HanziLiang: Modeling the Learning and Forgetting of Written Chinese Characters

Will Tsui
Advisor: Dana Angluin
Yale University
will.tsui@yale.edu

Abstract

This paper describes HanziLiang, a web application designed to help English-speaking learners of Chinese memorize written Chinese characters more efficiently. Using models of memory retention in humans, HanziLiang seeks to accomplish this by automatically identifying characters that a learner will likely fail to retain. We discuss the ACT-R model of learning and forgetting and how to extend it to the domain of Chinese characters. We then discuss the integration of our extended ACT-R model into HanziLiang, the challenge of using this model to make predictions based on actual recall test performance data, and future directions for developing computer-based study aids that enable more efficient memorization and retention of Chinese characters.

1 Introduction

The Chinese language uses a system of writing in which words are represented by unique logographs instead of by simpler phonetic characters, as in English. Each pictograph can range from one to tens of strokes, so for a student of Chinese who must learn 3,000 - 5,000 Chinese characters, called "hanzi", in order to read a Chinese newspaper, the task of learning and retaining all of these written forms is daunting and time-intensive.

HanziLiang is a web application designed to help students studying modern Mandarin Chinese learn hanzi with greater efficiency. HanziLiang provides students with a complete database of hanzi from which to test their knowledge. Additionally, HanziLiang aims to use a student's test performance data to generate estimates of a student's level of retention of characters over time. By modeling a student's level of retention, a student can easily determine which characters are being forgotten most quickly and thus which characters are deserving of more immediate study. This enables the student to skip review of characters that are well-retained in the student's memory and focus on characters whose study will yield the most benefit.

2 Related Research

For over a century there has been much research focused on understanding the processes of learning and forgetting, but still no consensus has been established as to what mechanisms drive these effects. There are several theories that try to explain how practice improves recall and how forgetting occurs over time. One theory proposes that elements of the context in which something is learned are encoded along with the memory record, so that the overlap between the learning context and the test context is not sufficiently large, recall failure will occur (Anderson, 2000). Another branch of thinking, called accessibility theories, argues that the memory strength contribution of a presentation depends on how accessible the memory is: less accessible memories result in more processing therefore greater retention in the long term (Cuddy & Jacoby, 1981).

Despite the lack of consensus in the underlying processes, there does seem to be consensus that learning and forgetting are governed by power functions, exponential functions, or some combination of the two (Anderson, 2000). In particular, it has been argued that because many biological processes can be described by exponential decay functions, it is plausible for forgetting, as an
aggregate of multiple exponential neural processes, to be modeled by a power function. (Pavlik & Anderson 2005).

Beyond modeling retention of Chinese characters in memory, other approaches to making Chinese character learning more efficient have been explored. Research has shown that the use of mnemonics, in particular self-generated mnemonics, is an effective means of improving recall of a written Chinese form (Kuo & Hooper, 2004; Heisig, 2001). In addition, because different Chinese characters often share similar components, applications have been developed to discover these similarities to help aid the student’s learning (Zucker & Mathieu, 1993; Lam et. al., 2001).

3 Application Design

HanziLiang was designed to complement CHNS 115, an elementary Mandarin Chinese course offered at Yale University. New characters introduced to a student in class each day were also automatically introduced to the student in HanziLiang the same day.

3.1 Main Screen

Once logged into the application, a student is presented with a graph displaying all characters he or she has taken responsibility for learning. By the end of CHNS 115, this graph displayed approximately 550 characters. Using the menu at left, the student can choose to study the characters from one or more chapters of the CHNS 115 character textbook, Character Text for Colloquial Chinese.

3.2 Test Interface

Once the student has selected characters, the student is presented a test trial for each character, one at a time in random order. In the character prompt screen, the character’s pronunciation, the character’s English meaning, and the pronunciation of associated compound words are displayed. The student is asked to write the Chinese written form on his or her own paper. When finished, he or she can either click “Incorrect” or “Correct” to indicate whether or not the character was written successfully, or, if the student is not sure, he or she can click the question mark to display the correctly rendered character.
To allow students to quickly display the characters and rate the correctness of their character recall, keyboard shortcuts were implemented using the arrow keys. If the student indicates that they recalled the character correctly, then the character is removed from the active character queue. If the character is recalled incorrectly, the character is appended to the active character queue and will reappear once all the selected characters have been tested. Once the student has correctly recalled each character in the active character queue, he or she can return to the main screen.

4 Implementation

HanziLiang’s main interface was developed as an Adobe Flash 8 application. The application receives Chinese character data as XML and transmits correctness ratings, character recall latencies, and exposure times to the web server using an HTTP POST request. Times are recorded in hundredths of a second. Simplified Chinese characters are rendered in the Flash application using the freely available KaiTi font.

HanziLiang uses a SQL Server 2005 database to store the following information for each character: written Chinese form, compound words in which the character appears, English meaning, pronunciation, number of strokes, middle 50% average recall latency, and metadata relating the character to the CHNS 115 class schedule. For each test trial, the database stores the correctness rating, the recall latency, and the exposure time along with the elapsed “psychological time” (see 6.4 for description) since the student first began using HanziLiang.

The middle tier connecting the Flash front-end with the SQL Server database is an ASP.NET web application written in C#. This C# application executes stored SQL procedures and outputs XML to the Flash application.

Students were required to login to HanziLiang using Yale’s Central Authentication System to keep individual performance data separate and secure.

5 Data Collection

Because students received so much exposure to the HanziLiang characters in real CHNS 115 class exercises, significant amounts of practice occurred external to the application that, of course, couldn’t be accounted for by HanziLiang’s memory model. As a result, the estimated retention and decay of many of the CHNS 115 characters would inevitably be significantly off. To obtain high quality data affected minimally by these external influences, a small study was completed by eight volunteer students from CHNS 115 over a period of five consecutive weekdays during the semester.

5.1 Procedure

Twenty characters of varying complexity were chosen from not yet studied chapters of the CHNS 115 character textbook. A separate Flash application based on HanziLiang was created to manage the exposure of these characters to the eight participants. The participants underwent one study trial on Day 1 followed by two test trials on each Day 1 through Day 5 for a total of 10 test trials. The participants were divided up into two conditions, one condition to be exposed to the characters for 30 seconds in the study trial, and one condition to be exposed to the characters for 45 seconds in the study trial.

5.2 Results

The data collected showed much more regularity than the ad-hoc data received through the main HanziLiang application. In general, students were able to successfully recall the presented characters after a day or two. Recall latencies decreased moderately over the five days in most cases. An example series of trials with a character is depicted here:
Recall latency (seconds x 0.01) is on the vertical axis, psychological time (see 6.4 for description) elapsed since the first exposure to the character is on the horizontal axis. The graph shows the 10 points representing test trials. In the first two trials, the student was unable to correctly recall the character. In the latter eight trials, on each day the student’s second recall was faster than his or her first recall.

Students in the 45-second condition did not perform any better on the 10 test trials than students in the 30-second condition.

This higher quality study data was used to determine additional relations in the memory model (see 6.6 for more information).

6 The Extended ACT-R Model of Learning and Forgetting

The ACT-R model of learning and forgetting is part of a larger theory of cognition primarily developed by John R. Anderson of Carnegie Mellon University. The ACT-R project is a cognitive architecture aimed at simulating various cognitive tasks on the computer.

6.1 Features Overview

ACT-R was chosen as a memory model because it accounts for two obvious and observable trends in Chinese character learning and retention, in particular, a student’s improvement in recall performance with practice and decrease in recall performance without practice. Both of these effects are modeled in ACT-R by power functions (Pavlik & Anderson, 2005). In addition, the HanziLiang memory model would need to account for (1) the spacing effect, where repeated practice over greater retention intervals yields greater long-term retention, (2) interference, where study of similar memory records produces negative effects on retention, and (3) the complexity of Chinese characters—the idea that more complicated written Chinese forms will be more difficult to retain. Finally, the model should be able to incorporate actual student performance on test trials with characters in order to produce more accurate estimates of retention in the future.

6.2 Integrated Practice and Forgetting

ACT-R calculates the strength of a memory item at the current time as the sum of each of the past individual memory strengthenings (corresponding to practice events). The strength of a memory item after \( n \) presentations is its activation:

\[
m_n(t_{1...n}) = \ln\left(\sum_{i=1}^{n} t_i^{-d}\right)
\]

This equation makes the assumption that each individual past practice of a memory item provides a uniform increment in its memory strength that decays at a constant rate \( d \) over time. The practice events can be in the form of a study trial (where the student is presented the memory item) or a test trial (where a student is asked to recall a memory item and then evaluated for correctness). Each \( t_i \) corresponds to how much time has elapsed since the \( i^{th} \) practice of the memory item occurred.

HanziLiang is mainly interested in communicating to the user which characters are more likely and less likely to be recalled, so estimated activation values are converted to probabilities of recall with the ACT-R equation:

\[
p_r(m) = \frac{1}{1 + e^{-\tau - s}}
\]

Here, \( \tau \) is the threshold parameter, equivalent to the activation value at which the probability of recall is 50%. \( s \) is the measure of noise—the higher the \( s \), the greater the difference between 0% and 100% probability of recall in terms of activation.
values have become standardized across memory models with $\tau = -0.704$ and $s = 0.255$.

6.3 The Spacing Effect
Because the decay rate of a unitary memory strengthening is constant, the widely observed spacing effect, in which the best retention occurs when the study interval matches the retention interval (Anderson 2000), is not accounted for. Pavlik & Anderson (2005) extend the standard ACT-R memory model to account for the spacing effect by replacing the standard ACT-R activation equation with the following equations:

\[
(2) \quad d_i(m_{i-1}) = ce^{-m_{i-1}} + a \\
(3) \quad m_n(t_{1-n}) = \ln\left(\sum_{i=1}^{n} t_i^{-d_i}\right)
\]

In this modification, the decay rate $d_i$ now depends on the activation at the time decay occurs. Higher activation yields faster decay; lower activation yields slower decay. If the spacing between two presentations is wider, activation will decrease between the presentations and the decay will be slower following the new presentation. In the decay function, $c$ is the decay scale parameter and $a$ is the intercept of the decay function. These values vary slightly between memory models, but Pavlik & Anderson’s $c = 0.217$ and $a = 0.177$ fell in the middle of the range and were used in HanziLiang’s model.

6.4 Interference
Interference occurs when the study of a new memory item can accelerate the forgetting of a previously studied memory item (retroactive interference). Additionally, interference occurs when a previously studied memory item impedes the learning of a new memory item (proactive interference).

Two ways of modeling interference effects were devised for use in HanziLiang. In the first model, each pair of characters in the database would be assigned a similarity rating that indicated how much the two characters have in common—graphically, semantically, or phonetically. Then, for the presentation of any character, retroactive interference would occur with any other character if the pair’s similarity rating is greater than zero. This interference would take the form of accelerated decay for the duration of the presentation—the higher the similarity rating, the faster the decay during that time. Because all unique pairs of characters would be assigned some positive (if low) similarity rating, storing these similarity ratings and applying them with each character presentation would have been computationally very intensive.

A simpler model, based on an Anderson model, was devised and implemented. Because there is significantly less destructive interference from events experienced outside of the character testing environment compared to that during character testing, the passage of time elapsed outside of HanziLiang was scaled down by a factor of 40. This time outside of the testing environment no longer reflects real time and is referred to as “psychological time” in Pavlik & Anderson (2005).

6.5 Complexity of Chinese Characters

Chinese characters appearing in HanziLiang can be simple, like the character meaning “one” (left), or very complicated, like the character meaning “to take” (right). The model used for HanziLiang must take into account the degree of complexity of each character in terms of practice and retention.

To do this, based on the assumption that more strokes in a character indicate accelerated forgetting, one modification to the extended ACT-R model would be to replace the constant decay function intercept $a$ with a value dependent on the number of strokes, as in:

\[ a = 0.177 \times k \times \text{strokes} \]

where $k$ is a strokes scale parameter. Based on student data obtained through HanziLiang, it did not appear that such a clear, linear relationship exists
between decay rate and number of strokes. This modification was not implemented in HanziLiang.

6.6 Incorporating Student Performance Data

The current, extended ACT-R model can make predictions for retention of a memory item given a series of presentations of it through time, but how do we use actual performance data obtained from character test trials to better understand the student and make more accurate predictions of the student’s character retention in the future? Also, because a student may have been exposed to HanziLiang’s characters outside of the application, how can HanziLiang discover and incorporate into the model these external influences? First, we needed to find a way of interfacing our performance data, in the form of correct/incorrect ratings and recall latencies (how long it takes an individual to remember and write down characters), with the activation values and probabilities of recall in our modified ACT-R equations.

Tarnow (2007) demonstrates that there exists a linear relationship between probability of recall of a memory item and the recall latency. Given recall latencies recorded by HanziLiang, we can calculate using a linear equation the probability that the written form was recalled correctly.

**Determining Recall Latency-Probability of Recall Equation.** By applying our extended ACT-R model to real student performance data, we discovered that the estimated probabilities of recall were roughly correlated with actual recall latencies recorded by HanziLiang. For example, the graph below shows the rough linear relationship for a particular set of 10 trials that a student had with the character 究. By looking at the best fit lines of several of these graphs across students and characters, we determined the form of the Recall Latency-Probability of Recall relation to be approximately:

\[
Pr = -1/(avg) * Lat + 1.75
\]

where \(Pr\) is the probability of recall and \(Lat\) is the recall latency. \(avg\) is a statistic unique to each character that is equal to the mean of the middle 50% recall latencies recorded by HanziLiang. Middle 50% averages were used to eliminate recall latency outliers caused, for example, by students quickly...
skipping over a character without writing it or by leaving HanziLiang mid-test session. If not enough trials have been recorded for a given character, this avg value can be estimated according to the equation:

$$\text{avg} = 50.917 \times \text{number of strokes} + 310.24$$

This was the best fit line of the plot of stroke count vs. middle 50% average of characters in the HanziLiang database. $R^2 = 0.5182$.

Matching estimated activations to actual activations. When a recall latency is recorded for a student with a particular character, the probability of recall of that character can be calculated using Equation 1. Finally, the initial decay $a$ can be manipulated such that the resulting predicted activation values across all past practice events are minimally different from the calculated actual activation values. This manipulation of $a$ was not implemented in HanziLiang.

7 Results and Evaluation

7.1 Application Usage

HanziLiang’s target audience was the roughly 100 students enrolled in CHNS 115 during the Spring 2007 semester at Yale. In the three months that the application was online during the semester, over 50 students used the application, 15-20 of which used the application regularly to prepare for written exams in CHNS 115. Over 30,000 individual character trials were submitted to the database in this period of time.

Usage of the application varied as a function of the CHNS 115 schedule. Days preceding written exams saw heavy usage of HanziLiang, while days over Spring Break saw little to no usage.

Using a comment form integrated into the HanziLiang Flash application, students regularly posted positive comments about how useful they found the application to be for their studies.

7.2 Character Retention Graph

The focus of the HanziLiang’s modeling of character retention was to communicate which characters are good candidates for review and which characters are not in need of review. To surface these predictions to the student, characters in the character graph on the main screen were colored along two axes. First, the character’s opacity was directly proportional to the character’s estimated probability of recall. The darker the character, the more likely the student is to recall the character. The second axis, a coloring of each character from green to red, was to be an indication of how quickly the student’s retention of any character was decaying. Red indicates faster decay while green indicates slower decay. A red coloring should indicate to the student that his or her retention of a character is not particularly stable in the long term—that further test trials are necessary in the future to reduce the decay rate.

Because the extended ACT-R memory model used in HanziLiang determines ongoing composite decay of memory strength by looking at the component decay of each individual past memory strengthening, an aggregate decay rate is calculated using a weighted average of each component decay, where each component decay is weighed in
according to the associated memory strengthening’s contribution to the current activation. While current probability of recall seemed to be reflected in a reasonable way in the opacity of each character in the graph, the green-red decay coloring based on the aggregate decay statistic did not produce the desired result. Instead, it appears the aggregate decay rate is a good measure of the current decay of a character and, therefore, fluctuates significantly by how a character’s exposure was.

8 Conclusions and Future Work

The final iteration of HanziLiang modeled practice and forgetting with Equations 1-3 describe above, which accounted for memory strengthening due to practice, memory decay over time due to forgetting, and the spacing effect. HanziLiang also accounted for interference through its treatment of elapsed time as psychological time. With this limited model, HanziLiang is still able to generate a character graph that indicated to a helpful degree of accuracy which characters are in need of review.

However, the final iteration of HanziLiang still was not able to incorporate into its retention estimates (1) decay rates that vary by complexity of character and (2) recorded performance data from student test trials. As discussed in 6.5, it would have been feasible to manipulate the decay function parameters to yield accelerated decay for more complex characters and slower decay for less complex characters, but no obvious relation between character complexity and decay rate given the data obtained in the eight-person study. If additional studies were conducted, it might be possible to further reduce sources of noise so as to find these subtle trends in decay.

Attempts such as those described in 6.6 were made to try to incorporate the student’s real past performance data into future estimates of retention, but the variation in calculated actual activation values almost always produced implausible future activation estimates. For example, because the extended ACT-R model calculates decay rate as being proportional to activation, just one very high or very low activation value inputted could result in devastating changes for future activation estimates. It is possible that by eliminating noise and variation in the incoming actual activation values, we could better use these indicators of actual performance to make predictions.

The method of aligning activation estimates with actual activation values described in 6.6 was not implemented in the final iteration of HanziLiang because the net result of this process was often insignificant. Often, the discrepancy between actual activations and estimated activations could be reconciled by only a few hundredths using this process.

With external exposure to characters, the variability in complexity between characters, and the relationships that characters can have through shared meanings, pronunciations, and radicals, modeling retention of Chinese written forms is quite a formidable task. HanziLiang is able to generate crude estimates of retention based on the extended ACT-R model, but there is room for further refinements to be made if the existing model is extended even more.

It would also be very worthwhile to explore alternative means of making character-learning easier and more efficient. Suggestions include:

1. Visually breaking down characters into their component parts, allowing students to see which components are shared between what characters.
2. Allowing students to generate their own English mnemonics based on the English meanings of component parts.
3. Creating a more portable version of HanziLiang that could run on a mobile phone, enabling students to review characters when they have a moment of spare time during the day.
4. Allowing students to input their own memory items.
5. Changing the direction of memory association: prompting students for character pronunciation and meaning by presenting written character forms. As computers are increasingly removing the need to handwrite Chinese characters, it is becoming increasingly important to know the pronunciation of each character so that it can be converted to the written form on the computer.
9 References


