Exploring Cognitive Diversity and Dynamics for Effective Language Memory Retention

Anonymous CogSci submission

Abstract

Spaced repetition, key for long-term memory retention through optimized review schedules based on predicted memory retention, is increasingly vital for effective language learning. Traditional methods, however, often fail to account for individual cognitive variations and material difficulty, resulting in a lack of high adaptability and effectiveness. To address this, our study introduces the Multidimensional Cognition Regression (MCR) model. MCR incorporates the Difficulty Engineering (DE) module, which integrates both objective and subjective factors to evaluate the intricacy of the content. Moreover, MCR further leverages a variety of user memory and cognitive characteristics, combined with psychological insights and machine learning techniques, to predict the memory "half-life" of material. This approach transcends methods like Half-Life Regression proven effective on Duolingo, reducing prediction errors demonstrated by lower Mean Absolute Error. Based on the predictive modeling of memory's halflife and corresponding biological memory patterns, we opt to schedule reviews at the juncture when the memory decays to its halflife point. Empirical validation in real-world settings showed enhanced retention efficiency.

Keywords: cognitive neuroscience; memory; language learning; human-computer interaction; machine learning

Introduction

The significance of language learning, especially English, in the global context cannot be understated. Central to language learning is the process of memory, which plays a crucial role in the acquisition and retention of new language skills, particularly in the repetitive recall of new vocabulary and grammatical structures. However, memory is characterized by its complexity and multifaceted nature (Klein, 2015), leading to a growing call for efficient learning and memorization methods.

One method that has shown promise in optimizing memory retention is spaced repetition. Spaced repetition, a learning technique that involves increasing intervals of time between subsequent reviews of previously learned material, leverages the psychological spacing effect (Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008) and lag effect (Melton, 1970) to enhance long-term memory retention. This approach has found applications for learners across various domains: law (Teninbaum, 2016), medicine (Shaw et al., 2012), statistics (Maass, Pavlik, & Hua, 2015), history (Carpenter, Pashler, & Cepeda, 2009), and particularly notably in language learning: advanced models like Half-Life Regression (HLR) (Settles & Meeder, 2016) and Difficulty-Halflife-Recall Probability (DHP) (Ye, Su, & Cao, 2022) have revolutionized the spaced repetition. HLR, effectively utilized in Duolingo, merges psycholinguistic theories with machine learning to accurately predict the "halflife" of words in memory, it employs a regression approach that factors in both the past performance of the learner and attributes of the words being learned. On the other hand, the DHP model, deployed in MaiMemo, also a popular online English learning application, employs time-series analysis and Markov properties, combining the learning material difficulty, memory halflife, and recall probability into a comprehensive framework, it create a more dynamic and responsive learning environment. Both models demonstrate significant advancements in personalized and efficient language learning, validated in real-world applications.

To implement an effective spaced repetition method for language learning, two components are essential: the memory prediction model, and the scheduling algorithm. Specifically, the memory prediction model accurately gauge the retention capacity of each learner, taking into account individual differences in memory strength and decay rates. It should also consider contextual factors such as prior knowledge, cognitive load, and material complexity. Meanwhile, the scheduling algorithm should adaptively schedule review sessions based on the predictions calculated by the model, ensuring maximum retention with minimal cognitive load, besides, it should balance the frequency of repetition with the learner's evolving proficiency, making efficient use of study time while avoiding both overlearning and underlearning.

However, existing aforementioned methodologies exhibit certain limitations. Primarily, they lack a comprehensive utilization of cognitive factors and temporal information, which are crucial for understanding individual learning patterns and memory dynamics according to spacing effect. Furthermore, the efficacy of memory is closely associated with the difficulty of the material (Sweller & Chandler, 1994), thereby constituting a critical aspect for consideration in spaced repetition. The existing models lack comprehensive consideration of multiple factors in the evaluation of material difficulty.

Our work introduces a comprehensive learning material difficulty evaluation system: Difficult Engineer (DE), which integrates both subjective and objective factors for learning materials. Besides, a novel and effective memory prediction model Multidimensional Cognition Regression (MCR) is proposed. This is complemented by Optimized Biological Semi-Memory Scheduler (OBS) to adjust learning pace. We compare MCR to HLR and DHP, besides, our schedule approach is validated in real-world learning environments to demonstrate its practical efficacy. The contribution aims to enhance the personalized and efficient application of spaced repetition, bridging gaps in current methodologies and fostering more adaptive language learning strategies.

Related Work

Memory Model

In the study of human memory modeling, several experiments have contributed significantly to our understanding. Ebbinghaus (1913) introduced the concept of the forgetting curve, illustrating how memory decays without review. Anderson (1996) expanded on this with the ACT-R theory, suggesting that each review generates a unique forgetting curve within the declarative memory module. Further, Pashler, Cepeda, Lindsey, Vul, & Mozer Pashler, Cepeda, Lindsey, Vul, and Mozer (2009) developed the multiscale context model, an amalgamation of two cognitive theories. This model differentiates between successful and unsuccessful recall, applying hand-picked weights in its analysis. However, these models tend to overlook the nuanced interplay of individual cognitive differences and the dynamic nature of memory formation and retention over time.

Scheduling Algorithm

SuperMemo (Woźniak & Gorzelańczyk, 1994), the original digital spaced repetition algorithm, updates its schedule based on user interactions to maintain a 5% user forgetfulness rate. Its primary limitation lies in its inflexible scheduling that does not adapt to individual learning variations. Reddy, Labutov, Banerjee, and Joachims (2016) proposed a queuing network model to maximize the Leitner system's learning speed; however, the model's rigidity and untested accuracy are significant drawbacks. Tabibian et al. (2019) introduced marked time-series points in a stochastic optimal control-based memory algorithm MEMORIZE, yet this approach suffers from overly complex implementation for practical use. Hochreiter and Schmidhuber (1997) enhanced Deep Reinforcement Learning (DRL) (Reddy, Levine, & Dragan, 2017) using Long Short-Term Memory (LSTM) networks for reward prediction. Upadhyay, De, and Gomez Rodriguez (2018) developed a policy gradient-based deep reinforcement learning algorithm, encoding all item reviews' histories into hidden states for multiple item sessions. However the complexity of deep learning models pose challenges for interpretability, high computational demands and practical deployment.

Methodology

Overview

In our study, we propose an innovative memory halflife prediction model, MCR. Central to memory halflife, represented by h, is the Ebbinghaus model, also known as the forgetting curve, positing that memory decays exponentially over time: $p = 2^{-\frac{\Delta}{h}}$, where p is the recall probability, Δ is the time since the last review, and h is the memory halflife of the word.

Specifically, this model leverages extensive user data derived from online English learning platforms, incorporating a multifaceted analysis of cognitive factors, such as frequency of learning sessions, learning intervals, and effectiveness of review, simultaneously, the model integrates a module which rigorously evaluates the inherent difficulty of the learning words. By fitting a large amount of the factors aforementioned, the model formulates a memory halflife transfer equation. This enables the prediction of changes in the halflife of a specific word under a given cognitive state of a user. Moreover, building on the predictive strength of our model concerning memory state transfer and biological principles, we propose a tailored spaced repetition scheduling algorithm: OBS, specifically designed for more efficient memory retention.

Dataset

Requirement Memory is dynamic and intricate, varying significantly among individuals and across different types of material. Consequently, the training data for a halflife prediction model should be adaptable to several key criteria. Firstly, there's the need for Expanse in User Data: the data should encompass a broad range of user profiles, providing the model with a profound understanding of diverse learning patterns and cognitive differences. Additionally, the Variability in Word Difficulty is crucial: the data should include English words of varying difficulty levels, aiding the model in mastering a wide distribution of vocabulary complexity. Finally, it's essential to capture the Diversity of the Learning Process: the data should reflect various learning processes of users, facilitating the model in assessing the effectiveness of different learning strategies and optimizing personalized learning pathways.

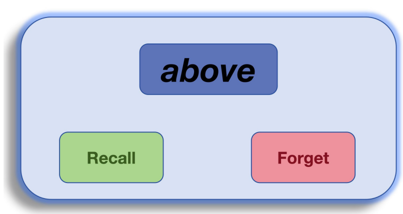

Figure 1: User interface of flashcard ("above" is the word to be reviewed)

Given the prerequisites, and in light of the burgeoning prevalence of English learning applications on mobile platforms, a substantial volume of user memory data can be feasibly acquired. In our research, we strategically selected a dataset from MaiMemo. The application should look like Figure 1, with the words presented as flashcards. Each interaction with a word requires the user to mark their memory status as recall or forget. This dataset comprising 220 million user records, serving the purpose to solve the Expanse of User Data and the Diversity of the Learning Process. Besides, we provide a new rating system to access the Variability in Word Difficulty.

Difficulty Engineering (DE) Considering MaiMemo's extensive lexicon of nearly 100,000 words, the sparsity of behavioral events collected across varying time series for each word renders the establishment of individual difficulty indices for each word impractical. Consequently, a compromise is necessitated, grouping words is essential to achieve a balance between differentiating individual words and reducing the problem of data sparsity.

To enhance the classification of word complexity and thereby refining the model's alignment with users' memory patterns, we aim to formulate an innovative evaluation system Difficulty Engineering (DE) for word difficulty. This approach entails an amalgamation of both subjective and objective factors associated with each word to categorize its difficulty.

Subjective. For the subjective factors, our analysis incorporates two critical aspects. The first aspect is p_{second}, which represents the probability of recalling a word on the subsequent second day. This is calculated as $\frac{n_{recall}}{N}$, where N is the total number of times the word is reviewed on the second day, and n_{recall} is the number of times the word is successfully recalled on that day. The second aspect is p_{recall}, which refers to the total probability of a particular word's record being recalled.

The p_{second} serves as an indicator of short-term retention and initial learning effectiveness, reflecting the ease with which a word is initially assimilated into memory. Conversely, p_{recall} provides a broader measure of long-term retention, highlighting the word's persistence in memory over extended periods and offering insights into its inherent memorability across a diverse user base.

To assess how hard a word can be subjectively remembered, we choose an average p_{avg} between long-term memory and short-term memory, The distribution of p_{second}, p_{recall} and p_{avg} is shown in Figure 2.

$$p_{avg} = 0.5 \cdot p_{second} + 0.5 \cdot p_{recall} \tag{1}$$

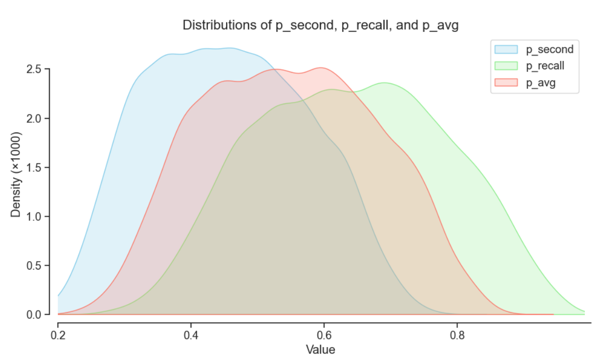

Figure 2: Distribution of subjective factors

Objective. In the context of objective factors, our attention converges upon the impact exerted by the intrinsic attributes of the lexical item on mnemonic retention. Attributable to the cognitive load associated with processing and internalizing linguistic information (Noroozi & Karami, 2022), longer words, characterized by increased phonological and orthographic content, place greater demands on cognitive resources during the processes of encoding, storage, and retrieval. Similarly, a higher syllable count enhances phonetic complexity, further influencing memorability. Consequently, we contemplate the syllable count (SC) and the length of the word (WL) which both emerge as critical determinants of the cognitive effort required for learning and recalling words. Empirical observations reveal a conspicuous linear correlation between SC and WL, accordingly, we quantify each lexeme's orthographic length and syllabic composition in dataset, then use linear regression to analyze their relationships:

$$SC = 0.387 \cdot WL - 0.399 \tag{2}$$

Definition of DE. Incorporating both subjective and objective factors, we introduce the novel word difficulty assessment framework denoted as DE, which is formally defined by the ensuing equation:

$$\mathcal{D} = \lfloor \alpha \cdot (1 - p_{\text{avg}}) + \beta \cdot (\text{WL} + \text{SC}) \rfloor \tag{3}$$

Here, the variables α and β serve as adjustable parameters that govern the magnitude of impact exerted by each respective factor on the overall assessment of word difficulty. The floor function, denoted as $\lfloor \cdot \rfloor$, is employed to ensure the discreteness of the difficulty $\mathcal{D}$, thereby facilitating the grouping of words based on their assessed difficulty levels.

Data Analysis and Process The comprehensive dataset, boasting over 220 million entries, is composed of several critical features, detailed in Table 1. Each entry includes the anonymized student user ID, denoted as u, representing the individual who reviewed the word. The word itself is identified by w, its spelling. The feature $t_{history}$ provides a sequence of intervals between the word's historical reviews, while $r_{history}$ offers insight into the word's retention by indicating whether it was recalled or forgotten in these past reviews. Additionally, Δt measures the time that has passed since the word's last review. Finally, the outcome of the most recent review is captured by r, specifying whether the word was recalled or not in the latest attempt.

Table 1: Example entries from the dataset

u	w	$t_{history}$	$r_{history}$	Δt	r
515b51	joke	0,1	0,1	3	1
6724cc	county	0,1,3	0,1,1	9	1
1cad68	lineage	0,1,10	0,0,1	3	1

To ensure that the calculation of halflife is both applicable and representative, we need to group a series of records with

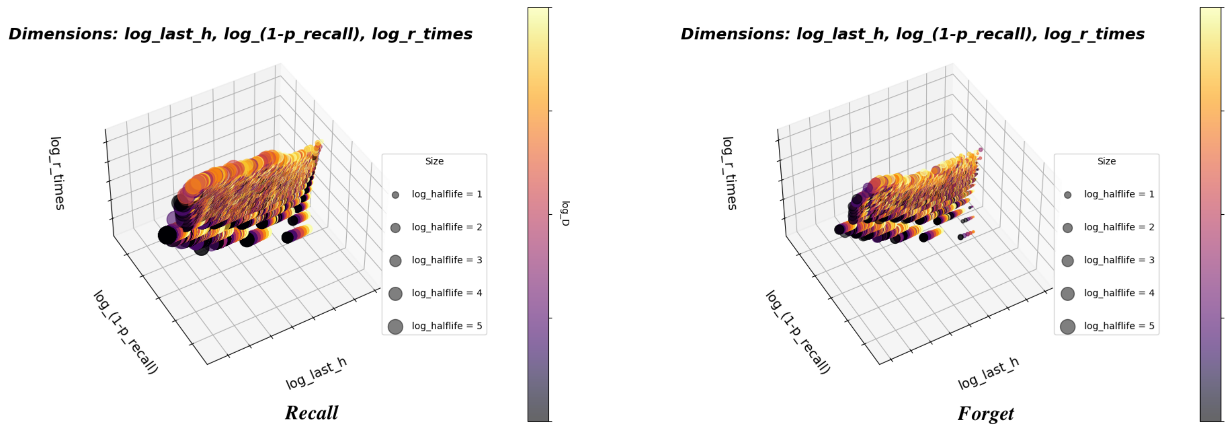

Figure 3: Visualization for $\mathbf{X} = (last\_h, \mathcal{D}, 1 - p_{recall}, r_{times})$ and halflife $(\alpha = 10, \beta = 5)$

the same cognitive states. In our approach, records that share the same word difficulty $\mathcal{D}$, and same review states $r_{history}$, and $t_{history}$ are considered to have a consistent halflife, then the term $total\_cnt$ is derived as a measure reflecting consistent behavioral patterns across these grouped states, representing the frequency of each unique combination.

The halflife is calculated using the following formula:

$$h = \frac{\log(0.5) \cdot \sum(\Delta t^2 \cdot \text{total\_cnt})}{\sum(\Delta t \cdot \log(p_{\text{recall}}) \cdot \text{total\_cnt})} \tag{4}$$

This equation is a weighted average that accounts for the varying intervals (Δt) between reviews and the differing recall probabilities (p_{recall}) for each cognitive state, weighted by $total\_cnt$. The numerator involves $\log(0.5)$ which is a constant reflecting the definition of halflife (the time when the recall probability drops to 50%). The denominator combines the impact of time intervals and recall probabilities across a wide range of the identical cognitive states, providing a comprehensive measure of memory retention over time.

Table 2: Examples after grouping and calculating the halflife$(\alpha = 10, \beta = 5)$

$\mathcal{D}$	$t_{history}$	$r_{history}$	Δt	p_{recall}	$total\_cnt$	halflife
24	0,1,3	0,1,1	6	0.9567	1754	98.21
24	0,1,3	0,1,1	8	0.9447	3455	98.21
24	0,1,3	0,1,1	22	0.8673	113	98.21
24	0,1,3,8	0,1,1,1	17	0.9431	580	244.71
24	0,1,3,8	0,1,1,1	21	0.9541	392	244.71

As shown in the Table 2, for record ($\mathcal{D}$=24, $t_{history}$=0,1,3, $r_{history}$=0,1,1, Δt=8, p_{recall}=0.9447, $total\_cnt$=3455), its corresponding halflife is 98.21 days. After a time interval of 8 days, the half-life is shifted to 244.71 days in the case where the result of the next review is recall.

Multidimensional Cognition Regression(MCR)

Goal Our target is to build a model that uses a series of features to fit the user's memory patterns and get the halflife transfer equation. That is, after a memorized state: recall or forget, the model predicts a change in the halflife. Specifically, considering a vector $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_n)$:

$$\hat{\text{h}}_i(\theta) = \mathcal{F}(\mathbf{X}; \theta) \tag{5}$$

where the $\mathcal{F}$ represents the fitness function mapping multidimensional cognitive features $\mathbf{X}$ to the user's memory halflife, and θ encompasses the model parameters. The $\hat{\text{h}}_i(\theta)$ is the predicted halflife for the i-th $\mathbf{X}$ obtaining through $\mathcal{F}(\mathbf{X}; \theta)$.

Choice of X and $\mathcal{F}$ In our model, the vector $\mathbf{X}$ incorporates four cognitive characteristics: $\mathbf{X} = (last\_h, \mathcal{D}, 1 - p_{recall}, r_{times})$. Each of these characteristics has been selected for its significant impact on memory retention. The first characteristic, $last\_h$, represents the previously measured halflife of memory for a word, indicating the storage strength of memory and reflecting the durability of the memory trace prior to the current recall attempt. Next, $\mathcal{D}$ denotes the difficulty level of the word, with more complex words often having shorter halflives due to the increased cognitive effort needed for encoding and retrieval. The third element, p_{recall}, is the probability of recall, providing a direct measure of memory strength; higher recall probabilities indicate a stronger memory trace and a longer halflife. Lastly, r_{times}, the number of times a word has been reviewed, acts as a gauge of retrieval practice. Frequent retrieval is known to strengthen memory, consequently extending the halflife of the word.

Shown in Figure 3, we employ a three-dimensional projection to visualize the feature space, selecting three of the quartet of features as the axes—designated X, Y, and Z respectively. The scatter plot's marker size is proportionally scaled to the distribution magnitude of the memory halflife, thereby representing the temporal extent of retention. Concurrently, the color gradient of each scatter point is mapped to the distribution range of the fourth feature. According to the visualization, compared with forget, the size of halflife in the case of recall increases significantly, which indicates that the case of successful recall reflects that the retention of memory is greater than the case of forget. Moreover, the distribution arrangement of the projected scatter points presents a multi-segment arrangement that is approximately linear, in particular, for each line segment, the size and color of the scatter point show the same increasing trend, which inspires us to use linear regression to fit the data:

$$\mathcal{F}(\mathbf{X};\theta) = \exp\left(\theta^\top \cdot \log(\mathbf{X}) + \theta_0\right) \tag{6}$$

In the model formulation, the $\log(\mathbf{X})$ implies the logarithmic transformation of each component in $\mathbf{X}$. θ_0 is a constant, the term $\theta^\top$ denotes the transpose of the parameter vector θ containing parameters $(\theta_1, \theta_2, \theta_3, \theta_4)$.

Memory Stage Transfer Equation The transfer of memory state is obtained by fitting a large number of user data:

$$\hat{\mathrm{h}}_i = \mathcal{F}(\mathbf{X};\theta_{recall}) \cdot I + \mathcal{F}(\mathbf{X};\theta_{forget}) \cdot (1-I) \tag{7}$$

In Equation 7, $\hat{\mathrm{h}}_i$ is the predicted halflife, where θ_{recall} and θ_{forget} are the parameters for the recall and forget cases. The application of the parameters is controlled by I, which is a binary indicator to distinguish between recall and forget states.

Scheduling Algorithm

Different spaced repetition scheduling strategies significantly influence the efficacy of memory retention, the most effective strategies are predicated upon the heuristic of reviewing material when predicted memory strength approaches a specific threshold according to spaced repetition algorithm (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Capitalizing on this heuristic, we present the OptiBioSemi Memory Scheduler (OBS). This algorithm is underpinned by the memory halflife transfer Equation 7 and principles of biological memory dynamics (Wang, 2013), positing that the most opportune moment for review is when memory retention descends to the 50% mark. The OBS is thus designed to trigger reviews at this critical retention threshold. Figure 4 illustrates examples of the scheduling facilitated by our algorithm.

Evaluation

Metrics

In evaluating the MCR model, we compare actual memory halflife h_i with the predicted $\hat{h}_i$. Our assessment uses several key metrics: Mean Absolute Error (MAE) $= \frac{1}{n}\sum|h_i - \hat{h}_i|$ measures the average error magnitude, Mean Squared Error (MSE) $= \frac{1}{n}\sum(h_i - \hat{h}_i)^2$ emphasizes larger errors, and Root Mean Squared Error (RMSE) $= \sqrt{\frac{1}{n}\sum(h_i - \hat{h}_i)^2}$ provides a sensitive error measure. Mean Absolute Percentage Error (MAPE) $= \frac{100\%}{n}\sum\left|\frac{h_i - \hat{h}_i}{h_i}\right|$ interprets error as a percentage.

Figure 4: Example of OBS($r_{history} = 0,1,0,1,0,1,0,1$)

Additionally, we calculate the coefficient of determination $R^2 = 1 - \frac{\sum(h_i - \hat{h}_i)^2}{\sum(h_i - \bar{h})^2}$, with $\bar{h}$ as the mean of h_i, to gauge the explanatory power of our model.

Balance of α and β

Table 3: MSE distribution of α and β (recall)

α / β	4.5	5.0	5.5
5.0	419.90	396.10	411.28
10.0	376.52	**348.72**	365.91
15.0	406.88	385.02	396.17

Owing to the more pronounced and perceptible distribution gap exhibited by the MSE, it has been selected as the metric for comparative analysis of the balance of α and β. As delineated in Table 3, the parameter values $\alpha = 10$ and $\beta = 5$ are associated with the minimal MSE. Consequently, $\alpha = 10$ and $\beta = 5$ are adopted as the definitive parameters for the DE.

Performance

The efficacy of the proposed model is benchmarked against two models: HLR and DHP. In the initial DHP experiment, instances featuring multiple forgotten memory within the $r_{history}$ were systematically excluded; for example, scenarios with $r_{history} = 0,0,1$ or $0,1,0,1$ were omitted. This exclusion constrains the scope of the investigation, rendering it misaligned with the complexities inherent in dynamic memory processes. A more encompassing experiment is proposed, wherein such instances will be retained. The resulting data are encapsulated in Table 4. The results show that our model provides better predictive power in the case of recall. Particularly, the model's incorporation of $\mathcal{D}$ and $r_{history}$ as primary features enables a more nuanced prediction of memory halflife. Our findings indicate that $\mathcal{D}$, representing the complexity of the learning material, significantly impacts memory retention, with more complex materials exhibiting shorter halflives. This aligns with cognitive theories suggesting increased cognitive load with more difficult content. Moreover, the $r_{history}$ feature, capturing the frequency and recency of interactions with the material, has proven critical in understanding memory dynamics. Frequent reviews are correlated

Table 4: Evaluation results

Method	Recall				Forget			
	MAE	RMSE	MAPE	R^2	MAE	RMSE	MAPE	R^2
HLR	41.70	86.99	45.75%	86.57%	7.08	30.96	112.70%	17.62%
HLR+DE	29.41	64.23	44.84%	85.38%	5.89	27.87	109.00%	17.53%
DHP	17.84	38.69	52.27%	95.51%	0.44	1.33	31.99%	67.78%
DHP+DE	14.02	31.68	52.86%	93.72%	0.37	1.22	31.01%	69.81%
MCR	**8.21**	**18.67**	**21.32%**	**98.97%**	0.37	1.25	30.46%	69.09%

with longer memory retention, emphasizing the importance of spaced repetition in learning.

User Experiment

To assess the application effectiveness of OBS, we developed an English word memorization program, which also features a flashcard-style user interface, as depicted in Figure 1. In this software, we implemented several classic spaced repetition algorithms: ANKI (Nguyen, 2021), MEMORIZE (Tabibian et al., 2019), and additionally conducted a comparison with the mainstream application MaiMemo. To ensure a balanced cognitive level among the users, our software's user base comprises undergraduate students, graduate students and English subtitle translation team members, with each method being utilized by 5 members. To minimize the influence of prior knowledge on memory retention, GRE vocabulary words are chosen as the study material. The experimental setup involves studying 10 new words daily and reviewing 20 words, spanning over a duration of two months. We employ the recall rate of all 600 words (60 * 10) on the final day as the evaluation criterion, and the results are detailed in the Table 5.

Table 5: User *Recall* comparison

	OBS	ANKI	MEMORIZE	MaiMemo
Recall	0.36	0.29	0.31	0.33

Discussion

Our study introduces the MCR model with DE method. The model adopts a comprehensive approach to cognitive processes, incorporating multiple cognitional features with previously underexplored variables in memory prediction models. It intersects psychological theory, machine learning techniques, leading to a significant improvement in memory fitting, as reflected in the enhanced metrics of MAE, RMSE, MAPE, and R^2. Additionally, the effectiveness of DE is validated through ablation experiments, demonstrating its accuracy and rationality in classifying material difficulty. These nuanced understanding of memory retention and decay, represents a advancement in language learning and memory retention methodologies. In addition, we tested the scheduling algorithm OBS based on MCR prediction in a real learning environment, ensuring that theoretical advancements are not just academically robust but also practically applicable.

However, it is important to acknowledge the limitations of our study. First of all, although the user experiment proves that the 50% threshold in the scheduling algorithm is practical and effective, its selection may need more consideration, and more controlled experiments can be carried out on this basis. Then, while our findings offer significant insights into language learning, the generalizability of these results beyond language acquisition remains to be fully explored. The applicability of the MCR model to other forms of learning, such as skill acquisition or conceptual understanding, is an area ripe for investigation. Additionally, the complexity of the MCR model might pose challenges in real-world settings, particularly in terms of the need for extensive data for accurate predictions.

Future research should prioritize key areas to address these limitations. Firstly, there is a need to identify additional features closely linked to memory, aiming to enhance the model's efficiency and accuracy while minimizing computational demands. Second, applying the OBS algorithm across various educational contexts, not limited to language learning, is essential. Investigating its effectiveness in diverse settings, such as STEM education and corporate training, could significantly extend its applicability. Further, integrating the MCR model with adaptive learning systems is vital. These systems, using real-time data, could continually tailor learning schedules, thus providing a customized and dynamic educational experience. Additionally, incorporating neuroscientific insights into the model is a promising avenue. Understanding the roles of different brain regions and neural pathways in memory formation and retention could improve the model's precision and biological relevance. Also, investigating long-term retention and the applicability of knowledge acquired through OBS-optimized schedules is crucial. It's important to assess not just initial learning success but also the enduring retention and practical application of knowledge over time. Such research would offer valuable insights into the long-term effects of personalized learning schedules on knowledge acquisition and application.

In conclusion, our study not only delivers innovative insights into the prediction of memory patterns in spaced repetition but also proposes efficacious strategies for the enhancement of language learning. By intertwining theoretical concepts with pragmatic applications, our research creates a confluence of psychological theories, machine learning techniques, principles of biological memory dynamics, and the pragmatic demands of real-world language acquisition. This synergistic approach marks a substantial advancement in the domains of cognitive science and memorial methodologies.

References

Anderson, J. R. (1996). Act: A simple theory of complex cognition. *American psychologist*, *51*(4), 355.

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of us history facts. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(6), 760–771.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin*, *132*(3), 354.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological science*, *19*(11), 1095–1102.

Ebbinghaus, H. (1913). A contribution to experimental psychology. *New York, NY: Teachers College, Columbia University*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Klein, S. B. (2015). What memory is. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(1), 1–38.

Maass, J. K., Pavlik, P. I., & Hua, H. (2015). How spacing and variable retrieval practice affect the learning of statistics concepts. In *Artificial intelligence in education: 17th international conference, aied 2015, madrid, spain, june 22-26, 2015. proceedings 17* (pp. 247–256).

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 596–606.

Nguyen, B.-P. T. (2021). Mobile-assisted vocabulary learning: A review of anki. *Journal of Educational Technology*, *18*(3).

Noroozi, S., & Karami, H. (2022). A scrutiny of the relationship between cognitive load and difficulty estimates of language test items. *Language Testing in Asia*, *12*(1), 1–19.

Pashler, H., Cepeda, N., Lindsey, R. V., Vul, E., & Mozer, M. C. (2009). Predicting the optimal spacing of study: A multiscale context model of memory. *Advances in neural information processing systems*, *22*.

Reddy, S., Labutov, I., Banerjee, S., & Joachims, T. (2016). Unbounded human learning: Optimal scheduling for spaced repetition. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1815–1824).

Reddy, S., Levine, S., & Dragan, A. (2017). Accelerating human learning with deep reinforcement learning. In *Nips workshop: teaching machines, robots, and humans.*

Settles, B., & Meeder, B. (2016). A trainable spaced repetition model for language learning. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 1848–1858).

Shaw, T. J., Pernar, L. I., Peyre, S. E., Helfrick, J. F., Vogelgesang, K. R., Graydon-Baker, E., ... others (2012). Impact of online education on intern behaviour around joint commission national patient safety goals: a randomised trial. *BMJ quality & safety*.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and instruction*, *12*(3), 185–233.

Tabibian, B., Upadhyay, U., De, A., Zarezade, A., Schölkopf, B., & Gomez-Rodriguez, M. (2019). Enhancing human learning via spaced repetition optimization. *Proceedings of the National Academy of Sciences*, *116*(10), 3988–3993.

Teninbaum, G. H. (2016). Spaced repetition: A method for learning more law in less time. *J. High Tech. L.*, *17*, 273.

Upadhyay, U., De, A., & Gomez Rodriguez, M. (2018). Deep reinforcement learning of marked temporal point processes. *Advances in Neural Information Processing Systems*, *31*.

Wang, B. (2013). *Study on english vocabulary memory model based on biological memory principles and its application on mobile devices*. Unpublished master's thesis, Fudan University.

Woźniak, P., & Gorzelańczyk, E. (1994). Optimization of repetition spacing in the practice of learning. *Acta neurobiologiae experimentalis*, *54*(1), 59–62.

Ye, J., Su, J., & Cao, Y. (2022). A stochastic shortest path algorithm for optimizing spaced repetition scheduling. In *Proceedings of the 28th acm sigkdd conference on knowledge discovery and data mining* (pp. 4381–4390).