Personalising and optimising online learning strategies by dynamic Weighted Probability Allocation (DWPA) and Logistic regression.

Liyanage M.L.A.P.
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it19120812@my.sliit.lk

Dinindu Koliya Harshanath Webadu Wedanage
Smart Infrastructure Facility
University of Wollongong
Wollongong, Australia
dkhww937@uowmail.edu.au

Archchana Kugathasan
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
archchana.k@sliit.lk

Samantha Thelijjagoda
SLIIT Business School
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
samantha.t@sliit.lk

***Abstract*—will be write at the EOF**

***Index Terms*—COVID-19, ADHD, Fuzzy-C means, Weight based indexing, Learning strategies, gamification**

I. INTRODUCTION

As a result of the widespread COVID-19 pandemic, the majority of day-to-day operations have been moved online. Of these areas, the provision of education via the internet has been receiving a substantial amount of attention. However, as a result of the rapid movement toward the provision of education online, it has brought to light that the vast majority of students and instructors are struggling with issues relating to their levels of productivity and their ability to deliver content in a reliable manner. When it comes to the factors that are affecting the productivity of online education, various health conditions of the student and the personal preference of the learning strategies of a student, types of devices students are using to learn in online education, and the literacy of usage of IT related methodologies can be highlighted as some of the most important factors. It has been said again that it is critical to pay attention to the consistent delivery of content because it is dependent on the type of device the student is using and how well they know how to use IT-related ways. Based on current technological trends and improvements, it's safe to say that one day, distance learning will replace in-class education and give every student the best and most productive education, no matter what their social or political problems are.

II. LITERATURE SURVEY

According to the previous case studies that has been conducted before, it can be identified that there is a rising amount of researches conducting the practices to provide a quality and productive online education to students. in accord with the literature survey, it was able to identify that, the above mentioned factors in the introduction are the main factors affecting to the quality and the productivity of the online education.

According to a case study conducted by Sindiani, Amer Mahmoud, et al on the topic of "Distance Education During the COVID-19 Outbreak: A Cross-Sectional Study Among Medical Students in Jordan" [1], 2212 out of 3700 students, or 55.8 percent, began to engage with online lecturers after 3 weeks of the module's start date. Another 31.4 percent learned from live lectures, and 22.8 percent gained the most from recorded lectures. The preceding case study concentrated mostly on medical students at Jordan University of Science and Technology (JUST). In conclusion, the survey found that the majority of students favored traditional face-to-face learning over online education, and they advised that online education be improved by personalizing the system with well-established infrastructure and learning methodologies.

Furthermore, it has been mentioned that there is a positive correlation between productivity and self-regulated learning strategies in a case study that was carried out by Daeyeoul Lee, Sunnie Lee Watson, and William R. Watson on the topic of "The Relationships Between Self-Efficacy, Task Value, and Self-Regulated Learning Strategies in Massive Open Online Courses" [2]. With the use of Pearson's correlation analysis, the case study was able to demonstrate the distinction between high levels of self-efficiency and low levels of self-efficiency.

Yang Tzu-Chi has done research on how observational learning and self-regulated learning strategies can affect the online learning performance of student [3]. The case study has focused on how observational learning (OL) and self-regulated

Identify applicable funding agency here. If none, delete this.

learning (SRL) can link up with the online learning strategies. With the research context, it was identified that, the learning performance can be positively altered with proper identification of behavioral patterns of the student. [3]. Furthermore, the implemented system has dual proposed mechanisms to support both OL and SRL [3]. It includes, setting and externalizing the goals, planning the learning strategies and the time student is willing/planning to study a particular subject area. The study was conducted in a classroom, with 2 main assessment tests as pre and post tests. All the students were made aware about the procedure which includes weekly online tests, after each lecture was delivered. Before the weekly assessments, the students were made to face a classic-type examination and after the online tutoring, there was another post-test to attend. With the experimental results, it was highlighted that there is no clear relationship between OL and SRL. Furthermore, it has identified that, students with SRL strategies have high performance in learning.

Jim B.J.Huang et al has done a research on exploring the learning strategies by sequence clustering and analyzing their correlation with student's engagement and learning outcome [5]. This case study has followed and iterative process of collecting and analyzing data to fine-tune the identification of the strategies. They have used, pre-defied questions and interviews to collect the data. With the test results, they were able to identify the learning strategy by the logs of the student test results, identify the relationship between learning strategy and learning outcome. The research was done with 53 college students with the data of python course. the questionnaire developed by the Elmaadaway [4] has made with 3 sectors covering behavioral, cognitive and emotional engagement.

Fidelia O. and Et al have done research [13] on how finding a relationship between students engagement and his/her performance. Thay have primareliy focused on Total time spent in MindTap (TimeOnTask), Number of logins, Percentage of Activities Accessed featues when building up the relationship. They have used both supervised learning (Random forests) and unsupervised learning (clustering) models to fine-tune the final output.

Furthermore, according to a review article which is referenced at [7] done by David Bueno, there are three main contributing factors that are influence to the satisfactory level of online education. They are, ”Online Learner Factors”, ”Online Instructor Factors” and ”Online Platform Factors”. Since this case study is mainly focused on optimising the learner experience, more attention has been paid to the factors such as gender, age, health problems, device types which were fallen under Online Learner Factors.

III. METHODOLOGY

To identify and assign necessary possibility elements to each of the aforementioned learning strategies, a novel Dynamic Weighted Probability Allocation (DWPA) method has been proposed in this case study. The new DWPA algorithm determines the necessary weights for the weighted function using logistic regression and a straightforward weighted arithmetic mean. The newly developed algorithm will take into account student's health characteristics, such as hearing and vision disorders, as well as other focus-related challenges, while creating the necessary weight elements. Additionally, it was primarily concerned with the kind of device the learner was using to access the educational materials. The above key characteristics have been identified among the features mentioned in the section 2 with the help of ”hyper-parameter tuning” concept. Each student receives a unique weight factor based on their health and other relevant factors.

A. Data Collection

Since the introduced DWPA algorithm is based on logistic regression, sample data has been collected from 388 students via an online survey conducted through Google forms. Majority of the sample were university undergraduates (95 percent) and the rest were school children from age 15 to 19. From the sample, 99 percent of the students are fully engaging in distance learning while the rest are engaging hybrid education.

B. Process Followed

Before dive into the creation of the hypothesis function, hyper-parameter tuning process was done in order to identify the contributing parameters. In the data collection phase, the audience was asked several questions to clearly identify the factors they are much concerned about in the online learning.

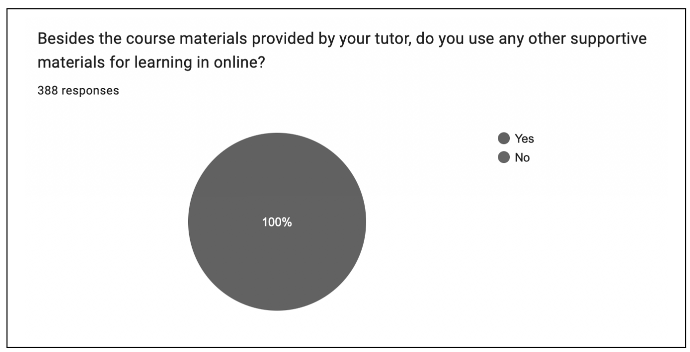


Fig 1. Feedback to the question asked regarding the usage of supportive materials in the online learning

According to the Figure 1. it has been identified that the entire audience is using supportive materials in the online education.

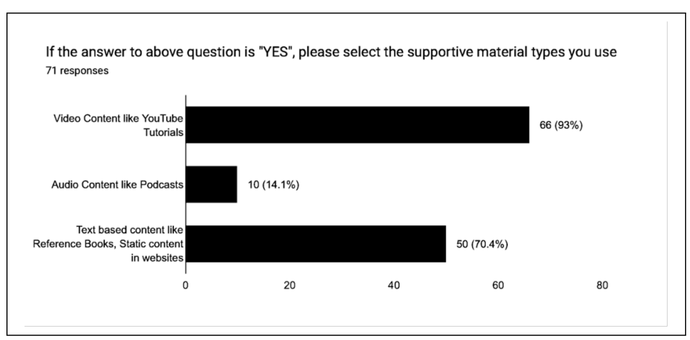


Fig 2. Feedback to the question asked regarding the usage of supportive materials in the online learning

Furthermore, 3 main learning strategies has been identified as Audio, Video and Text. it can be concluded that, all the supportive materials that the students are using is falls under these three main strategies. Therefore, the case study has mainly focused on categorizing and optimising each students learning curve in-align with the above three main strategies.

Additional to the above, it was identified that, the compatibility of adoption to the online learning of students have been limited to few of the other parameters as well. According to the survey results, and the information gathered form the literature survey, below factors were identified as contributing factors to have a smooth online learning experience.

- Age
- Gender
- Specific health conditions related to learning
- Device type - D

From the survey feed-backs, it was identified that considerable number of students have Hearing(H), Vision(V), Focus related health issues(F). Therefore, specific health conditions have been further divided in to above sub-sections.

After identification of the contributing factors, with the help of "hyper-parameter optimization" technique, the above factors have been narrowed down to have a fine-grained hypothesis function. The hyper-parameter tuning phase concluded that, the aforementioned health conditions and the device type has a higher impact on the final probability factor.

Therefore, the newly introduced DWPA algorithm is based on linear polynomial function with these 4 distinct parameters(H, V, F, D).

Below is the primary arithmetic equation which is used to generate the probability weight of a given learning strategy. It is also the core hypothesis function which is used to calculate the probability weights of a given learning strategy.

$$P(\alpha) = (H(\alpha)\theta^h + V(\alpha)\theta^v + F(\alpha)\theta^f + D(\alpha)\theta^d) / \sum_{i=h}^{d} \theta_i$$

Fig 3. Core hypothesis function

Here the α value represents the the 3 learning strategies namely video, audio and text.

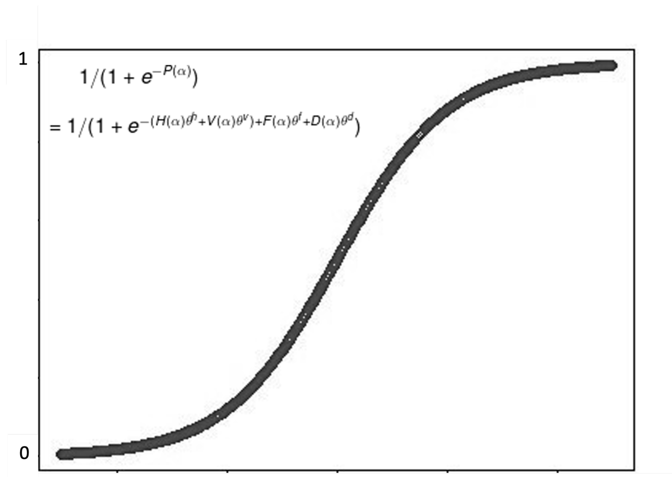
$$\alpha = \left\{ \begin{matrix} Audio \\ Video \\ Text \end{matrix} \right\}$$

Fig 2. Possible alpha values

But, given the fact that the P(α) value should a probability value and it should lie between 0 and 1, the aforementioned equation is being converted into a **sigmoid** function. This is where the principals of logistic regression is used in the DWPA algorithm. After the transformation of the above hypothesis function, below is the simplified version of the final equation.

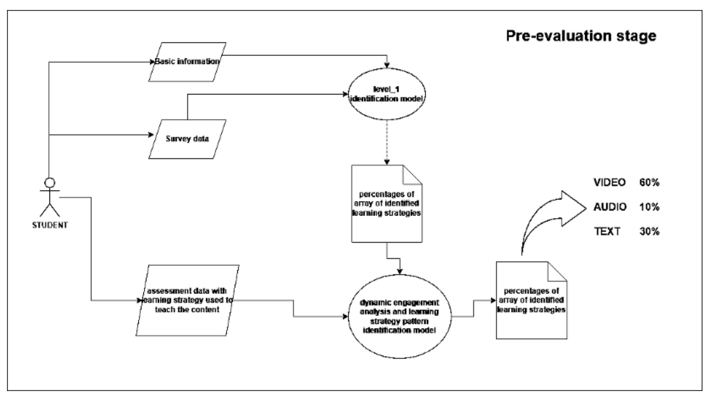
$$\begin{aligned} & 1/(1 + e^{P(\alpha)}) \\ &= 1/(1 + e^{(H(\alpha)\theta^h + V(\alpha)\theta^v + F(\alpha)\theta^f + D(\alpha)\theta^d) / \sum_{i=h}^{d} \theta_i}) \end{aligned}$$

For more clarity, the plot for the newly introduced function will look like below.



Here, in order to find the best values for θ, **gradient descendent** method has been used. With the help of multivariate gradient descendent, the function will converge into the best possible θ values. Since there is a requirement to fine-tune the weight factors (θ) values of the student, the proposed algorithm will re-run the workflow with the updated values for H, V, F and D.

Below is the high level workflow diagrams of the newly proposed algorithm.



C. *Workflow Accuracy and Justifications*

Focusing of the accuracy of the workflow, it has been optimised with the help of multivariate gradient decedent algorithm. The motivation behind the workflow is to fine-tune the assigned learning strategy probability weights each after the student is assigned with new H, V, F, D values. Therefore, the model will explore unique ways of reaching the global optima of the given data set more frequently resulting the perfect adaptation to the data set of the model Furthermore, since the proposed core equations is based on weighted average principals, the error impact on a given scenario will be minimized.

IV. Conclusion

Students engagement in the education is a major contributing factor towards the best performance no matter whether is distance learning or in-class learning. Yet in the distance learning, given the fact that the tutor/teacher do not have full closure with his/her students, boosting the engagement of the students is a must. Withing this case study, a novel algorithm called Dynamic Weighted Probability Allocation (DWPA) is being introduced. The intention of this algorithm is to fetch the students with personalized learning materials in order to increase the productivity of education. The output of aforementioned function will give fine-tuned probability weights for each learning strategy. Apart from that, with the help of gradient descendent algorithm, the values has been fine-tuned.

References

[1] Sindiani, Amer Mahmoud, et al. 'Distance Education during the COVID-19 Outbreak: A Cross-Sectional Study among Medical Students in North of Jordan'. Annals of Medicine and Surgery, vol. 59, Nov. 2020, pp. 186–94. ScienceDirect, https://doi.org/10.1016/j.amsu.2020.09.036.

[2] Lee, Daeyeoul, et al. 'The Relationships Between Self-Efficacy, Task Value, and Self-Regulated Learning Strategies in Massive Open Online Courses'. International Review of Research in Open and Distributed Learning, vol. 21, no. 1, 2020, pp. 23–39. www.erudit.org, https://doi.org/10.19173/irrodl.v20i5.4389.

[3] Y. Tzu-Chi, "Impacts of Observational Learning and Self-regulated Learning Mechanisms on Online Learning Performance: A Case Study on High School Mathematics Course," in 2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT), Tartu, Estonia, Jul. 2020, pp. 194–197. doi: 10.1109/ICALT49669.2020.00063.

[4] M. Elmaadaway, "The effects of a flipped classroom approach on class engagement and skill performance in a Blackboard course: Effects of the flipped classroom approach," British Journal of Educational Technology, vol. 49, Mar. 2017, doi: 10.1111/bjet.12553.

[5] J. B. J. Huang, A. Y. Q. Huang, O. H. T. Lu, and S. J. H. Yang, "Exploring Learning Strategies by Sequence Clustering and Analysing their Correlation with Student's Engagement and Learning Outcome," in 2021 International Conference on Advanced Learning Technologies (ICALT), Jul. 2021, pp. 360–362. doi: 10.1109/ICALT52272.2021.00115.

[6] "Gradient Descent for Multivariable Regression in Python — by Hoang Phong — Medium." https://medium.com/@IwriteDSblog/gradient-descent-for-multivariable-regression-in-python-d430eb5d2cd8 (accessed Jun. 30, 2022).

[7] "Frontiers — Factors Influencing Online Learning Satisfaction." https://www.frontiersin.org/articles/10.3389/fpsyg.2022.852360/fullB23 (accessed Jul. 29, 2022).