Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 17 No. 22 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijim.v17i22.45303

PAPER

Student Behavior Simulation in English Online Education Based on Reinforcement Learning

Wenjing Wang(✉)

Zhejiang Conservatory of Music, Hangzhou, China

wangwenjing20160225@gmail.com

ABSTRACT

In class, every student's action is not the same. In this era, most courses are taken online; tracking and identifying students' behavior is a significant challenge, especially in language classes (English). In this study, Student Behaviors' Simulation-Based on Reinforcement Learning Framework (SBS–BRLF) has been proposed to track and identify students' online class behavior. The simulation model is generated with various trained sets of behavior that are categorized as positive and negative with Reinforcement Learning (RL). Reinforcement learning (RL) is a field of machine learning dealing with how intelligent agents act in an environment for cumulative rewards. With a web camera and microphone, the students are tracked in the simulation model, and collected data is executed with RL's aid. If the action is assessed as good, the pupil is praised, or given a warning three times, and then, if repeated, suspended for a day. Hence, the pupil is monitored easily without complications. The research and comparative analysis of the proposed and the current framework have proved that SBS-BRLF works efficiently and accurately with the behavioral rate of 93.2%, the performance rate of 96%, supervision rate of 92%, reliability rate of 89.7 % for students, and a higher action and reward acceptance rate of 89.9 %.

KEYWORDS

reinforcement, simulation model, behavior, rewards and framework

1 INTRODUCTION TO STUDENT SUPERVISION AND LEARNING IN ONLINE EDUCATION

The current and most famous method of distance education nowadays is online communication. For the last decade, the movement has had an extensive effect on post-secondary schooling. Online Education has changed the way people communicate, shop, interact socially, do business, and care about education and understanding [1][2]. Online training transforms conventional schools' faces and makes learning more available than before, far better than just a recent trend on distance learning education [3]. Online Education is for students to access their individual computers

Wang, W. (2023). Student Behavior Simulation in English Online Education Based on Reinforcement Learning. *International Journal of Interactive Mobile Technologies (iJIM)*, 17(22), pp. 136–151. https://doi.org/10.3991/ijim.v17i22.45303

Article submitted 2023-08-26. Revision uploaded 2023-10-05. Final acceptance 2023-10-10.

on the network. In recent years, online graduation ceremonies and classes have been common for many non-traditional participants, including all those who want to pursue work full time or raise a family [4]. The online education platform of the chosen institution typically offers online grades and classes, some of which have been carried out using emerging technology.

Online Education is digital training and is based on the Internet, talking with teachers/students, and sharing teaching material [5–7]. This basic description is taken from the almost infinite variety of teaching methods and studying outside schools and school campuses. Students can transform online Education into a school anywhere with power and internet connectivity [8]. It can provide audio, video, email, graphics, interactive training settings, and live instructor chats [9]. It is a more diverse and versatile learning experience than a typical classroom.

The online curriculum in English is a modular teaching method covering all aspects of learning across the Internet [10]. Online Education enables teachers to meet students who would not be ready to interact in a typical classroom and encourages students who could function through their own time and speed [11]. In most domains, the sum of distance education and online certificates is rising exponentially. A growing number of schools and universities deliver learning activities based on English Online Education [12]. Students who graduate from the online method must be cautious to guarantee that they do their courses via a reputed and accredited organization.

Through the growing usage of Internet and computing infrastructure, students may access any material for English learning traditionally accessible in a classroom setting at any period and any location [13]. Studies have found that children understand in a virtual class almost as well as in typical classrooms. Numerous students choose online courses for English because their aggressive schedules provide consistency. With the abundance of Education and understanding, students in the modern world have to become learners and more concentrated and self-determined.

Additional online education opportunities for English would be accessible with upgraded applications, equipment, and Internet access [14]. The future of Online Education would continue to expand with student enrolment exponentially faster than workplaces are possible. The students are becoming far more technologically advanced, and the students seek an education that fits their needs. Internet degree courses, which are more traditional activities, would be approved. Learners' personalities and awareness levels grow over time in online learning situations. In light of real-time adaptation, the current approaches focused on stable metrics are not successful. Indeed, online learning platforms should be versatile in recommending quality information based on actual learners' background. Moreover, the latest contents are added progressively in online learning frameworks, which tend to be new, without any information about previous experiences and online learning components in English [15]. It may be costly to access details, and customer satisfaction may be reduced.

Reinforcement learning (RL) is a well-known approach to these problems. It provides an efficient decision-making algorithm for scalable and dynamic settings [16]. Unlike other methods of advising, RL does not use superiors but prefers compensation signals. The investigator produces the perfect series choice in optimizing the accumulating charge, depending on the advantages of checking the error. Besides, since RL is focused on signal input, the worker may benefit from the job if they have no environmental knowledge. This allows the agent to investigate when the world has a new element. At the same time, the operator still makes the right choices by logic and studying from previous experience or present understanding.

It analyses the world to collect knowledge and to take advantage of information to determine or forecast. In addition to finding maximal results in a particular state, the operator controls the setting. RL teaching strategies are built for a complex and persistent agent acting by progressively proposing appropriate learner routes.

SBS-BRLF has been proposed for improving student monitoring and assessment in online classes. RL is a robust model that helps individual systems determine agents act and their corresponding online English learning rewards. It is the perfect way to understand actions in chosen activities.

The remaining sections of the article are organized as follows: Section 2 comprises the literature survey about various behavior analysis on Online Education and their student assessment issues. Section 3 described the infrastructure details and elaborated the working process of the proposed SBS-BRLF. Section 4 explains the results of the conducted experiments and the corresponding discussions compared with the traditional techniques. Section 5 concludes the research process with future perspectives.

2 BACKGROUND STUDY ON STUDENT SUPERVISION AND LEARNING IN ONLINE EDUCATION

This section discusses several works that various researchers have carried out; Gaurav Kumar et al. [17] developed an optimal monitoring program (OPM), which is searched to display instances during learning provided a training sample of sentence-level functionality such as interference. The education structure allows unlimited numbers of representatives to emerge and generalize data measurement, sorting, and completion systems. Refinement learning is used simultaneously with the system in a single training program instead of depending on previous experience to create a framework.

Jing Zhang et al. [18] proposed Graded Reinforcement Learning for online education supervision (GRLOES). The explosion of extensive open online courses demands an efficient form of customized course suggestions. The new attention-based recommendation models may discern the influence of various background classes when proposing specific target classes. However, while a consumer is involved in several different subjects, the focus function would fail as other components dilute the responding fields' impact. Experimental findings suggest that the GRLOES model exceeds the styles of the new recommendations substantially.

William Villegas-Ch et al. [19] introduced the Assessment of Digital Education (ADE) Model. ADE suggests the introduction into learning management frameworks of approaches such as artificial intelligence and data processing. The goal is illustrated in a new standard that seeks strong educational models in which such activities take place—Internet mode, accompanied by technology that allows learners to be directed by interactive advisors. ADE involves Education that has taken bets on the use of communication technology to meet students. Learning analytics, which are becoming perfect platforms to handle resources and build programs, is described.

Waleed Al-Rahmi et al. [20] discussed Immense Online Assessment Courses (IOAC). All of the criteria used to analyze and identify the IOAC were the journal, country of birth, scholars, release results, conceptual method, templates, methods, and sample respondents. These details add to the documents sought by readers involved in various facts of IOAC literature in Education. Five factors in optimizing IOAC were analyzed concerning the use, contact, commitment, motives, and fulfillment. Students' academic success can be affected by the IOAC, which provides materials and facilitates knowledge exchange.

Muhammad Adnan et al. [21] narrated OLA Model. OLA model looks at the Pakistani improvement of students' reactions to mandatory multimedia and distance education programs in coronavirus. Bachelor and postgraduate students were asked to find their views on online learning in Pakistan. The report indicates that online learning cannot achieve the required outcomes in developing nations such as Pakistan. Due to technological and financial problems, a substantial number of students seem unable to browse the website.

Osama Isaac et al. [22] discussed the teach -learn-knowledge (TLK) model. Survey data was collected in nine universities from 448 students in Yemen. Structural equation simulation using PLS 3.0 has been assessed centrally. These findings contained six main results. Compatibility significantly affects user satisfaction and practical use; functionality facilitates relations between overall efficiency and practical help; real service and user content substantially affect the overall results, customer satisfaction, and user functionality.

Jingliang Duan et al. [23] developed a hierarchical reinforcement learning framework. It provides a hierarchical approach for improving self-driving car decision-making that doesn't focus on a massive volume of branded riding data. This approach takes high-level into account, both the transverse and vertical direction of a collection of movements and low-level action regulation.

Saman Rizvi et al. [24] proposed a hierarchical decision tree-based approach (HDT). HDT aims to analyze the various effects on academic performance of demographic factors in the online education sense. The complex impact of six demographic properties on online education results were studied and compared using a survey. Online students from four different disciplines at Open University are considered and observed. Inequality in the community and prior schooling were good predictors of overall academic achievement. Nevertheless, this effect evolved at a grain stage, both as the course advanced and between classes.

Based on the survey, SBS-BRLF strengthens student engagement and promotes monitoring and assessing student behaviors during learning in online courses. Reinforcement learning is a robust model that allows autonomous structures to make choices in Online English Education. Reinforced learning enables effective behavior study in activities that need to be chosen.

3 STUDENT BEHAVIORS' SIMULATION-BASED ON REINFORCEMENT LEARNING FRAMEWORK (SBS–BRLF)

This section elaborates some structural information concerning student behavioral analysis using Student Behaviors' Simulation-Based on Reinforcement Learning Framework (SBS–BRLF), which has been provided in this segment. A brief reference to a standard online education model based on the continuous supervision and the action taken with warnings and punishments using web camera and mic in SBS–BRLF models are given in the following subsections as shown in the Figure 1.

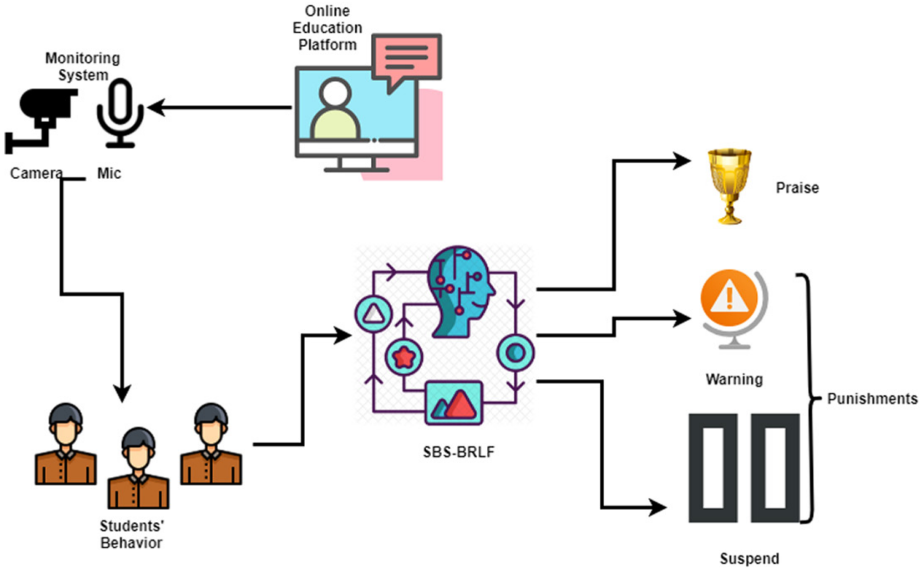

Fig. 1. SBS-BRLF model

3.1 Student behavior supervision analysis

A student behavior supervision model using reinforced learning focused on students' behaviors required in online education and examination architecture. A reinforcement learning network encodes the entire source term $a = a_1, a_2, \ldots, a_{T_a}$ in a vectors chain, and then a neural decoder uses these vectors to produce the desired supervision behavior $b = b_1, b_2, \ldots, b_{T_b}$.

The supervision section repetitively translates the phrase and the process mainly based on the procedural structure's attention sequence. Likewise, the learning has been done in the same way in the reverse condition choosing another attention data sequence, elaborately shown in Figure 2.

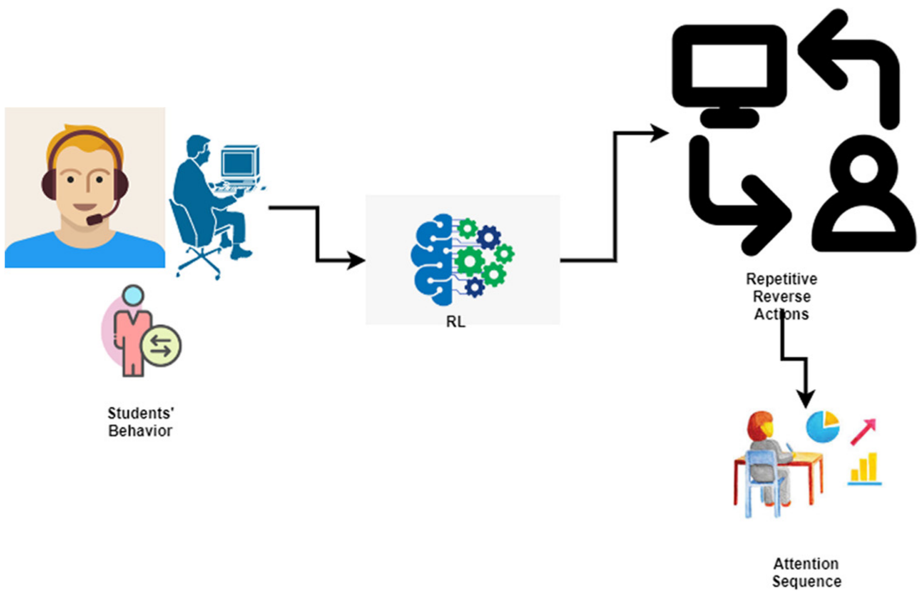

Fig. 2. Student behavior supervision

The behavior supervision usually is used as a two-way repetitive network utilizing forward and reverse repetition for the observation of the source function in the front one to one communication, which uses the Equation (1)

$$s_j = \overrightarrow{s_j} : \overleftarrow{s_j} \tag{1}$$

Where, the source data a for the data from left to right, computing a forward state $\vec{s} = \left[\overrightarrow{s_1}, \overrightarrow{s_2}, \ldots, \overrightarrow{s_{T_a}}\right]$. The backward source b states of the initial sentence $\overleftarrow{s} = \left[\overleftarrow{s_1}, \overleftarrow{s_2}\overleftarrow{s_{T_a}}\right]$, the concatenated to shape annotations. The supervision is essentially a repeated and conditional analyzing pattern that model the desired behavior by splitting it into every model acceptance rate generation as in Equation (2),

$$logp(b \backslash a) = \sum_{t=1}^{T} logp(b_t \backslash b_{<t}, a) \tag{2}$$

In particular, after the partial sequence $b_{<t} = b_1, b_2, \ldots, b_{T_b}$ is generated, the interacted data generates the b_t term. The likelihood $logp(b \backslash a)$ of b_t is determined with the new state s_t of $b_{<t}$, a.

3.2 Data collection analysis

The normal-linear structure describes data collection, analysis models. Throughout the case of a source term a, the data structure looks for optimal collection b dependent on various functions as in Equation (3), (4).

$$\hat{b} = argmaxp(b \backslash a) \tag{3}$$

$$\hat{b} = argmax\left\{\sum_{t=1}^{T} \beta_t s_t(b, a)\right\} \tag{4}$$

Where the data collection over online English teaching functions $s_t(b,a)$ and the corresponding weights of recurring actions are the β_t. Inverted action likelihood of words, inverted behavior weighting, the possibility of a direct statement, the language model, the gap model, the term penalty with the features are decoded.

The most common feature recognition of behavior analysis based on data collection analysis is sequential in nature. The English language network architecture for data collection analysis is as shown in Figure 3. It may be seen as external functions that typically improved. Each data collection in the source block has been chosen in a recurrent learning sequence with sufficient reinforcement conditions. The final suggested data content is obtained in the resultant active recurring block by the reverse state.

Pattern tuning is aimed at finding optimal weights for the linear model. Weights may be modified. The data collection analysis selects a goal to extend the partial $b_{<t} = b_1, b_2, \ldots, b_{T_b}$, for an exposed source section during data collection. Once all source fragments are converted, the localization cycle stops.

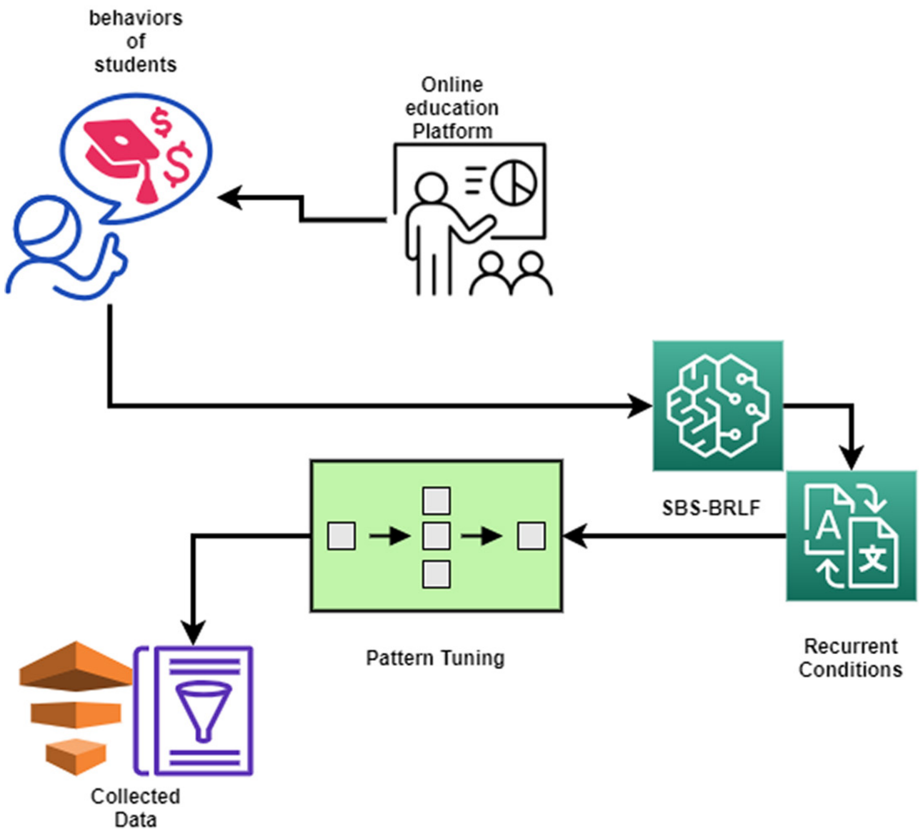

Fig. 3. Data collection analysis architecture

3.3 Behavior analysis model

This segment explains our structure, which involves the knowledge and behavior analysis process, and two concept variants developed to incorporate the system. The term prediction mechanism is the most distinct aspect from the traditional to the proposed one. The prediction method model suggested it predict terms focused not only on the traditional term estimation but also on supervision smart English learning terms. Throughout the following chapters, the online instruction on the individual steps is first defined. The inclusion in the supervision and the data collection analysis architecture of the behavior analysis recommendations will now be identified.

Behavior over supervision suggestions. For this analysis, a recent time-by-time simplification of the general concept-based model is followed to render English teaching suggestions consistent with the action and reward formation. The model uses the supervision details for effective suggestions, praising and warning or suspending at any stage. In the light of the terms $b_{<t} = b_1, b_2, \ldots, b_{T_b}$ previously developed in an analysis method, it extends the $b_{<t}$ partial action and produces the following suggestions based on the supervision rate values.

$$int_{rate}(b_t,\ a) = \sum_{t=1}^{T} \beta_t s_t(b_t, a_t) \tag{5}$$

Where in Equation (5) of int_{rate}, b_t is an action cum reward generated and a_t is the correct source time. The functions are $s_t(b_t, a_t)$, and β_t is the equivalent of the weight.

Ideally, the model is used to direct term previews for the highest ratings. According to the different communication processes between the two versions, however, there are two resultant reward suggestions.

Since SBS–BRLF produces goal rewards precisely before collected action coverage details, the model cannot document which terms to be protected. When the sentence symbol end is created, the conversion cycle is complete. In comparison, a coverage function is directly used by the supervision behavior to show whether or not a source action and reward is interpreted. When all source terms are under the action, the behavior analysis cycle is complete.

There are no coverage details since supervision is unable to acquire clear information for source coverage from decoding, which results in incorrect suggestions being created. A coverage vector $V = v_1, v_2,..., v_T$ for decoder has been added to resolve the problem. Actually, $V_T = 1$ applies to the conversion of the source xi from the current layout and the decoder from the source term $V_T = 1$.

When the term is included in current term guidelines, the associated vector element should be changed to a zero vector during each decoding process step. It first uses the explicit assignment to find the source that aligns with the reward suggestion, then updating the vector by setting the related item to one.

Differences in actions during assessment order are between language types, and the reordering model, therefore, plays a decisive role in the current approaches to supervision. The supervision model is usually used for calculating the reordering score using RL action and reward knowledge. For instance, it calculates a remote rearrangement score accordingly as in Equation (6),

$$d(upb_t) = -\,|\,upb_t - upb_{t-1} - 1\,| \tag{6}$$

When upb_t shows the origin location, which is matched with the goal b_t.

The supervision and associated behavior assessment use weight to emphasize the similarity between the goal term and the source phrase. The importance of emphasis can be viewed in strategies as term alignments. It splits the focus into the different source, far from standard alignment. Preferably a single source expression uses generalized focus probabilities over the whole source term to provide alignment information. Therefore, the modern rearrangement model cannot operate with the generalized probability of focus provided by the model for treatment. To solve the problem in Equation (6), Equation (7) is used to measure the reordering scoring using normalized probabilities for the entire paragraph. Like the approximate likelihood of bi-directional reward and action, the adjusted reorganization score is determined as follows:

$$d(b_t) = -\sum_{j=1}^{J} \mu_{i-1,j} \left|upb_t - J - 1\right| \tag{7}$$

Where, $\mu_{i-1,j}$ are the focus pattern alignment volumes. Likewise, the linguistic reordering model should work. Considering the second drawback, as stated, the suggestion filters out stop words to preserve the information word recommendations that convey the meaning of the behaviors.

Integration of assessment and rewards in English online education. In this analysis, the possibilities of action guidance and prediction were measured and merged across the two definitive behavioral versions, a correlated mechanism and a resulting system. The proof presented the calculation and variation of probabilities in the following section.

Figure 3 demonstrates the architecture of the correlation mechanism. The Figure shows that it is observed by comparing with the source, each action w1, w2, w3, to the correlation mechanism procedure followed by the behavior-based reward and improving it with the supervision-based suggestion process. The knowledge correlation mechanism produces the following expression (8). The standard mechanism uses the supervision and condition present, the target behavior previously created, and the current background vector to predict characters as in Equation (8):

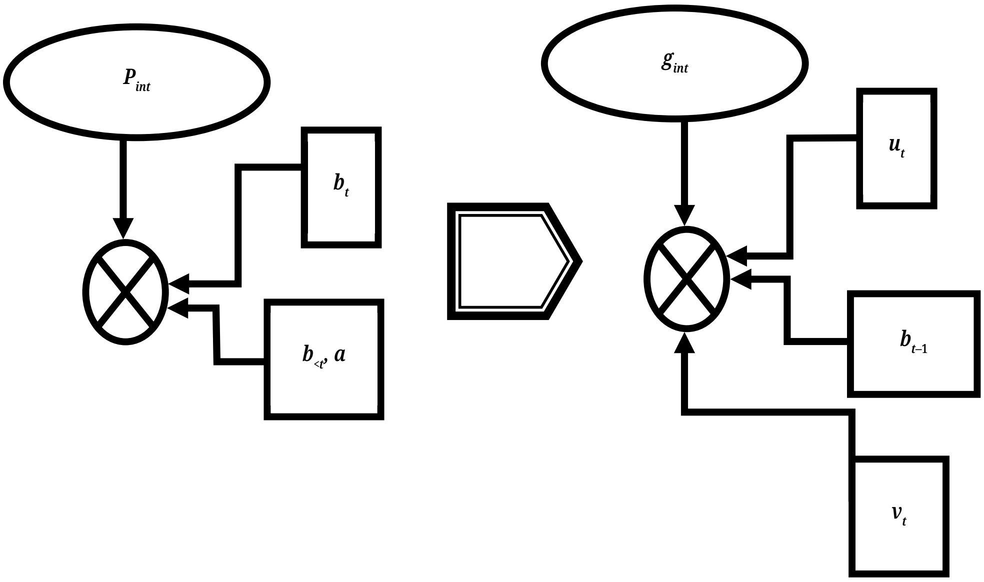

Fig. 4. Correlation mechanism of SBS–BRLF

$$P_{int}(b_t \backslash b_{<t}, a) = g_{int}(u_t, b_{t-1}, v_t) \tag{8}$$

In Figure 4 and equation (8) correlation mechanism of the SBS-BRLF model has been illustrated, where g_{int} is the non-linear, status correlation function u_t, previously created b_{t-1} and the current background vector v_t is computed as a hidden status b_{t-1}. The RLSBSM classification method predicts the probabilities for behavioral terms that the auxiliary model produces. The current model uses the descending framework. At stage t of the correlation mechanism, the behavioral suggestion vocabulary beh_{int} estimate probabilities of the oriented classifier in (9):

$$P_{int}(b_t \backslash b_{<t}, a) beh_{int}(b_t \backslash b_{<t}, a) \tag{9}$$

Where, b_t is the recommendation of behaviors and $beh_{int}(b_t \backslash b_{<t}, a)$ and the resulting system function defined as follows in (10):

$$resulting_{int}(b_t \backslash b_{<t}, a) = f_{int}(u_t, b_{t-1}, b_t, v_t) \tag{10}$$

Where, f_{int} is either an attribute or a non-linear structure activation structure. Eventually, to measure the likelihood after estimation for the standard behavior v_t, it has been added the correlation function. In particular, to update the prediction probability and to reduce the error rate, a linear interpolation is carried out between the two probabilities of generation and the recommendations of the model term in (11):

$$P(b_t \backslash b_{<t}, a) = (1-\delta t)P_{int}(b_t \backslash b_{<t}, a) + (\delta t)\, P_{int}(b_t \backslash b_{<t}, a) \tag{11}$$

The behavior for $P_{int}(b_t \backslash b_{<t}, a)$ is global, which is smaller than the $P_{int}(b_t \backslash b_{<t}, a)$ a shared of δt as for the behavior term suggestion in stage $P(b_t \backslash b_{<t}, a)$ is as described.

By matching probabilities, the proposed model updates the behavior analysis choosing likelihood using a resulting system mechanism. The term awareness decoder produces the following expression (12):

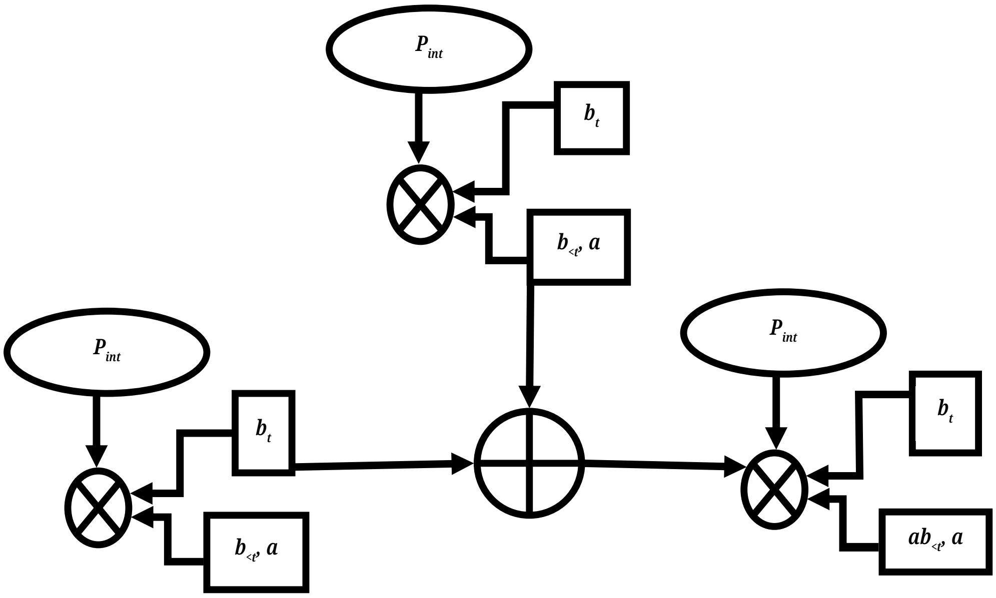

Fig. 5. Resulting system of SBS–BRLF

Figure 5 displays the resulting system in online education behavior analysis utilizing the framework for resulting comparison. In this behavior analysis method, the same process of source comparison is followed by applying the suggestion algorithm to the sequence w1, w2, w3, …, and the layer of the supervision and data collection model is directly compared with the sequence obtain the proper teaching behavior.

$$P(b_t \backslash b_{<t}, a) = P_{int}(b_t \backslash b_{<t}, a) + P_{int}(b_t \backslash b_{<t}, a) \tag{12}$$

$$P_{int}(b_t \backslash b_{<t},\ a) = \begin{cases} \dfrac{1}{R} e^{beh_{int}(b_t|b_{<t},\, a)} & b_t \epsilon\ S^{int} \\ 0 \quad otherwise \end{cases} \tag{13}$$

$$P_{int}(b_t \backslash b_{<t}, a) = \begin{cases} \dfrac{1}{R} e^{beh_{int}(b_t|b_{<t},\, a)} & b_t \epsilon\ S^{int} \\ 0 \quad otherwise \end{cases} \tag{14}$$

The standardization behavior analysis term is determined for the correlation and resulting models using the above Equation as a resultant condition. The teaching behavior guidelines are given by interacting with SBS–BRLF generations through the vector feature due to the popular standardization.

Despite signs, the SBS–BRLF concept allows suggestions. However, the reinforcement also maps the rating terms into a symbol with the behavior suggestions. To resolve the problem, the symbol should be substituted by an appropriate suggestion. The proposed model switches the behavior suggestion to replace the symbol, when created, with the highest value. The parameters exchanged with the standard correlation and resulting model centered on focus would be initialized by utilizing the advanced reference models. Random initialization of the latest parameters of the proposed model. Ultimately, the cost-efficient function is reduced by all the parameters of the model suggested as follows.

4 RESULTS AND DISCUSSION

SBS–BRLF has been validated based on student supervision, data collection, and behavior. The students from various schools are randomly selected, and an online simulation is taken regarding student behavior in English Online Education. SBS–BRLF approach is evaluated based on qualitative bandits improving learners' ability to include appropriate Education for students in online learning systems. The method proposed would provide complex and ongoing suggestions depending on user learning actions, from which the corresponding contextual knowledge is determined over time. SBS–BRLF solution benefits from situational RL because it is quick and easy to apply compared with the entire RL issue in a high degree of ambiguity. Besides, agents can analyze complexity more efficiently by using qualitative knowledge. The findings of the tests suggest that SBS–BRLF has the highest performance rate among students in Online Education. In versatile conditions, the proposed approach can enhance optimum decision-making. When considering the averaging view count, it exceeded well-known methods and solved all customized online services. The performance and supervision rate of students is shown in Figure 5a and b.

4.1 Performance rate and supervision rate

The performance rate of students is an integral part of all learning experiences, as shown in the Figure 6a. This exchange usually arises in a classroom environment as students respond to each other's thoughts, ask questions, and identify connections through repeated communication. Teachers should promote contact with students in an online environment, although they can involve formal and informal possibilities like the course. Developing a high degree of student engagement on the Internet is critical that accreditation institutions need proof of the overall issue of customized Online Education for situational supervision among the students worked better than measuring the experimental findings' approaches. The supervision among student's behavior simulation in English Online Education is evaluated based on reinforcement learning. The supervision among students is shown in Figure 6b. Supervision between students is necessary for creating a culture that promotes effective and satisfactory training in an online environment and encourages students to improve reasoning and problem-solving capabilities.

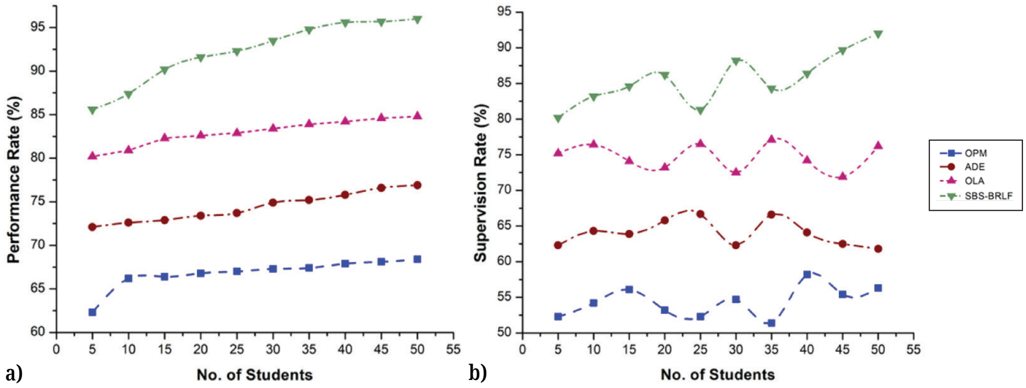

Fig. 6. (a). Performance rate, (b). Supervision rate

4.2 Behavior rate and reliability rate

Compared to the other online education evaluation scales outlined, the student behavior rate makes a particular attempt to build supervision among students and educators. In reality, the quality of behavior among students is produced over multiple iterations, during which the authors concentrated on teacher observation material. Observing students' actions to avoid diversion is necessary to establish a healthy atmosphere in online education classrooms. Even so, solely online-based supervision of students promotes respect for society's values without delay. Teachers often express the student behavior rate in Online Education by depending on the standards about students' actions. Suppose teachers are using a particular method to evaluate the students, in which they are predominantly interested in fixing the issues. In that case, the behavior rate of students in Online English Education must be identified. The behavior rate of students is shown in the Figure 7a.

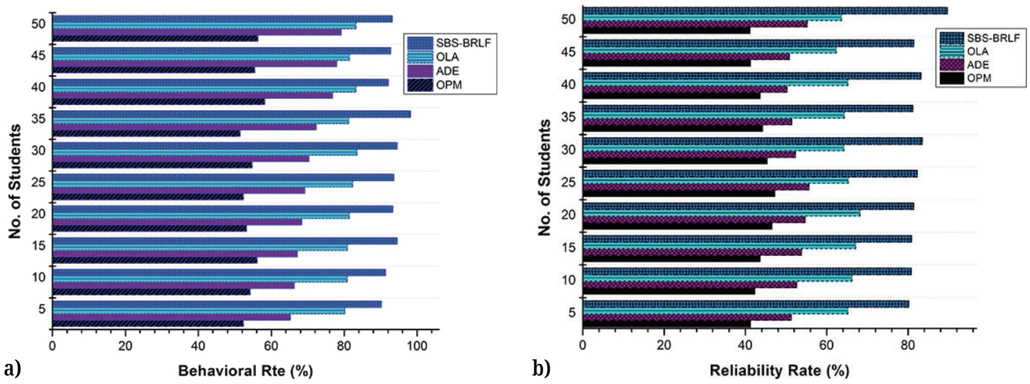

Fig. 7. (a). Behavior rate, (b). Reliability rate

The suggested implementation of the SBS–BRLF coverage mechanism in students' reliability is to resolve the over and under action and reward online education issues. The decision-making model supports the target reward generation to learn students' reliability in Online Education, as shown in Figure 7b. Reinforcement learning is a robust framework that helps self-employed programs to determine in English online. Strengthened learning facilitates efficient behavioral learning in offering recommendations. All bench marking strategies have the same efficiency in the first place in the decision-making system. The reliability of the students is focused on the behavior simulation in English online education-based learning algorithm. However, the SBS–BRLF approach must use historical knowledge to analyze the student state and improve it over time.

4.3 Action and reward acceptability rate

Student participation occurs as students indulge socially in schooling and aspire to figure out what it provides, not to receive the standardized metrics of achievement but to get to know and integrate content in their lives. The rate of commitment to Online Education quantifies the degree of commitment and student behavior in Online Education based on reinforcement learning. The engagement level with supporters created by user-created material is supported by all students and educators in online learning. The acceptance rate is commonly used to assess the success of brand promotions in English Online Education. The acceptance rates have included subset indicators such as share metrics that emphasize Education's influence in all ways of online teaching. The acceptance rate of students is illustrated in Figure 8. Student participation requires the student's desire, ability, and willingness to participate in the learning process, fostering a deeper understanding of sustainable comprehension of problems and barriers.

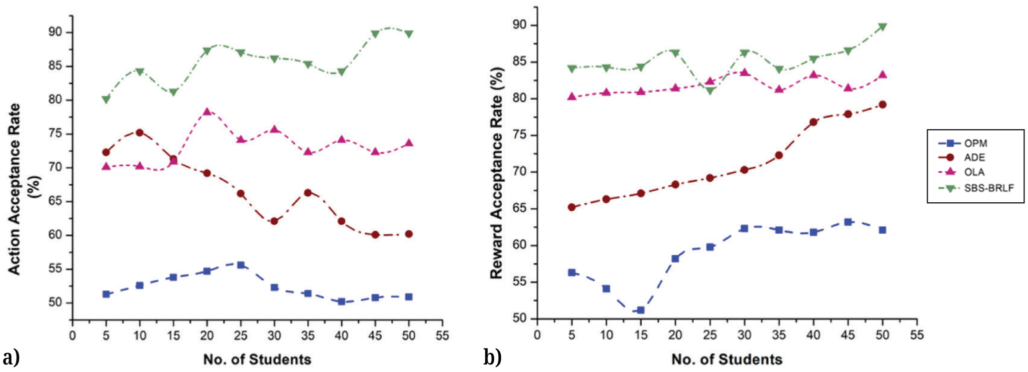

Fig. 8. (a). Action acceptance rate, (b). Reward acceptance rate

Thus, the proposed SBS–BRLF achieves the highest behavior rate and proper action and reward rate compared to other existing methods. It is the best and practical, a suitable method of assessing students' performance in online teaching and evaluation.

4.4 Mean square error (MSE) rate

The various properties and structure of the online education platform have been validated using SBS-BRLF with RL tools. Data collected are with minimum MSE from diverse modalities. It is clear how descriptive consistency can be achieved, and it can directly impact machine learning algorithms in online education performance. A linear association is found between the number of depending actions and rewards. In Figure 9, it has been shown that the method achieves a very low MSE as compared to the existing methods of online education behavior analysis by adequate supervision.

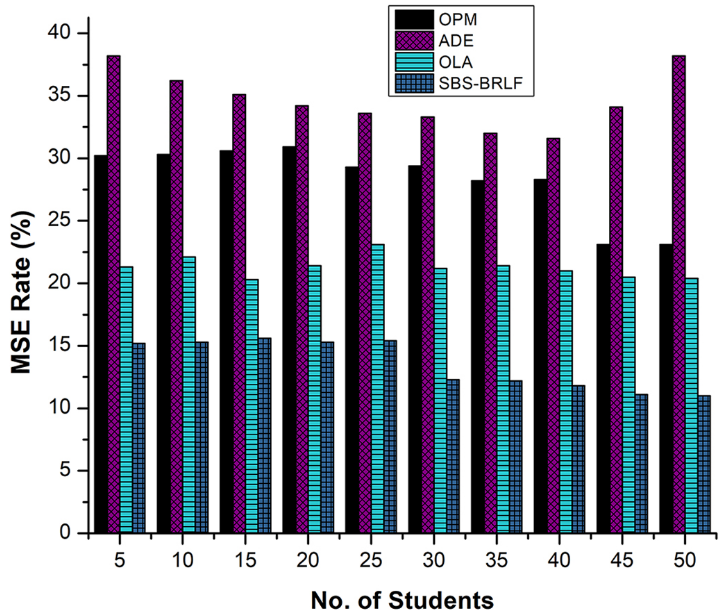

Fig. 9. Mean square error rate

Thus, the proposed SBS-BRLF achieves a good impact on the various techniques, and it broadly encourages smart online education processes in an efficient manner. It paves the way for the effective implementation of the model in real-time too.

5 CONCLUSION

The SBS–BRLF is presented to enhance student behavior supervision and enhance learning in Online Education; the reinforcement learning paradigm is based on student compliance simulation. Reinforcement learning is thus a strong paradigm that helps self-employed programs to determine practical actions and appropriate rewards in English Online Education. Increased learning provides the perfect way to learn actions in activities to be chosen. An RL learns as a previous practice of creating decision-making rules for suggestions. SBS–BRLF has adopted a plan to increase supervision in English Online Education. SBS–BRLF can efficiently guide students in English Education to determine the behavioral, action, and acceptance rates. Experimentally, SBS–BRLF has demonstrated positive changes in the online platform of Education with a behavioral rate of 93.2%, performance rate of 96%, supervision rate of 92%, reliability rate of 89.7 % for students, and a higher action and reward acceptance rate of 89.9 % each in English online teaching, learning, and assessment.

6 REFERENCES

[1] D. Shawky and A. Badawi, "Towards a personalized learning experience using reinforcement learning," in *Machine Learning Paradigms: Theory and Application*, Springer, Cham, 2019, pp. 169–187. https://doi.org/10.1007/978-3-030-02357-7_8

[2] I. Mustapha, M. Ali, N. Khan, and H. Sikandar, "The impact of industry 4.0 on innovative organisations, a thematic review using the PRISMA statement 2020," *International Journal of Interactive Mobile Technologies*, vol. 17, no. 9, pp. 88–105, 2023. https://doi.org/10.3991/ijim.v17i09.39465

[3] N. R. Clayton and H. Abbass, "Machine teaching in hierarchical genetic reinforcement learning: Curriculum design of reward functions for swarm shepherding," in *2019 IEEE Congress on Evolutionary Computation (CEC), IEEE*, 2019, pp. 1259–1266. https://doi.org/10.1109/CEC.2019.8790157

[4] G. Zhou, H. Azizsoltani, M. S. Ausin, T. Barnes, and M. Chi, "Hierarchical reinforcement learning for pedagogical policy induction," in *International Conference on Artificial Intelligence in Education,* Springer, Cham, 2019, pp. 544–556. https://doi.org/10.1007/978-3-030-23204-7_45

[5] W. E. Frankenhuis, K. Panchanathan, and A. G. Barto, "Enriching behavioral ecology with reinforcement learning methods," *Behavioural Processes*, vol. 161, pp. 94–100, 2019. https://doi.org/10.1016/j.beproc.2018.01.008

[6] Y. Vaicondam, H. Sikandar, S. Irum, N. Khan, and M. I. Qureshi, "Research landscape of digital learning over the past 20 years: A bibliometric and visualisation analysis," *International Journal of Online and Biomedical Engineering*, vol. 18, no. 8, pp. 4–22, 2022. https://doi.org/10.3991/ijoe.v18i08.31963

[7] N. T. Van, S. Irum, A. F. Abbas, H. Sikandar, and N. Khan, "Online learning—Two side arguments related to mental health," *International Journal of Online and Biomedical Engineering*, vol. 18, no. 9, pp. 131–143, 2022. https://doi.org/10.3991/ijoe.v18i09.32317

[8] C. Qiu, S. Cui, H. Yao, F. Xu, F. R. Yu, and C. Zhao, "A novel QoS-enabled load scheduling algorithm based on reinforcement learning in software-defined energy internet," *Future Generation Computer Systems*, vol. 92, pp. 43–51, 2019. https://doi.org/10.1016/j.future.2018.09.023

[9] D. Shawky and A. Badawi, "A reinforcement learning-based adaptive learning system," in *International Conference on Advanced Machine Learning Technologies and Applications,* Springer, Cham, 2018, pp. 221–231. https://doi.org/10.1007/978-3-319-74690-6_22

[10] S. Cunningham, L. Foote, M. Sowder, and C. Cunningham, "Interprofessional education and collaboration: A simulation-based learning experience focused on common and complementary skills in an acute care environment," *Journal of Interprofessional Care*, vol. 32, no. 3, pp. 395–398, 2018. https://doi.org/10.1080/13561820.2017.1411340

[11] J. Källström and F. Heintz, "Tunable dynamics in agent-based simulation using multi-objective reinforcement learning," in *Adaptive and Learning Agents Workshop (ALA-19) at AAMAS,* Montreal, Canada, May 13–14, 2019, pp. 1–7.

[12] E. H. Ha, "Effects of peer-led debriefing using simulation with case-based learning: Written vs. observed debriefing," *Nurse Education Today*, vol. 84, p. 104249, 2020. https://doi.org/10.1016/j.nedt.2019.104249

[13] T. Kobayashi, "Student-t policy in reinforcement learning to acquire global optimum of robot control," *Applied Intelligence*, vol. 49, no. 12, pp. 4335–4347, 2019. https://doi.org/10.1007/s10489-019-01510-8

[14] M. P. Gustin, M. Abbiati, R. Bonvin, M. W. Gerbase, and A. Baroffio, "Integrated problem-based learning versus lectures: A path analysis modelling of the relationships between educational context and learning approaches," *Medical Education Online*, vol. 23, no. 1, p. 1489690, 2018. https://doi.org/10.1080/10872981.2018.1489690

[15] D. D. Prasetya, A. P. Wibawa, and A. S. Ahmar, "Design of web-based lightweight interactive multimedia for distance learning," *Journal of Physics: Conference Series*, IOP Publishing, vol. 1028, no. 1, 2018, p. 012059. https://doi.org/10.1088/1742-6596/1028/1/012059

[16] P. Shaw, "A practice orientated framework to support successful higher education online learning," in *European Distance and E-Learning Network (EDEN) Conference Proceedings*, European Distance and E-Learning Network, no. 1, 2018, pp. 361–369.

[17] X. Hu, "Design of automatic response system for handheld mobile learning equipment based on augmented reality," *International Journal of Continuing Engineering Education and Life Long Learning*, vol. 31, no. 1, pp. 119–132, 2021. https://doi.org/10.1504/IJCEELL.2021.111850

[18] S. L. Savilova, A. A. Kropotkina, and Y. V. Kokhanovskaya, "Globalization of education: Online teaching Russian as a foreign language in a pandemic condition," in *SHS Web of Conferences*, *EDP Sciences*, vol. 92, 2021. https://doi.org/10.1051/shsconf/20219201044

[19] P. Bhargava, "Blended learning: An effective application to clinical teaching," in *Preparing the Next Generation of Teachers for 21st Century Education, IGI Global*, 2018, pp. 302–320. https://doi.org/10.4018/978-1-5225-4080-9.ch018

[20] C. McGuinness and C. Fulton, "Digital literacy in higher Education: A case study of student engagement with e-tutorials using blended learning," *Journal of Information Technology Education: Innovations in Practice*, vol. 18, pp. 001–028, 2019. https://doi.org/10.28945/4190

[21] F. Ahmadi, S. E. Mulyono, B. A. A. R. Al-Ghamdi, and D. Harjunowibowo, "Comparative study of the development of android-based flipped classroom model between Jeddah and Indonesia," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 9, no. 5, 2019.

[22] H. S. Laird-Fick, D. J. Solomon, C. J. Parker, and L. Wang, "Attendance, engagement and performance in a medical school curriculum: Early findings from competency-based progress testing in a new medical school curriculum," *PeerJ*, vol. 6, p. e5283, 2018. https://doi.org/10.7717/peerj.5283

[23] R. N. Aisyah and W. Budijastuti, "The feasibility development of android-based application media (Skeletonpedia) in biological learning on Skeletal systems submaterial of senior high school in XI grade," *BioEdu*, vol. 9, no. 1, pp. 124–130, 2020.

[24] C. V. de Carvalho, R. Butkiene, D. Gudoniene, O. Zubikova, and T. Bruegmann, "Work in progress: A MOOC-based innovative instructional approach for curriculum design," in *2020 IEEE Global Engineering Education Conference (EDUCON), IEEE*, 2020, pp. 1844–1847. https://doi.org/10.1109/EDUCON45650.2020.9125394

7 AUTHOR

Wenjing Wang, Zhejiang Conservatory of Music, Hangzhou, China (E-mail: wwj@zjcm.edu.cn).