VOCABULARY LEARNING APP

蔡欣樺, 張瑄容, 盧敬元

Abstract:

This project is a app that allows users to learn TOEIC vocabulary and help them to memorize more efficiently through a review algorithm, which is related to forgetting curve. We use streamlit as front-end and postgreSQL as database.

There are four tables used in our project's database: user, learning_history, vocab and meaning. The first one stores each user's basic data, the second one stores what a user has learned, therefore, each user has a learning_history, the third and the last one store all the vocabulary information.

After finish constructing this app, we did an experiment to test the review algorithm, the result shows that using the algorithm can decrease the time a learner spend to learn a fixed number of vocabularies.

Keywords:

Forgetting Curve;

1. Introduction

TOEIC has is one of the largest English exams. Many schools demend students to take this exam to qualify graduation. However, memorizing vocabulary takes a lot of time. Therefore, this project is design for those who needs to learn vocabulary in a faster way. We design a method referring to Forgetting Curve to help user to review more efficiently.

We will show the database structure in section 2 and explain the implemented review algorithm in section 3. The simulation method and result will be shown in section 4 and section 5.

2. Algorithm

2.1. Forgetting Curve

The concept of forgetting curve is that how the remain things that you learned will decrease through times. The memory will retain shorter when learning a new word, however, it will retain longer if the user review it in the right timing.

Below is a graph showing the forgetting curve example, the horizontal axis represents the time and the vertical axis represents the remain memory.

2.2. Algorithm 1

After a user learn a new word, he or she will need to review it 3 times. We calculate the review timing through forgetting curve. When the remain memory drops under 80%, the program will give the word to user to review. In this way, users will save time from learning words that they already know.

2.2. Algorithm 2

Our app updates review time constantly according to learner's learning status. A vocabulary will occur only when it reaches the review time. If the vocabularies incorrectly answered are less than 15%, we decrease the time between reviewing. On the other hand, if the vocabularies incorrectly answered are more than 25%, we decrease or increase the time between reviewing. In this case whether we decrease or increase the time between reviewing depends on the effect of last update. If last update we increased the time between reviewing, yet caused negative effect on the percentage of vocabularies incorrectly answered, we will calibrate our last update and decrease the time between reviewing since the cause of high incorrect answers percentage can derive from both too short or too long time range between reviewing.

3. Database

3.1. User Table

	username [PK] characte	name_1 character varying	email character varying	password character varying	gender character varying	birth date	prepare_time integer	level character var	user_id integer
1	danny	Danny Lin	789@gmail.com	$2b$12$RBZTh.pxW2...	male	1996-05-18	70	3	5
2	jennie	Jennie Lin	123@gmail.com	$2b$12$RBZTh.pxW2...	female	1998-10-22	80	1	1
3	jsmith	John Smith	jsmith@gmail.com	$2b$12$RBZTh.pxW2...	male	1998-10-28	60	2	2
4	lizzie	Lizzie Tsai	456@gmail.com	$2b$12$RBZTh.pxW2...	female	1998-10-23	60	2	4
5	rbriggs	Rebecca Briggs	rbriggs@gmail.com	$2b$12$ZrI76M3Wm...	female	1996-05-03	100	4	3

This table stores users' basic data. The primary key is "user_id". When they registered to this website, they need to give these information.

3.2. Learning History Table

	vocab_id integer	mean_id integer	user_id integer	time timestamp without time zone	is_right boolean	learn_id [PK] integer	learn_date date
1	1	1	1	2023-05-28 00:00:00	true	1	2023-05-28
2	5	5	1	2023-05-31 12:05:24.850776	true	47	2023-05-31
3	2	2	1	2023-05-31 12:05:43.596461	false	48	2023-05-31
4	3	3	1	2023-05-31 12:06:02.309705	true	49	2023-05-31
5	4	4	1	2023-05-31 12:08:03.856024	true	50	2023-05-31
6	5	5	2	2023-06-01 13:45:17.3581	true	51	2023-06-01
7	1	1	2	2023-06-01 13:45:45.595805	false	52	2023-06-01

This table stores what a user has learned. The primary key is “learn_id”. “vocab_id” is the word that has been learned, it is Vocab table’s foreign key. “mean_id” is the word’s meaning, it is Meaning table’s foreign key. “user_id” means which user learned it, it is User table’s foreign key. “learn_date” and “time” represent when it was learned. “is_right” means whether the user answered it right or not.

3.3. Vocab Table

	vocab_id [PK] integer	name_1 character varying	level_1 integer
1	1	devotion	2
2	2	reasonable	1
3	3	conviction	2
4	4	domain	2
5	5	executive	2

This table stores all the vocabulary. The primary key is “vocab_id”. “name” represent the vocabulary itself(eg. “cat”...). “level” represents the difficulty of this word.

3.4. Meaning Table

	vocab_id [PK] integer	meaning [PK] character varying	part_of_speech character varying	example character varying
1	1	忠誠，忠實；奉獻	noun	He inspired respect and devotion from his pupils. 他贏得了學生們的尊敬和愛戴。
2	2	合情理的；講道理的；公平的	adjective	If you tell him what happened, I'm sure he'll understand - he's a reasonable man. 如果你把
3	3	定罪；判罪	noun	Since it was her first conviction for stealing, she was given a less severe sentence. 由於她
4	4	領域，領地	noun	She treated the business as her private domain. 她把公司看作是自己的私人勢力範圍。
5	5	（尤指商業機構中的）行政主管，經理	noun	She is now a senior executive, having worked her way up through the company. 她現在是

This table stores all the meaning for vocabulary. The primary key are “vocab_id” and “meaning”, it matches the word and its meaning. “part_of_speech” represent whether the word is a noun, adjective or verb. “example” is a sentence using the word.

4. Simulation

In order to conduct experiment, we use computer to simulate learners’ behaviors to test our Algorithm 2, to prove the effectiveness of this algorithm for assisting learners to remember vocabularies in long terms while spending as little time as possible. Thus, in this section, we define what behaviors we aim to simulate and how we simulate it.

4.1. Learning

We decide whether a word is learned through the correctness of a user answering the question. If the user answered the known vocabulary correctly twice, we marked it as “learned”. If the user practiced the newly learned vocabulary correctly three times and answered it correctly twice, we also marked it as “learned”. Here, we define a word as “known” or “newly learned” according to whether the learner answer the vocabulary correctly the first time. If the word is correctly answered at the first time, we marked the word as “known”, otherwise, we marked it as “newly learned”.

4.2. Effort

A learner use 1 unit of effort to learn a “known vocabulary”, 3 unit of effort to learn a “newly learned vocabulary”, 0.5 unit of effort to review a “remembered vocabulary” and 1.5 unit of effort to review a “forgotten vocabulary”.

4.3. Forget

A learner forgets according to his forgetting curve which is not known by our learning app. Besides, if a learner review a vocabulary before the suitable reviewing time, we define the review as a meaningless review. In this case, the forgetting rate of this word will only decrease slightly. On the other hand, if a learner review a vocabulary after the suitable reviewing time, the forgetting rate of this word will remain on the steeper forgetting curve and thereby leads to higher forgetting rate, which implies higher possibility that the learner will incorrectly answer this word next time.

5. Experiment

5.1. Functions

For our experiments, we use five functions and seven initialized variables to run the entire simulation. The five functions and their functionalities are listed as follow:

a. recite_new_vocab(.....): This function is used to simulate learner reciting new vocabularies.

b. create_record_after_first_day(.....): This function is used to simulate learner reciting and reviewing vocabularies after the first day.

c. forget(.....): This function is used to simulate learner forgetting vocabularies according to his forgetting curve.

d. update_review_time(.....): This function is used to update the our app's review time using Algorithm 2.

e. check_if_complete(.....): This function is used to check if the learner finish learning a fixed number of vocabularies. If the learner has finished learning, the function returns "True". Otherwise, it returns "False".

5.2. Variables

After defining these functions, we also need initialize a few variables listed below:

a. vocab_num: This variable is used to set the number of vocabularies the learner aims to learner. Here, we initialized it to 300.

b. review_time: This variable is a list that set the app's review time. We initialize this list to [0, 1, 3, 6], which means the learner will review a vocabulary the first, the third and the sixth day after the learner learned a vocabulary. This list will update every time the function update_review_time(……) being called.

c. forget_curve: This variable is a list that set the learner's most suitable review time. The learner will forget 20% of the learned vocabularies every time it reaches a suitable review time in this list. For example, if we set this list to [0, 2, 4, 8], it means the learner will forget 20% of leaned vocabularies the second day after he learned the vocabularies, and forget 20% of reviewed vocabularies two days after his first review, and forget 20% of reviewed vocabularies four days after his second review.

d. last_update: This variable is a list that record how the review_time variable is updated last. [0, 1, -1, 0] means the second index of the review_time list increase last time, the third index of the review_time list decrease last time and the forth index remain the same after the update. This variable along with last_per variable are used to check the effect of last update.

e. last_per: This variable is a list that record the percentage of incorrectly answered vocabularies last time. This variable along with last_update variable are used to check the effect of last update. If the percentage of incorrectly answered vocabularies this time is larger than last time and according to last_update, the second index had been increased. We know increasing the time gap between reviewing has negative effect on learner's learning performance. Thus, review_time will be calibrate back to its origin state.

f. quota: This variable is used to simulate the unit of effort a learner used to learn vocabularies each day. Here, we initialize it to 40.

g. day: This variable is used to record the number of day a learner use to learn the fixed number of vocabularies (vocab_num).

5.3. Method

The functions will be called in the following manner until the fixed number of vocabularies had been learned, and the number of day spent will be returned.

```
1  initilaize variables
2  recite_new_vocab(.....)
3  while incomplete:
4      forget(.....)
5      create_record_after_first_day(.....)
6      update_review_time(.....)
7      check_if_complete(.....)
8      day += 1
```

5.4. Experiment Result

From our experiment, we can see that using our Algorithm 2 to dynamically update our review time can decrease the time required for learner to learn all 300 vocabularies for about 7-9 days.

User's Forgetting Curve	Fixed Review Time	Dynamic Review Time
[0, 2, 4, 8]	50	40
[0, 1, 3, 6]	53	41
[0, 2, 5, 8]	49	42
[0, 1, 2, 5]	56	45
[0, 1, 3, 8]	51	44
[0, 1, 3, 7]	55	49
[0, 1, 3, 8]	53	46
[0, 2, 4, 6]	52	47
[0, 2, 5, 9]	48	41