AN INTELLECTUAL APPROACH TO DESIGN PERSONAL STUDY PLAN VIA MACHINE LEARNING

Shiyuan Zhang¹, Evan Gunnell², Marisabel Chang², Yu Sun²

¹Arnold O. Beckman High School, Irvine, CA 92620, USA ²Department of Computer Science, California State Polytechnic University, Pomona, USA

ABSTRACT

As more students are required to have standardized test scores to enter higher education, developing vocabulary becomes essential for achieving ideal scores. Each individual has his or her own study style that maximizes the efficiency, and there are various approaches to memorize. However, it is difficult to find a specific learning method that fits the best to a person. This paper designs a tool to customize personal study plans based on clients' different habits including difficulty distribution, difficulty order of learning words, and the types of vocabulary. We applied our application to educational software and conducted a quantitative evaluation of the approach via three types of machine learning models. By calculating cross-validation scores, we evaluated the accuracy of each model and discovered the best model that returns the most accurate predictions. The results reveal that linear regression has the highest cross validation score, and it can provide the most efficient personal study plans.

KEYWORDS

Machine learning, study plan, vocabulary.

1. INTRODUCTION

Vocabulary is essential in reading, writing and essay passages of standardized tests such as the Scholastic Assessment Test (SAT) and American College Test (ACT). Higher-level vocabulary may appear in reading section since diverse variety of authors from historical to modern period wrote these reading passages. Similarly, questions of sophisticated vocabulary may appear in the writing passage; though it would not ask for the synonym of a word instead it would ask for the proper diction of the word. Building vocabulary will be efficient in improving scores, and learning new terms requires memorization skills.

Knowing the definition of complex words help students to answer the section of sentence completion and word definition correctly. However, reading, writing and essay sections cannot be completed without using memorization.[13] Since memorization is a common and vital process of developing a sophisticated level of vocabulary, learners spend many hours study to retain and recall vocabulary for their standardized test; but most of the time memorizing the definition of vocabulary can be very challenge. Although learners may know the definition of a new complex words, they may forget the definition a month later. Thus, memorizing the definition of a vocabulary may not be a satisfactory choice.

Computer Science & Information Technology (CS & IT)

Open Problem: Although there are many strategies to learn English vocabulary, these memorization strategies cannot bring quality result. [9][11] Learners may use books that provide the vocabulary that they must know before the standardized test. [10] However, they need to spend a lot of time memorizing the definition. Also, these books may be a old edition and contain a few words that may appear in the standardized test. Other technique is using word-search-puzzle games for improving vocabulary, but this game may not word for everybody since different learners think in different way. [12] Other strategies is using flashcards to recall the definition of words. This technique cannot be too sophisticated since learner need to memorize the definition. Furthermore, they would spend a lot of time when they need to search a word in multiple stacks of flashcard.

Solution: A pictorial method and automated personal study plan using machine learning and mobile computing. In this paper, we presented a new approach for recalling and retaining vocabulary. We have developed a mobile application that enable learners to customize their vocabulary by adding picture that assist them to remember the meaning of a vocabulary. Furthermore, this application can enable learner to input their desirable ACT/SAT score for the standardized test to obtain a personal study plan. In order to get the personal study plan learners must take an assessment. Once a learner finishes the assessment, the system would predict the group of words that is appropriate it for him/her.

In three experiment scenarios, we demonstrated the accuracy of different machine learning algorithms to predict our vocabulary. First, we showed the accuracy of Support Vector Machine (SVM), Linear regression and polynomial regression algorithms to predict difficulty order of learning words. Second, we demonstrated the accuracy of machine learning models to predict Distribution in Terms of Problem Difficulty. Finally, the accuracy of the models to predict Optimal Typing Distribution for Studying. In all case, Linear regression model has a better performance than other two training model. We obtained that the linear regression algorithm performs better in Distribution in Terms of Problem Difficulty model with an accuracy of 41%.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5. Finally, Section 6 gives the conclusion remarks, as well as pointing out the future work of this project.

2. CHALLENGES

48

Nowadays, machine learning has been used to develop mobile application to make people life easier.[17][18] Although machine learning can analyze and interpret pattern of structure of a data to learn, reason and solve an issue, we would encounter some challenges during the training process. We will use machine learning to provide a learning plan which can be used during the SAT/ACT study process. In our project, there are a few limitations at the time to train different machine learning models.

2.1. Challenge 1 - Deciding student study methods which can fit for most of learners.

Using an appropriated study design may impact learners' memorization. Many learners memorize definition of words by reading word definition, using flashcards, and playing games. Others may group their vocabulary by its type, level of difficulty, alphabetic order etc. However, if learners

use the inappropriate study method, they may forget the meaning of the words after a period. Thus, choosing the correct study plan methods is important to improve learners' vocabulary. In this project, we plan to use existing survey to analyze and obtain the better learning vocabulary methods that suit most of the learners.

2.2. Challenge 2 – Effectively design a study plan

Designing a study plan may be difficult if a learner does not know what to expect and what study methods are more efficient than others. Learners may ask others to share their study methods and try them. Nerveless, trying the learning methods may be costly and inefficient since learners would spend a lot of time by trying them out and not getting satisfactory result. We proposed an intelligence approach using different machine learning algorithm to elaborate an appropriate learning plan. In this approach, learners would get a study plan based on their performance of an assessment and their desirable SAT/ACT score.

2.3. Challenge 3 – Training a machine learning model with small dataset

The volume of the dataset is one of the most important portions of training a model. When a dataset is small the training model might not be very accurate to predict a specific feature. Thus, the accuracy of a training model might be a little lower than we expected. We plan to use a dataset that contains over 500 data to investigate the accuracy of our training model. In our experiment, we will use three different machine learning algorithms which are SVM, linear regression and polynomial regression

3. SOLUTION

3.1. Machine Learning Model and Prediction

We used open source machine learning library calls scikit-learn to predict the difficulty distribution, difficulty order of learning words, and the types of vocabulary in our model [1]. Scikit-learn provides state-of-the-art machine learning algorithms for supervised and unsupervised problems. We utilized 3 different machine learning algorithms and compare them to obtain the most accurate approach to our problem. [2]In our experiment, we used SVM, linear regression and polynomial regression algorithm. Linear regression is a machine learning algorithm in which two variables is linearly correlated [3]. Polynomial regression is also considering a linear regression with higher order term [4]. SVM is a supervised machine learning algorithm that is utilized for classification, outliner detection and regression.

3.2. Mobile Application

As shown in Figure 1, our app, Memelish, contains a home screen that provides 4 different options which users can choose during their learning process. The button in top left is a search bar, in which the user can type words and search for definitions and memes of the words. The button on top right can direct the user to the personal settings page after being clicked. It contains a banner on the top area which can tap on. After tapping it, the client will be directed to the latest unfinished study set in history. If there is no unfinished set in history, the banner will be titled with "Start a new set?" and automatically suggest a new set based according to the client's personal study plan. There are four buttons in the middle area which connect to the Learn, DIY, and Me pages. The Random button will randomly generate a list to test the user's skill level.

Computer Science & Information Technology (CS & IT)



Figure 1: Home page screen

In order to create and set up the home screen buttons, we used UIBarButtonItem and UIStackView (see Figure 2). We used UIBarButtonItem to create the search and setting buttons. In this UIBarButtonItem, we passed icon image as a title and set the corresponded action that the button needs to execute when the button is pressed. For "My Set", "Create New", "Random" and "SAT/ACT", we used UIStackView. These buttons are created and set up in the stackview() function. In order to locate the button in the desirable position, we used 2 different lists that contains 2 buttons and passed it to the UIStackView.

```
Let leftButton = UIBarButtonItem(title: "\", style: .done, target:
                                                                     self.
action: #selector(handleLeftButton))
navigationItem.leftBarButtonItem = leftButton
action: #selector(handleRightButton))
navigationItem.rightBarButtonItem = rightButton
func stackview() {
let list = [MySet, CreateNew]
  MySet.heightAnchor.constraint(equalToConstant: 150).isActive = true
  MySet.widthAnchor.constraint(equalToConstant: 150).isActive = true
  CreateNew.heightAnchor.constraint(equalToConstant: 150).isActive = true
  CreateNew.widthAnchor.constraint(equalToConstant: 150).isActive = true
  let stackView1 = UIStackView(arrangedSubviews: list)
  stackView1.axis = .horizontal
  stackView1.translatesAutoresizingMaskIntoConstraints = false
  stackView1.distribution = .equalSpacing
  stackView1.spacing = 10
  view.addSubview(stackView1)
  stackView1.topAnchor.constraint(equalTo: view.topAnchor, constant:
300).isActive = true
  stackView1.leadingAnchor.constraint(equalTo: view.leadingAnchor, constant:
30).isActive = true
```



Figure 2: Code Segment to create and set up button in homescreen

Figure 3 shows the search engine mentioned in the home screen. The user needs to click the search bar and enter a word. The algorithm can find the words in the database based on user inputs. Also, it will return a list of previewed outcomes with icons and words. The user can tap on one cell of the list for details. Details include the original images, words, and definitions. There is a button on the bottom left corner for users to add the word into personal study sets.



Figure 3: Searching page contains all vocabulary in alphabetic order. (left) Searching page that contains words based on user

Input. (middle) Vocabulary page that contains word, definition and image. (right)

```
Extension SearchWord{
  Override func tableView( tableView: UITableView, didSelectRowAt
indexPath: IndexPath) {
     let explanation = ExplanationVC()
      if inSearchMode{
         explanation.memeString = searchData[indexPath.row].listNameData
         explanation.imageString=UIImage(named:
      searchData[indexPath.row].listNameImage) ?? UIImage(named: "1")!
/send image + if no image then use image1
         self.present(explanation, animated: true,
                                                    completion: nil)
      }else{
         explanation.memeString = data[indexPath.row].listNameData
         explanation.imageString = UIImage(named:
data[indexPath.row].listNameImage) ?? UIImage(named: "1")!
            self.present(explanation, animated: true, completion: nil)
   }
```

Figure 4: Algorithm uses for searching word in the search engine.

Figurer 5 is the editing page for users to make their own memes to help memorize the words. Users can click on the plus button to import their own images from their phones or to select templates from online sources. The text button allows users to add text on the image. Draw button can let users paint with various colors by using fingers. If users make a mistake while drawing, they can choose to undo the last step by clicking the undo button or to erase the drawing by tapping the eraser button. Below the tool buttons, there is a rectangular area that will demonstrate the word and its definition. On top right, there is a save button that allows users to save DIY memes to their study sets.



Figure 5: Editing page for adding memes to a vocabulary.

```
Override func draw( rect: CGRect) {
    super.draw(rect)
    guardlet context = UIGraphicsGetCurrentContext() else {return}
      //exception
    context.setStrokeColor(UIColor.systemBlue.cqColor) //control the color of
    context.setLineWidth(10) //control the width of the pen
    context.setLineCap(.round) // pen shape
    lines.forEach { (line) in
       for (i,j) in line.enumerated() {
         ifi==0{
            context.move(to: j)
          }else{
            context.addLine(to: j)
        }
     }
    context.strokePath()
//2D coordinate set
Fileprivate var lines = [[CGPoint]]()
Override func touchesBegan (_ touches: Set<UITouch>, with event: UIEvent?) {
  lines.append([CGPoint]()) //所有点连起来成线
Override func touchesMoved(_ touches: Set<UITouch>, with event: UIEvent?) {
   guardlet point = touches.first?.location(in: nil) else {return}
   guardvar endPoint = lines.popLast()else{return}
   endPoint.append(point)//pen up, append the end point when finished drawing
   lines.append(endPoint) //collect the first&last points to form a line
   setNeedsDisplay()
```

Figure 6: Code segment for creating the editing and setting up page

Figure 7 is a list of view of study sets. Users can scroll up and down to view the entire list. This page lists out all words in the set with words and meme previews. The preview images might be adjusted to a low resolution, but users are able to view the original image by clicking on the words. After tapping, users will enter a detailed view page that contains word, image, definition, and an explanation of how the word and the image relate to each other.

Computer Science & Information Technology (CS & IT)



Figure 7: Study slist page contains words and memes. (left). Vocabulary page contains word, definition and image. (right)

```
Override func tableView ( tableView: UITableView, cellForRowAt indexPath:
IndexPath) ->UITableViewCell {
   let cell = tableView.dequeueReusableCell(withIdentifier: "cell", for:
                                               indexPath) as! ListCell
   cell.NewLabel.text = vocab[indexPath.row]
   cell.backgroundColor = CCFavColor
   cell.textLabel?.font = UIFont.boldSystemFont(ofSize: 28)
   cell.textLabel?.textColor = UIColor.systemGreen
   cell.coolerImageView.image = UIImage(named: imageList[indexPath.row])
    return cell
Override func tableView( tableView: UITableView, numberOfRowsInSection
                                section: Int) -> Int {
     return vocab.count
Override func tableView( tableView: UITableView, didSelectRowAt indexPath:
IndexPath) {
       let dtvc = DetailedVC()
       dtvc.imageString = UIImage(named: imageList[indexPath.row])!
        dtvc.dataString = vocab[indexPath.row]
        self.present(dtvc, animated: true, completion: nil)
```

Figure 8: Code segment for displaying study list page

As show Figure 9, there is a personal settings page where users can edit their profiles. The gray block on top illustrates the user's profile picture. Users can change their profile pictures by tapping on the gray block and selecting an image from a local photo library. The username will be demonstrated below the profile picture. If users click on My Account button, their account

information including name, email address and total number of memorized words will appear. My collection button will direct users to a list of the study sets they collected from others. The View History button can show the recent records of study sets users viewed. Lastly, My Plan button generates a personal study plan by utilizing machine learning.[15] [16] The user only needs to take an evaluation test and enter the goal. The program will recommend study sets corresponding to the user's skill level and record performance.



Figure 9: Personal setting page



Figure 10: Code segment for adding image to the profile

4. EXPERIMENT

Memory retention for students through standardized studying is always a potential misstep for students as they attempt to figure out what system works best for them. By highlighting common areas of change, we can better test to see what changes will have a lasting impact. Using a study method dataset, we check for changes in vocabulary ordering, difficulty distribution, and problem typing.

4.1. Comparison of Vocabulary Ordering through Multiple Machine Learning Models

As each sub area has a largely different impact, we separate our experiments individually. When trying to retain new words such as those needed for the SAT, word groupings can often impact the level at which students remember. Testing for these involved gathering over 500 samples of various student study methods and using Linear Regression, support vector machine, and linear regression to predict the optimal study route.

Figure 11 illustrates the accuracy of each model's prediction. From the results, Linear Regression contains the highest predictive accuracy. The "line of best fit" regressive algorithm generates an algorithm that more accurately matches the general data of students that was created. Simply swapping out the ordering of vocab words allows for different groupings to emerge that mimic a linear relationship. By comparison, Polynomial Regression and SVM did not fare as well in this experiment. The sample data does not follow the polynomial algorithm nor can it be organized via support-vector machines.



Figure 11: Accuracy of prediction of vocabulary ordering through SVM, Linear regression and Polynomial regression

4.2. Predicting the best Distribution in Terms of Problem Difficulty

Each of the following experiments follows the same formatting of over 500 generated samples for the given area, with the accuracy test for each individual algorithm: SVM, Linear Regression, Polynomial Regression. For this specific experiment we aimed to see which algorithm could predict the best distribution in terms of problem difficulty: all very hard, all very easy, variation, etc.

As seen in Figure 12, the various models all have a marginal impact on difficulty distribution, with Linear Regression and Polynomial regression both being the standout methods. Difficulty distribution is a much more noticeable area of change as even small shifts, such as from mostly easy words to mostly difficult words, can have a big impact on future memory retention. Both linear and polynomial algorithms work quite well as, while a singular line of straight fit covers a majority of the changes, there is a large swing in difference as simply studying different words has a large impact for students.



Figure 12: Accuracy of prediction of distribution in term of problem difficulty through SVM, Linear regression and Polynomial regression

4.3. Predicting the Optimal Typing Distribution for Studying

With our final experiment we once again used a sample size of 500 generated data points for testing whether the selected machine learning models can predict the optimal typing distribution for studying. This compared the memory retention of each of our study sets with the main variance being which type of vocabulary words are grouped together: (all from the same subject, even split between related topics, random, etc.). Oftentimes, learning multiple related topics can not only increase understanding, but drastically increase how long afterwards a student can remember.

Figure 13 showcases the results of the three tests and aligns well with the previously completed experiments. Linear Regression continues to be the model with the highest accuracy. This is primarily due to our common study data and its lack of polynomial factors. The primary pieces of data in this experiment as well showcase that our gathered information has a linear relationship that is impactful during the training process. SVM continues in this experiment to be the model with the lowest resulting accuracy, going so far down as to be almost completely inconsequential.





4.4. Analysis

Overall, the experiment results support our goal of using machine learning algorithms to test study plan changes that impact student memory retention. Across all three experiments, Linear Regression consistently stands out with a much higher level of accuracy compared to other methods, peaking at 41%. Given the topic and organization of data this result makes sense as a major portion of the data follows a linear algorithmic relationship. For problem ordering and difficulty distribution in particular, linear regression was able to have a much stronger impact.

5. RELATED WORK

Nguyen, T and Khuat, T. proposed a method to learn English vocabulary in a classroom. [6] Even though the research was conducted for 2 weeks, they showed that using fun games can help Vietnamese students improve their English vocabulary. However, they suggested that vocabulary games need to be designed according to the number of students, proficiency level, cultural context, timing and learning topic. Thus, we decided to use machine learning model to personalize the study plan. As different of Nguyen, T and Khuat, T. method, our learning method does not have time and place restriction since it can be used anytime and anywhere.

Merriam, A. presented a comparison between translation and pictorial methods to recall and retain English vocabulary. [7] In his research, he concluded that there is not difference between translation and pictorial methods for recalling and retaining English vocabulary. He stated that pictorial method can produce misunderstanding since learners can interpret a picture in different way. In our mobile application, we used similar method; but as different of this method, learners can add their own picture that help them to remember vocabulary.

Azadeh, R., Mai, N., and Pesaranghader, A. conducted a research about the effect of using mobile application on learning English vocabulary [8]. Their study showed that using mobile application can bring positive impact on learners' performance since learners improved their English vocabulary. We took similar approach, but the difference is that we use customize vocabulary that can be showed in SAT or ACT test.

6. CONCLUSION AND FUTURE WORK

Vocabulary is vital to communicate our thought and perform satisfactory in standardized test such as ACT and SAT. In order to be successful is standardized test, learners use appropriately study plan methods. However, different learning plan can be suitable for some learners but not for others. Some of them may use flashcards, vocabulary games, pictorial method etc. Using flashcards and memorizing the definition may seem efficient, but learners may forget the definition after a while, so all the work would be wasted.

In this project, we proposed an intelligent study method to solve memorization issue using mobile computing and machine learning; and a memorization method to recall and retain vocabulary by using picture and memes. A mobile application has been developed to design an efficient study plan. To obtain a study plan, learners must enter the desirable ACT/SAT score and take an assessment. Once a learner finished the assessment, the system predicts a study plan that is suitable for the learner and return a study plan based on his/her performance.

Our experiments show that the linear regression model stand out in our three experiments. The highest accuracy of the linear regression model was 41% when the vocabulary is grouping based on level of difficulty of vocabulary.

As a future work, we plan to gather a group of high school students to use our mobile application during the SAT study period and observe if our mobile application improved learners' memorization.

One of the limitations that we encounter during the training process was that the size of the study method dataset used in our machine learning model was not enough to get a satisfactory accuracy. We plan to handle a survey to high school students who already took the SAT and input to our dataset. In this survey, we will ask students to pick the best study method and their SAT score.

REFERENCES

- [1] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." the Journal of machine Learning research 12 (2011): 2825-2830.
- [2] Goodfellow, Ian, Y. Bengio, and A. Courville. "Machine learning basics." Deep learning. Vol. 1. MIT press, 2016. 98-164.
- [3] John, Ralph C. St. "Applied linear regression models." (1983): 201-202.
- [4] Ostertagová, Eva. "Modelling using polynomial regression." Procedia Engineering 48 (2012): 500-506.
- [5] Soman, K. P., R. Loganathan, and V. Ajay. Machine learning with SVM and other kernel methods. PHI Learning Pvt. Ltd., 2009.
- [6] Huyen, Nguyen Thi Thanh, and Khuat Thi Thu Nga. "Learning vocabulary through games." Asian EFL Journal 5.4 (2003): 90-105.
- [7] Al Nassir, Merriam. "Meaning recall and retention: comparison between translation method and pictorial method in learning vocabulary in Saudis' school." 2000-2019-CSU Theses and Dissertations (2012).

- [8] Rezaei, Azadeh, Neo Mai, and Ahmad Pesaranghader. "The effect of mobile applications on English vocabulary acquisition." Jurnal Teknologi 68.2 (2014).
- [9] Mountain, Lee Harrison. Ten essential vocabulary strategies: Practice for success on standardized tests. Educators Publishing Service, 2004.
- [10] Nadler, Burton Jay, Jordan Nadler, and Justin Nadler. Words You Should Know In High School: 1000 Essential Words To Build Vocabulary, Improve Standardized Test Scores, And Write Successful Papers. Simon and Schuster, 2004. [11]
- [11] Lessard-Clouston, Michael. "Strategies and success in technical vocabulary learning: Students' approaches in one academic context." Indian Journal of Applied Linguistics 34.1/2 (2008): 31.
- [12] Vossoughi, Hossein, and Marzieh Zargar. "Using word-search-puzzle games for improving vocabulary knowledge of Iranian EFL learners." (2009): 79-85.
- [13] Webb, Stuart A. "The effects of pre-learning vocabulary on reading comprehension and writing." Canadian Modern Language Review 65.3 (2009): 441-470.
- [14] Kornell, Nate. "Optimising learning using flashcards: Spacing is more effective than cramming." Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition 23.9 (2009): 1297-1317.
- [15] Alpaydin, Ethem. Introduction to machine learning. MIT press, 2020.
- [16] Michie, Donald, David J. Spiegelhalter, and C. C. Taylor. "Machine learning." Neural and Statistical Classification 13.1994 (1994): 1-298.
- [17] Moran, Kevin Patrick, et al. "Machine learning-based prototyping of graphical user interfaces for mobile apps." IEEE Transactions on Software Engineering (2018).
- [18] Aggarwal, Vaneet, et al. "Prometheus: Toward quality-of-experience estimation for mobile apps from passive network measurements." Proceedings of the 15th Workshop on Mobile Computing Systems and Applications. 2014.

© 2020 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

60