specific contents (e.g., (Abhilash et al, 2021; Pramudianto et al, 2016; Baralis and Cagliero, 2018)).
- *Learning By Doing support* (LBD): Provide learners with new stimuli by exploiting innovative educational tools and contents (e.g., interactive maps, videos, images) (e.g., (Goularte et al, 2019)).

Existing applications cover specific use cases and learning data types at various levels of detail. However, the progresses of educational technology systems and Learning Analytics frameworks leaves room to numerous future research directions and applications. The prospects of extension will be thoroughly discussed in Section 4.

The methodology presented in this paper focuses on summarizing video lectures. Previous attempts to summarize video lectures have been made by Garg (2017), Lv et al (2021), Yoo et al (2021), Abhilash et al (2021), and Saini et al (2022). Specifically, the systems proposed by Garg (2017) and Yoo et al (2021) summarize the video contents using subtitles and provide students with specific links and keywords for content retrieval. As a drawback, the manual annotation of video subtitles is very time-consuming. Conversely, Lv et al (2021), Abhilash et al (2021), and Saini et al (2022) apply extractive summarization to video lecture transcripts, i.e., the summary consists of existing portions of the lecture's speech transcription. Extractive approaches show inherent limitations in the readability and fluency of the generated summaries. Hence,

Table 6 Prior work on learning content summarization

Paper	Summarization task	Learning goal
Choudary and Liu (2007)	Extract a subset of video frames conveying overall content	Content curation, Accessibility
Fujii et al (2008)	Extractive text summarization of videolectures	Content curation
Pramudianto et al (2016)	Feedback summarization in peer review systems	Decision making
Baralis and Cagliero (2016)	Scientific papers summarization	Content curation
Garg (2017)	Video lecture summarization using subtitles	Accessibility, Learning-by-doing support
Shimada et al (2018)	Lecture slides summarization	Accessibility, Learning-by-doing support
Cagliero et al (2019)	Personalized summarization of teaching materials based on topic	Personalization, Learning-by-doing support
Gottipati et al (2019)	Topic based summarization from class discussion forums	Content curation, Learning-by-doing support
Goularte et al (2019)	Summarizing text assessment	Decision making
Miller (2019)	Extractive text summarization of videolectures content	Content curation
Rahman et al (2020)	Identifying keyframes conveying the overall content	Accessibility, Content curation
Abhilash et al (2021)	BERT-based extractive text summarization of lecture notes and videolecture transcripts	Content curation
Lee et al (2021)	Creating personalized video summaries leveraging student's attention	Personalization
Yoo et al (2021)	Video lecture summarization using subtitles	Accessibility, Learning-by-doing support
Lv et al (2021)	Extractive text summarization of videolectures content	Content curation
Saini et al (2022)	Generation of lecture minutes using extractive summarization techniques	Content curation

the usability of the generated outputs remains limited. In this paper, we overcome the aforesaid issue by adopting an abstractive summarization approach on top of the extracted speech content.

4 Summarization in education: current limitations and prospects

When the interaction between learners and teachers is supported by educational technology systems the use of Learning Analytics (LA) solutions can provide end-users with data-driven insights. Hereafter, we will examine the extent to which summariza-

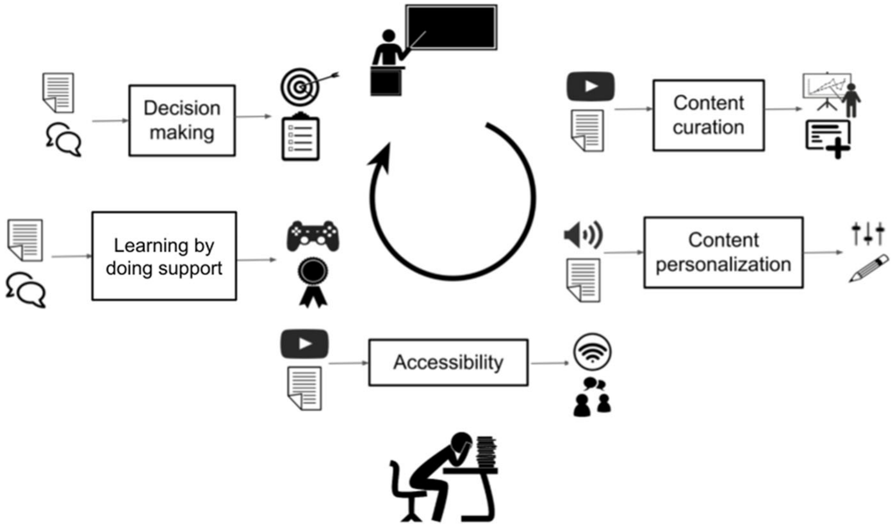

Fig. 2 The learning process: prospects of integration of summarization techniques at different stages

tion techniques could support either teachers or learners in their activities in the near future.

We envision various steps of the learning process in which we can profitably exploit state-of-the-art summarization techniques. Specifically, Fig. 2 depicts a sketch of the learning process, where we highlight the steps in which summarization techniques are deemed as actionable. Notice that the highlighted steps recall the paper categorization given in Table 6 because our purposes are to (1) Clarify the limitations of current approaches, and (2) Position the future research directions in the existing literature. More details on each learning step are given below.

4.1 Content curation

This step concerns the preparation, annotation, and update of the teaching materials. Summaries of teaching materials can be considered as a kind of automatically generated annotations. They can be either provided as additional materials to bridge a learning gap or used to generate new teaching content in support of the frontal lectures (e.g., presentation slides).

The types of available teaching content are rather diversified. They include not only textual data but also videos, audio tracks, images, and code. Furthermore, depending on the learning context the structure of the content and related metadata are highly variable. To address this issue, we envision the following directions of extension:

- The integration of multimodal summarization techniques, e.g., (Zhu et al, 2018), to enrich textual annotations with multimedia content. Examples of multimodal content are the transcript of the learner's questions, the images of the blackboard, and the presentation slides' content.

- The extension of current summarization methods to handle heterogeneous data sources such as educational books, video lectures, slides, scientific papers, and news articles.
- The use of deep learning models, e.g., (Borsos et al, 2022), to generate abstracts consisting not only of written text but also of audio speeches, images, code, etc.

4.2 Decision making

Summarization techniques can be also helpful in docimology, providing teachers with automated tools to extract the salient concepts from learner-generated data such as lecture notes and assignments. For example, to assess the level of knowledge of a learner on a specific topic the summarizer can be used to automate formative assessment procedures. This allows teachers to early identify critical situations and prevent course dropout or exam failures. To this end, we envision the application of query-based summarization techniques (e.g., (Litvak and Vanetik, 2017)) to learner-generated data in order to generate not only generic summaries but also abstracts tailored to specific topics.

4.3 Content personalization

Summaries of teaching documents can be either generic or personalized according to the learners' needs. To generate personalized summaries the use of text summarizers has already been explored (see, for example, (Cagliero et al, 2019). However, each summary provides a high-level description a given topic rather than a more specific insight into specific aspects.

We envision the application of more advanced aspect-based summarization techniques, e.g., (Tan et al, 2020)), which allow us to both condense the key information about a topic by recognizing and well separate the underlying aspects. Contents related to different aspects can be selectively recommended to learners according to their actual needs.

4.4 Content accessibility

Summarization has been recommended as an effective tool for improving accessibility in various domains (Parmanto et al, 2005). Access to educational materials can be hampered by technological, cultural, or linguistic barriers. For example, disabilities may hinder access to particular content types (e.g., text written using small font sizes for visually impaired learners, audio podcasts for deaf students).

Improving content accessibility encompasses (1) The generation of multimodal versions of the teaching content to overcome technological barriers such as the lack of adequate equipment in the labs/rooms. (2) The generation of multilingual content to support foreign learners, foster student exchange, and promote cross-cultural interactions between learners and between teachers and learners. (3) The selection of a reduced amount of teaching content to be shared and use due to, for instance, the presence of limitations in the network bandwidth (e.g., in low-connectivity regions). In this field, we envision the application of recent cross-modal summarization techniques to handle multiple data types, languages, and modalities at the same time. Another open research direction is the application of multilingual and cross-lingual summarization techniques to overcome linguistic barriers. The recent advances in self-supervised learning can be profitably exploited to design more advanced solutions to cross-lingual summarization.

4.5 Learning-by-doing support

Mitchell et al (2017) has highlighted the benefits of *learning by summarizing* lesson content. We envision the integration of automatic summarization techniques in a *learning-by-doing* approach to teach. Specifically, the proposed approach entails the following steps: (1) We apply a text summarizer to a reference textbook to automatically generate topic-specific summaries. (2) The lecturer checks the correctness and completeness of the automatically generated summaries (with a relatively limited human effort). (2) After the lecture on a given topic, we ask learners to write a short summary of the lecture content. (3) We compare the learner-generated summaries with the corresponding ground truth. The cost of the assessment procedure is very limited as relies on automatic tools such as ROUGE-score Lin (2004) and BERTScore Zhang et al (2019b).

5 Conclusions and future work

The paper presented an abstractive method to video lecture summarization. It overcomes the limitations of extractive models, previously used to summarize the lecture's speech, by generating more readable summaries. It also proposes to reuse models pre-trained for meeting summarization under a *transfer learning paradigm*.

The main takeaways from the experimental results on benchmark data are given below.

- Abstractive summarization techniques produce human-readable summaries that are alternative to handwritten descriptions whenever manual annotations are missing.
- Despite there is currently a lack of human-annotated data, transferring summarization models trained on video meetings to the learning analytics context appears to be effective in capturing the core aspects of the video lecture.
- The performance of the proposed summarization pipeline is superior to that of existing abstractive speech transcript summarization models (e.g., HMNet). Integrating an automatic punctuation restoration step into the summarization pipeline appears to be helpful for improving the syntactic and semantic relevance of the generated summaries.

The paper also discussed the prospects of use of summarization techniques in education. The main takeaways are enumerated below.

- Learner-oriented solutions are currently under-using the multimedia content available through education learning systems, in particular MOOCs.
- Most summarization-based approaches to learning analytics are focused on content curation, especially for higher education.
- Content accessibility and personalization have received little attention, even if there is an increasing need to support learners with special needs.
- Although learner-generated content has received increasing attention by the Learning Analytics community, the application of cross-modal summarization techniques to improve learning by doing activities is still open[8].

Author Contributions All authors contributed equally to this research work. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability The datasets analyzed during and/or analysed during the current study are available at https://ocw.mit.edu/

Declarations

Informed consent The paper is an extended version of the preliminary work presented in Benedetto et al. (2022). Unlike the prior work, the current manuscript contains
• An overview of the existing benchmark datasets for video lecture summarization (see Section 2.1 of the current manuscript).
• A more thorough description of the presented methodology (see Section 2.2).
• A validation of the summaries generated from the open-source video lectures available in the MIT OpenCourseWare repository (see Sections 2.3, 2.4, and 2.5).
• A more extensive overview of the related works on summarization in education (See Section 3).
• A discussion of the future prospects of use of summarization techniques in education (See Section 4).

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

Abhilash, R. K., Anurag, C., & Avinash, V. (2021). Lecture video summarization using subtitles. In A. Haldorai, A. Ramu, & S. Mohanram (Eds.), *2nd EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing* (pp. 83–92). Cham: Springer International Publishing.

Alam, T., Khan, A., & Alam, F. (2020). Punctuation restoration using transformer models for high-and low-resource languages. In: *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020). Association for Computational Linguistics, Online* (pp.132–142). https://doi.org/10.18653/v1/2020.wnut-1.18. https://aclanthology.org/2020.wnut-1.18

Atapattu, T., & Falkner, K. (2018). Impact of lecturer's discourse for students' video engagement: Video learning analytics case study of moocs. *J Learn Anal, 5*(3). https://doi.org/10.18608/jla.2018.53.12

Baralis, E., & Cagliero, L. (2016). Learning from summaries: Supporting e-learning activities by means of document summarization. *IEEE Transactions on Emerging Topics in Computing, 4*(3), 416–428. https://doi.org/10.1109/TETC.2015.2493338

Baralis, E., & Cagliero, L. (2018). Highlighter: Automatic highlighting of electronic learning documents. *IEEE Trans Emerg Top Comput, 6*(1), 7–19. https://doi.org/10.1109/TETC.2017.2681655

[8] https://ocw.mit.edu/about/ Latest access: August 2022

Benedetto, I., Canale, L., Farinetti, L., Cagliero, L., & Quatra, M. (2022). Leveraging summarization techniques in educational technology systems. 46th IEEE Annual Computers, Software, And Applications Conferenc, COMPSAC 2022, Los Alamitos, CA, USA, June 27 - July 1, 2022 (pp. 415-416). https://doi.org/10.1109/COMPSAC54236.2022.00068

Borsos, Z., Marinier, R., & Vincent, D. (2022). Audiolm: a language modeling approach to audio generation. https://doi.org/10.48550/arXiv.2209.03143. arXiv:2209.03143

Cagliero, L., Farinetti, L., & Baralis, E. (2019). Recommending personalized summaries of teaching materials. *IEEE Access, 7*:22,729–22,739. https://doi.org/10.1109/ACCESS.2019.2899655

Carletta, J., Ashby, S., & Bourban, S. (2005). *The ami meeting corpus: A pre-announcement.* https://doi.org/10.1007/11677482_3

Chandrasekaran, D., & Mago, V. (2021). Evolution of semantic similarity-a survey. *ACM Computing Surveys, 54*(2), 1–37. https://doi.org/10.1145/3440755

Choudary, C., & Liu, T. (2007). Summarization of visual content in instructional videos. *IEEE Transactions on Multimedia, 9*(7), 1443–1455. https://doi.org/10.1109/TMM.2007.906602

Devlin, J., Chang, M.W., & Lee, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805

El-Kassas, W. S., Salama, C. R., & Rafea, A. A. (2021). Automatic text summarization: A comprehensive survey. *Expert Systems with Applications, 165*(113), 679. https://doi.org/10.1016/j.eswa.2020.113679

Fujii, Y., Yamamoto, K., & Kitaoka, N. (2008) Class lecture summarization taking into account consecutiveness of important sentences. 2438–2441

Garg, S. (2017). Automatic text summarization of video lectures using subtitles. In: Patnaik, S., & Popentiu-Vladicescu, F. (Eds.) *Recent Developments in Intelligent Computing, Communication and Devices* (pp. 45–52). Springer Singapore, Singapore

Gliwa, B., Mochol, I., & Biesek, M. (2019) Samsum corpus: A human-annotated dialogue dataset for abstractive summarization. arXiv:1911.12237

Gottipati, S., Shankararaman, V., & Ramesh, R. (2019). TopicSummary: A Tool for Analyzing Class Discussion Forums using Topic Based Summarizations. In: *2019 IEEE Frontiers in Education Conference (FIE)* (pp. 1–9). IEEE, Covington, KY, USA. https://doi.org/10.1109/FIE43999.2019.9028526. https://ieeexplore.ieee.org/document/9028526/

Goularte, F. B., Nassar, S. M., & Fileto, R. (2019). A text summarization method based on fuzzy rules and applicable to automated assessment. *Expert Systems with Applications, 115*, 264–275. https://doi.org/10.1016/j.eswa.2018.07.047, . https://www.linkinghub.elsevier.com/retrieve/pii/S0957417418304743

Hermann, K.M., Kocisky, T., & Grefenstette, E. (2015). Teaching machines to read and comprehend. In: *NIPS*

Janin, A., Baron, D., & Edwards, J. (2003). The icsi meeting corpus. I–364. https://doi.org/10.1109/ICASSP.2003.1198793

Khalil, M., Prinsloo, P., & Slade, S. (2022). A comparison of learning analytics frameworks: A systematic review. In: *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 152-163). Association for Computing Machinery, New York, NY, USA, LAK22. https://doi.org/10.1145/3506860.3506878

Lee, H., Liu, M., & Riaz, H. (2021). Attention based video summaries of live online zoom classes. https://dblp.org/rec/journals/corr/abs-2101-06328.bib

Lewis, M., Liu, Y., & Goyal, N. (2019) Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. https://aclanthology.org/2020.acl-main.703/

Lin, C.Y. (2004). Rouge: A package for automatic evaluation of summaries. p 10

Litvak, M., & Vanetik, N. (2017). Query-based summarization using MDL principle. In: *Proceedings of the MultiLing 2017 Workshop on Summarization and Summary Evaluation Across Source Types and Genres* (pp. 22–31). Association for Computational Linguistics, Valencia, Spain. https://doi.org/10.18653/v1/W17-1004. https://aclanthology.org/W17-1004

Lv, T., Cui, L., & Vasilijevic, M. (2021). Vt-ssum: A benchmark dataset for video transcript segmentation and summarization. arXiv:2106.05606

Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into text. In:*Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (pp. 404–411). Association for Computational Linguistics, Barcelona, Spain. https://aclanthology.org/W04-3252

Miller, D. (2019). Leveraging BERT for extractive text summarization on lectures. arXiv:1906.04165

Mitchell, A., Petter, S., & Harris, A. (2017). Learning by doing: Twenty successful active learning exercises for information systems courses. *Journal of Information Technology Education : Innovations in Practice, 16*:21–46. https://doi.org/10.28945/3643

Page, L., Brin, S., & Motwani, R. (1999). *The pagerank citation ranking: Bringing order to the web*. Stanford InfoLab: Tech. rep.

Parmanto, B., Ferrydiansyah, R., & Saptono, A. (2005) Access: Accessibility through simplification and summarization. In: *Proceedings of the 2005 International Cross-Disciplinary Workshop on Web Accessibility (W4A)* (pp. 18–25). Association for Computing Machinery, New York, NY, USA, W4A '05. https://doi.org/10.1145/1061811.1061815

Pedrotti, M., & Nistor, N. (2014). Online lecture videos in higher education: Acceptance and motivation effects on students' system use. In: *IEEE 14th International Conference on Advanced Learning Technologies* (pp. 477–479). ICALT 2014, Athens, Greece, July 7-10, 2014. IEEE Computer Society. https://doi.org/10.1109/ICALT.2014.141

Pi, Z., Zhang, Y., & Xu, K. (2022). Does an outline of contents promote learning from videos? a study on learning performance and engagement. *Education and Information Technologies, 28*, 3493–3511. https://doi.org/10.1007/s10639-022-11361-5

Pramudianto, F., Chhabra, T., & Gehringer, E. (2016). Assessing the quality of automatic summarization for peer review in education. In: *EDM*

Rahman, M. R., Shah, S., & Subhlok, J. (2020). *Visual summarization of lecture video segments for enhanced navigation.* https://doi.org/10.1109/ISM.2020.00033

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery 10(3):e1355. https://doi.org/10.1002/widm.1355. https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1355. arXiv:10.1002/widm.1355

Saini, M., Arora, V., & Singh, M. (2022). Artificial intelligence inspired multilanguage framework for note-taking and qualitative content-based analysis of lectures. *Education and Information Technologies*, 1–23

Shimada, A., Okubo, F., & Yin, C. (2018). Automatic summarization of lecture slides for enhanced student preview-technical report and user study. *IEEE Transactions on Learning Technologies* 11(2):165–178. https://doi.org/10.1109/TLT.2017.2682086, funding Information: This research was partially supported by "PRESTO", Japan Science and Technology Agency (JST) Japan, and "Research and Development on Fundamental and Utilization Technologies for Social Big Data" (178A03), the Commissioned Research of the National Institute of Information and Communications Technology (NICT) Japan. Publisher Copyright: 2008-2011 IEEE.

Tan, B., Qin, L., & Xing, E.P. (2020). Summarizing text on any aspects: A knowledge-informed weakly-supervised approach. arXiv:2010.06792

Tilk, O., & Alumäe, T. (2016). Bidirectional recurrent neural network with attention mechanism for punctuation restoration. In: *INTERSPEECH*

Wang, F., & Chen, Z. (2018). Self-attention based network for punctuation restoration. In: *2018 24th International Conference on Pattern Recognition (ICPR)*, 2803–2808. https://doi.org/10.1109/ICPR.2018.8545470

Yoo, T., Jeong, H., & Lee, D. (2021). Lectys: A system for summarizing lecture videos on youtube. In: *26th International Conference on Intelligent User Interfaces - Companion* (p 90–92). Association for Computing Machinery, New York, NY, USA, IUI '21 Companion. https://doi.org/10.1145/3397482.3450722

Zhang, J., Zhao, Y., & Saleh M. (2019a). PEGASUS: pre-training with extracted gap-sentences for abstractive summarization. arXiv:1912.08777

Zhang, T., Kishore V., & Wu, F., (2019b). Bertscore: Evaluating text generation with BERT. arXiv:1904.09675

Zhu, C., Xu, R., & Zeng, M. (2020). A hierarchical network for abstractive meeting summarization with cross-domain pretraining. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. https://www.microsoft.com/en-us/research/publication/end-to-end-abstractive-summarization-for-meetings/

Zhu, J., Li, H., & Liu, T. (2018). MSMO: Multimodal summarization with multimodal output. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 4154–4164). Association for Computational Linguistics, Brussels, Belgium. https://doi.org/10.18653/v1/D18-1448. https://aclanthology.org/D18-1448