Investigating Learner’s Online Learning Behavioural Changes during the COVID-19 Pandemic

Kok Khiang Lim  
Nanyang Technological University, Singapore  
w200004@e.ntu.edu.sg

Chei Sian Lee  
Nanyang Technological University, Singapore  
leecs@ntu.edu.sg

ABSTRACT
The objective of this study is to understand the pandemic’s impact on online learning behaviour based on learners’ comments collected from online video tutorials hosted on YouTube. The topic modelling approach is employed to uncover the changes in topic prevalence during the pre- and post-pandemic timeframe to infer learning behaviour. Ten topics were uncovered, and each exhibited a varying degree of changes over time. Overall, the study identified two learning behavioural changes, (1) learners were more active in learning relevant skillsets through sharing their experiences; and (2) learners had altered their help-seeking behaviour to aid in their learning.

KEYWORDS
Online learning, topic modelling, pandemic, YouTube, user-generated content.

INTRODUCTION
On March 12, 2020, the World Health Organization declared COVID-19 as a global pandemic after assessing the severity of the virus (World Health Organization, 2020). Many countries have taken unprecedented measures to lockdown the cities, including the closure of schools to curb the virus transmission. As a result, learning has changed drastically, seeing a sharp increase in online learning (Shah, 2020; YouTube, n.d.), whereby learners access the learning content remotely on a digital platform such as massive open online courses and video tutorial hosting sites. Amid the high participation rate, there are concerns about the pandemic’s impact on online learning (Mheidly et al., 2020). Studies have since focused on areas, for example, learners’ engagement, satisfaction, and perception of online learning (Alawamleh et al., 2020; Baber, 2020). There are also indications to suggest that the pandemic could affect behavioural changes in online activities (Popa et al., 2020; Rosa et al., 2020).

To date, there is limited study on the pandemic’s impact on learners engaging in online video tutorials, in particular, on the use of learners’ online comments to uncover changes in learning behaviour. Posting online comments allows learners to share their learning experiences publicly or even seek guidance or clarifications when they encounter learning difficulties. Past studies have shown that it is valuable to analyse comments to understand commenters’ concerns or satisfaction (Chen et al., 2020; He et al., 2020; Lee et al., 2017), and feasible to use the computational technique to analyse user behaviour in an online environment (Kim & Cho, 2020). Thus, this study aims to address the current gap by proposing a topic modelling approach to analyse learners’ online comments from the content hosting site, YouTube, to understand learners respond to online learning before and after the declaration of the pandemic.

METHOD
According to a report by World Economic Forum, many people have picked up coding as a new skill or reskill themselves as technology jobs were seen to provide better job security amid the COVID-19 pandemic (Marchant, 2021). As such, the dataset was built from a popular YouTube channel, the freeCodeCamp, that curated coding-related video tutorials and have a large number of subscribers. A total of 52,431 comments were collected from 24 video tutorials (M = 2,185, SD = 3,437) that had more than a million views and published before 2019. Structural topic modelling (STM) was used for analysis as it had the advantage to incorporates documents’ metadata such as date into the model in the form of covariates to estimate the covariates’ moderation effect on the topics (Roberts et al., 2013).

First, the comments were filtered to focus on the seven months before and after the pandemic announcement to assess for potential changes in the two periods. Words were extracted from the comments and pre-processed by removing stopwords, high-occurrence words, numbers, punctuations, and followed by auto-correction and lemmatisation of the remaining terms. Next, STM was used to uncover latent topics and compute the topic prevalence. Before building the STM model, the number of topics was evaluated using three metrics: held-out likelihood estimation, semantic coherence, and residuals (Naab & Sehl, 2016; Roberts et al., 2013). The STM model was then fine-tuned to obtain a good model fit determined by exclusivity and semantic coherence (Naab & Sehl, 2016).

84th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 3, 2021 | Salt Lake City, UT. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.
Topics generated were manually assigned with meaningful and representable labels to describe the topics based on the uncovered keywords and their associated comments.

RESULTS
The STM model uncovered ten topics and broadly grouped into two distinct categories: text-based topics and coding-based topics. Text-based topics consist of descriptive comments, while coding-based topics comprise primarily programming languages. For example, Topic 2 (free, exam, pass), categorised under text-based topics, reflected how learners engaged in the learning activities. Similarly, in Topic 6 (work, error, create), learners posted coding questions and codes related to technical errors, and thus categorised as coding-based topics. The text-based topics comprised Topic 2, 3, 8 and 10, were associated with sharing learning practices, sharing compliments, sharing general feedback, and sharing specific feedback on video content, respectively. The remaining six topics were coding related comments that constituted about 50.7% of the overall expected topic proportion. These topics included Topic 1 (sharing learning achievement), Topic 4 (sharing clarifications), Topic 5 (sharing coding errors), Topic 6 (sharing technical errors), Topic 7 (sharing coding queries), and Topic 9 (sharing programming codes).

The combined topic prevalence plot in Figure 1 showed that nine topics converged towards a similar topic proportion during $T_2$, suggesting a shift in learning behaviours. Meanwhile, Topic 3 remains relatively high and stable throughout $T_1$ and $T_2$ compared to other topics. Figure 2 showed an example of the detailed topic prevalence plots for Topic 2 and 6 that exhibited a sharp rise and significant drop in the topic proportion leading towards $T_2$, respectively.

DISCUSSION AND CONCLUSION
This study presents a preliminary analysis on the use of topic modelling to study the pandemic’s impact on learners engaging in online video tutorials and infer learning behaviour changes. Latent topics were first identified by the STM model, followed by computing the topics prevalence over the pre- and post-pandemic timeframe.

The topic prevalence plots showed that there were differences in learning behaviour between the pre- and post-pandemic. The convergence of the nine topics and the high fluctuation of the topics observed during $T_2$ reflected the pandemic’s volatility and effects. This phenomenon could be attributed to the different lockdown periods among the various countries during $T_2$ and a higher take-up and completion rate of learning (Hirsch, 2020; YouTube, n.d.). In particular, Topic 2 and 6 exhibited the most significant changes in topic prevalence. Topic 2 offered a perspective that active learning occurred during $T_2$ when more learners shared their learning experiences during the lockdown period. The drop in Topic 6 suggested a possible change in the type of questions posted during $T_2$. In sum, these changes highlighted events such as the pandemic could alter learners’ concerns and learning focus. Thus, it is important for educators and content creators to quickly identify these changes to better address learners’ needs and support them.

This preliminary study may be limited by the homonyms used among the comments. Thus, it required additional effort to ensure that the keywords generated were interpreted correctly against its associated comments. To address this, we plan to explore the use of semantic-based topic modelling to improve topic identification. In addition, this study focuses on coding-related learning and could not be generalised to other types of courses. Hence, we also plan to investigate other learning domains (e.g., learning languages). Against the backdrop of interrupted schooling from the pandemic where online learning will become more common and increasingly important, this study is critical and contributes to better understanding of how learners responded to online learning during the pandemic.
REFERENCES
Alawamleh, M., Al-Twait, L. M., & Al-Saht, G. R. (2020). The effect of online learning on communication between instructors and students during Covid-19 pandemic. *Asian Education and Development Studies*. Advance online publication. https://doi.org/10.1108/AEDS-06-2020-0131

Baber, H. (2020). Determinants of students' perceived learning outcome and satisfaction in online learning during the pandemic of COVID19. *Journal of Education and e-learning Research, 7*(3), 285-292. https://ssrn.com/abstract=3679489

Chen, T., Peng, L., Yin, X., Rong, J., Yang, J., & Cong, G. (2020). Analysis of user satisfaction with online education platforms in China during the COVID-19 pandemic. *Healthcare, 8*(200). https://doi.org/10.3390/healthcare8030200

He, L., Han, D., Zhou, X., & Qu, Z. (2020). The voice of drug consumers: Online textual review analysis using structural topic model. *International Journal of Environment Research and Public Health, 17*(3648). https://doi.org/10.3390/ijerph17103648

Hirsch, C. (2020, March 21). *Europe’s coronavirus lockdown measures compared*. Politico. https://www.politico.eu/article/europes-coronavirus-lockdown-measures-compared/

Kim, S. H., & Cho, H. G. (2020). User-topic modeling for online community analysis. *Applied Sciences, 10*(3388). https://doi.org/10.3390/app10103388

Lee, C. S., Osop, H., Goh, D. H. L., & Kelni, G. (2017). Making sense of comments on YouTube educational videos: A self-directed learning perspective. *Online Information Review, 41*(5), 611-625. https://doi.org/10.1108/OIR-09-2016-0274

Marchant, N. (2021, March 21). *These are some of Coursera’s most popular coding courses*. World Economic Forum. Retrieved from https://www.weforum.org/agenda/2021/03/popular-coding-courses-coursera/

Mheidly, N., Fares, M. Y., & Fares, J. (2020). Coping with stress and burnout associated with telecommunication and online learning. *Frontiers in Public Health, 8*(574969). https://doi.org/10.3389/fpubh.2020.574969

Naab, T. K., & Sehl, A. (2016). Studies of user-generated content: A systematic review. *Journalism, 18*(10), 1256-1273. https://doi.org/10.1177/1464884916673557

Popa, D., Repanovici, A., Lupu, D., Norel, M., & Coman, C. (2020). Using mixed methods to understand teaching and learning in COVID 19 times. *Sustainability, 12*(8726). https://doi.org/10.3390/su12208726

Roberts, M. E., Stewart, B. M., Tingley, D., & Airoldi, E. M. (2013). The structural topic model and applied social science. *Neural Information Processing Society*. http://scholar.harvard.edu/dtingley/node/132666

Rosa, R. L., De Silva, M. J., Silva, D. H., Ayub, M. S., Carrillo, D., Nardelli, P. H. J., & Rodriguez, D. Z. (2020). Event detection system based on user behavior changes in online social networks: Case of the COVID-19 pandemic. *IEEE Access, 8*, 158806-158825. https://doi.org/10.1109/ACCESS.2020.3020391

Shah, D. (2020, August 16). *By the numbers: MOOCs during the pandemic*. The Report by Class Central. Retrieved from https://www.classcentral.com/report/mooc-stats-pandemic

World Health Organization. (2020). *Coronavirus disease (COVID-19) pandemic*. Retrieved from https://www.who.int/emergencies/diseases/novel-coronavirus-2019

YouTube (n.d.). *Watching the pandemic: What YouTube trends reveal about human needs during COVID-19*. YouTube Culture & Trend. Retrieved from https://www.youtube.com/trends/articles/covid-impact/