ISSN 2522-6398 Vol. 9, Issue 8: 121-128, DOI: 10.25236/FER.2026.090816

Research on Algorithms for Improving English Vocabulary Teaching Effectiveness Using Machine Learning

Huizhuo Zhang*, Yajun Zhao, Wen Bai

Shenyang Open University, Shenyang, China
**zhz78@sina.com*

Abstract: *Addressing the poor adaptability of traditional static vocabulary teaching models and the limitations of existing machine learning schemes such as fixed forgetting models and single feature dimensions, this paper proposes a vocabulary learning optimization algorithm (DF-MMFL algorithm) that integrates dynamic forgetting perception and multimodal feature fusion. This algorithm comprises three core modules: a multimodal data preprocessing module that constructs a 12-dimensional feature vector covering four categories of information, including vocabulary difficulty and answering speed, through improved Min-Max normalization and mutual information entropy filtering; a dynamic forgetting model module that, based on Ebbinghaus forgetting theory, designs a forgetting coefficient correction factor using real-time answering data and iteratively updates the forgetting coefficient through gradient descent; and a reinforcement learning recommendation module that optimizes the recommendation strategy using a short-term and long-term weighted composite reward function and an improved Q-learning algorithm. A controlled experiment was conducted with 500 first-year non-English major students, divided into four groups. Results showed that the experimental group achieved memory retention rates of 82.3% and 71.5% after one and two weeks, respectively, representing increases of 28.1 and 30.3 percentage points compared to the control group. The overall accuracy rate of vocabulary application was 68.9%, the recommendation accuracy rate was 90.1%, and the model converged after only 120 iterations. This study demonstrates that the DF-MMFL algorithm can effectively improve vocabulary memory retention and application capabilities, optimizing algorithm performance.*

Keywords: *Vocabulary teaching; personalized recommendation; dynamic forgetting model; multimodal feature fusion; reinforcement learning; DF-MMFL algorithm; memory retention rate*

1. Introduction

English vocabulary, as a core element in language competence construction, directly impacts language acquisition efficiency through its teaching effectiveness. Traditional vocabulary teaching systems often employ fixed teaching schedules and content delivery logics. Essentially, they are static teaching models built on the average level of the group, failing to adapt to individual differences among learners in terms of memory capacity, decay rate, and information processing efficiency. This leads to a mismatch between teaching resources and learners' actual memory states, ultimately manifesting as low long-term memory retention and weak knowledge transfer abilities.

With the popularization of online education technology, vocabulary learning platforms can now reliably collect massive amounts of multi-dimensional data, including learners' answer results, reaction times, interaction click sequences, and periodic test scores [1]. This structured and semi-structured data provides a crucial data foundation for the integration of machine learning technology into personalized vocabulary teaching, making the transition from "batch teaching" to a teaching model that "precisely adapts to individual memory patterns" technically feasible.

However, current machine learning-based vocabulary teaching optimization schemes still have significant technical limitations [2]. On the one hand, most schemes use fixed-parameter forgetting models (such as fixed-period review strategies based on the Ebbinghaus forgetting curve), failing to adjust the memory decay coefficient based on dynamic data such as learners' real-time answer accuracy and reaction time, resulting in insufficient accuracy in depicting individual forgetting patterns [3]. On the other hand, feature dimensions are often limited, focusing primarily on a two-dimensional feature space of "vocabulary difficulty - learner's historical performance," neglecting multimodal features

ISSN 2522-6398 Vol. 9, Issue 8: 121-128, DOI: 10.25236/FER.2026.090816

reflecting learning status such as attention interaction frequency and answer hesitation index, thus restricting the accuracy of personalized recommendations.

To address these technical bottlenecks, this paper proposes a vocabulary learning optimization algorithm that integrates dynamic forgetting perception and multimodal feature fusion (DF-MMFL algorithm), aiming to improve the adaptability of machine learning models to individual memory patterns and enhance teaching effectiveness [4]. The specific research framework of this paper is as follows: First, the core shortcomings of existing algorithms in terms of memory model dynamism and feature dimension are identified; second, the three core modules of the DF-MMFL algorithm—multimodal data preprocessing, dynamic forgetting model construction, and reinforcement learning recommendation—are designed in detail, and the model training parameters are optimized; subsequently, comparative experiments are conducted to verify the performance advantages of the algorithm in terms of memory retention rate, application accuracy, and other indicators; finally, the technical value and application boundaries of the algorithm are summarized, and optimization directions for different learning groups and scenarios are proposed [5].

2. Design of Dynamic Forgetting Perception and Multimodal Feature Fusion Algorithm (DF-MMFL)

2.1. Multimodal Data Preprocessing Module

Data preprocessing begins with normalization to address dimensional differences, employing an improved Min-Max normalization method: For features with fixed preset intervals (levels 1-5), such as vocabulary difficulty, the formula $x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$ maps them to the [0, 1] interval, ensuring comparability of feature values across different difficulty levels. For features susceptible to outliers (such as excessively long durations due to sudden interruptions), such as answer duration, a dynamic interval parameter Δt (the standard deviation of the last 3 answer durations) is introduced [6]. The formula $t_i' = \frac{t_i - (\bar{t} - \Delta t)}{\max(t_i) - (t - \Delta t)}$ compresses the impact of outliers, where $\bar{t}$ is the average answer duration of the last 3 times. When a particular duration significantly exceeds the average, this formula can control its normalized value to within 1.2, preventing interference with subsequent models.

Subsequently, strongly correlated features were selected using mutual information entropy, and the mutual information value between each feature and the memory accuracy after one week was calculated. The formula is as follows:

$$I(X;Y) = H(Y) - H(Y \mid X) \tag{1}$$

Where $H(Y)$ is the information entropy of memory accuracy, and $H(Y \mid X)$ is the conditional entropy given feature X. Experimental verification shows that features with mutual information values below 0.3 (such as login time and device type) have extremely weak correlation with memory retention rate. Retaining them would increase the computational load of the model and reduce accuracy [7]. Therefore, only 12 highly correlated features were ultimately retained to construct a multimodal feature vector $\boldsymbol{X} = [x_1, x_2, \ldots, x_{12}]$, covering four core categories of information: vocabulary difficulty (2 indicators), answering speed (3 indicators), attention interaction (4 indicators), and emotional state (3 indicators).

2.2. Dynamic Forgetting Model Construction Module

This module first constructs an initial memory retention rate model, introducing a learner's initial memory ability coefficient m_0 and a vocabulary difficulty factor d_i (normalized value) based on Ebbinghaus's theory of memory decay. m_0 is determined through a basic test, calculated as "average score of 3 basic tests / 100 " (out of 100). If the learner's average score is 85, then $m_0 = 0.85$, quantifying their initial memory ability [8]. The initial memory retention rate model is: $R_0(t) = m_0 \cdot e^{-k_0 \cdot d_i \cdot t}$, t, where t is the time interval after learning (in days), and k_0 is the initial forgetting coefficient (default value 0.2). This model reflects the principle that "the higher the difficulty and the weaker the initial memory ability, the faster the memory retention rate decays," which aligns with real-world learning scenarios.

To achieve dynamic adaptation of the forgetting model, a forgetting coefficient correction factor α is designed, which combines the real-time answer accuracy A ($0-1$ interval) and the normalized reaction time value t'($0-1$ interval). The formula is as follows:

ISSN 2522-6398 Vol. 9, Issue 8: 121-128, DOI: 10.25236/FER.2026.090816

$$\alpha = 1 + 0.2A - 0.2t' \tag{2}$$

When learners have a high accuracy rate ($A = 1$) and a fast response ($t' = 0$), $\alpha = 1.2$, indicating a solid grasp of the current vocabulary and a need to slow down the rate of forgetting. When the accuracy rate is low ($A = 0$) and the response is slow ($t' = 1$), $\alpha = 0.8$, indicating weak mastery and a need to accelerate the updating of the forgetting coefficient to trigger review.

The gradient descent method is used to iteratively update the forgetting coefficient k based on the correction factor α. First, the loss function L is defined as the "mean squared error between the predicted memory retention rate and the actual accuracy rate," and the formula is:

$$L = \frac{1}{n}\sum_{i=1}^{n}\left(R_{\text{pred},i} - A_i\right)^2 \tag{3}$$

n is the number of words learned in a single learning session, $R_{\text{pred},i}$ is the retention rate predicted by the model, andA_i is the actual correct answer rate. Taking the partial derivative with respect to k and substituting it into the gradient descent formula, the paper obtained the forgetting coefficient update formula:

$$k_{t+1} = k_t - \eta \cdot \frac{\partial L}{\partial k_t} = k_t - \eta \cdot \frac{2}{n}\sum_{i=1}^{n}\left(R_{\text{pred},i} - A_i\right) \cdot \left(-d_i \cdot t \cdot R_{\text{pred},i}\right) \tag{4}$$

Where η is the learning rate (set to 0.01 after experimental optimization). After each learning session, k is updated based on the answer data to ensure that the model always adapts to the learner's current memory state.

2.3. Reinforcement Learning Recommendation Module

The reward function adopts a composite form of short-term and long-term weighting, the formula is:

$$R = 0.3R_{\text{short}} + 0.7\beta \cdot R_{\text{long}} \tag{5}$$

Where R__short is the immediate answer accuracy (0-1), with a weight of 0.3 to consider short-term learning effects; R_long is the vocabulary memorization accuracy after one week (0-1), with a weight of 0.7 to highlight long-term memory goals; β is the long-term reward decay factor (set to 0.95), used to reduce the impact of uncertainty in long-term memory data—if there are errors in the test data after one week, the decay factor can reduce its interference with the reward function.

Q-value updates use an improved Q-learning algorithm, the formula is

$$Q(S,A) = Q(S,A) + \alpha_q \cdot \left[R + \gamma \cdot \max_{a'} Q(S',a') - Q(S,A)\right] \tag{6}$$

Where γ is the discount factor (set to 0.8), controlling the influence of future rewards; α_q is the Q-value learning rate, which adopts a dynamic adjustment strategy, with an initial value of 0.1 to quickly update the Q-value, decreasing by 0.01 every 100 iterations, stabilizing after reaching 0.02 to avoid model oscillation in the later stages. The exploration strategy adopts the ε-core algorithm, with the formula $\epsilon = \epsilon_0 \cdot e^{-\tau/500}$, where ε_0 is the initial exploration rate (0.3), and τ is the number of iterations. As training progresses, ε gradually decreases, decreasing to $0.3 \times e^{\wedge}(-1) \approx 0.11$,after 500 iterations, and to ≈0.04 after 1000 iterations. This ensures both the effectiveness of exploring different words in the early stages and the priority recommendation of high-value words in the later stages, balancing the relationship between exploration and utilization.

3. Experimental Simulation and Result Analysis

3.1 Experimental Design

1) Experimental Subjects

500 first-year non-English major students from a university were selected for the experiment. The selection criteria strictly controlled variables: College Entrance Examination (Gaokao) English scores of 110-125 (out of 150) to exclude differences in basic knowledge caused by excessively high or low scores; the initial vocabulary test used the "College English Test Band 4 Vocabulary Table" (containing 500 core words, divided into 5 difficulty levels), with scores all falling within the 3000-3500 word range; and all students had no overseas study experience and had not attended any Band 4 preparation training courses to avoid additional learning intervention. Stratified random sampling was used to divide the students into

ISSN 2522-6398 Vol. 9, Issue 8: 121-128, DOI: 10.25236/FER.2026.090816

4 groups, with 125 students in each group [9]: Experimental Group (DF-MMFL algorithm-assisted teaching), Control Group 1 (collaborative filtering recommendation algorithm, recommending vocabulary based on student answer similarity), Control Group 2 (fixed forgetting curve recommendation algorithm, using fixed review cycles of 1 day, 3 days, and 7 days), and Control Group 3 (random recommendation algorithm, randomly pushing vocabulary based on word ID).

To verify the rationality of the grouping, a one-way ANOVA was performed on the initial vocabulary test scores of the four groups. The results showed that the mean of the experimental group was 3256±128 words, the mean of control group 1 was 3248±132 words, the mean of control group 2 was 3261±125 words, and the mean of control group 3 was 3253±130 words. The inter-group difference test showed an F-value of 0.82 and a P-value of 0.48>0.05, indicating that there was no statistically significant difference in the initial vocabulary ability among the four groups, and the influence of the basic level on the experimental results could be ruled out. During the experiment, all students were required to maintain a fixed learning frequency each week. Data were excluded if there were more than two absences. The final effective sample size was 492 (123 in the experimental group, 1124 in the control group, 2122 in the control group, and 3123 in the control group), with a sample validity rate of 98.4%.

2) Experimental Environment

The experimental server adopts a dual-node cluster architecture. The master node is configured with an Intel Xeon Gold 6330 CPU (24 cores and 48 threads), an NVIDIA A100 40GB GPU (to accelerate model training and real-time recommendation calculation), 128GB DDR4 memory, and a 2TB SSD (to store experimental data and logs). The slave nodes are configured with the same CPU and memory for load balancing, ensuring that the recommendation response latency is less than 1 second when 500 people access the server concurrently. Student terminals are ordinary PCs (CPU i5-10400/AMD Ryzen 5 5600, 8GB memory, Windows 10/11 system) or smartphones (Android 12/iOS 16 and above, screen size ≥ 5.5 inches), meeting the requirements for web platform adaptation. The algorithm model is developed based on Python 3.9, relies on TensorFlow 2.8 to build the reinforcement learning Q-learning network layer, uses Scikit-learn 1.1.3 for feature selection and normalization, Pandas 1.5.3 to process structured answer data, and Matplotlib 3.6.2 to generate visualization charts. The self-developed web learning platform adopts the Django 4.1 + Vue 3.0 architecture, and its core functions include: 1) Real-time data collection (answer accuracy, reaction time accurate to milliseconds, interactive behaviors including the number of pronunciation playbacks, example sentence viewing time, and analysis dwell time); 2) Personalized learning interface (vocabulary is displayed in the order recommended by the algorithm, and supports the collection of wrong questions); 3) Automatic scoring system (objective questions in memory tests are scored instantly, and writing questions are initially scored by an NLP semantic similarity model and then manually reviewed to ensure scoring consistency).

3) Evaluation Indicators

The experiment employed two evaluation indicators: teaching effectiveness and algorithm performance. Both were tested three times and the average was taken to reduce random error. For the teaching effectiveness indicator, the memory retention rate after one week was the accuracy rate of the test on the 7th day after vocabulary learning (question types: synonym substitution and part-of-speech conversion, difficulty coefficient 0.55), reflecting the short-term memory consolidation effect; the memory retention rate after two weeks was the accuracy rate on the 14th day, with the question types consistent with the one-week test to avoid discrepancies, reflecting the long-term memory effect; the vocabulary application accuracy rate included a cloze test adapted from a CET-4 exam (containing 20 target words) and a writing exercise on the theme of "University Life" (requiring at least 5 target words, scored 25% each for content completeness, vocabulary accuracy, grammatical correctness, and text coherence) [10]. The two scores were converted into an accuracy rate and then averaged. In the algorithm performance metrics, recommendation accuracy is the proportion of weak words recommended by the algorithm among students with an error rate ≥ 40%, further subdivided into 5 difficulty levels; model convergence speed is the number of iterations when the loss value drops below 0.1 and the fluctuation is ≤ 5% for 10 consecutive iterations, while the loss value at convergence is also recorded; average recommendation response time is the average time taken for the algorithm to generate a single learning list (20 new words + 10 review words) for a single student, calculated as the average of 500 times from server logs.

3.2 Experimental Results and Analysis

1) Comparison of Teaching Effectiveness Indicators

ISSN 2522-6398 Vol. 9, Issue 8: 121-128, DOI: 10.25236/FER.2026.090816

Table 1 summarizes the memory retention rate and vocabulary application accuracy of the four groups at 1 week and 2 weeks (including mean and standard deviation; * indicates P<0.05 compared with the experimental group, independent samples t test). All indicators of the experimental group were significantly better than those of the three control groups, and the advantage increased over time: the memory retention rate at 1 week was 82.3%±3.2%, which was 15.6, 13.2, and 28.1 percentage points higher than that of control groups 1 (66.7%±4.1%), 2 (69.1%±3.8%), and 3 (54.2%±5.3%), respectively; the retention rate at 2 weeks was 71.5%±2.8%, which was 18.2, 16.4, and 30.3 percentage points higher than those of control groups 1 (13.4%), 2 (14.0%), and 3 (13.0%), respectively. The core reason is the dynamic forgetting model of the DF-MMFL algorithm: the forgetting coefficient is adjusted according to the real-time answer accuracy and reaction time, and the vocabulary with "slow answering and low accuracy" is reviewed 1-2 days in advance, while the control group 2 misses the best review window due to the fixed 7-day cycle, resulting in a faster decline in long-term retention rate; the 2-week retention rate of Level 5 vocabulary in the experimental group was 65.2%±3.5%, which was 16.9 percentage points lower than that of control group 1 (48.3%±4.2%), verifying the protective effect of the model on the memorization of high-difficulty vocabulary. In vocabulary application, the experimental group achieved an overall accuracy rate of 68.9% ± 2.7% (cloze test 70.2% ± 2.5%, writing 67.6% ± 3.1%), representing improvements of 11.9 and 12.7 percentage points respectively compared to the best control group 1 (cloze test 58.3% ± 3.6%, writing 54.9% ± 4.2%). This improvement was attributed to the algorithm's reward function, which weighted "long-term memory retention rate" by 0.7, recommending frequently used vocabulary. In contrast, control group 1 relied solely on "group similarity errors," making it difficult for students to apply the vocabulary flexibly.

Table 1 Comparison of Teaching Effectiveness Indicators among the Four Groups (%)

Group	Memory retention rate after 1 week	Memory retention rate after 2 weeks	Vocabulary Application Accuracy - Cloze Test	Vocabulary usage accuracy - Writing	Overall accuracy in vocabulary application	2-week retention rate decline (percentage points)	Level 5 vocabulary retention rate in 2 weeks
Experimental group (DF-MMFL)	82.3±3.2	71.5±2.8	70.2±2.5	67.6±3.1	68.9±2.7	10.8	65.2±3.5
Control group 1 (collaborative filtering)	66.7±4.1*	53.3±3.5*	58.3±3.6*	54.9±4.2*	56.6±3.8*	13.4	48.3±4.2*
Control group 1 (collaborative filtering)	69.1±3.8*	55.1±4.0*	59.5±3.9*	56.1±3.7*	57.8±3.5*	14	49.7±3.9*
Control group 3 (randomized)	54.2±5.3*	41.2±4.8*	46.8±4.5*	44.2±5.1*	45.5±4.7*	13	38.5±4.5*

Figure 1 is a line graph comparing the memory retention rates of the four groups at 1 week and 2 weeks. The horizontal axis represents the "experimental group" (experimental group and control groups 1-3), and the vertical axis represents the "memory retention rate (%)". The two lines represent the data at 1 week and 2 weeks, respectively. The curves clearly show that the two lines of the experimental group are always at the top, and the gap is the smallest (only 10.8 percentage points), indicating that its long-term memory stability is the best. Although the initial retention rate of control group 2 is higher than that of control group 1, the margin of the lead narrows after 2 weeks, confirming that the fixed forgetting cycle is not well adapted to long-term memory. Figure 2 is a bar chart of the detailed indicators of vocabulary application accuracy for the four groups. The horizontal axis represents the "experimental group", and the vertical axis represents the "accuracy rate (%)". Each group includes two bars: "cloze test" and "writing". Both bars of the experimental group are higher than those of the other groups, and the difference in writing accuracy between the experimental group and the control group is slightly larger than that in cloze test, highlighting the effect of the algorithm on improving vocabulary application ability.

ISSN 2522-6398 Vol. 9, Issue 8: 121-128, DOI: 10.25236/FER.2026.090816

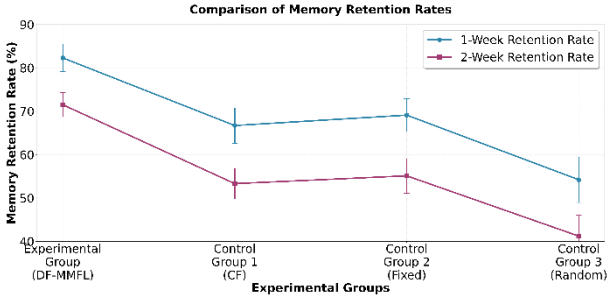

Figure 1. Line chart comparing memory retention rates at 1 week and 2 weeks for 4 groups.

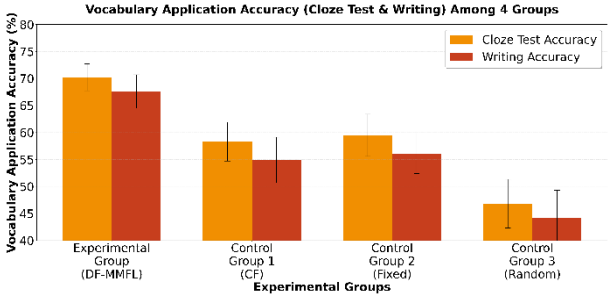

Figure 2. Bar chart of detailed indicators for the accuracy of word application in 4 groups.

2) Comparison of algorithm performance indicators

Table 2 compares the performance differences of the 4 groups of algorithms from three dimensions: recommendation accuracy (divided into 5 difficulty levels), model convergence speed, and average recommendation response time. The experimental group achieved an overall recommendation accuracy of 90.1%±2.1%, significantly higher than control group 1 (81.3%±3.4%), control group 2 (78.5%±3.7%), and control group 3 (52.7%±4.9%). It also demonstrated balanced high adaptability across different difficulty levels: the recommendation accuracy for Level 1 basic words was 94.5%±1.8%, with a small difference (4.3 percentage points) compared to control group 1 (90.2%±2.3%), indicating limited differences between groups due to high mastery of basic words; however, the recommendation accuracy for Level 5 difficult words reached 85.7%±2.9%, significantly higher than control group 1 (72.1%±3.8%) and control group 2 (69.8%±4.1%) by 13.6 and 15.9 percentage points respectively, indicating a significantly widening gap. Regarding model convergence speed, the experimental group only required 120 convergence iterations, a reduction of 33.3% and 20% compared to control group 1 (180 iterations) and control group 2 (150 iterations), respectively. Furthermore, the convergence loss value (0.078±0.005) was lower than that of control group 1 (0.121±0.008) and control group 2 (0.105±0.007). This is attributed to the synergistic optimization of gradient descent and dynamic forgetting coefficients: real-time updates of the forgetting coefficients provide more accurate state feedback for reinforcement learning Q-values, accelerating Q-value convergence to the optimal solution. In contrast, control group 2's fixed forgetting parameters require more iterations to adapt to the population average state, resulting in slower convergence.

Table 2 Comparison of Performance Indicators of the Four Algorithms

Group	Overall recommendation accuracy (%)	Number of convergence iterations (n)	Loss value at convergence	Average recommended response time (seconds)
Experimental group (DF-MMFL)	90.1±2.1	120	0.078±0.005	0.32±0.03
Control group 1 (collaborative filtering)	81.3±3.4*	180	0.121±0.008	0.45±0.04
Control group 2 (fixed amnesia)	78.5±3.7*	150	0.105±0.007	0.38±0.03
Control group 3 (randomized)	52.7±4.9*	- (No training)	- (No training)	0.15±0.02

Figure 3 shows the model convergence curves of the three algorithms (control group 3 had no training process). The horizontal axis represents the number of iterations (0, 50, 100, 150, 200), and the vertical axis represents the loss value (0-0.4). The curves show that the experimental group's loss value decreased rapidly from the initial 0.32, dropping below 0.1 after 100 iterations, and stabilizing at around 0.08 after

ISSN 2522-6398 Vol. 9, Issue 8: 121-128, DOI: 10.25236/FER.2026.090816

120 iterations. Control groups 1 and 2, on the other hand, only dropped below 0.1 after 160 and 130 iterations, respectively, and their final loss values were higher, verifying the dual advantages of the experimental group in terms of convergence speed and accuracy. Figure 4 shows a bar chart comparing the recommendation accuracy of four groups of words with different difficulty levels. The horizontal axis represents the "vocabulary difficulty level (levels 1-5)" and the vertical axis represents the "recommendation accuracy (%)". Each group contains 5 bars. The experimental group's 5 bars are all higher than those of the other groups. The higher the difficulty level, the greater the gap with the control group (level 5 has the largest gap, ≈13-35 percentage points). This intuitively reflects the DF-MMFL algorithm's ability to provide personalized recommendations for high-difficulty words and provides technical support for improving the depth of vocabulary teaching.

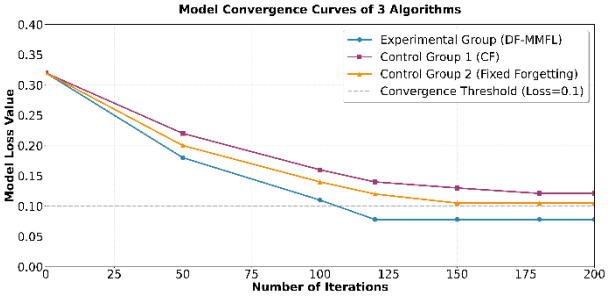

Figure 3. Convergence curves of the three algorithms.

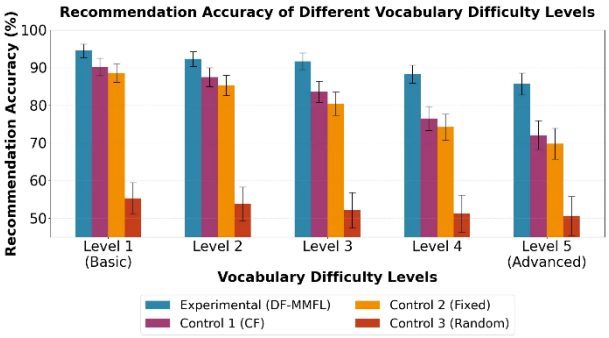

Figure 4. Comparison of recommendation accuracy for four groups of words with different difficulty levels.

4. Conclusion

The DF-MMFL algorithm designed in this paper effectively solves the core pain points of traditional vocabulary teaching and existing machine learning solutions. Through the synergy of multimodal feature fusion, dynamic forgetting coefficient update, and reinforcement learning recommendation, it achieves a technological breakthrough in personalized vocabulary teaching. Experimental verification shows that the algorithm is significantly better than the control group in both teaching effectiveness and algorithm performance: In terms of teaching effectiveness, the experimental group's memory retention rate at 1 week and 2 weeks was 13.2 and 16.4 percentage points higher than the best control group, respectively, and the vocabulary application accuracy was improved by 11.9-12.7 percentage points, with outstanding memory protection effect for high-difficulty words; in terms of algorithm performance, the total recommendation accuracy reached 90.1%, which was 37.4 percentage points higher than the highest control group, the number of model convergence iterations was reduced by 20%-33.3%, and the response latency met the concurrency requirements. The study still has limitations. The experimental subjects were only first-year students who were not majoring in English, and the scenario focused on basic vocabulary learning, without involving adaptation to different academic levels and language skills. In the future, the experimental group can be expanded to include junior and senior high school students and professional English learners. By combining multimodal data such as eye movement and speech to optimize feature dimensions, the algorithm's adaptability in complex learning scenarios can be further improved, providing more comprehensive technical support for intelligent vocabulary teaching.

ISSN 2522-6398 Vol. 9, Issue 8: 121-128, DOI: 10.25236/FER.2026.090816

References

[1] Ye Junyao, Su Jingyong, Wang Yaowei, & Xu Yong. A long-term memory prediction model for language learning based on LSTM. Journal of Chinese Information Processing, vol. 36, pp. 133-138, December 2022.
[2] Wang Aiqing. Intelligent evaluation technology: data-driven analysis in personalized teaching. Artificial Intelligence Education Research, vol. 1, pp. 41-56, January 2025.
[3] Hao Yongbin, Zhou Lanjiang, & Liu Chang. An end-to-end multi-task Lao word segmentation method based on LSTM. Journal of Chinese Information Processing, vol. 35, pp. 75-81, September 2021.
[4] Yin Zhige, & Lu Jianfeng. Research on chemical named entity recognition with fusion of radical features. Computer and Digital Engineering, vol. 51, pp. 809-816, April 2023.
[5] Chen Enhong, Liu Qi, Wang Shijin, Huang Zhenya, Su Yu, Ding Peng, ... & Zhu Bo. Key technologies and applications of adaptive learning for intelligent education. Journal of Intelligent Systems, vol. 16, pp. 886-898, May 2021.
[6] Huang Lihe, Qu Huiyu, & Yang Jingjing. Automatic text analysis of elderly discourse by computers: progress and prospects. Language Strategy Research, vol. 7, pp. 88-96, March 2022.
[7] Huang Zhenhua, Yang Shunzhi, Lin Wei, Ni Juan, Sun Shengli, Chen Yunwen, & Tang Yong. A review of knowledge distillation research. Chinese Journal of Computers, vol. 45, pp. 624-653, March 2022.
[8] Wu Yajuan, Niu Jiakui, Xie Hongtao, & Ma Ning. Named entity recognition in well control domain based on dictionary and character vector fusion. Journal of Hainan University (Natural Science Edition, Chinese and English), vol. 40, pp. 125-133, February 2022.
[9] He Xuejian, Chen Anqi, Guo Zhiqiang, Wang Zhiru, & Chen Qun. A Chinese Question Matching Method Based on Progressive Machine Learning. Journal of Engineering Science, vol. 47, pp. 79-90, January 2025.
[10] Liu Ying, Guo Yingying, Fang Jie, Fan Jiulun, Hao Yu, & Liu Jiming. A Review of Deep Learning Cross-Modal Image and Text Retrieval Research. Computer Science and Exploration, vol. 16, pp. 489-511, March 2022.