InPath Forum: A Real-Time Learning Analytics and Performance Ranking Forum System

ADELINE CHEW YAO YI1, TEY KAI YING1, SIANG JO YEE1, WEE MEE CHIN1, AND TING TIN TIN2

1School of Information Technology, Monash University Malaysia, Subang Jaya, Selangor 47500, Malaysia
2Faculty of Data Science and Information Technology, INTI International University, Nilai, Negeri Sembilan 71800, Malaysia

Corresponding author: Wee Mee Chin (wee.meechin@monash.edu)

ABSTRACT Numerous studies have been conducted on the influence of peers on students’ learning processes and their participation in online forums. However, these studies are limited in terms of system functionality and lack real-time analysis. In this study, we present InPath, a real-time analytics forum system, to rank and provide feedback on the performance of online participation of students. We used the other students of the online discussion forum performance as an effective reference point to inspire forum participation. A set of learning metrics was generated to analyze the contributions to online forums. The K-means clustering method was used to classify the students into three broad levels: Hall of Fame, All Star, and Rookies. The results showed that students with higher badge levels were more likely to spend more time on forums. In summary, this study highlights the implications of this state-of-the-art system for learning analytics on online forums, including supporting instructors and students in determining overall and individual performance on the forum.

INDEX TERMS Real-time learning analytics, online discussion forums, feedback system, ranking system, visualisation.

I. INTRODUCTION

In 2020, the COVID-19 pandemic greatly impacted the educational landscape by affecting 94% of the world’s student population in more than 190 countries [1], [2]. According to the researchers, the number of students forced to stay home due to educational institution closures reached a high of 1.598 billion from 194 countries on April 1, 2020 [3]. One year after the pandemic, almost 50% of the world’s students are still affected by partial or full school closures [1]. Based on research conducted to study how delivery of education is adapted to online learning in response to the pandemic, the author discussed online discussion forums as an online asynchronous platform that has improved student content, student education, and student-student interaction [5], [6]. Participation in the online discussion group ensures that students become intimately involved with the course material and understand the meaning of the content through interaction with their peers [7].

There is a lack of existing studies that investigate student participation in online discussion forums by tracking their performance and providing feedback in real time. Table 1 highlights the gap in existing research with respect to system functionality, such as real-time performance analysis, performance classification using machine learning algorithms, performance ranking, performance feedback mechanisms, and a dashboard for data visualisation. All features listed in Table 1 were incorporated into our proposed system (InPath).

We investigate how the students participated and contributed to the online discussion forum. Our main goal is to develop a real-time system that allows both students and educators to evaluate and reflect on student participation in the online discussion forum. In Section II, we present the proposed learning metrics as a measurement of the of students of participation level in online forums based on the literature and availability of data. Then, we explain the data flow and machine-learning approach used in the proposed system. We used peer success as a reference point for students to compare their participation. We introduce a hierarchical badge system that categorises students into three badge levels.
TABLE 1. Comparison of existing research with our proposed features in the online discussion forum.

| Source | Real-time Analysis | Machine Learning | Ranking | Feedback | Data Visualization |
|--------|-------------------|-----------------|---------|----------|-------------------|
| [5]    | No                | No              | No      | No       | No                |
| [9]    | No                | No              | Yes     | Yes      | Yes               |
| [11]   | No                | Yes             | No      | No       | No                |
| [12]   | No                | No              | No      | Yes      | No                |
| [16]   | No                | No              | No      | No       | No                |
| [17]   | No                | Yes             | No      | No       | No                |
| [18]   | No                | No              | No      | No       | No                |
| [19]   | Yes               | Yes             | Yes     | Yes      | Yes               |
| InPath | Yes               | Yes             | Yes     | Yes      | Yes               |

by applying machine learning clustering algorithms based on the participation level of each student. We also provide a feedback mechanism for student performance based on the proposed metrics to promote self-efficacy and self-regulated learning skills [8]. As a result, we present a dashboard that displays the visualisation of the analysed data and the ranking based on the students’ participation metrics in the online forum, as explained in Section III. In Section IV, we discuss the relationship between the proposed learning metrics and the three badge levels, as well as the current limitations of the InPath system. Finally, in Section V, we conclude our study by summarising the features and implications of the InPath system for students and instructors.

The main contributions of this research are as follows:
- Propose learning metrics to measure and assess student participation in the online discussion forum.
- Apply a machine learning clustering algorithm to categorise different participation levels.
- Introduce a hierarchical badge system based on the level of participation.
- Use the feedback mechanism to improve the student’s performance.
- Introduce a real-time automated dashboard that visualises the participation metrics.

II. MATERIALS AND METHODS

A. DATASET

The dataset consists of five months of anonymized data obtained from the Ed Discussion Forum for an Information Technology course at an Australian research university in the 2020/21 academic year during the COVID-19 pandemic. In total, 423 undergraduate students and 16 tutors were registered users of the online forum. After obtaining ethical clearance from the university, we loaded this dataset into an online discussion forum for system testing.

B. LEARNING METRICS

Based on prior research [4], [9], the authors used the users’ contributions as metrics to analyse their participation in the online discussion forum. By taking these metrics into consideration, we proposed a set of learning metrics based on the objective of this research and the availability of forum data. The learning metrics proposed are as shown in Table 2, where the number of words, posts, replies, and posts for each category are considered the students’ contributions to the online forum. Meanwhile, the total number of days a student visited the forum and the total number of posts read by the student were used to measure the activity of the students participating in the online forum.

Existing research models reported in the literature still have room for improvement. They can be refined using a mechanism to assess the quality of user-generated content [4], [9]. Cheng and Vassileva stated that the quality of user contributions declined as the number of user contributions increased [9]. This could be due to the urge of individuals to increase their contribution scores without considering the quality of content, as quality was not taken into account in the motivation strategy proposed. Therefore, in this study, we decided to use the total number of likes received as a qualitative metric that determines the quality and usefulness of a student’s posts or replies.

C. DATA FLOW

In this research, we propose InPath, a real-time automated system built on open-source software that collects student activities on an online student discussion forum and generates instant feedback. The InPath system predicts students’ contributions to the online student discussion forum via machine learning clustering, generates dashboard visualisations that can be viewed as comparisons to other peers, and obtains auto-generated feedback indicating the gap between the student’s current performance (indicated by the badge level) and a higher badge level. Fig. 1 illustrates the standard data flow of the InPath system. To begin, the database associated with the online discussion forum stores all user information and activity; any updates to the users are passed through the Kafka Connector and read by Kafka Consumer Group 1; additionally, the application filters out non-student users and extracts only the seven learning metrics specified in the previous section. Following that, the Kafka Producer writes each processed record to the Kafka topic belonging to consumer group 2. The streaming data is then passed to the Scala-based Spark streaming application, which forecasts each student’s current badge level using the trained machine learning model. The results were then stored in a database. Finally, the learning analytics dashboard displays student performance in comparison.
D. MACHINE LEARNING APPROACH

The K-means method is a popular unsupervised clustering algorithm that effectively produces clusters in n-dimensional space [10]. The InPath system employs the same clustering algorithm as [11], which is a three-level K-means algorithm that performs the best in grouping students based on the learning metrics in their research. As demonstrated in the pseudocode in Fig. 2(a), the K-means algorithm takes the number of clusters and the number of data points as input and initialises K centroids randomly. The algorithm proceeds by repeatedly assigning each data point in D to the nearest cluster and recalculating the centroids’ positions until there are no further changes in the positions. Our machine learning approach consists of two major steps: (1) training the K-means clustering model and (2) prediction of data points. As shown in the pseudocode in Fig. 2(b), we utilised Apache Spark and implemented a Scala-based Spark streaming application that constructs a Kafka consumer to receive data from the database via a stream. Subsequently, the application initialises the K value to be 3 to obtain three ranks and uses the Apache Spark Machine Learning library to load a K-means model. The three-rank mechanism was chosen in our application as it adequately classifies the students based on their participation level on the forum, which can be described as low, moderate and high participation level. In comparison, the two-rank clustering model performed poorly since it resulted in many students being flagged as unengaged due to over-clustering of points into unnecessary extra clusters [10].

Next, the dataset was fitted to a model with three clusters based on the learning metrics we proposed. After transforming the students’ attributes from the Kafka stream into vectors, we performed predictions on the vectors one by one. The outcome is a tuple consisting of the student’s ID and predicted cluster ID, which we will use to categorise the students as either Hall of Fame, All Star, or Rookies badge. Table 3 shows the three badge levels and the characteristics of the students who will be classified into each level. The outcomes were then updated in the database.

E. RANKING SYSTEM

To understand and evaluate the performance of students in the online discussion forum under each K-means cluster, we proposed a ranking system on our learning analytics website. Let \( F = f_1, f_2, f_3, \ldots, f_7 \) be the set of contributions for each student, similar to the learning metrics we proposed. The formula to calculate the score of each student by taking the average of the scores of the seven selected features is presented in (1). As a result, the top 10 students with the highest scores will be displayed on the Learning Analytics website. This ranking system will act as a powerful reference point for all students and will motivate user participation, as stated by [12], and will help achieve the objective of creating a better web-based learning environment.

\[
Score = \frac{\sum_{i=1}^{7} f_i}{7}
\]
TABLE 3. Badge levels and their characteristics.

| Badge        | Characteristics                                      |
|--------------|------------------------------------------------------|
| Hall of Fame | High participation in the forum; High contribution in useful contents |
| All Star     | Medium engagement in the forum; Spent average time on forum; Contribute several useful contents |
| Rookies      | Low to no participation in the forum; Rarely post or reply; Contents are unlikely to be useful |

real-time feedback for students to understand their personal performance in terms of their participation level in the online discussion forum, realise the gap between the current badge level and the next upper badge level (Table 3), and perhaps become motivated to act and improve learning participation. First, Type 1 Feedback: A graphical representation of student-to-student comparisons on learning measures, average Hall of Fame performance, and the highest performance is provided by InPath. Second, Type 2 Feedback: InPath emphasises the difference between current personal statistics and a higher badge level. Student A, for example, with a Rookies badge, is 10 times behind the average in the All Star category in terms of likes received. As a result, student A realised that his posts/questions may be excessively extensive and unproductive in terms of post content for his peers.

III. RESULTS

This section presents the generated visualization of InPath and the results computed based on the methods applied in the implementation stage of our work. An open-sourced data visualisation tool was used to construct the interactive charts on our learning analytics website.

A. DASHBOARD

The dashboard is the main page of the learning analytics website. It displays a graphical examination of the performance of a student compared to others. The dashboard interface is depicted in Fig. 3. At the top of all the charts are statistics on the total number of students who have earned a badge level. The dashboard includes six charts that allow for comparisons between students and the data obtained based on the performance of other students, corresponding to the six learning metrics presented in this study. The second section of the dashboard delves into the first type of feedback mechanism discussed in Section 2.6. The radar chart shown in Fig. 4 compares the user with the highest statistics and the average statistics of the highest badge level (Hall of Fame).

B. INDIVIDUAL STATISTICS

The Individual Statistics Page, as shown in Fig. 5, is a page where students can view their forum statistics based on their respective participation levels in the forum. The metrics displayed on this page are the number of posts, replies, words posted, reads, days the student visited the forum, and likes received by the student. We also included a summary of feedback based on these metrics with reference to the average statistics calculated from the metrics of the badge, which is one level higher, as shown in Fig. 6. For students at the highest badge level (Hall of Fame), we will compare them with the highest value obtained for that metric.

C. A RANKING SYSTEM

As demonstrated in Section III (C), we present an equation to rank students in each cluster based on their level of participation in order to evaluate the powerful effect of peer pressure and encourage participation. Fig. 7 shows the results of the ranking system on the learning analytics website. This page is designed to function as a leaderboard, which allows tutors and students to discover the top 10 most contributing students on the system, along with their badge. For example, David Hayes had the highest engagement score among the students based on the calculation results. He/she is awarded the Hall of Fame badge and is rated first on the leaderboard.

IV. DISCUSSION

In this study, we present the application of the InPath system, a real-time analytics forum system, to rank and provide feedback on the performance of student online participation. It is an innovative and revealing way to motivate students to engage in self-reflection, improve student interaction in
online discussion forums, and allow instructors to track students’ learning behaviour in real time. As previously noted in Section II (A), the dataset from the five-month forum that we used as the experimental data for our InPath system corresponded to the course of the COVID-19 pandemic. There were 423 students enrolled in this course, and the three K-Mean clusters computed by InPath are presented; the centroids reflect the seven learning metrics, and the centroid results are displayed in Table 4.

In Fig. 8, it appears that students who earned a Hall of Fame write longer posts and receive more likes than their peers. This shows that students in the Hall of Fame were more likely to ask relevant questions and earn better forum interactions than others at a lower badge level. Figure 9 also shows the relationship between the total number of days spent on the online discussion forum and the total number of posts read. Students with higher badge levels, it goes without saying, are more likely to spend more time on online discussion forums and read more posts. In total, there were 423 students who earned the following badges by the end of the semester: 18 were inducted into the Hall of Fame, 75 were named All Star, and 330 were Rookies; the simplified ratio of the three badges is 6:25:110.

| Learning Metrics                          | Hall of Fame | All Star | Rookies |
|------------------------------------------|--------------|----------|---------|
| Number of words                          | 1151         | 94       | 29      |
| Total posts read                         | 898          | 639      | 105     |
| Total days spent on forum                | 51           | 44       | 11      |
| Number of posts                          | 8.00         | 1.00     | 0.37    |
| Average number of posts per category     | 1.50         | 1.50     | 0.52    |
| Number of replies                        | 10.00        | 1.00     | 0.33    |
| Number of words posted                   | 5.94         | 2.52     | 0.84    |

FIGURE 5. First part of individual statistics.

FIGURE 6. Second part of individual statistics.

FIGURE 7. Ranking system.

FIGURE 8. Diagram of three clusters against number of words and number of likes.

FIGURE 9. Diagram of three clusters against total post read and total days spent on forum.
Owing to time constraints, we were unable to collect and analyse data from a variety of different sample forum sizes. Although InPath results only reflect the conclusion of a single forum dataset, they provide critical insight into the real-time learning analytics system implemented in the online discussion forum environment. Some datasets from more comprehensive courses will be used in our future improvements, and we will integrate the InPath system into an ongoing course to track and analyse the influence of this system on students’ attitudes regarding the online learning environment. In addition to this work, different machine learning approaches, such as classification and quality-based methods, can be used as an improvement approach to improve the performance of our system.

V. CONCLUSION

With the advancement of technology, e-learning has grown in popularity. However, e-learning faces a barrier in allowing instructors to assess student performance and improve the e-learning environment. Therefore, we set a research question for this paper: Are we able to assess student participation on discussion forums in real time? As a result, this study proposes a learning analytics tool, InPath, to assess student participation in educational online discussion forums. This research was motivated by the need to understand how to analyse individuals’ performance in comparison to other students’ performance as a reference point, and how an analytics dashboard can assist students in learning their own participation level. As discussed in previous sections, we designed a set of learning metrics and a real-time Spark Streaming K-means system to monitor and rank students based on their participation levels in the discussion forum. As a result of this study, students are categorised into three ranks: Hall of Fame, All Star and Rookies. Furthermore, this paper presents the implementation of a web-based forum analytics website to illustrate the real-time results generated by the InPath system. Ranking mechanisms and feedback systems are also deployed on the website to encourage other learners to become more involved in the forum community and to enhance the quality of the e-learning environment. We believe that InPath will not only motivate students to participate in the online forum but will also support instructors in understanding students’ behaviours and improving their teaching methods.

REFERENCES

[1] Education: From School Closure to Recovery, UNESCO, Paris, France, 2022.
[2] K. Chai, V. Potdar, and E. Chang, “User contribution measurement model for web-based discussion forums,” in Proc. 3rd IEEE Int. Conf. Digit. Ecosyst. Technol., Jun. 2009, pp. 347–352, doi: 10.1109/DEST.2009.5276787.
[3] A. Aristovnik, D. Keršič, D. Veselj, N. Tomazević, and L. Umea, “Impacts of the COVID-19 pandemic on life of higher education students: A global perspective,” Sustainability, vol. 12, no. 20, p. 8438, Oct. 2020, doi: 10.3390/s12208438.
[4] E. Zarzycka, J. Krasodomska, A. Mazurczak-Maka, and M. Turek-Radwan, “Distance learning during the COVID-19 pandemic: Students’ communication and collaboration and the role of social media,” Cogenti Arts Humanties, vol. 8, no. 1, Jan. 2021, Art. no. 1953228, doi: 10.18080/3311983.2021.1953228.
[5] M. Adraoui, A. Retbi, M. K. Idrissi, and S. Bennani, “Social learning analytics to describe the learners’ interaction in online discussion forum in Moodle,” in Proc. 16th Int. Conf. Inf. Technol. Based Higher Educ. Training (I THEET), Jul. 2017, pp. 1–6, doi: 10.1109/I THEET.2017.8067817.
[6] J. P. Saru, “Rapid adaptation of a traditional introductory lecture course on catalysis into content for remote delivery online in response to global pandemic,” J. Chem. Educ., vol. 97, no. 9, pp. 2590–2597, Sep. 2020, doi: 10.1021/acs.jchemed.0c00786.
[7] S. L. Markel and E. E. Eci, “Technology and education online discussion forums: It’s in the response,” Online J. Distance Learn. Assist. Vol., Apr. 2011, pp. 511–517, doi: 10.1145/3448139.3448212.
[8] Y.-S. Tsai, R. F. Mello, J. Jovanović, and D. Gašević, “Student appreciation of data-driven feedback: A pilot study on OrTask,” in Proc. LAK2114th Int. Learn. Anal. Knowl. Conf., Apr. 2021, pp. 511–517, doi: 10.1145/3448139.3448212.
[9] R. Cheng and J. Vassileva, “User motivation and persuasion strategy for peer-to-peer communities,” in Proc. 35th Annu. Hawaii Int. Conf. Syst. Sci., Jan. 2005, p. 193, doi: 10.1109/HICSS.2005.653.
[10] S. A. Hudli, A. A. Hudli, and A. V. Hudli, “Identifying online opinion leaders using K-means clustering,” in Proc. 12th Int. Conf. Intell. Syst. Design Appl. (ISDA), Nov. 2012, pp. 416–419, doi: 10.1109/ISDA.2012.6416574.
[11] A. Moubayed, M. Injadt, A. Shami, and H. Latifiyaha, “Student engagement level in an e-learning environment: Clustering using K-means,” Amer. J. Distance Educ., vol. 34, no. 2, pp. 137–156, Apr. 2020, doi: 10.1080/08923647.2020.1696140.
[12] A. Wise, Y. Zhao, and S. Hausknecht, “Learning analytics for online discussions: Embedded and extracted approaches,” J. Learn. Anal., vol. 1, no. 2, pp. 48–71, Aug. 2014, doi: 10.18608/jla.2014.12.4.
[13] J. Hattie and H. Timperley, “The power of feedback,” Rev. Educ. Res., vol. 77, no. 1, pp. 81–112, 2007, doi: 10.3102/0034654307304948.
[14] L.-A. Lim, S. Gentili, A. Pardo, V. Kovanoivić, A. Whiellock-Wainwright, D. Gašević, and S. Dawson, “What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course,” Learn. Instruct., vol. 72, Apr. 2021, Art. no. 101202, doi: 10.1016/j.learninstruc.2019.04.003.
[15] A. Pardo, J. Jovanović, S. Dawson, D. Gašević, and N. Miriria, “Using learning analytics to scale the provision of personalised feedback,” Brit. J. Educ. Technol., vol. 50, no. 1, pp. 128–138, Jan. 2019, doi: 10.1111/bjet.12592.
[16] J. B. Pena-Shaff and C. Nicholls, “Analyzing student interactions and meaning construction in computer bulletin board discussions,” Comput. Educ., vol. 42, no. 3, pp. 243–265, Apr. 2004, doi: 10.1016/j.compedu.2003.08.003.
[17] C. Romero, M.-I. Lopez, J.-M. Luna, and S. Ventura, “Predicting students’ final performance from participation in on-line discussion forums,” Comput. Educ., vol. 68, pp. 458–472, Oct. 2013, doi: 10.1016/j.compedu.2013.06.009.
[18] W. Chin and A. Abrizah, “An analysis of an assessment model for participation in online forums,” Comput. Sci. Inf. Syst., vol. 9, no. 1, pp. 121–140, 2011, doi: 10.2298/CSIS100113036C.
[19] T. Owatari, A. Shimada, T. Minematsu, M. Hori, and R.-I. Taniguchi, “Real-time learning analytics dashboard for students in online classes,” in Proc. IEEE Int. Conf. Teaching, Assessment, Learn. Eng. (TALE), Dec. 2020, pp. 523–529, doi: 10.1109/TALE48869.2020.9368340.

ADELINE CHEW YAO YI is currently pursuing the bachelor’s degree in computer science with the School of Information Technology, Monash University, Malaysia. She is mainly interested in software development, machine learning, and artificial intelligence. She is also participating in an exchange program at The University of British Columbia, Canada, from September 2022 to December 2022, majoring in computer science.
TEY KAI YING is currently pursuing the degree in computer science with the School of Information Technology, Monash University, Malaysia. She is also having an enjoyable learning experience at The University of British Columbia, Canada, as an Exchange Student, majoring in computer science. She has a strong interest in software development and exploring the world.

SIANG JO YEE is currently pursuing the bachelor’s degree in computer science with the School of Information Technology, Monash University Malaysia. She is also undergoing an exchange program with the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, specializing in computer science.

TING TIN TIN received the B.S. and Ph.D. degrees in computer sciences from the University of Science, Penang, Malaysia, in 2003 and 2012, respectively.

In 2005, she joined Gemplus Technologies Asia Pte. Ltd., Singapore, as a Telecommunication Software Engineer. After her Ph.D. graduation, she started her career as an educator and a researcher. She is currently a Research Track Educator mainly supervising postgraduate’s projects in INTI International University, Malaysia. At the same time, she serves as a Freelance Lecturer at Monash University, the Tunku Abdul Rahman University of Management and Technology, and the Methodist College Kuala Lumpur. She has received her professional certification in project management from PMI and data analytics from SAS. She has more than 12 years of lecturing, supervising projects, and research experience. Her research interests include big data analytics, information systems engineering, educational data mining, psycho-academic research, and software engineering.

WEE MEE CHIN received the B.Sc. degree in computer science and mathematics from Campbell University, Buies Creek, NC, USA, in 1996, the M.S. degree in vision, visualization, and virtual environment from the University of Leeds, U.K., and the Ph.D. degree in information systems from the University of Malaya, Malaysia.

She was with the University of Malaya, from 2010 to 2016, and Tunku Abdul Rahman College, from 1997 to 2003. She is currently a Senior Lecturer at the School of Information Technology, Monash University Malaysia. Her research interests include education/learning technology, joint emotional response and education technology, information visualization, and ICT in healthcare.

****