So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Characterising Video Segments to Support Learning

Abrar Mohammed[a] and Vania Dimitrova[a]
[a] *School of Computing, University of Leeds, UK*

Abstract: Videos provide opportunities for engagement and independent learning and are widely used in various learning contexts. However, there are challenges with using videos for learning, e.g. long videos can reduce the concentration span, learners may become bored, not everyone can be able to detect the main points in the video, and not all parts in a video will be relevant to the learner. To address these challenges, our research aims to develop automatic ways to generate narratives by combining short video segments and tailoring this to the learner's needs. As a first step, this paper is proposing an original framework to characterise video segments for learning by combining video content and audience attention. The input for the framework includes the video transcripts, past user interactions with the videos, and an ontology defining the core domain concepts. The output is a set of patterns that are associated with the video segments, describing the focus topic and concepts of the segment. We have applied the framework on a dataset from user studies with the AVW space for presentation skills learning, including 49 video segments that are high attention intervals from past user interactions. The video segment characterisation provides useful insights to inform recommendations and segment combinations to support informal learning.

Keywords: Video-based learning, Video characterisation, Ontologies, Presentation skills.

1. Introduction

Videos have been widely used in various learning settings to facilitate independent learning and are becoming a key platform for digital learning (June et al., 2014; Hsin & Cigas, 2013). However, there are major challenges that affect user engagement with videos. Learners prefer watching short videos as their concentration span is reduced over time (Meseguer-Martinez et al., 2017; Risko et al., 2012). Also, video content complexity could affect the engagement with videos and may cause confusion or boredom (Mongkhonvanit et al., 2019). 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 (called hereafter video segments) that 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. Therefore, *manual solutions would not scale up* - using experts to analyse the video segments and identify how they will be used by potential learners can be costly and ineffective.

This calls for *computational means to automate the characterisation of video segments* in order to identify what concepts are covered and whether learners will grasp these concepts. To address this challenge, we are proposing a data-driven approach inspired by crowdsourcing where the interaction of past learners with a video provides an indication of the main points noted by learners in a video. To represent the domain, we use an ontology defining the main concepts in the domain and their relationships. This also allows linking segments to support learning. Our approach is developed within a PhD project, which aims to automatically characterise video segments, identify optimal video segmentation and create narratives by combining different segments.

This paper presents the first step of our approach, addressing the following **research question**: *How to characterise video-segments for learning using learners' interaction and video content*?

To address this research question, we will present a framework for characterising video segments which specifies the domain knowledge covered in a segment. The learner comments and the video transcript of the same segment will be linked to a domain ontology to identify the main topics and concepts to characterise the video segment. By comparing the video transcript and the learner comments, we gather further indication about the usefulness of the segments for learning. The video characterisation framework is applied on a dataset from using videos for informal learning of presentation skills (Mitrovic et al, 2016). Video segments which capture high attention intervals where

learners from a previous study have noted points in the videos and have generated comments accordingly are used. The characterisation of each video segment identifies what domain aspects have been covered which can show the suitability of the segment for learning. Furthermore, we illustrate how the ontology-based characterisation allows combining segments through aggregation and linking.

The rest of the paper is organised as follows. Section 2 positions our work in relevant literature on video characterisation and points at the key contribution. Section 3 outlines the video characterisation framework, including the (a) pedagogical underpinning, (b) preliminaries, (c) main definitions for characterising video segment by identifying domain topics, concepts, their coverage and level of abstraction, and (d) the video segment characterisation pipeline. Section 4 presents the application context and dataset, while Section 5 presents and discusses the results of applying the video characterisation framework in this context. Section 6 concludes and points at future work.

2. Related Work

Video characterisation has been a major goal of research to enhance the use of videos for different purposes including using videos for learning. **Manual video characterisation** (Chiu et al., 2018; Colasante et al., 2016) involves asking teachers and students to characterise and highlight video-content for learning to help find video-content quickly. Similarly, Dutta & Zisserman (2019) use manual characterisation by human annotators to describe video content. To speed up the video characterisation process, Benkada & Moccozet (2017) use a constrained tracker to choose which video frame a user should characterise, and also allow collaborative characterisation by a group of people. These approaches characterise the whole video considering different purposes, including learning. Learners' experience (captured via student annotations) are used to characterise the videos. However, manual characterisation is laborious and does not scale.

Automatic video characterisation approaches have focused on analysing the visual content of the video. For example, multiple video files that are captured by calibrated imaging devices have been characterised based on a manual characterisation of a single image frame of one of the video files (Goldenberg et al., 2019). The video characterisation presented in (Xue et al., 2017) collects metadata about the video using an assembly of computer vision techniques (shoot boundary detection, a tensor based compact representation, player detection and tracking and optical character recognition). These automated video characterisation approaches have not been applied in learning, which would require not only object detection in the video but also linking objects to a learning domain.

Semantic video characterisation which adds meaning by linking objects from the videos to a specific domain has been proposed in several works. To promote the re-use of the learning videos, Rezazadeh Azar (2017) utilises a probabilistic graphical model that shows a set of variables and their dependencies in a directed acyclic graph to exploit the semantic relationships between domain concepts. Ashangani et al. (2016) use an ontology to add semantic meaning to video characterisation to facilitate video retrieval. Videos are automatically analysed to detect shot boundaries, to extract features, and to conduct automatic text annotation. The user's query and video content are matched by using an ontology. A recent review of semantic video characterisation (Sikos, 2017) points at key challenges, including the wide variety of video codecs, the lack of standardised vocabularies, the vast number of video resources, the inherent ambiguity of audio-visual contents, and the unstructured nature of user-generated content. As a way to address these limitations, Sikos recommends semi-automatic or automatic video annotation using ontologies and Linked Data combined with semantic tagging tools.

The research presented here performs semantic video characterisation. Similar to the existing approaches, it uses an ontology to represent the domain of interest and utilises natural language processing and semantic tagging. However, there are some crucial differences. While current approaches advance video retrieval, we propose semantic characterisation *to support the use of video segments for learning* which is underpinned by a pedagogical model - Ausubel's subsumption theory – to allow recommendation and linking of video segments. We primarily deal with textual data (video transcript and the experience of the learners captured in their comments), which relates to the learning content introduced in the video. Finally, we provide a formal description that allows the utilisation of our approach in a broad range of domains and illustrate its application in a practical learning context.

3. Video Characterisation Framework

3.1 Pedagogical Underpinning

The main goal of our video characterisation framework is to help select and combine video segments to support learning. Therefore, we need an appropriate pedagogical underpinning to inform what to include in the video characterisation and how to combine segments to support people to learn meaningfully from material presented to them (video segments in this case). Ausubel's subsumption theory for meaningful learning (Ausubel et al., 1968) has been selected for this purpose. 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. Ausubel argued that new material should be integrated with previously presented information through comparisons and cross-referencing of new and familiar concepts. Successful adoption of the subsumption theory for meaningful learning includes using concept maps (Katagall et al., 2015; Liu et al., 2018) that allow learners to group information in related modules making the connections between modules more apparent. While concept maps provide domain structure, they do not allow reasoning and automation. Similarly to concept maps, ontologies define the main concepts and relationships in a domain. However, in addition, ontologies allow reasoning to automate the generation of instruction paths through the domain. Therefore, instead of concept maps, we use an ontology to represent the domain.

Similarly to Al-Tawil et al. (2019), who operationalises the subsumption theory to generate information exploration paths through a large knowledge graph, we explore hierarchical relationships between ontology concepts to identify how to link content. The fundamental difference is that we first need to characterise video segments by mapping them to an ontology and to do this in a way that will allow supporting the subsumption processes proposed by Ausubel. We identify focus topics and concepts of video segments using reasoning over the hierarchical links in the ontology (as presented below). This enables subsumption links (derivative, correlative, super-ordinate, and combinational subsumption) to combine video segments for meaningful learning (as illustrated in Section 5).

3.2 Preliminaries

A **video segment** V is a video interval $[V_s,V_e]$ that has a start point V_s and an end point V_e. A video can include several video segments. We assume that each video has an audio transcript (text) and a set of textual comments made by users who have interacted with the video. Consequently, each video segment V is associated with a **transcript** V_t representing the audio in this segment and a set of **user comments** V_u which have come from users U who have interacted with this video.

We assume that the domain is defined with an **ontology** $\Omega = \{C,H\}$ which includes the relevant **domain concepts** $C \neq \emptyset$ and a concept hierarchy H linking these concepts. We use $c_i \subseteq C$ to denote that c_i is a **subclass** of C. Each concept c has a set of **immediate sub-classes** $c_i \subseteq c$ which are directly linked in the concept hierarchy. The main **domain topics** are the top level concepts in the concept taxonomy $T_H = \{C_1,...,C_m\}, m > 0$, i.e. $C_i \in T$ where T is the top of the concept hierarchy.

Using semantic tagging, we *can* **link text to concepts** in an ontology. The **mentions** of a concept c in text t accumulates all mentions of c and its sub-classes, i.e. $mentions_t(c)=\sum f_i$ where f_i indicates the number of times that the concept c_i has been mentioned in the text t, for all $c_i \subseteq c$. For each video segment V we identify the **domain concepts mentioned in the transcript** V_t as $M_t(c) = \{c_1,c_2,...,c_p\}$ where $mentionsV_t(c_i) > 0$, i.e. these concepts have been mentioned at least once in the video transcript. Similarly, for each video segment V we identify the **domain concepts mentioned in the user comments** V_u *as* $M_u(C) = \{c_1,c_2,...,c_q\}$ where $mentions_{Vu}(c_i) > 0$, i.e. c_i mentioned at least once.

3.3 Focus Topics, Focus Concepts and Coverage

Based on the preliminaries, we can define focus topics and concepts presented in the video segment transcript and the user comments. A **focus topic in the transcript** $F_t(C)$ of a video segment V is a top concept $C \subseteq T$ in the ontology that has a notable number of mentions in the video transcript V_t, which we indicate with a parameter θ (i.e. $\text{mentions}V_t(C) \geq \theta$). Depending on θ, a video segment can have several focus topics (for example, the application presented in the next section uses $\theta = 1/3$). In a similar

way, we define the **focus topic in the user comments** $F_u(C)$ of a video segment by considering the mentions of the top concepts in the user comments V_u. See example focus topics in tables 3 and 4.

Within a focus topic, we can identify corresponding focus concepts. A **focus concept** $F_t(C,c)$ **within the topic** C **as expressed in the transcript** of a video segment V_t is identified if mentions$V_t(c) \geq \theta$. Depending on θ, a video segment can have several focus concepts within the same focus topic; for example, the application presented in the next section uses $\theta = 1/3$, see example focus concepts in tables 3 and 4. Similarly, we define **focus concept** $F_u(C,c)$ **within the topic** C **as expressed in the user comments** of a video segment counting the mentions of the sub-classes of the focus topic in V_u.

It is important to know not only that a domain topic or concept are in the focus of a video segment, but also to identify how they have been covered which can help in deciding how to use them for learning. Using the ontology enables us to define the concept coverage. A domain concept is **covered broadly** if most of its immediate sub-classes have been mentioned, otherwise it is **covered narrowly**. We use a parameter θ to set the threshold for defining the coverage type (the application presented in the next section uses $\theta = 1/3$). Accordingly, we can define the type of coverage for both focus topics and focus concepts. Examples of focus topics and focus concepts' coverage are shown in tables 3 and 4.

3.4 Video Segment Characterisation Pipeline

In this section, we describe the video segments characterisation pipeline (shown in Figure 1) which provides the main computation steps to apply the video characterisation framework.

Input data. The characterisation process begins with the use of predefined video-segments (this can be video intervals at different length, including whole videos when short). The input data required is start and the end time of the video segments, user comments associated with the segment (if available), video transcript text, and an ontology representing the domain.

Data processing. The text from the input data (video transcript and user comments) is processed in two steps - text processing (which extracts the main words in the text) and semantic tagging (which maps words/phrases to concepts in the ontology). To perform these steps, existing natural language processing tools can be used. The next section provides example tools used in a practical application.

Video segment characterisation. Using the ontology, the definitions provided in Section 3.3. can be applied to extract the domain knowledge represented in each video segment (focus topics, focus concepts, coverage). Section 5 shows the results of domain characterisation of video segments.

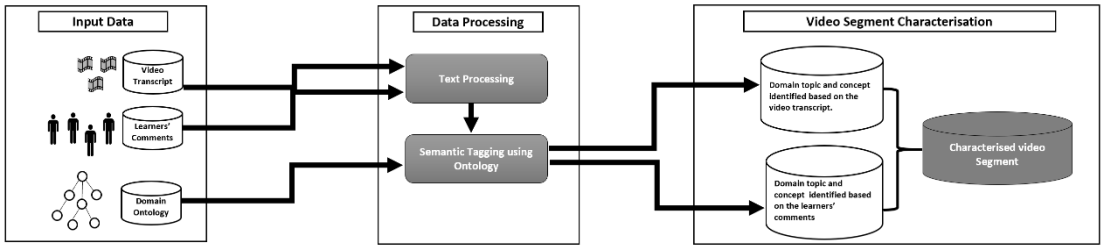

Figure 1. Video Segment Characterisation Pipeline.

4. Application Context and Dataset

4.1 AVW User Studies Datasets

The video characterisation framework presented in Section 3 has been applied to characterise video segments used in the Active Video Watching (AVW-Space) in the context of learning to give pitch presentations (Mitrovic et al., 2016). We use past interactions from several studies with AVW-Space with undergraduate and postgraduate university students (Mitrovic et al., 2016; Dimitrova et al. ,2017; Hecking et al., 2017), which used 8 YouTube videos - 4 tutorials and 4 examples of pitch presentations. We use the interaction of 460 learners (aggregated from past user studies) who made 3532 comments.

The video segments used here are the high attention intervals that indicate continuous stretches of videos where learners have noted something interesting in the video (Dimitrova et al., 2017). We use 49 video segments - 24 segments from tutorial videos and 25 segments from example videos. High

attention intervals are derived by aggregating user comments. Aggregation of a set of comments Com, is performed as follows: $A(Com) \equiv \forall(com_i \in Com)\exists(com_j \in Com)[(com_i \neq com_j) \wedge distance(com_i, com_j) \leq \theta]$. The granularity of continuity is determined by θ indicating the interpolation gap between adjacent comments. To take into account the time required to start writing a comment, 5 seconds adjustment has been made to the start and end of the high attention intervals. We have chosen the segments of high attention of self-regulated learners who were most engaged and generated domain related comments (Hecking et al., 2017). Comments within these predefined video segments are used as input for the video segment characterisation, comments that are not within these segments have been discarded.

4.2 Data Processing

The processing of the video transcript and user comments is done in two steps (see Section 3.4.) Text processing (tokenisation and stop words removal) uses NLTK[1]. Then, the resumed words are matched with the domain ontology terms, using WordNet[2] for synonym check (if there is no direct match found, synonymous of the text words will be found). We first use lemmatisation to get the root of the words, then use WordNet to get synonymous, then use the SequenceMatcher in Python[3] to find the closest word in meaning to the ontology terms (similarity threshold of 0.85 which was identified with several tests). Figure 2 illustrates the data processing conducted in a short video transcript.

An existing ontology with key concepts related to delivering pitch presentations (Abolkasim, 2019) is used. It is organised in a hierarchy with four main classes: Structure (70 concepts), Visual Aid (95 concepts), Delivery (106 concepts), and Presentation Attribute (27 concepts). The ontology is represented in OWL and is available online[4]; example concepts are shown in tables 3-4.

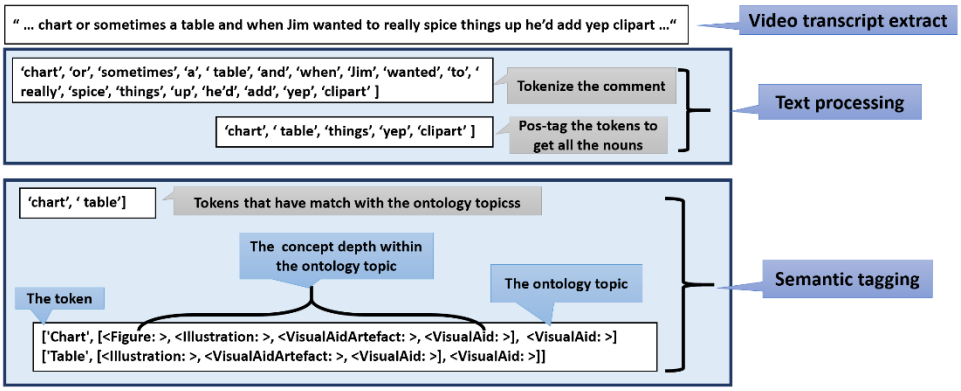

Figure 2. Example Data Processing for Video Segment Characterisation Using the Video Transcript.

5. Video Segment Characterisation: Results and Discussion

5.1 Overview

By applying the video segment characterisation framework, we are able to get an overview of the domain topics and concepts in the video segments in our data set (see tables 1 and 2). Two domain topics are prevalent: Visual Aid (63% in the tutorials and 52% in the examples) and Delivery (58% in the tutorials and 52% in the examples). Therefore, this set of video segments can be used for learning these topics. Although tutorial videos refer to Structure and this was picked by learners when watching tutorials, learners did not pick Structure in the examples. This indicates that Structure would be a 'difficult topic' which learners could miss to see in examples. Indeed, structure is a challenging domain topic in similar soft skill domains, e.g. writing, argumentation, negotiation. Based on the characterisation, we can note another 'difficult topic' - Presentation Attribute - which was not noted by learners in any of the

[1] https://www.nltk.org/
[2] https://wordnet.princeton.edu/
[3] https://docs.python.org/3/library/difflib.html
[4] Link to open: http://www.semanticweb.org/sc10ena/ontologies/2017/10/PresentationOntologyV1

video segments with example presentations (although these examples were selected as highly engaging presentations). Therefore, additional scaffolding (e.g. prompts) would be needed to draw the learner's attention to Structure and Presentation Attributes when watching examples. Note that the main ontology topics which are Delivery (D), Presentation Attribute (P), Structure (S) and Visual Aid (V) hereafter will be mentioned in the tables with the abbreviation D, P, S and V, respectively.

Table 1. *Summary of Focus Topics Identified in the 24 Segments of the Tutorial which Considers both Video Transcripts(T) and User Comments(C)*

	D	**P**	**S**	**V**
Focus Topic in *T* and *C*	38%	8%	33%	17%
Focus Topic in *T*	8%	21%	4%	46%
Focus Topic in *C*	13%		4%	
All Focus Topics	58%	29%	42%	63%

The identified focus concepts give further detail about the representation of the domain topics in the data set. Because the tutorials explicitly refer to domain concepts, many of these concepts were also picked by the learners. Our video characterisation framework assigned focus concepts for all tutorial video segments, apart from one. However, although each of the 25 segments with examples had a focus topic, only 8 are with specific focus concepts. This indicates that the learners have noticed domain topics in the examples but have not articulated specific concepts within these topics.

Table 2. *Summary of the Focus Topics identified in the 25 Segments of the Example Videos which considers only on User Comments*

	D	**P**	**S**	**V**
Focus Topic without Focus Concept(s)	40%			28%
Focus Topic with Focus Concept(s)	12%		4%	24%
All Focus Topics	52%		4%	52%

5.2 Characterisation Based on User Comments

In both example and tutorial videos, we have user comments which indicate how the domain was noticed by the users when watching these videos. Tables 3 and 4 show the characterisation of the segments from example and tutorial videos, which are used as illustrations in the discussion below. The broad coverage of the topic means the segments can be used to give an overview at an abstract level of the focus topic or the focus concept. While the narrow coverage of the topic means the segment can be used to illustrate in depth the focus topic or the focus concept. We identified several patterns that indicate the usefulness of the video segments for learning, as follows.

Table 3. *Example Focus Topics, Focus Concepts and Coverage in Segments from Example Videos(E)*

SegmentID	**Focus Topics and Coverage**	**Focus Concepts and Coverage**
E11	S(broad)	S->PresenterIntro(narrow)
E13	D(broad)	D->RhetoricalDevice(narrow)
E14	D(broad)	
E31	D(narrow) V(broad)	D->AudienceEmotion(narrow) V->RecordedVoice(narrow)
E32	V(broad)	
E33	V(broad)	
E34	V(narrow)	V->VisualArtefact(narrow)
E35	V(narrow)	
E41	V(narrow) D(broad)	D->BodyMotion(broad)

Table 4. *Example Focus Topics, Focus Concepts and Coverage in Sample Segments from Tutorial Videos(T). Note the use of short terms for the concept names, e.g. StrCom=StructureCompnent, VisArt=VisualArtefact, etc. Concepts full name will be used through the paper*

SegmentID	Focus Topics and Coverage	Focus Concepts and Coverage
T11	Transcript: S(broad),V(narrow) Comments: S(broad),V(narrow)	Transcript: V->Text (narrow) Comments: V->Text(narrow)
T12	Transcript: S(broad),V(broad) Comments: S(broad),V(broad)	Comments: S->PreIntro(narrow) Comments: V->Text(narrow)
T21	Transcript: S(narrow) Comments: S(narrow)	Transcript: S->StrCom(narrow) Comments: S->StrCom(narrow)
T22	Transcript: S(broad),V(narrow) Comments: S(narrow)	Transcript: V->VisArt(narrow) Comments: S->StrCom(narrow)
T26	Transcript: S(narrow),V(narrow) Comments: S(narrow),D(broad)	Transcript: S->StrCom(narrow) Comments: D->AudEmo(broad)
T33	Transcript: V(narrow),S(narrow) Comments: S(broad)	Transcript: V->Audio(narrow) Transcript: V->Text(narrow) Transcript: S->ConInt(broad) Transcript: S->PreIntro(broad)
T42	Transcript: D(narrow) Comments: D(broad)	Transcript: D->Para(narrow) Transcript: D->LanEle(narrow)
T43	Transcript: V(narrow),D(narrow) Comments: D(broad)	Transcript: V->RecVoi(broad) Transcript: V->Illust(broad) Transcript: D->BodMot(broad)
T45	Transcript: D(narrow),V(broad) Transcript: S(narrow) Comments: D(narrow)	Transcript: D->SpeEmo(narrow) Transcript: D->Para(narrow) Transcript: S->StrCom(broad) Comments: D->Para(narrow)

The focus topic is broadly covered. Video segments with such pattern can be used to introduce the topic to learners in an abstract level without focusing on specific details. For instance, **E14** shows an introductory example for Delivery and there is no focus on specific concepts within the focus topic. This pattern (there is a focus topic but there is no focus concept) is observed in 39% of the segments. Further 18% of the segments have a broadly covered focus topic and a focus concept within it. These segments can be used to introduce the focus topic at an abstract level but also to draw the learner's attention to the specific focus concepts. For instance, **E11** can be used for introducing Structure by focusing on Presenter Introduction.

The focus topic is narrowly covered. The segments following this pattern are useful for illustrating aspects of a focus topic. 12% of the segments have a narrowly covered focus topic and no focus concept, e.g. **E35** is good to illustrate Visual Aid. Further 22% of the segments have narrowly covered focus topics and specific focus concepts, which can be useful for elaborating the focus topic. For instance, **T21** can be used for elaborating Structure by focusing on Structure Component.

Having Two Focus Topics. The characterised video segments with this pattern can be recommended to the learners to show the link between topics. For instance, **T33** shows links between use Visual Aid (focusing on use of Audio and Text) and Structure (focusing on ContentIntro and PresenterIntro). This pattern was observed rarely in our dataset (8.2% of the segments), which is expected given the short duration of the video segments. With longer video segments, or when applied to whole videos, this pattern can identify useful video content that can illustrate relationships between topics/concepts.

Segments not suitable to use on their own. From our observation of the characterisation patterns mentioned above, we found 46% of the example segments (e.g. **E13**) and 36% of the tutorial segments (e.g. **T11** or **T12**) did not fall in any of the above patterns, and could not be used on their own. However,

they can be aggregated to create learning materials that cover the learning presented in them in a more coherent way. The aggregation of segments is discussed in Section 5.4.

5.3 Comparing the Video Transcript and Learner Comments

For tutorial videos, the segment characterisation allows comparing the learner attention (learner comments) and the video content (video transcript) to understand whether the learners have picked the points mentioned in the tutorial. Examining the 24 tutorial segments, we have identified three patterns.

Full focus topic alignment. This shows that the focus topic in the video has been noticed by the learners as shown in their comments. Segments characterised with this pattern will be good to illustrate the focus topic. In 8% of the tutorial segments, there is a full alignment on the focus topic and focus concept among the video transcript and the learner comments. For instance, **T21** is a good segment to illustrate the topic Structure in depth as it is narrowly covered and there is a focus on one of its concepts- Structure Component and this is picked by the learners. In 8% of the tutorial video segments the focus topics aligned but the learners additionally focused on a concept which is not a focus in the transcript. For instance, in **T12** both the video transcript and the user comments are focusing on Structure and Visual Aid and the learners are also focusing on specific concepts - Presenter Introduction and Text. This segment is useful to show the relationship between these concepts. In 8% of the segments, while the focus topics aligned, learners have missed the focus concepts in the transcript; hence, learners' attention should be directed to notice the focus concepts. For instance, in **T42** the transcript and the comments agreed that the segment is illustrating the topic Delivery but the learners' attention would have to be directed to the specific concepts - Paralanguage and Language Element.

Partial focus topics alignment. The segments characterised with this pattern indicate that learners focus on topics and concepts but miss other topics and concepts mentioned in the transcript. Hence, when recommending these segments, the learner's attention should be directed to the relationship between the focus topics and the elaboration of some focus topics. In 13% of the tutorial segments learners have missed some of the focus concepts in the transcript. For instance, in **T45** the transcript and comments are focusing on the topic Delivery but the transcript is also relating Delivery to Structure by illustrating the effect of the Structure Component and Paralanguage on Audience Emotion. This learning is missed in the user comments which focus only on Paralanguage. There are 25% of the tutorial segments where the learners missed all the focus concepts in the transcript. For instance, in **T42** the transcript is elaborating on the Delivery by focusing on Paralanguage and LanguageElement; however, this connection is missed in the comments which did not focus on any concept. In 17% of the tutorial segments the learners and the transcript have different focus concepts within the same focus topic. For instance, in **T45** the transcript and learners are focusing on Delivery but the transcript is showing the relationship between this topic and the use of StructureComponent from Structure, while the learners are focusing on Paralanguage on Delivery. The discrepancies in focus concepts can indicate aspects of the video that may be missed by the learners, which should be considered when recommending the video segments (on their own or in combination with other).

Misalignment of focus topics. The cases characterised with this pattern, which shows that there is a clear deviation between the transcript and the comments, represent 21% of the tutorial segments. Consequently, the learning content in these segments is not articulate enough to make them useful learning materials. For instance, in **T26** the transcript and the comments are illustrating specific concepts within two different topics, which indicates that this segment cannot be recommended.

5.4 Combining Video Segments

The video segment characterisation can be used to combine segments in order to provide a more effective way to use video segments for learning. Combining video segments includes aggregating adjacent segments and linking segments from different parts within one video or from different videos.

Aggregating segments from the same video. Individual segments can be too short or may not provide good enough coverage to be recommended on their own. Hence, when adjacent segments share focus topics/concepts, they can be aggregated in longer segments. The coverage allows us to further indicate how the aggregated segment can support learning. For instance, following the sample segment characterisation from Table 3 and Table 4, we can aggregate:

<**E31,E32,E33,E34,E35**>, Visual Aid → Visual Artefact - all segments have the same focus topic Visual Aid. **E31**, **E32** give examples to introduce Visual Aid, **E33** then illustrates a specific Visual Aid concept – Visual Artefact, followed by **E34** and **E35** to complete the elaboration on the topic Visual Aid.

<**T11,T12**>, Structure, Visual Aid → Text - both segments have the same focus topics. **T11** links Text from Visual Aid to Structure, while **T12** links Text and Presenter Introduction.

Linking segments based on subsumption relationships. We can use the video segment characterisation to operationalise the subsumption processes proposed by Ausubel to foster meaningful learning (Ausubel et al., 1968). This linking can be done automatically, following the focus topics and concepts of video segments, and using the hierarchical links in the ontology.

Linking through Derivative subsumption. Derivative subsumption occurs when new material is learned as an illustration or example of an existing construct in the human cognitive structure. For instance: Structure → Presenter introduction → example, < **T12,E11** > - both segments have the same focus topic and focus concept; **T12** describes how presenters should introduce themselves and **E12** gives an example. Visual Aid → Visual artefact →{Text,Audio,Illustration}, < **T22,T33,T43** > - the segments have the same focus topic - Visual Aid; **T22** focuses on a higher level concept Visual Artefact, which is illustrated with its sub-classes Text and Audio (in **T33**) and Recorded Voice and Illustration (in **T43**).

Linking through Correlative subsumption. Correlative subsumption occurs when new material is learned as an elaboration of existing concepts within the same class. For instance: Delivery → Non-verbal communication →{Paralanguage,Bodymotion}, < **T42,T45,E41** > - Non-verbal Communication is elaborated with two sub-classes - Paralanguage (in **T42** and **T45**) and Body Motion (in **E41**).

Linking through Super-ordinate subsumption. Super-ordinate learning occurs by linking several learned concepts with their super-class concept. For instance: {Language,RhetoricalDevice}→ Verbal communication → Delivery, < **T42,E12** > - **T42** and **E12** focus on Language and Rhetorical Devices after introducing these concepts, a link to their super-class Verbal Communication can be made.

Linking through Combinational subsumption. Combinational subsumption occurs when new material presents relevant links but is not subsumed through a subordinate relationship or a super-ordinate relationship. In our case, this will allow linking concepts from more than one topic to give a broader view of the domain. This enables the use of common focus concepts across segments. For instance: {Structure,Delivery}→{StructureComponent,Speakeremotion,Paralanguage}, < **T21,T45** > - both segments have a common focus concept Structure Component, based on which they can be linked; **T21** will show concepts from the topic Structure and **T45** will link these concepts to the topic Delivery.

6. Conclusion

This paper proposes a semantic approach to characterise video segments to enable their use for learning. A generic video segment characterisation framework is presented which is underpinned by Ausubel's subsumption theory for meaningful learning and utilises an ontology to represent the domain. The framework considers both video transcripts and user comments when interacting with the videos (when available). It can be applied on any video segments and is fully independent from the segmentation. The transcript-based domain characterisation is applicable to any settings, assuming that the video content discusses the specific domain. The characterisation using past learner comments and the patterns comparing transcript and comments can be used if past learners' interactions are available. The availability of user generated interaction data is becoming highly popular, e.g. most video sharing platforms allow user comments, MOOCs integrate user interactions, such as comments, annotations, forums, to enhance the effectiveness of videos for learning. User interactions are also captured in bespoke video-based learning platforms like the AVW-Space which was used in this paper.

The video characterisation framework is applied on a dataset of 49 video segments derived from user interactions in a video-based learning system for presentation skills training. This allowed us to characterise video segments by identifying their focus topics, focus concepts, and coverage using video transcripts and comments of a fairly large user group. Based on the domain characterisation, we derive patterns which allow recommending and combining video segments to support meaningful learning.

In future work, we conduct evaluation to examine the added pedagogical value of the patterns for recommending video segments to learners. We will automate the generation of patterns and the generation of sequences of video segments based on subsumption links. Finally, we will utilise the

video characterisation to improve video segmentation based on the identified focus topics and concepts. To ensure the generality of the approach, we will apply to another context of using videos for soft skills learning, e.g. medical communication.

Acknowledgements. The authors wish to thank Prof. Tanja Mitrovic and her colleagues at the University of Canterbury New Zealand for sharing the user interaction data from past AVW-space user studies with university students at their university. The data used to derive the video segments was collected in a user study with postgraduate students using an earlier version of the AVW system at the University of Leeds, UK. The authors are grateful to Dr. Lydia Lau and Dr. Amali Weerasinghe for their contribution in conducting the user study.

References

Abolkasim, E.N.A.: Semantic Approach to Model Diversity in a Social Cloud.Ph.D. thesis, Univ of Leeds (2019)

Al-Tawil, M., Dimitrova, V., Thakker, D.: Using knowledge anchors to facilitate user exploration of data graphs. Semantic Web (Preprint), 1–30 (2019).

Ashangani, K., Wickramasinghe, K., De Silva, D., Gamwara, V., Nugaliyadde, A.,Mallawarachchi, Y.: Semantic video search by automatic video annotation using tensorflow. In: 2016 Manufacturing & Industrial Engineering Symposium (MIES).pp. 1–4. IEEE (2016).

Ausubel, D.P., Novak, J.D., Hanesian, H., et al.: Educational psychology: A cognitive view (1968).

Benkada, C., Moccozet, L.: Enriched interactive videos for teaching and learning. In: 2017 21st International Conference Information Visualisation (IV). pp. 344–349. IEEE (2017).

Chiu, P.S., Chen, H.C., Huang, Y.M., Liu, C.J., Liu, M.C., Shen, M.H.: A video annotation learning approach to improve the effects of video learning. Innovation in Education and Teaching **55**(4), 459–469 (2018)

Colasante, M., Douglas, K.: Prepare-participate-connect: Active learning with video annotation. Australasian Journal of Educational Technology **32**(4) (2016).

Dimitrova, V., Mitrovic, A., Piotrkowicz, A., Lau, L., Weerasinghe, A.: Using learning analytics to devise interactive personalised nudges for active video watching. In: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization. pp. 22–31. ACM (2017)

Dutta, A., Zisserman, A.: The via annotation software for images, audio and video. In: Proceedings of the 27th ACM International Conference on Multimedia. pp. 2276–2279. ACM (2019)

Goldenberg. R., Medioni, G.G., Meidan, O., Rivlin, E.B., Kumar, D.: Multi-video annotation (Mar 5 2019), US Patent 10,223,591

Hecking, T., Dimitrova, V., Mitrovic, A., Ulrich Hoppe, U.: Using network-text analysis to characterise learner engagement in active video watching. In: ICCE 2017 Proceedings. pp. 326–335 (2017)

Hsin, W.J., Cigas, J.: Short videos improve student learning in online education. Journal of Computing Sciences in Colleges **28**(5), 253–259 (2013)

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. Intern. Education Studies **7**(8), 56–67 (2014)

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)

Liu, C., Kim, J., Wang, H.C.: Conceptscape: Collaborative concept mapping for video learning. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. p. 387. ACM (2018)

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)

Mitrovic, A., Dimitrova, V., Weerasinghe, A., Lau, L.: Reflective experiential learning: Using active video watching for soft skills training. In: Proceedings of the 24th international conference on computers in education. Asia-Pacific Society for Computers in Education (2016)

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)

Rezazadeh Azar, E.: Semantic annotation of videos from equipment-intensive construction operations by shot recognition and probabilistic reasoning. Journal of Computing in Civil Engineering **31**(5), 04017042 (2017)

Risko, E.F., Anderson, N., Sarwal, A., Engelhardt, M., Kingstone, A.: Every day attention: Variation in mind wandering and memory in a lecture. Applied Cognitive Psychology **26**(2), 234–242 (2012)

Sikos, L.F.: Rdf-powered semantic video annotation tools with concept mapping to linked data for next-generation video indexing: a comprehensive review. Multimedia Tools and Applications **76**(12), 14437–14460 (2017)

Xue, Y., Song, Y., Li, C., Chiang, A.T., Ning, X.: Automatic video annotation system for archival sports video. In: 2017 IEEE Winter Applications of Computer Vision Workshops (WACVW). pp. 23–28. IEEE (2017)