

PREDICTING FOREIGN VOCABULARY RECALL

Jordan West (Supervised by Dr. Mark Schulz)

INTRODUCTION

The human mind is mysterious, but with modern computer algorithms it is possible to analyse the workings of the mind in many novel ways. This study aims to predict memory recall of foreign vocabulary at any given point in time. This could potentially lead to uses in education to predict future test scores, identify struggling students and isolate difficult material.

METHODOLOGY

To predict vocabulary recall, software was built to record student reviews of vocabulary in a flash-card like interface, where students are asked to recall a word, view the answer then rate their recall for accuracy.

In order to space out reviews, a “Spaced Repetition” algorithm is employed – a method of scheduling facts for review exploiting the psychological “Spacing Effect” of memory^[1]. Knowing that each review of a fact solidifies the fact in the person’s memory, one can review the fact at increasingly longer periods to retain that fact. This is actually beneficial when compared with massed repetitions in immediate succession.

As a useful side-effect, spaced repetition algorithms must keep a database with several variables pertaining to each fact. This database essentially becomes a model of the user’s knowledge of all facts.

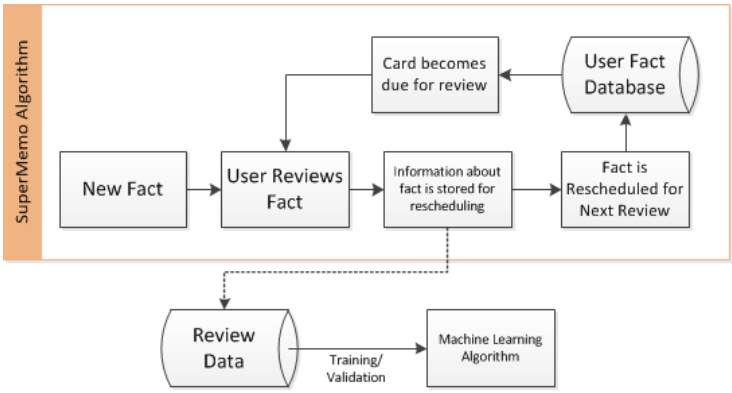


Figure 1: Spaced repetition algorithm as used in this project to gather data on reviews which is later used in machine learning algorithms.

The software built implements the popular SuperMemo^[2] spaced repetition scheduling algorithm for the nearly 100 students in JAPN1023 at the University of Queensland to review vocabulary while recording every review. The software was populated with 240 Japanese words from the course textbook along with associated English meanings.

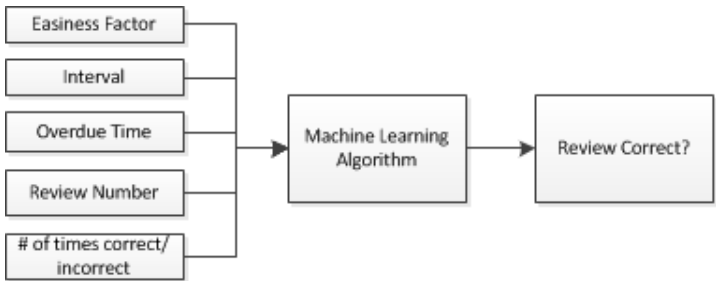


Figure 2: Review data used for training and validation of a machine learning model. The goal is to predict whether a user will correctly recall a given word using only these variables

RESULTS

As of week 12, 28 users have made over 7,500 total reviews with the software.

The recorded review data was grouped to produce forgetting curves as shown in Figure 3.

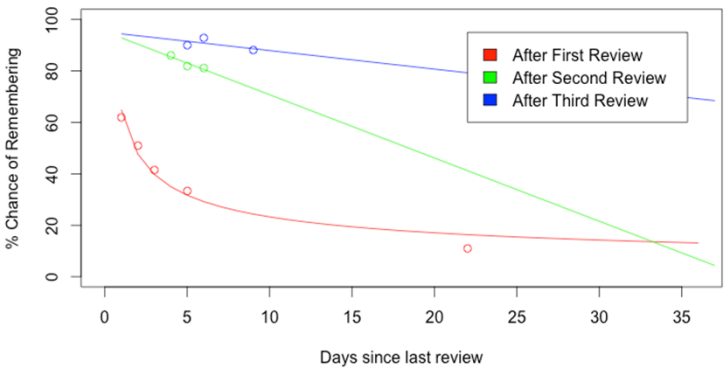


Figure 3: Forgetting curves fitted to experimental data points. Although data is limited, these curves suggest that chance of recall remains higher for longer after more reviews of a word as expected. Data points with less than 50 reviews were removed and considered too noisy.

Training various machine learning algorithms on a subset of the review data and validation on the remainder yielded accuracies shown in Table 1.

Algorithm	Training Accuracy (%)	Validation Accuracy (%)
SVM (Linear Kernel)	68.7	68.4
SVM (Radial Kernel)	71.3	70.1
Neural Net work (5 hidden units)	71.2	70.1

Table 1: Accuracy of machine learning algorithms at predicting whether a user will correctly recall a word. Values averaged over 5 runs with a 60% training set size

DISCUSSION

The experimental forgetting curves show a close similarity in shape to the forgetting curves proposed by

Ebbinghaus^[3] and may support the theory that chance of recall improves with each review, although more review data is needed to further define these curves.

The accuracy of the machine learning algorithms at predicting whether a word will be correctly recalled is significant at 70%.

Unfortunately the number of users actively reviewing was relatively low, thus the data tended to be biased towards a few individuals. Additionally a large portion of the reviews were newly introduced words which contain no useful data and were filtered out.

Future work would likely run over a longer time span with more participants in order to gather much more review data. Other spaced repetition algorithms might also be considered.

REFERENCES

[1] T. C. Toppino and L. C. Bloom, “The spacing effect, free recall, and two-process theory: A closer look.,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 28, no. 3, pp. 437–444, 2002.

[2] P. Wozniak, “Optimization of learning,” University of Technology in Poznan, 1990.

[3] H. Ebbinghaus, “Memory; a contribution to experimental psychology”, 1913.



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

School of Information Technology & Electrical Engineering

INNOVATION EXPO