Proceedings Paper:
Mohammed, A (2022) Generating Narratives of Video Segments to Support Learning. In: Lecture Notes in Computer Science. AIED 2022: Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium, 27-31 Jul 2022, Durham, UK. Springer Nature , pp. 22-28.

https://doi.org/10.1007/978-3-031-11647-6_4

Generating Narratives of Video Segments to Support Learning

Abrar Mohammed (Supervised by Vania Dimitrova)

School of Computing, University of Leeds, UK

Abstract. The predominance of using videos for learning has become a phenomenon for generations to come. This leads to a prevalence of videos generating and using open learning platforms (Youtube, MOOC, Khan Academy, etc.). However, learners may not be able to detect the main points in the video and relate them to the domain for study. This can hinder the effectiveness of using videos for learning. To address these challenges, we are aiming to develop automatic ways to generate video narratives to support learning. We presume that the domain for which we are processing the videos has been computationally presented (via ontology). We are proposing a generic framework for segmenting, characterising and aggregating video segments VISC-L which offers the foundation to generate the narratives. The narrative framework designing is in progress which is underpinned with Ausubel's Subsumption theory. All the work is being implemented in two different domains and evaluated with people to test their awareness of the domains-aspects.

Keywords: Learning videos · Domain ontology · Video segmentation · Video characterisation · Video aggregation · Video narratives.

1 Problem Addressed

Videos have been widely used in various learning settings to facilitate independent learning and are becoming a key platform for digital learning [9, 10]. However, there are major challenges that affect user engagement with videos. Learners' concentration span is reduced over time, which makes it hard to follow long videos [13, 16]. Also, video content complexity could affect the engagement with videos and may cause confusion or boredom [15]. Consequently, learners may have to watch videos many times and may not be able to identify the most relevant key points in a video. This calls for finding new ways to identify the main points in a video and to direct learners to the corresponding parts in the video, and crucially, create narratives from these video parts to elaborate specific key points. These challenges are experienced at a scale with the increase of both the amount of video footage available and the number of learners who use videos for learning. Previous researches have different attempts to address this issue by manually annotation important parts in the videos by teachers or learners [5, 12]. To maintain the quality of the annotation some researchers use ontologies [8, 17]. However the manual attempts did not scale the process; hence

an automated approaches are required to facilitate how to characterise segments of videos, especially if the domain of the videos is represented with an ontology (or a knowledge graph) [4, 6]. Existing automatic ways for characterising videos do not offer domain related annotation, not linking to the domain hierarchy of the concepts mentioned in the videos. Moreover, existing studies have not evaluated the impact of segmenting and characterising videos on learning.

To address these challenges, this PhD project poses the following research questions: **RQ1:** How to characterise video transcripts for learning by using domain ontology and past users' comments? **RQ2:** How to automatically segment videos to identify segments which are suitable for learning? **RQ3:** How to generate narratives from the characterised videos segments to support learning?

For **RQ1**, we were able to characterise predefined video segments by using the video's transcript, past users' comments and the domain ontology to apply semantic tagging . The output was characterised video segments with the focus topic/concept mentioned in both the transcript and the users' comments. This work was published in [14]. When there are no predefined video segments, we are proposing our generic framework for video segmentation, characterisation and aggregation to support learning (VISC-L), which addresses **RQ2** (see Figure 1). The outcome of VISC-L, together with the domain ontology, will be used to video generate narratives following the subsumption theory for learning (**RQ3**).

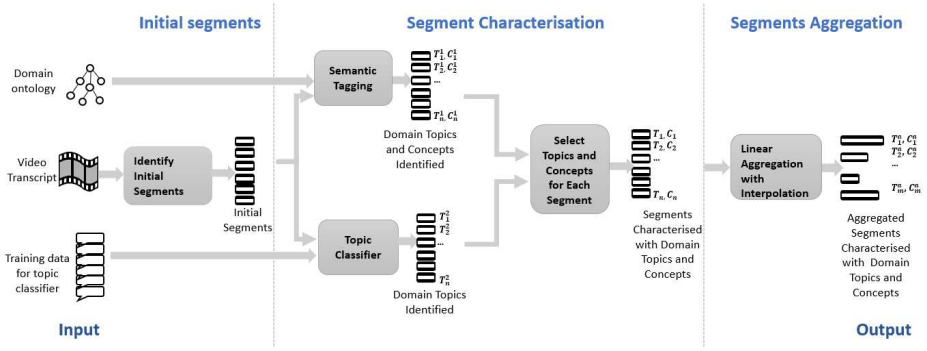

Fig. 1. Framework of Videos Segmentation and Characterisation for Learning VISC-L.

2 Framework Outline and Methodology

2.1 Framework Outline and Theoretical Underpinning

Input. VISC-L is based on two assumptions. Firstly, it is assumed that the video transcript relates to the domain to be learned is providing a description of what aspects of the domain are covered. The second assumption is that there is a domain ontology $\Omega = \{C, H\}$ which includes the relevant domain concepts C

linked in a concept hierarchy H. We use $c_i \subset c_j$ to denote that c_i is a subclass of c_j. The top level concepts in the concept hierarchy define the main domain topics $\{T_1, ..., T_m\}$. In order to identify the main topics in the video, as part of the characterisation step, training data with domain topics as labels are needed. This can either be created with expert annotators or collected from past user interactions. when we applied our framework. We have used past user interactions in one domain, and we will explore expert annotations in the other domain.

Output. The output of VISC-L is a set of aggregated video segments with a start and end time in the corresponding video. Each video segment i is characterised with a set of domain focus topics (top concepts in Ω) and a set of concepts from the focus topics mentioned in the transcript of the video segment. **Initial Segments.** Our video segmentation approach is inspired by text-tilling in text segmentation - starting with smaller units (e.g. sentences) and aggregating them to get larger coherent units (e.g. paragraphs). Hence, we include an initial segmentation step where the video transcripts are cut into small segments that are used as a starting point for aggregation. Initial segments can be done by using certain number of text lines (e.g. we are using 6 lines) or by using pre-defined segments (e.g. as when we applied our semantic tagging algorithm in one of the domain where we have used high attention intervals from past interactions).

Segment Characterisation. In order to aggregate the initial segments, we need to identify what domain content is presented in each segment. This is done during the segment characterisation step which links each video segment i with a set of focus topics T_i and a set of concepts C_i. To do so, we propose to use two algorithms: semantic tagging and topic classification. The **semantic tagging** algorithm links each video segment to focus topics and concepts by mapping the terms from the ontology to the text in the video transcript. The algorithm first pre-processes the transcript, including: (a) tokenisation; (b) cleaning from stop words and punctuation; (c) selecting nouns and noun phrases from the transcript; (d) matching the ontology terms to the noun phrases. If there is a match between the transcript noun phrases and the ontology, the ontology concept c_i will be identified (tagged to the text), noting also the path to reach a top-level concept. As a result, each segment i is linked to a set of focus topics and their corresponding concepts; we denote this as $< T_i^1, C_i^1 >$ (where 1 indicates that this is an output from the first segment characterisation algorithm). A key challenge for this algorithm is word sense disambiguation - we need to disambiguate the topics based on the context, which is done with the second algorithm.

The second algorithm is a **topic classifier** which identifies a domain topic based on the context of that topic. Following the latest development in natural language processing, we use Bidirectional Encoder Representations from Transformers (BERT) [7] as a topic classifier. BERT embeds pre-trained deep bidirectional representations from unlabelled text by jointly conditioning on both left and right context in all layers. Accordingly, it can be fine tuned with just one additional output layer to create state-of-the-art models for different language tasks, topic classification in this case. First, the BERT model is fine-tuned using training data with domain topic labels. The fine-tuned model is used as classi-

fier to link each segment i to domain topics T_i^2 (2 indicates an output from the second segment characterisation algorithm). The last step in segment characterisation is to **combine the outputs from both algorithms**. For each segment i, the outcomes from both algorithms $< T_i^1, C_i^1 >$ and T_i^2 are combined by intersecting the focus topics $Ti = T_i^1 \cap T_i^2$ and selecting the concepts C_i from C_i^1 that belong to T_i. Each segment is characterised by $< T_i, C_i >$ - a set of focus topics and their concepts.

Segments Aggregation. Following the text-tilling approach, small segments will be aggregated in larger segments. To maintain the flow of information within adjacent segments, we have developed an aggregation algorithm based on **Thematic Progression Theory** [3]. This theory has been widely used for creating coherent text, and states that a good written text should have a relation between theme (which is the main clause) and rheme (which is the remainder of the text used to develop the theme). Three patterns for coherent text are suggested: Constant theme (when the first theme in one sentence is carried on and used at the beginning of the second sentence); Linear theme (the important message in a rheme of one sentence is carried on a theme in the second sentence), and Split theme (a development of a rheme with important information is used as themes in the subsequent sentence).

We adapt the Thematic Progression Theory theory when we aggregate adjacent segments to indicate coherent parts within videos. We associate the focus topic with the segment's theme and the focus concepts with the segment's rheme. We propose a **linear aggregation with interpolation** algorithm. The linear theme pattern was selected as the most appropriate, as it allows to keep a continuous focus topic and at the same time to take into account the specific concepts within that topic. Some segments can be without characterisation (i.e. it is not possible to link the video transcript to domain concepts), which can be because the speaker is silent or is digressing from the domain. If we look strictly for adjacent segments, these gap segments which break the topic flow will lead to starting a new aggregate. To smoothen the aggregation, we use interpolation. If the segments before and after a gap segment have common focus concepts, it is assumed that the common concepts spread across the three segments. Hence, the gap segment will be interpolated in the aggregated segment.

To generate video narratives, the video segments are combined following the **Ausubel's Subsumption Theory** for meaningful learning [2]. According to this theory, a primary process in learning is subsumption in which new material is related to relevant ideas in the existing cognitive structures derived from learning experiences. According to the subsumption theory, there are four types of subsumption: Derivative, Correlative, Super-ordinate and Combinational. We are aiming to automate the linking of video segments to generate narratives by following the focus topics and concepts in them using the hierarchical of the concepts in the domain ontology. Our narratives work using Ausubels' theory is motivated by its successful adoption for meaningful learning by using concept maps [1, 11] that allow learners to group information in related modules making the connections between modules more apparent.

2.2 Methodology

We have adopted a data-driven approach to generate narratives from videos by first: segmenting videos, characterising video segments, aggregating adjacent segments and creating narratives from these segments. Our data set is either available videos collected by other researchers or to be collected using the search schema we have designed by utilising ontology terms to search for videos available on social learning platforms (i.e. YouTube) as follows: <*Domain Name, Topic Name, Concept Name*>. The input to our work is the video transcript and the domain ontology. Additionally, we need training data labeled with domain topics. Based on this, we can apply our segmentation, characterisation and aggregation framework (VISC-L). The output segments will provide the foundation to generate video segments narratives. The narratives will be generated by applying the narratives framework in two different domains and evaluate them with people to test their awareness of the domain aspects and their possible effect on their life.

2.3 Progress to Date

The work on characterising video segments using domain ontology and videos-past users' comments and videos transcript has been published in [14]. Additionally, we have applied VISC-L framework in a domain (Presentation Skill). The result have been evaluated with learners by comparing the usability, perceived usefulness, mental demand and the learning impact of the characterised video segments generated by our work and by using the Google outcome. This work is submitted to another conference and is currently under review.

The next step in this PhD research is to design, apply and evaluate the narrative framework in one domain (Presentation Skill). After that VISC-L framework and the narrative framework will be be applied and evaluated in another domain (i.e. health domain-COPD).

3 Expected Contribution

We propose a novel way to create narratives from video segments to support learning which is underpinned by pedagogical theories and utilises natural language processing. Our main contribution is the designing of two generic frameworks - one for videos segmentation, characterisation and aggregation for learning (VISC-L) and the other one for generating narratives from video segments. The work is being applied in two soft skills domains - presentation skills (giving pitch presentations) and healthcare (patient's quality of life needs assessment).

Acknowledgements. The application in the healthcare domain is funded by the the European Union's Horizon 2020 research and innovation programme under grant agreement No 825750 (InADVANCE project).

References

1. Al-Tawil, M., Dimitrova, V., Thakker, D.: Using knowledge anchors to facilitate user exploration of data graphs. Semantic Web **11**(2), 205–234 (2020)
2. Ausubel, D.P.: A subsumption theory of meaningful verbal learning and retention. The Journal of general psychology **66**(2), 213–224 (1962)
3. Bloor, T., Bloor, M.: The functional analysis of English. Routledge (2013)
4. Cagliero, L., Canale, L., Farinetti, L.: Visa: A supervised approach to indexing video lectures with semantic annotations. In: 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC). vol. 1, pp. 226–235. IEEE (2019)
5. Castro, M.D.B., Tumibay, G.M.: A literature review: efficacy of online learning courses for higher education institution using meta-analysis. Education and Information Technologies **26**(2), 1367–1385 (2021)
6. Das, A., Das, P.P.: Semantic segmentation of mooc lecture videos by analyzing concept change in domain knowledge graph. In: International Conference on Asian Digital Libraries. pp. 55–70. Springer (2020)
7. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
8. Dias, L.L., Barrére, E., de Souza, J.F.: The impact of semantic annotation techniques on content-based video lecture recommendation. Journal of Information Science **47**(6), 740–752 (2021)
9. Hsin, W.J., Cigas, J.: Short videos improve student learning in online education. Journal of Computing Sciences in Colleges **28**(5), 253–259 (2013)
10. June, S., Yaacob, A., Kheng, Y.K.: Assessing the use of youtube videos and interactive activities as a critical thinking stimulator for tertiary students: An action research. International Education Studies **7**(8), 56–67 (2014)
11. Katagall, R., Dadde, R., Goudar, R., Rao, S.: Concept mapping in education and semantic knowledge representation: an illustrative survey. Procedia Computer Science **48**, 638–643 (2015)
12. Lagrue, S., Chetcuti-Sperandio, N., Delorme, F., Thi, C.M., Thi, D.N., Tabia, K., Benferhat, S.: An ontology web application-based annotation tool for intangible culture heritage dance videos. In: Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents. pp. 75–81 (2019)
13. Meseguer-Martinez, A., Ros-Galvez, A., Rosa-Garcia, A.: Satisfaction with online teaching videos: A quantitative approach. Innovations in Education and Teaching International **54**(1), 62–67 (2017)
14. Mohammed, A., Dimitrova, V.: Characterising video segments to support learning. In: Proceedings of the 28th International Conference on Computers in Education (2020)
15. Mongkhonvanit, K., Kanopka, K., Lang, D.: Deep knowledge tracing and engagement with moocs. In: Proceedings of the 9th International Conference on Learning Analytics & Knowledge. pp. 340–342 (2019)
16. Risko, E.F., Anderson, N., Sarwal, A., Engelhardt, M., Kingstone, A.: Everyday attention: Variation in mind wandering and memory in a lecture. Applied Cognitive Psychology **26**(2), 234–242 (2012)
17. Schulten, C., Manske, S., Langner-Thiele, A., Hoppe, H.U.: Bridging over from learning videos to learning resources through automatic keyword extraction. In: International Conference on Artificial Intelligence in Education. pp. 382–386. Springer (2020)