Chen, W. et al. (Eds.) (2017). Proceedings of the 25th International Conference on Computers in Education.
New Zealand: Asia-Pacific Society for Computers in Education

Using Network-Text Analysis to Characterise Learner Engagement in Active Video Watching

Tobias HECKING[a]*, Vania DIMITROVA[b], Antonija MITROVIC[c] & H. Ulrich HOPPE[a]
[a]*COLLIDE Research Group, University of Duisburg-Essen, Germany*
[b]*School of Computing and Leeds Institute of Medical Education, University of Leeds, UK*
[c]*University of Canterbury, Christchurch, New Zealand*
*hecking@collide.info

Abstract: Video is becoming more and more popular as a learning medium in a variety of educational settings, ranging from flipped classrooms to MOOCs to informal learning. The prevailing educational usage of videos is based on watching prepared videos, which calls for accompanying video usage with activities to promote constructive learning. In the Active Video Watching (AVW) approach, learner engagement during video watching is induced via interactive notetaking, similar to video commenting in social video-sharing platforms. This coincides with the JuxtaLearn practice, in which student-created videos were shared on a social networking platform and commented by other students. Drawing on the experience of both AVW and JuxtaLearn, we combine and refine analysis techniques to characterise learner engagement. The approach draws on network-text analysis of learner-generated comments as a basis. This allows for capturing pedagogically relevant aspects of divergence, convergence and (dis-) continuity in textual commenting behaviour related to different learner types. The lexical-semantic analytics approach using learner-generated artefacts provides deep insights into learner engagement. This has broader application in video-based learning environments.

Keywords: video-based learning, learning analytics, network-text analysis

1. Introduction: Video-based Learning and Analytics

Learning by watching videos (Yousef, Chatti and Schroeder, 2014; Vieira, Lopes and Soares, 2014) is becoming more and more popular, especially in new learning contexts, such as flipped classrooms (Kurtz, Tsimerman and Stainer-Lavi, 2014), MOOCs (Guo, Kim and Rubin, 2014; Koedinger et al., 2015), or informal learning. Video watching per se is a passive activity, and therefore it is desirable to provide additional support for learner engagement for better educational benefit (Koedinger et al., 2015, Yousef, Chatti and Schroeder, 2014; Vieira, Lopes and Soares, 2014; Pardo et al., 2015). There are strong indications that increased engagement is more effective for learning. The ICAP framework of Chi and Wylie (2014) can serve as a theoretical reference for this hypothesis. Engagement during video watching can be supported by embedding interactive activities such as quizzes into videos (Giannakos, Sampson and Kidziński, 2016; Kleftodimos and Evangelidis, 2016; Kovacs, 2016; Wachtler et al., 2016), which requires additional effort from the teacher, or by video-annotation features (cf. Chatti et al., 2006) providing opportunities for students to annotate and discuss videos.

Inspired by these approaches, the Active Video Watching (AVW) system (Mitrovic et al., 2016) supports engagement during video watching to facilitate informal learning. The support includes providing micro-scaffolds to facilitate the commenting on videos and the reviewing of comments made by others. Our approach is primarily aimed at informal learning of soft (or “transferable”) skills, such as communicating, negotiating, collaborating, etc. Videos have been shown to be useful for teaching soft skills (Cronin and Cronin, 1992; Conkey et al., 2013), requiring that the learner reflects on his/her own experience and is able to see different perspectives. Another perspective on learning with videos has been adopted by the European project JuxtaLearn (Hoppe et al., 2016): here learners create “dramatised videos” that combine the explanation of science concepts with active storytelling. These learner-generated videos are then shared and discussed on a social media platform.

Drawing on the experience of both AVW and JuxtaLearn, we further explore and refine analytic techniques through which we can characterise the learning benefits, particularly learner engagement around videos. Previous studies performed with the AVW platform focusing on presentation skills show that only constructive learning results in increased conceptual understanding of the chosen soft skill (Mitrovic et al., 2017). Further analysis of constructive learning behaviour revealed that not all constructive learners increased their domain knowledge, and hence user-adaptive engagement support is needed. Consequently, we characterised constructive learners in three clusters, which informed adaptive support in the form of personalised "nudges" (Dimitrova et al. 2017).

Whereas the studies in the AVW context used data mining techniques related to activity parameters, learner-generated artefacts in the form of textual comments have not been considered. In contrast, the analyses in JuxtaLearn used as a main data source the comments as learner-generated textual artefacts (Daems et al., 2014). This approach was based on network-text analysis (NTA) as introduced by Carley (1997). Here core concepts (terms) are extracted from a given text (a video comment in this case) and arranged in a network. Connections between concepts are introduced based on co-occurrence of the corresponding terms in a window that slides over a normalised version of the text. The central claim of NTA is that the extracted networks are representations of the mental models underlying the texts (Carley, 1997). The transformation of textual artefacts into networks allows for applying network measures and network analysis techniques (Wasserman and Faust, 1994) for further analysis and interpretation of the given data.

Although the cluster analysis of AVW data revealed useful insights for the identification of learner types and the design of adaptive support (see Section 2), it did not reveal much detail regarding the actual differences of the learner types in terms of their commenting behaviour on the content level. Therefore, using only structured interaction log data did not provide sufficient insight into learner engagement with the videos. For such insights, we utilise the learner-generated artefacts. In this paper, we applied NTA to gain deeper insights into the lexical and semantic features of the learner comments. This allowed us to capture aspects of divergence/convergence of vocabulary related to the different learner types. Including temporal aspects, we can also identify continuity vs. variation in textual utterances. This allows for a deeper interpretation of the profiles associated with the learner types, which is a relevant premise to identifying pedagogical challenges and remedial actions.

The rest of the paper is organised as follows. Section 2 elaborates on the AVW system and findings from earlier analyses, including a categorization of learner types. Section 4 outlines the analysis approach and presents the finding of the network-text analysis of learner comments. Section 5 discusses the findings and points at implications for video-based learning environments.

2. Learning through Active Video Watching: the AVW System

The AVW system is a controlled video-watching environment that supports engagement during video watching via interactive notetaking, tapping into learners' familiarity with commenting on videos in social networking sites. AVW is customised by the teacher, who selects videos for students to watch, and defines mini-scaffolds for reflective learning. The AVW is particularly aimed at informal learning of soft skills; two studies have been conducted so far focusing on presentation skills and involving university students from engineering subjects (Mitrovic et al., 2016; Mitrovic et al., 2017, Dimitrova et al., 2017). Presentation skills, and transferable skills in general, are highly sought by employers and are crucial for employability (National Research Council, 2012; Walsh and Kotzee, 2010). Teaching soft skills to tertiary students in technical disciplines is challenging, as they are time-consuming and difficult to document (Anthony and Garner, 2016). Learners need to practice under various conditions, receive feedback, reflect on it and do more practice. Teachers typically do not have enough resources to provide such support to each individual student. AVW was developed to address these challenges by providing a video watching space for reflective learning.

Learning in AVW consists of two phases. In Phase 1, students watch and comment on videos individually, using aspects to tag their comments made anytime during the viewing. We selected eight videos from YouTube: four tutorials on how to give presentations, and four example presentations (two TED talks and two 3-minue PhD pitch presentations). The student can stop a video at any time, enter a comment and specify an aspect, which indicates the intention of the comment. For the tutorial videos, aspects aimed at stimulating reflection included: "*I didn't realise I wasn't doing it*" (TA2), "*I am rather good at this*" (TA3), "*I did/saw this in the past*" (TA4). There was one additional aspect, "*I like this point*" (TA1), to encourage the learner to externalize relevant learning points. For the example videos, the aspects corresponded to presentation skills covered in the tutorials, which included "*Delivery*" (EA1), "*Speech*" (EA2), "*Structure*" (EA3), and "*Visual aids*" (EA4).

In Phase 2, students review and rate each other's' anonymised comments, and can click on 'view video snippet' to watch the part of the video to which the comment refers. In such a way, the student can compare his/her own comments to those of others, and further reflect on their experience. The AVW instantiation for presentation skills included five categories for rating comments: "*This is useful for me*", "*I hadn't thought of this*", "*I didn't notice this*", "*I don't agree with this*", and "*I like this point.*"

Figure 1 presents a screenshot from the AVW instantiation for presentation skills, which was used in two studies. The overarching goal of the studies was to investigate whether AVW is beneficial for teaching soft skills. Both studies used the same set of videos, aspects and rating categories, as well as three surveys: (i) prior using the system, participant profiles were collected including demographic information, background experiences, motivation and attitudes using the Motivated Strategies for Learning Questionnaire (MLSQ) (Pintrich and de Groot, 1990) as well as domain knowledge about presentations; (ii) at the end of Phase 1, participants' knowledge of presentations was checked again, together with questionnaires measuring the users' cognitive load and perceived usefulness of the system (Hart, 2006; Davis, 1989); (iii) at the end, knowledge of presentations was tested again, and the system's cognitive load and perceived usefulness assessed again.

Figure 1. A screenshot illustrating browsing and rating comments in AVW

One of the AVW studies was conducted with undergraduate students at the University of Canterbury. It compared an experimental condition with reflection mini-scaffolds to the control condition that followed free YouTube-like video watching (Mitrovic et al, 2017). We found a significant increase in conceptual knowledge in the experimental group participants using mini-scaffolds when constructive learning behaviour was followed (i.e. active video watching by making comments and rating others' comments). There was no significant increase in conceptual knowledge in the control group. The other study, performed with postgraduate students from the Universities of Leeds and Canterbury, looked in depth into constructive learning behaviour (Dimitrova et al, 2017). Thirty-eight out of 48 participants completed all surveys and commented on videos. The initial analyses showed relatively high level of engagement: the participants made a total of 744 comments, and 2,706 ratings (Mitrovic et al., 2016). There were no significant differences between participants based on their gender, age or whether or not they were native English speakers.

Although all 38 participants were constructive learners, not all of them increased their conceptual knowledge after interacting with AVW. Clustering the participants based on their profiles, using k-means clustering, revealed three distinctive types of behaviours generated using the following variables: experience with giving presentations, using YouTube for learning, six MSLQ variables (self-efficacy, extrinsic motivation, academic control, rehearsal, self-regulation, organization), and conceptual knowledge score (Dimitrova et al., 2017). The significant differences between the learner clusters were identified using the 2-sided Kruskal-Wallis test (pairwise comparison with a Bonferroni correction). The seventeen learners in the **Cluster 1** exhibited **Parochial Learning** behaviour. They made relatively high numbers of comments/ratings, had the least presentation experience overall, and had generally low self-regulation and learning skills (they had the lowest MSLQ scores for self-efficacy, extrinsic motivation, rehearsal, self-regulation and organization). Surprisingly, they found AVW the most useful, yet there was no significant improvement of their conceptual knowledge. **Cluster 2** exhibited **Habitual Video Watching** behaviour. These learners were confident, self-regulated students who made fewer comments that the other clusters. At the same time, their conceptual knowledge at the start of the study was the lowest, and there was only a slight increase after using the system. We hypothesised that these participants might be used to watching videos in a passive way so they did not engage sufficiently. **Cluster 3** exhibited **Engaged Self-regulated (SR) Learning** behaviour. This was the "ideal" cluster illustrating the target user behaviour with AVW. The participants were actively engaged while watching the videos, making the highest number of comments and receiving the highest number of ratings on their comments. This cluster was the highest on previous experience and conceptual knowledge on the pretest, and lowest on using YouTube for learning. They significantly improved their conceptual knowledge scores after using the system.

Although the clusters enable characterizing constructive learning behaviour, they do not provide sufficient insights to understand what might be the users' attention while interacting with videos. E.g., do learners notice relevant points in the videos, are there any notable differences in attention between the three clusters, does engagement change with time, are there any notable links between learners? To answer these questions, and inform the design of intelligent scaffolding to facilitate active video watching, we used the leaners' artefacts generated during the interaction (i.e. the textual comments) and employed networked text analysis. The results are presented in the next section.

3. Computational Analysis of AVW Comments

3.1 Approach

As the first step, a **taxonomy of domain keywords** in comments was derived using a semi-automatic ontology engineering process. A middle-up ontology authoring approach (Uschold and Gruninger, 1996) was followed, starting from the learners' answers to questions about conceptual knowledge (bottom-up) and using key categories from several university guides on presentation skills (top-down). Three main categories of domain terms, related to the domain knowledge captured in the tutorial videos, were identified: structure, delivery and speech, and visual aids. For each category, the relevant terms were identified manually by three annotators working independently. As a start, a subset of comments was marked, the disagreements were discussed, and a unified approach for term selection was agreed. The domain terms in all answers to conceptual knowledge (38 students multiplied by 3 surveys for each student) were marked independently by each marker. The majority voting was used to select the relevant terms (a term in a participant's answer was seen as relevant if it was selected by at least two markers). All cases when there was no majority were examined by a fourth marker, who made changes to the term list. The list of terms was then used in text analysis of the user comments generated during the interaction in the AVW system (Dimitrova et al., 2017). The final refinement of the term list was made by adding the most frequent unigrams in the user comments which were missing from the original term list. This process resulted in a taxonomy of domain terms, including three upper level categories (structure, delivery and speech, visual aids), and the list of frequent domain terms.

Taking the taxonomy as a controlled vocabulary, the learner comments can be transformed into a **bipartite learner-keyword network**, similar to the approach used by Hoppe et al. (2016). Each learner is connected to all vocabulary terms occurring in at least one of his/her comments. Furthermore, the edges of the resulting network are annotated with further information, in particular, the time when the relation between the learner and the keyword was established, the learner's cluster (see Section 2), and a list of video types (example video or tutorial) where the learner used the keyword. The edge-annotated learner-keyword network can then easily be sampled into sub-networks. For example, the sub-network corresponding to the keyword affiliations of parochial learners to the first minute of tutorial videos can be derived by deleting all the edges from the original network that are not annotated by the corresponding cluster, video type, and timestamp. After that, nodes that became isolated through the edge deletions are deleted as well.

The network representation of vocabulary usage during video watching has several advantages. First, it is easy to identify central concepts that are used by many learners, or learners with a broad range of vocabulary terms by calculating the degree centrality of keyword or learner nodes respectively. Furthermore, structural properties of the network, such as the emergence of densely connected regions (or network subgroups) of learners and keywords indicate differences in the vocabulary usage, and thus, attention on different video aspects of subsets of learners.

3.2 Usage of Domain Vocabulary

Table 1 gives the most frequent keywords for each learner cluster. The values for the keywords correspond to the fraction of cluster members who used the word at least once, in particular the average degree of a keyword in the corresponding user-keyword network. The threshold was set to 0.5 meaning that all terms in the table column corresponding to a learner cluster were used by at least half of the cluster members. Terms like "presentation" and "story" are frequently used across different clusters, which is not surprising given the topic of the videos. In contrast, differences can be seen in the number of terms and their semantic orientation. Habitual video watchers do not have many shared terms. There are 12 terms used by more than the half of all Cluster 3 participants and the top 6 terms were used by at least 65%, which indicates that these learners tend to use a common vocabulary when commenting on videos. This observation will be further explained in Section 3.3.

Table 1: Keyword usage of different types (with distinguishing concepts in bold face)

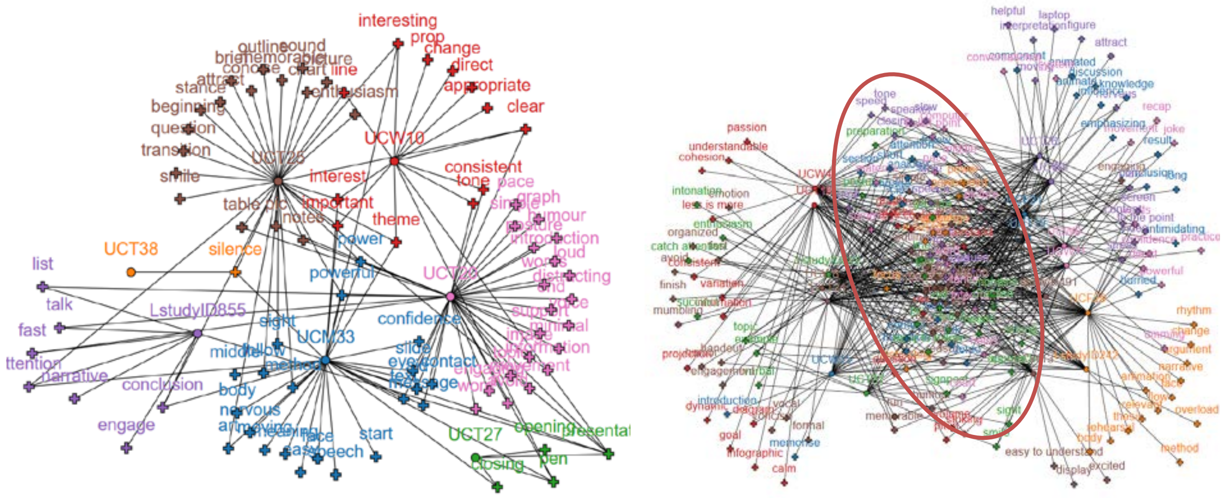
Parochial learners (14)		**Habitual video watchers (7)**		**Engaged SR learners (17)**	
Keyword	**Usage**	**Keyword**	**Usage**	**Keyword**	**Usage**
presentation	0.71	presentation	0.86	presentation	0.94
story	0.71	pen	0.57	slide	0.76
end	0.64			story	0.76
clear	0.64			**line**	0.65
beginning	0.57			pen	0.65
talk	0.57			art	0.65
speech	0.5			beginning	0.59
pen	0.5			end	0.59
art	0.5			**interest**	0.59
slide	0.5			**interesting**	0.53
				text	0.53
				eye contact	0.53

Table 1 gives a general view of the usage of domain vocabulary for all videos. A significant difference with respect to the used domain terms between example and tutorial videos could be found by a g-test ($G = 318.4$, $df = 216$, $p << 0.005$). The g-test is an alternative to the well-known chi^2 test which has become popular for computational linguistics since it can also be applied to compare two sparse term frequency vectors (Dunning, 1993). This difference can be especially observed within the learner clusters of parochial learners ($G=189.87$, $df=162$, $p=0.066$) and self-regulated learners ($G=225.7$, $df=177$, $p=0.007$).

By taking a deeper look into the comments made on example/tutorial videos, one can see that affirmative comments such as "Good idea" or "Very interesting" are more frequent in example videos, while especially SR learners post more comments regarding the concrete video content in tutorial videos. Furthermore, in tutorial videos SR learners show the most agreement on the vocabulary, having 6 terms used by more than the half of all learners in this cluster.

3.3 Agreement on Vocabulary and Shared Attention

The existence of a shared vocabulary in the absence of direct interactions, especially found in the cluster of SR learners (described in the previous section), is investigated in more detail in the following. Again, the analyses are based on the bipartite learner-keyword subnetworks described earlier. Bipartite modularity optimisation (Hecking, Steinert, Gohnert and Hoppe, 2014) was applied to identify densely connected regions (modules) in these networks. Learners and terms are assembled to modules such that within one module the learners are densely connected to a set of vocabulary terms, while the number of edges between modules is minimised. Each module, consequently, represents a set of learners who share a set of terms in their comments. Examples of such network partitions are depicted in Figure 2.



Little vocabulary overlap | **High vocabulary overlap**

Figure 2. Bi-partite sub-network of learners and vocabulary terms with high modularity (left: habitual video watchers) and low modularity (self-regulated learners).

By definition, modularity optimisation methods create a network partition, i.e. they assign each node to exactly one module, even though possibly this separation may not be very strong for the given network structure. The bipartite modularity (Barber, 2007) is a quality function, which measures how separated the modules of a given partitioning are. It takes the values -0.5 at minimum, 0 for a random partitioning, and 1 in case of perfect separation. Consequently, a low modularity for the identified modules in the learner-keyword networks indicates that a high number of high number of learners and keywords cannot be clearly separated into different modules. Positively speaking, this means that there a certain degree of common ground in terms of shared vocabulary between these actors. This can especially be observed for the SR learner cluster (right-hand side of Figure 2). The corresponding modularity values can be found in Table 2. Here, the 17 learners were split into 7 clusters for all videos and for the tutorial videos respectively. It can be seen that each module has some characteristic terms that are not used by the learners in other modules.

However, there is also a high share of terms that have many connections to learners of different modules, and thus, cannot be clearly assigned to a particular module. This gives further evidence that the SR learners have a certain agreement on the vocabulary used in their comments. The reason can be that the SR learners follow the videos thoughtfully and take up concepts from the videos in their postings. Particularly in video tutorials, it could be observed that these learners tend to post comments on the actual video content. In contrast, the habitual video watchers and parochial learners show a different behaviour. Since these learners tend to post more affirmative comments on the general style of the videos, their corresponding learner-keyword networks can be split into more separated modules, which results into higher modularity values (Table 2). The habitual video watchers denote an extreme case where each of the seven members of this cluster forms an own module in the networks (Table 2 and left-hand side of Figure 1).

Table 2: Characteristics of partitioned networks extracted from all videos vs. tutorials for different learner clusters

Cluster: Video type	No. modules	Modularity	Keywords / user	Users / keyword
Parochial: all videos	5	0.37	28.6	28.57
Parochial: tutorials	6	0.46	18.1	18.08
Habitual: all videos	7	0.45	18.43	18.43
Habitual: tutorials	7	0.46	14.43	14.48
Self-reg.: all videos	7	0.3	34.7	34.7
Self-reg.: tutorials	6	0.35	22.41	22.41

3.4 Attention Shift During Video Watching

In this section, the joint attention of the three learner clusters is analysed on a fine-grained level. The domain vocabulary terms used by the students in each minute of the particular video are investigated in order to identify different patterns that further characterise the engagement of different learner clusters. On the one hand, periods of high attention indicated by a high number of vocabulary terms per learner have to be identified. On the other hand, it is of interest how the used vocabulary changes during the course of the video.

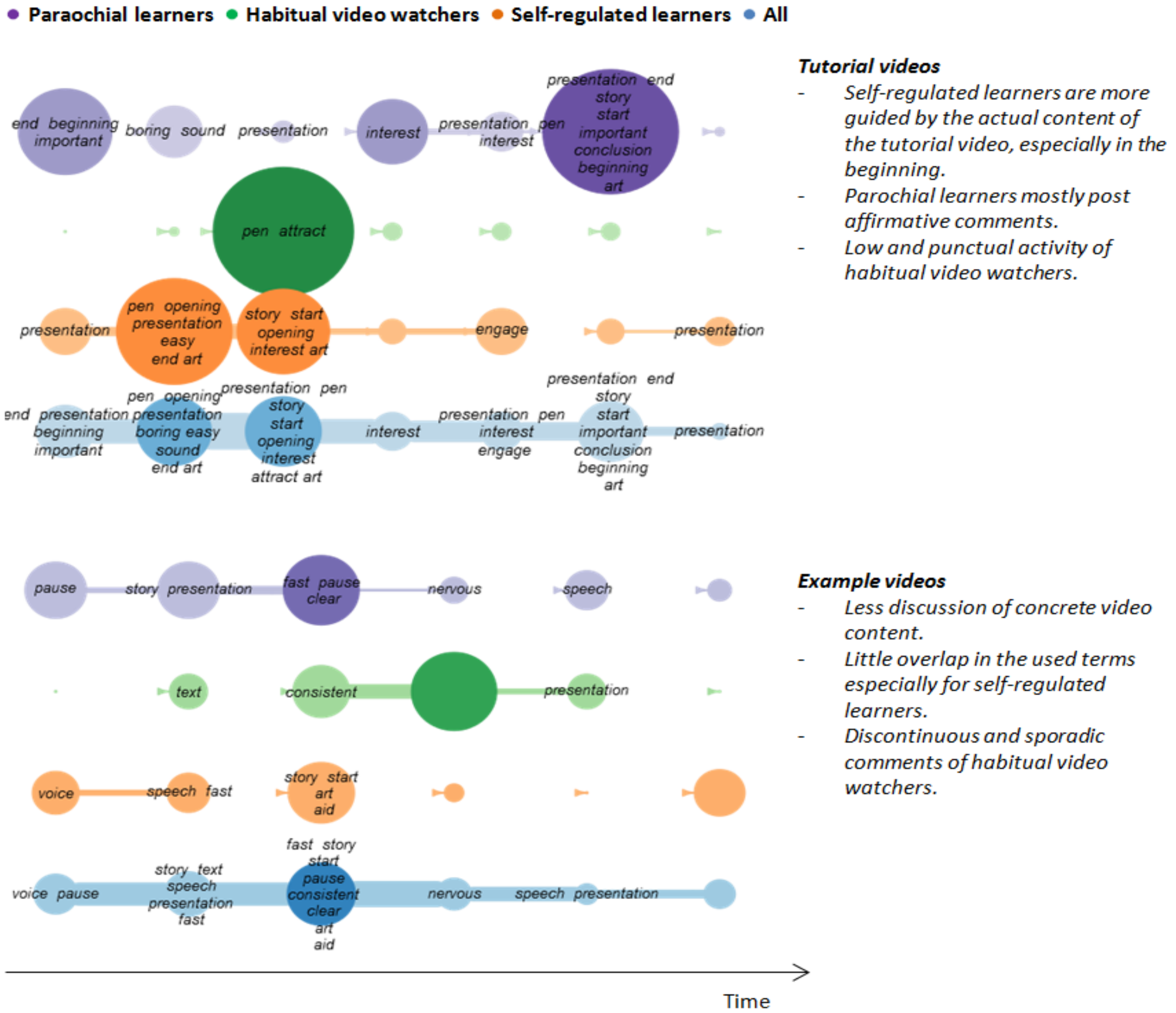


Figure 3. Attention diagrams for two example videos for different learner types

We developed an integrated visualisation that captures these aspects (Figure 3). The number of vocabulary terms per learner in every minute of a video is represented by the size of the circles on the horizontal axis. The more vocabulary terms have been used (indicating higher attention), the bigger the diameter of the circle is. The width of the horizontal arcs depicts the overlap of vocabulary terms in two consecutive minutes so that attention shift becomes visible. The terms in the circles are those that were used by at least two learners in the corresponding minute.

Figure 3 shows the attention diagrams for two videos (one tutorial and one example), to illustrate different patterns of attention shift for different learner clusters. The typical pattern for parochial video watchers is the high frequency of general and affirmative comments, as already stated above. These types of comments cannot be clearly attributed to a part of the video since video content is not mentioned. This explains the relatively high overlap in the used terminology. It can also be seen that the use of vocabulary terms is more or less evenly spread over the course of the videos. Habitual video watchers do not only have a low usage of domain terms, but also show discontinuous posting activity in certain minutes of the videos. Here it is important to mention that this cluster comprises of only seven learners, which can also partially explain the discontinuation of the posting activities. The SR learners show activity throughout the video, similar to the parochial learner cluster. However, there is a general tendency to write comments at the beginning of a video. In the tutorials, the usage of vocabulary terms is more oriented towards the content of the video indicating that the commenting activity of self-regulated learners is more guided by the topics discussed by the tutorials, and consequently, there is less continuity (or overlap) in the used terms in consecutive video minutes.

4. Discussion and Conclusion

In this paper, we applied the network-text analysis of video comments to investigate engagement related to attention and thematic focus of learners in active video watching tasks, where learners were supposed to watch examples and tutorials on giving presentations, and additionally post comments on specific parts of the videos. Earlier analysis (Dimitrova et al., 2017) showed that learners were subdivided into three different clusters with respect to their constructive learning behaviour and video watching habits. The overall goal of the lexical semantic analysis conducted in this paper was to gain further insights by having a closer look into actual content produced by these clusters, in particular, the usage of domain vocabulary in video comments. We believe that only such a combination of behaviour and content analysis for revealing characteristic patterns of engagement and attention can support tutors and designers of video-based learning scenarios in designing good videos. Furthermore, this can give indicators which guidance mechanisms could be established to improve the learning experience in active video watching tasks. Network-text analysis is especially suited to achieve these goals since it allows for extracting the overall relational structure of learners.

Insights gathered from NTA. Applying NTA, we were able to extract a lexical-semantic structure from the learner-generated artefacts in the form of unstructured textual comments. This provided an interpretable model analysed with well-established network analysis methods. There were several important insights which were not identified by the earlier analysis using only interaction logs.

Firstly, it was shown that the *usage of vocabulary terms differed significantly between example videos and tutorials*. Tutorial videos seemed to be better suited for engaging learners in reasoning about specific concepts, while examples trigger more comments referring to presentations in general, like “interesting”, or “speech”. The highest agreement on a common set of terms, and therefore, the existence of joint attention could be found for self-directed learners, while parochial and habitual video watchers only have very few keywords in common.

Secondly, the observation of the *emergence of a shared vocabulary was confirmed* based on a fine-grained analysis. The usage of domain vocabulary was investigated by identifying densely connected modules of learners and used terms in learner-keyword networks derived from the video comments. In particular, the identified network modules corresponding to the self-directed learners could be characterised by some unique terms, but in addition, there was also a high share of keywords that were used by almost all modules. This finding is interesting for further research on video-based learning, since it indicates that engaged self-directed learners are able to recognise important concepts from the videos and use them in their comments. Thus, it will be easier for these learners to find a common ground or at least common vocabulary in possible post-video group discussions. For other types of learners, this phenomenon was not very salient, especially for habitual video watchers with little background knowledge. This indicates that additional scaffolding would be needed for these learner types to point at important aspects in the videos.

Thirdly, we observed that *the attention of learners shifted during the course of a video*. It could be shown that the commenting behaviour and attention for particular learners differed between the three learner clusters. The parochial and self-directed learners tend to post comments throughout the entire duration of the videos. However, while parochial learners mainly write general affirmative comments and opinions, self-directed learners were more guided by the actual video content since their comments were more closely aligned to concepts discussed in specific moments of the video, especially in tutorials. Habitual video watchers show a very different pattern - little activity and posting comments for much narrower periods of the video.

Implications to video-based learning environments. As a possible consequence of the presented findings, future active video watching tasks could be enriched by scaffolding mechanisms. This can help learners to focus on important aspects and to find a vocabulary to express themselves in video comments more precisely. Thereby, the emergence of a shared vocabulary is desirable since this would be beneficial for post-video discussions and facilitate conceptual framing. This kind of lexical-semantic support can be achieved, for example, by presenting a list of important domain vocabulary terms to the learners. This would, on the one hand, guide their attention to important aspects discussed in the video (as identified in the comments), and on the other hand, help learners to find the wording for dedicated comments (by pointing at example comments).

Since the more specific comments could be found in tutorial videos that had to be watched before the example videos, it can be assumed that some of the learners had difficulties to apply the concepts mentioned explicitly in the tutorials when they commented on examples. By using NTA, relevant terms extracted from tutorial comments could be presented to the learner in example videos, which can support learners to put example video content in a conceptual framework. This can lead to more specific comments, helpful for triggering reasoning about the learning objective than simple affirmative comments.

The automatic analysis of learner-generated artefacts to gain an understanding of learner engagement with videos is applicable in a broader context. Starting from textual comments, it will be possible to derive notable links between learners and learner behaviour, to identify the areas of attention for a group of learners or for a specific learner, and to depict how attention changes during the video. We intend to apply NTA on a recently completed large user study with engineering students at the University of Canterbury. Furthermore, we will investigate the application of AVW in other domains where videos are used as part of soft skill training, e.g. communication skills in Medicine. This will allow us to validate/tune the learner clusters and derive lexical semantic characteristics for each cluster.

Acknowledgements

This research was supported by the EU-FP7-ICT-257184 ImREAL grant, a teaching development grant from the University of Canterbury, and an Ako Aotearoa grant. We thank Lydia Lau and Amali Weerasinghe for their contribution to designing and conducting the user study, and we thank all participants in the experimental study. Regarding the analysis methodology, we acknowledge to have drawn on results from the EU project JuxtaLearn (EU-FP7-ICT-317964).

References

Anthony, S., Garner, B. (2016). Teaching soft skills to business students: An analysis of multiple pedagogical methods. *Business and Professional Communication Quarterly*, *79*(3), 360-370.

Barber, M. J. (2007). Modularity and community detection in bipartite networks. *Physical Review E*, 76(6), 066102.

Carley, K. M. (1997). Extracting team mental models through textual analysis. *Journal of Organizational Behavior* 18, 533-558.

Chatti, M. A., Marinov, M., Sabov, O., et al. (2016). Video annotation and analytics in CourseMapper. *Smart Learning Environments*, *3*(1), 10.

Chi, M. T., Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219-243.

Conkey, C. A., Bowers, C., Cannon-Bowers, J., Sanchez, A. (2013). Machinima and video-based soft-skills training for frontline healthcare workers. *Games for health*, *2*(1), pp. 39-43.

Cronin, M. W., Cronin, K. A. (1992). Recent empirical studies of the pedagogical effects of interactive video instruction in “soft skill” areas. *Computing in Higher Education*, *3*(2), 53.

Daems, O., Erkens, M., Malzahn, N., Hoppe, H. U. (2014). Using content analysis and domain ontologies to check learners’ understanding of science concepts. *Journal of Computers in Education*, 1 (2-3), 113-131.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.

Dimitrova, V., Mitrovic, A., Piotrkowicz, A., Lau, L. Weerasinghe, A. (2017). Using learning analytics to devise interactive personalised nudges for active video watching. Paper accepted for UMAP 2017.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational linguistics*, 19(1), 61-74.

Giannakos, M., Sampson, D., Kidziński, Ł. (2016). Introduction to smart learning analytics: foundations and developments in video-based learning. *Smart Learning Environments*, *3*(1), 1-9.

Guo, P., J., Kim, J., Rubin, R. (2014). How video production affects student engagement: An Empirical Study of MOOC Videos. Proc. 1st ACM Conf. *Learning at Scale*, pp. 41-50.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In Proc. the Human Factors and Ergonomics Society annual meeting, 50(9), pp. 904-908, Sage Publications.

Hecking, T., Steinert, L., Gohnert, T., Hoppe, H. U. (2014). Incremental clustering of dynamic bipartite networks. In Network Intelligence Conference (ENIC), pp. 9-16.

Hoppe, H. U., Müller, M., Alissandrakis, A., Milrad, M., Schneegass, C., Malzahn, N. (2016). “VC/DC”-Video versus Domain Concepts in Comments to Learner-generated Science Videos. In *24th Int. Conf. on Computers in Education, (pp. 172-181)*. Asia-Pacific Society for Computers in Education.

Kleftodimos, A., Evangelidis, G. (2016). Using open source technologies and open internet resources for building an interactive video based learning environment that supports learning analytics. *Smart Learning Environments*, *3*(1), pp. 1-23.

Koedinger, K.R., Kim, J., Jia Z., McLaughlin E., Bier, N. (2015). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. Proc. 2nd *Learning @ Scale,* pp. 111-120.

Kovacs, G. (2016). Effects of in-video quizzes on MOOC lecture viewing, Proc. 3rd Learning @ Scale, pp. 31-40.

Kurtz, G., Tsimerman, A., Steiner-Lavi, O. (2014) The Flipped-Classroom Approach: The Answer to Future Learning? *European J. Open, Distance and E-learning*, 17(2), pp. 172-182.

Lau, L., Mitrovic, A., Weerasinghe, A., Dimitrova, V. (2016). Usability of an Active Video Watching system for soft skills training. Proc. 1st Int. *Workshop on Intelligent Mentoring Systems, ITS 2016*, Zagreb, Croatia.

Mitrovic, A., Dimitrova, V., Weerasinghe, A., Lau, L. (2016) Reflexive experiential learning using active video watching for soft skills training. *Proc. 24th Int. Conf. Computers in Education* (pp. 192-201). APSCE.

Mitrovic, A., Dimitrova, V., Lau, L., Weerasinghe, A., Mathews, M (2017) Supporting constructive video-based learning: requirements elicitation from exploratory studies. In: Baker, R. et al. (Eds.) Proc. 19th Int. Conf. on *Artificial Intelligence in Education* (accepted for publication).

National Research Council (2012) Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. Washington, DC: The National Academies Press.

Pardo, A., Mirriahi, N., Dawson, S., Zhao, Y., Zhao, A., Gašević, D. (2015). Identifying learning strategies associated with active use of video annotation software. In Proc. 5th Int. Conf. *Learning Analytics and Knowledge*, ACM, pp. 255-259.

Pintrich, P.R., De Groot, E.V. (1990) Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*, *82*(1), 33.

Uschold, M., Gruninger, M. (1996). Ontologies: Principles Methods and Applications. *Knowledge Engineering Review*, 11(2), 93-136.

Vieira, I., Lopes, A. P., Soares, F. (2014). The potential benefits of using videos in higher education. In *Proceedings of EDULEARN14 Conference* (pp. 0750-0756). IATED Publications.

Wachtler, J., Hubmann, M., Zöhrer, H., Ebner, M. (2016). An analysis of the use and effect of questions in interactive learning-videos. *Smart Learning Environments*, *3*(1), 13.

Walsh, A., Kotzee, B. (2010). Reconciling ‘graduateness’ and work-based learning. *Learning and Teaching in Higher Education*, (4-1), 36-50.

Wasserman, S., Faust, K. (1994). Social Network Analysis: Methods and applications. Cambridge University Press.

Yousef, A. M. F., Chatti, M. A., Schroeder, U. (2014). The state of video-based learning: A review and future perspectives. *Int. J. Adv. Life Sci*, *6*(3/4), 122-135.