

Article

Analysis of User Needs on Downloading Behavior of English Vocabulary APPs Based on Data Mining for Online Comments

Tinggui Chen ^{1,2,*} , Lijuan Peng ¹, Jianjun Yang ³ and Guodong Cong ⁴

¹ School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018, China; Cherrylijuanpeng@163.com

² Collaborative Innovation Center of Statistical Data Engineering, Technology & Application, Zhejiang Gongshang University, Hangzhou 310018, China

³ Department of Computer Science and Information Systems, University of North Georgia, Oakwood, GA 30566, USA; Jianjun.Yang@ung.edu

⁴ School of Tourism and Urban-Rural Planning, Zhejiang Gongshang University, Hangzhou 310018, China; cgd@mail.zjgsu.edu.cn

* Correspondence: ctgsimon@mail.zjgsu.edu.cn

Abstract: With highly developed social media, English learning Applications have become a new type of mobile learning resources, and online comments posted by users after using them have not only become an important source of intellectual competition for enterprises, but can also help understand customers' requirements, thereby improving product functionalities and service quality, and solve the pain points of product iteration and innovation. Based on this, this paper crawled the online user comments of three typical APPs (BaiCiZhan, MoMoBeiDanCi and BuBeiDanCi), through emotion analysis and hotspot mining technology, to obtain user requirements and then the K-means clustering method was used to analyze user requirements. Finally, quantile regression is used to find out which user needs have an impact on the downloads of English vocabulary APPs. The results show that: (1) Positive comments have a more significant impact on users' downloads behavior than negative online comments. (2) English vocabulary APPs with higher downloads, both the 5-star user ratings and the increase of emotional requirement have a negative effect on the increase in APP downloads, while the enterprise's service requirement improvement has a positive effect on the increase of APP downloads. (3) Regarding English vocabulary APPs with average or high downloads, improving the adaptability and Appearance requirements have significant negative impact on downloads. (4) The functional requirements to improve products will have a significant positive impact on the increase in downloads of English vocabulary APPs.

Keywords: English vocabulary APPs; user requirements; APP downloads; online comments; quantile regression



Citation: Chen, T.; Peng, L.; Yang, J.; Cong, G. Analysis of User Needs on Downloading Behavior of English Vocabulary APPs Based on Data Mining for Online Comments. *Mathematics* **2021**, *9*, 1341. <https://doi.org/10.3390/math9121341>

Academic Editor: Ignacio Javier Perez Galvez

Received: 12 May 2021

Accepted: 7 June 2021

Published: 9 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid development of the mobile Internet, the utilization of smart mobile devices is becoming more popular. A new learning method (mobile learning) has come into being, and it comes with more and more mobile learning APPs. Compared with traditional classroom teaching manners, mobile learning has stronger autonomy and individuality. Learners can independently arrange learning content and learning schedules according to their own learning level and learning environment [1]. Compared with English learning of other aspects, English vocabulary learning has the characteristics of high quantity of information to be learned and memorized, and "memory fragmentation, and easy to forget" [2]. Therefore, it is very suitable to carry out mobile learning. This is also the reason for the rapid development of English vocabulary APPs. In recent years, many IT companies have successively launched their own English vocabulary APPs and made money from users' downloads and utilization. However, the APPs in the current

market are facing a more common phenomenon of mutual imitation and homogeneity. For companies, it is important to know how to enhance the attractiveness of their products in the increasingly competitive market environment, with accurate and comprehensive insight into the user's evaluation feedback and needs, and implementing product development and innovation [3]. With the e-business arising in recent decade, the reviews with a user-oriented content play an important role instead of the word-of-mouth in the offline world. According to Deloitte and Touche USA LLP's data collected in a survey, more than 80% of those who read online reviews said that their purchasing decisions were directly influenced by the reviews [4]. Therefore, it is of great practical significance to study how companies can use online comments to extract information about the quality of products and services, user emotions and attitudes, and identify changes in user requirements in the changing market environment, and adjust product and service competitive strategies in a timely way.

At present, academia mainly adopts methods such as text mining, machine learning, and questionnaire surveys to extract users' views and concerns about products in online comments in order to recognize user requirements, and propose corresponding improvement suggestions accordingly. In fact, as APPs develop from unknown to widespread use, the user requirements change accordingly. Especially for APPs that are less popular, that is, with fewer downloads, user requirements may focus on basic performance, while for APPs that are more popular, that is, with higher downloads, user requirements are not only about the users' satisfaction with the product. Functions, and demands for product intelligence and adaptability are becoming more diversified. In the existing literature, there are relatively few studies on mining user requirements based on the popularity of APPs, and most of the research literature on mobile APPs uses the conditional mean linear regression model, which requires the random disturbance term of the model to meet the normal distribution, but in reality, this assumption is often not satisfied. Bassett and Koenker [5] extended the median regression to general quantile regression, which can effectively deal with the error of non-normal distribution and achieve more robust parameter estimation results. As such, this paper uses quantile regression model to characterize the user requirements that APPs need to focus on at different popularity levels, that is, different quantiles. Based on this, this paper takes English vocabulary APPs as the study object. First, it collects user reviews of three free English vocabulary APPs that are typical and widely used in the Application market, and extracts product features from comments about each one. Emotion analysis and visual analysis, and the statistical method of text similarity are used to extract out the user requirements, which are classified by the K-means method. Secondly, based on the integration of the previous document variable system, APP downloads are selected as a reflection of APP popularity, combined with the information mined from users' online comments, a user behavior variable system that affects the downloads of English vocabulary APPs is constructed. Finally, through quantile regression, this paper finds out the focus of attention of English vocabulary APPs with different downloads so as to make suggestions for the improvement of the product in Appearance, function, adaptability and so on. The innovation of this paper lies in mining the user demand points that need to be improved for the same APP in different popularity periods, rather than generally discussing the user demand that needs to be iterated in the APP.

The structure of the paper is as follows: Section 2 is a literature review. Section 3 selects three typical English vocabulary APPs and collects user comments. Section 4 extracts product features through mining the comments about those APPs so as to dig out user requirements and classify them. Section 5 constructs a variable system that affects English vocabulary APPs downloads and uses quantile regression model to test it. Section 6 is the conclusions and the prospects of the work in the future.

2. Literature Review

Due to the popularity of the Internet and the diversification of similar products, consumers are more inclined to check online comments before downloading an APP.

Consumers can extract important information such as APP performance and satisfaction from a large amount of comments, and use this as an important reference for downloading English vocabulary APPs. For an enterprise, the ultimate goal of product production and service provision is to meet the requirements of users. User requirements are the internal motivation that drives innovation of mobile APP products, and it is also a decisive factor for a company's sustainable and healthy development in the face of fierce market competition. How to accurately and efficiently grasp the real requirements of users and use them as a guide to realize the innovative design of mobile APP and enhance the competitiveness of products has become a key step for enterprises to gain and maintain a foothold in the market. In summary, this section analyzes the literature from the following two aspects: one is the study of user requirement mining based on online reviews, and the other is the study of the relationship between online reviews and APP consumer decision-making behavior.

One stream is the research on mining user requirements based on online comments. Online comments are the personal feelings or opinions expressed by users after experiencing a certain product or service. Some of them describe the user's experience of using the product [6,7], and others express the user's views on various aspects of the product [8]. Based on the information adoption model, Hussain et al. [9] explored the behavioral motivations of online review users from the perspective of online food purchase. The research showed that consumers' demands for social interaction, economic incentives, and self-value reinforcement were the main driving forces for online review participation. Kang and Zhou [10] proposed rule-based method (RUBE), a rule-based unsupervised learning method, and used it to extract the subjective and objective features of searchable products from online user reviews. Wong and Qi [11] used text mining techniques to investigate the evolution of online reviews about Macau on Trip Advisor's platform between 2005 and 2013 and compared the evolution of user-generated online content with official platform forecasts, showing similar trends. Yang [12] proposed a dynamic Kano model construction method and a method of automatic identification of fine-grained user needs based on the related theories of text mining and machine learning. Tirunillai and Tellis [13] used the Latent Dirichlet Allocation (LDA) model to mine the key dimensions of consumer satisfaction in online product review data and, and monitor how the importance of these dimensions changed over time by using a dynamic perspective. Yu [14] proposed a research on the evaluation of mobile phone product attributes based on online shopping review text mining. By mining and analyzing online review, Yu found out the user's consumption evaluation of each attribute of a specific model of mobile phone, which is convenient for consumers to understand the advantages and disadvantages of the various attributes of mobile phones, providing suggestions for merchants to improve their products. In the development of software products, Kasiviswanathan and Ramalirgam [15] used the Software User Review Defect Corrective (SURDCM) model for requirements analysis and prioritization, combined with Software failure mode and effects analysis (SWFMEA), a technique to ensure that the software development process was defect-free, to detect possible defects and their respective impacts. Xu et al. [16] used text mining and other methods to connect customers' online text comments with customers' perceptions, helping business managers better understand customers' needs through User Generated Content (UGC). Wang et al. [17] conducted sentiment analysis on online comments and established a regression model to measure how product attributes affected customer satisfaction, and to help enterprises analyze user needs. Chen et al. [18] briefly analyzed the impact of COVID-19 on the national cultural and tourism industry and selected several representative types of tourism policies, crawled the comment data of Weibo users, analyzed users' perception and emotional preference to the policy, and thus mined the social effect of various policies. The aforementioned studies mainly focus on extracting the content of online review texts. The most important thing for enterprises is how to increase the downloads of APP to obtain the maximum profit. However, the research on quantitative description of the relationship between user requirements and download behavior decisions is relatively rare.

The second stream is the research on the relationship between online reviews and APP consumer decision-making behavior. At present, most of the relevant research on APP is about APP background development, or about the macro trend of APP's future development, but rarely focuses on the research of micro APP downloads. Individual literature only analyzes the factors that affect APP downloads or APP user stickiness. Kim et al. [19] found that both external and internal benefits motivated individuals to use a technological device. That was, individuals continued to use information technology because they perceived the possible benefits of obtaining utility (extrinsic) and playfulness (intrinsic) from it. Gommans et al. [20] presented an integrated framework of e-loyalty (see figure below) and its underlying drivers in terms of (a) Website & Technology (b) Customer Service & Logistics (c) Trust & Security (d) Product & Price and (e) Brand Building Activities. The nature of these factors in building customer loyalty was discussed with examples of current practices. Managerial and future research implications from the proposed framework were also presented. Cai et al. [21] pointed out that the probability of consumers seeing review information about products is largely affected by online review data, that is, the number of online reviews affects the self-consciousness of other consumers who have not yet made purchasing decisions. Cognition has a certain impact. By studying the relationship between hotel occupancy rate and online reviews, Shan and Lu [22] found that consumers' self-perception of hotels can be improved through online reviews, and the hotel's occupancy rate is positive. Positive online reviews have a more significant impact on the occupancy rate than negative online reviews. Cai [23] used the classic Technology Acceptance Model (TAM) model to build a mobile phone application store user adoption model and introduced the context awareness theory to derive the factors that influence users to download applications in the mobile phone store. Duan et al. [24] aimed to explore the persuasion effect and awareness effect of online user comments on the daily box office performance of films. After considered endogenous factors, this paper found that online user comments had no significant impact on the box office revenue of films, and the box office of films was significantly affected by the number of online posts. Chen [25] showed that the box office was related to the emotional factors. When the emotional tendencies of the current period and the later period of reviews were different, it had a significant impact on the box office, showing a negative correlation. Based on the text analysis of online review data captured from movie websites, Moon et al. [26] found that there was a significant positive correlation between the number of reviews and the box office, while consumers did not pay much attention to the content of online reviews. This showed that the more online reviews there were, the more popular the film would be; and because of the herd effect, viewers tended to make viewing decisions. Li [27] studied the influence of the quality of online negative reviews on consumers' purchase intention. The empirical results show that the degree of involvement can effectively adjust the impact of negative evaluation quality on consumers' purchase intentions. Yoo et al. [28] explored how different features of the retail environment influenced consumers' emotional responses in the shopping environment and how these emotions, in turn, the features influenced consumers' attitudes toward stores. Their research also used ethnographic interviews to identify emotions that arose in retail shopping Settings. Data collected from a sample of 294 Korean consumers indicated that store characteristics had a significant impact on consumers' in-store emotions, and these emotional experiences played a key mediating role in the relationship between store characteristics and store attitude. The above literature mainly focuses on the study of the influence of online reviews on consumers' purchasing decisions. Among them, the number of online reviews, text length, timeliness, positive and negative reviews, emotions, and other factors impact on consumers' purchase intentions. However, the research on willingness to buy for downloading mobile phone APPs is rare.

In summary, in the current research on consumer purchase behavior intentions, less consideration is given to the impact of user needs on consumer behavior. Existing research results also show that mining user needs in online reviews is conducive to extracting product and service characteristics, and helps companies develop products that are more

in line with user needs in the future. In addition, many users are more inclined to trust the online reviews of other users during the purchase process and refer to user reviews to understand whether the product can meet their own needs. When consumers find that the product cannot meet their own needs through the comments, it will accordingly reduce their willingness to buy. Based on this, this paper uses online comments of English vocabulary APPs, text analysis, product feature extraction and other technical means to dig out user requirements, emotional tendencies, satisfaction and other information, and selects APP downloads as the indicator of its popularity. At the same time, since quantile regression can effectively deal with the errors of non-normal distribution and achieve more robust parameter estimation results [29], the quantile model of the downloads of English vocabulary APPs is finally constructed according to the variable system, and the quantile model is described in detail. Through the constructed model, the user requirements and concerns of different quantiles can be mined, and companies can improve them, thereby increasing downloads for more benefits.

3. Data Acquisition and Preprocessing

This paper focuses on online English vocabulary APPs. After developing and putting those APPs into distributed software, companies can obtain users' views and opinions on products through online comments. Therefore, companies can find out the user's concerns and needs for the product, so as to improve the product quality and better satisfy users. Therefore, from the user's perspective, this paper collects user reviews based on the APPstore download platform as a basis for analyzing the pros and cons of products and digging out user requirements and concerns.

3.1. Acquisition of Online Comments for English Vocabulary APPs

3.1.1. Choice of English Vocabulary APPs and Mobile Terminals

English vocabulary APPs refer to arrange English vocabulary and use technical means to present vocabulary information. Learners can learn vocabularies by using this kind of APPs, which can translate and interpret words, and provide query and the functionalities of querying and memorizing words. Qimai Data is a domestic professional mobile promotion data analysis platform launched by Beijing Qimai Technology Co., Ltd. This platform can provide data queries with IOS, Android application market, as well as WeChat, and small programs. This paper selects data samples of English vocabulary APPs on the Qimai platform, and uses the APP's download ranking, review rankings and the characteristics of English vocabulary APPs as the representative measurement criteria for selecting APPs. BaiCiZhan, MoMoBeiDanCi and BuBeiDanCi are selected. The specific information is shown in Table 1.

It can be seen from Table 1 that, excluding the paid English vocabulary APPs, these three APPs rank in the top free software. Their outstanding characteristics are different. Qingyuan Mo Mo Education Technology Co., Ltd. (Qingyuan City, Guangdong Province) only developed the software "MoMoBeiDanCi", while the other two companies developed the same type of products. On the basis of user requirements, the main focus of those enterprises is also different. BaiCiZhan mainly emphasizes "Picture back word", MoMoBeiDanCi mainly highlights the "forgetting curve", and BuBeiDanCi mainly highlights "the sentence situation to understand the different meaning of the word and the use of the word".

In order to facilitate the detailed analysis of product features in the following text, the detailed sections in the three types of English vocabulary APPs are briefly introduced below, as shown in Figures 1–3.

Table 1. Selected English Vocabulary APP information.

Application Name	BaiCiZhan	MoMoBeiDanCi	BuBeiDanCi
APP Classification Ranking	7 (Education Ranking)	17 (Education Ranking)	5 (Education Ranking)
Keyword Coverage	18,190 (Comment Ranking First)	10,519 (Comment Ranking Third)	15,360 (Comment Ranking Second)
Number of comments from 22 January 2019 to 22 March 2019	19,687	12,016	16,839
Affiliated Companies	Chengdu super you love technology Co., Ltd.	Qingyuan Mo Mo Education Technology Co., Ltd.	Beijing Aiskoo Technology Co., Ltd.
Online Time	25 September 2012	13 July 2014	22 February 2014
Outstanding Feature	“Picture back word” software, by carefully selecting interesting pictures and example sentences for each word, improve the way of memorizing words, make it fun to remember words	An anti-forgetting memorizing software, through the big database and intelligent algorithm technology, according to different users’ forgetting curve to plan the learning content every day, to achieve the efficient anti-forgetting strategy	Focus on the sentence situation to understand the different meaning of the word and the use of the word, the word is associated with a large number of all kinds of real exam data over the years, the design is more inclined to deal with the exam

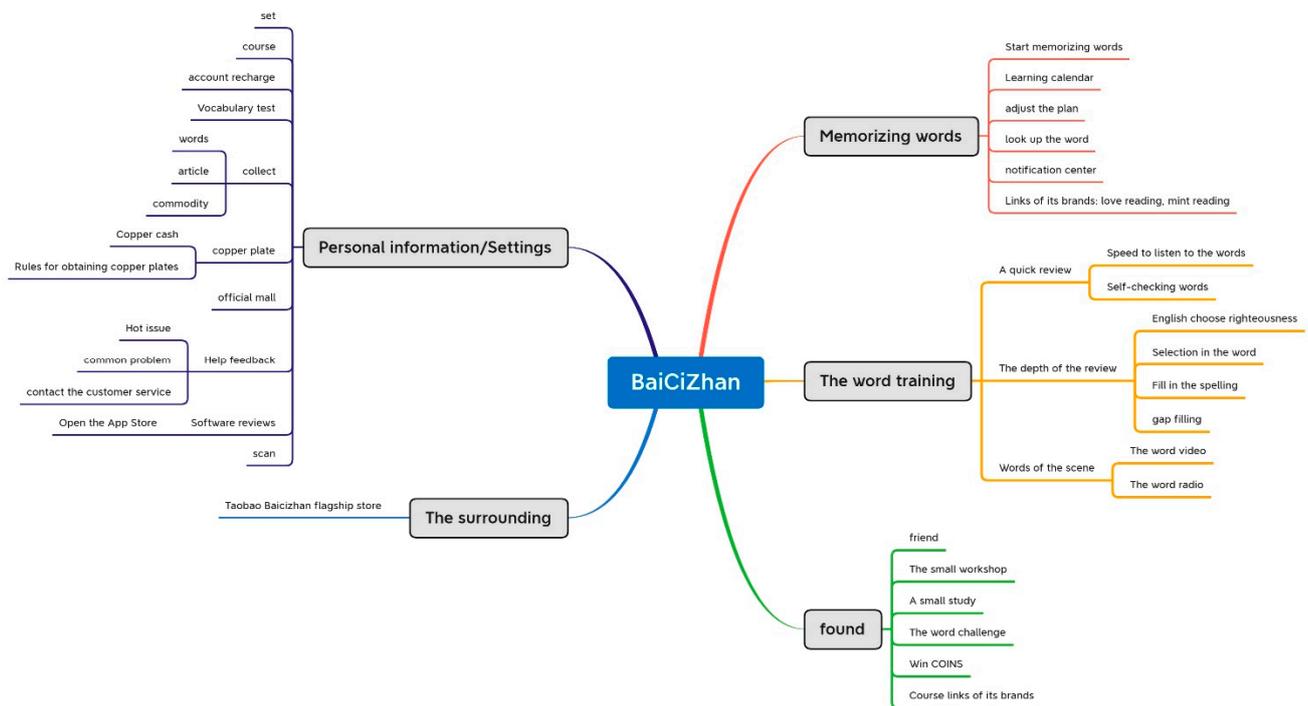


Figure 1. Distribution of BaiCiZhan APP modules.

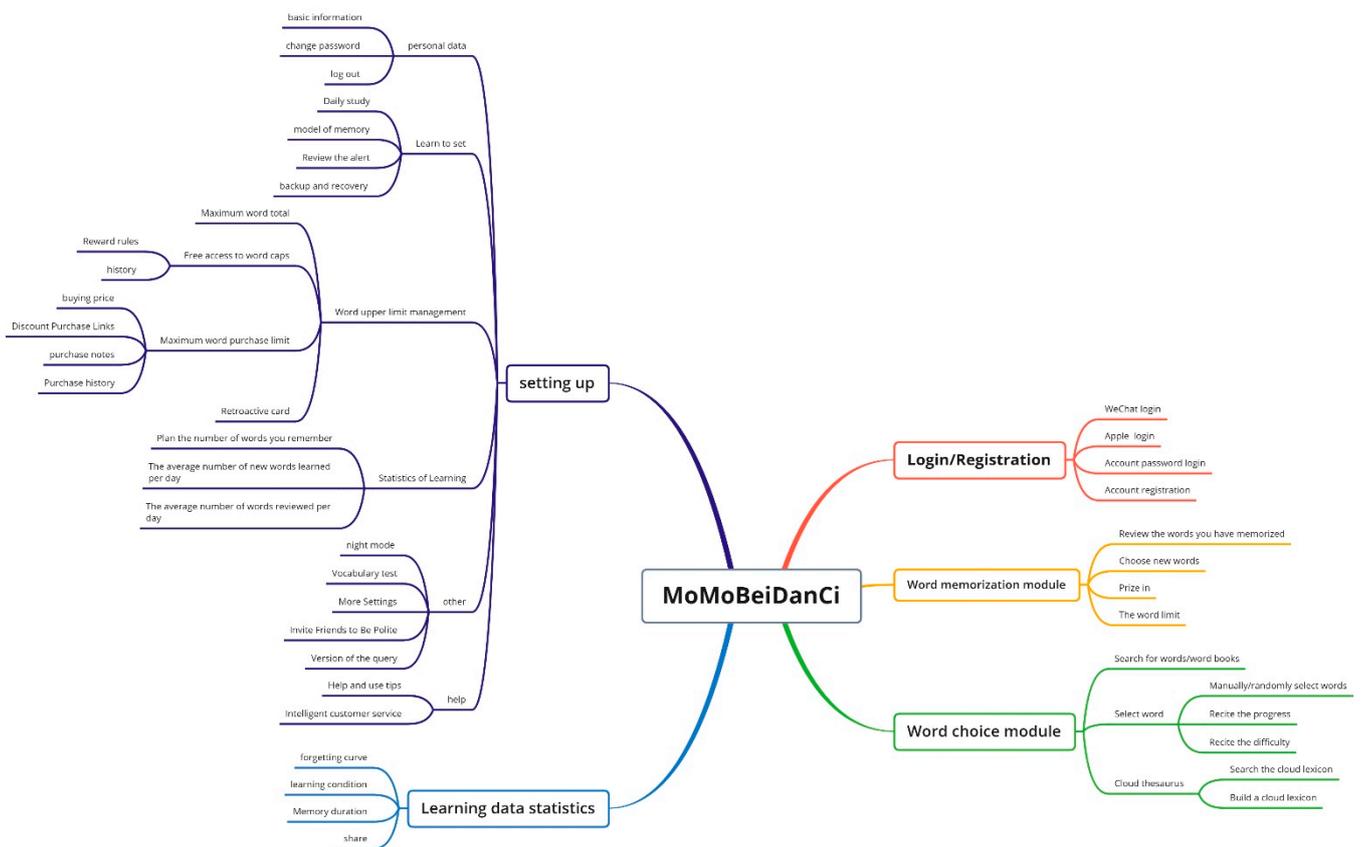


Figure 2. Distribution of MoMoBeiDanCi APP modules.

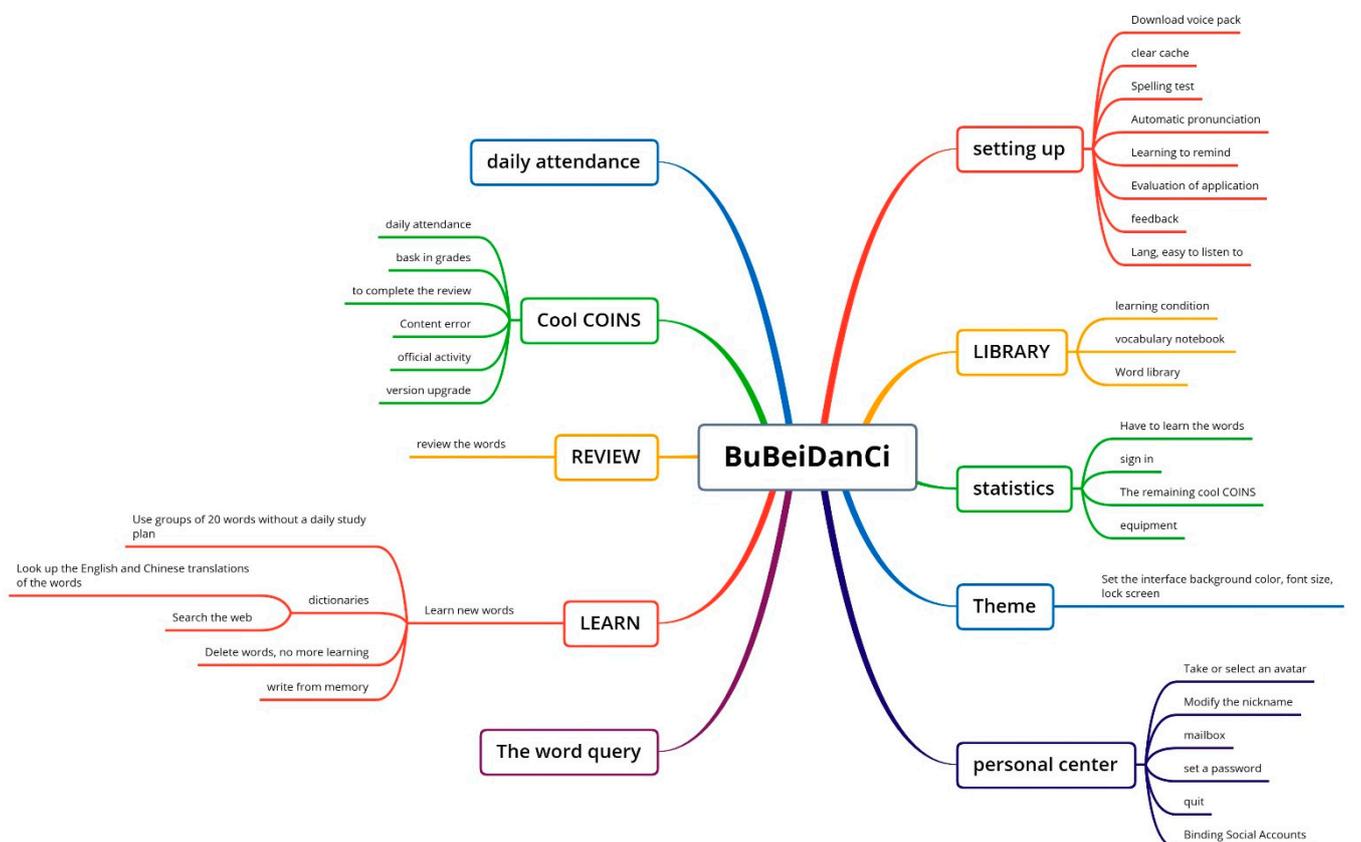


Figure 3. Distribution of BuBeiDanCi APP modules.

3.1.2. Obtaining Online User Comments

For the collection of those comments, two aspects need to be considered: one is the time node of the comment collection, and the other is the selection of the mobile terminal. The functions of those APPs are updated frequently. Due to the frequent update of function and the release time of the latest version, user comments in the past two months (22 January 2021–22 March 2021) are collected. According to data from the research organization Counterpoint, Huawei's mobile phone market share in China reached 46% in the first half of 2021, followed by Vivo's 16%, OPPO's 15%, Xiaomi and Apple 9% respectively. In order to make online user comments as comprehensive as possible to cover different types of mobile phones, this paper selects the comments of different users on these four mobile phones.

3.2. Preprocessing of Online Comments

In order to improve the credibility of data analysis, the user comments are preprocessed as follows: (1) Data cleaning. Remove the repetitive data in the original data and meaningless ultra-short texts and other worthless data, such as "words", "software" and other words that are not related to product evaluation but appear frequently, and finally a total of 48,560 comments on Huawei, Vivo, OPPO, and Apple mobile phones are obtained. (2) The Jieba Chinese analysis library in the Python software is used for word segmentation. Perform word segmentation operations on each comment data and convert it into a list of words. (3) The screening basis of stop words. This paper uses a general stop word list containing 600 stop words as the basis for screening and removes some words and phrases that have no practical significance for the analysis of comment data, as well as punctuation marks.

4. Text Mining of User Online Comment

Based on the information content reflected in the online comments after the above preprocessing, text analysis is performed on the comments of different types of English vocabulary APPs used on Huawei, Vivo, OPPO, and Apple mobile phones, to discover the key points that companies need to improve, and to make suggestions on product service quality.

4.1. Analysis of Popular Words in English Vocabulary APP General Reviews

For the analysis of popular words in three APPs, the Word cloud library of Python software is used to draw general comment word cloud map. The size of different word fonts in the word cloud map is intuitive to reflect the frequency of appearance in the overall comment data. By word cloud map, the user's popular comments on the English vocabulary APP's can be obtained. By analyzing the popular comment words, we can get the user's attention points and the advantages of the three types of English vocabulary APPs. In BaiCiZhan's comments, the frequency of high-frequency words such as picture, good reputation, convenience, amusing, vocabulary, intensity, useful, five-stars, function, practical, and so on were 422, 406, 356, 341, 296, 291, 256, 195, 182, 175. In MoMoBeiDanCi's comments, the frequency of high-frequency words such as memory, review, curve, forgetting, upper limit, vocabulary, Ebinho, and so on were 4163, 2733, 2637, 1990, 1205, 1079, 994. In BuBeiDanCi's comments, the frequency of high-frequency words such as example sentence, function, study, memory, concise, interface, spell, page, and so on were 2160, 1822, 1780, 1670, 1582, 1298, 1058, 852. The results are shown in Figures 4–6.

of memorizing words in a graphic mode, aiming to play the role of image memory and make the meaning of words more specific. The selected pictures are more vivid and add the fun of memorizing words. Later, it can gradually expand the personalized learning mode of the product, such as word TV, word radio and so on. In addition, terms such as “free” and “advertisement” are mentioned more frequently. However, the payment and advertisements during use also directly affect user satisfaction.

The words “memory”, “Ebinho”, “curve”, and “forgetting” in Figure 5 are also prominent features of MoMoBeiDanCi. It applies the Ebinho memory curve with characteristics of anti-forgetting using big data technology and intelligent algorithms to achieve an efficient anti-forgetting strategy according to the forgetting curve law of different users. Compared with the words presented in the form of images in BaiCiZhan, MoMoBeiDanCi provides a large number of mnemonic association entries, such as comparison of synonyms, comparison of similar words, Chinese homophonic stalks, etc., to help review words step by step. At the same time, vocabularies such as “upper limit”, “vocabulary”, and “sign in” are the unsatisfactory aspects that users often mention while using. The free vocabularies in this software are limited. If the free word quota is used up later, you will need to pay for more words. You can buy it with money, and each vocabulary book cannot be re-learned. Only by participating in the daily check-in activities and sharing the software can you get the free words. This is also a way used by company to increase the downloads. In addition, “simplicity” and “interface” mainly highlight the characteristics of the concise interface design.

In Figure 6, the word “pronunciation” will be given every time. This type of software is more suitable for auditory memorizers, because English itself is a pinyin text. However, if you are a visual memorizer, it may be more suitable for the graphic mode of BaiCiZhan. At the same time, vocabulary such as “example”, “root”, “affix” emphasizes that the characteristic of BuBeiDanCi is the root of the word and choice of the meaning of the word, supplemented by the root and original example sentence memory. In addition, “interface”, “simplicity”, “design”, and “advertising” mean that the user interface of this software is simple, with more white space and no advertisements.

4.2. Extraction of User Requirement Features of English Vocabulary APPs

The user’s emotional tendency for English vocabulary APPs will be expressed through a series of emotional words. Choosing an appropriate emotional dictionary can significantly improve the effect of user requirement analysis. Based on this, ROST CM6.0 is used to calculate the emotional scores of the comments on the four mobiles, the proportion of positive, neutral and negative comments of three English vocabulary APPs is obtained. The calculation principle is mainly based on the BosonNLP sentiment dictionary illustrated as follows: firstly, the texts are segmented using the Jieba word segmentation here. Secondly, the segmented texts are matched by the BosonNLP dictionary corresponding word segmentation good list data one by one, and the score values of the matched emotion words are recorded. Finally, all the emotional score values are counted. In addition, in this paper, negative words are defined as negative emotion words, such as “not good”, “not so good” and “need to be improved”, etc.; positive words are defined as positive affective words, such as “love”, “like” and “great”. If the word “dislike” appears in a user’s review, it is counted as a negative emotion in the calculation”. The results are shown in Table 2.

It can be seen from Table 2 that the overall emotional tendency is positive, followed by neutral and negative. In order to further dig out the user’s requirements and concerns for English vocabulary APPs, the following similarity statistics method is used to analyze the similar information in the positive, neutral, and negative comments, and the information is ranked according to the number of occurrences of similar information to obtain users comment on high-frequency feature words.

Table 2. The proportion of positive, neutral and negative comments of the three English vocabulary APPs.

APP	Mobile Terminal	Percentage of Positive Comments	Percentage of Neutral Comments	Percentage of Negative Comments
BaiCiZhan	HUAWEI	53.17%	21.77%	25.05%
	Vivo	67.43%	10.75%	21.82%
	OPPO	50.55%	31.21%	18.24%
	Apple	49.64%	11.81%	38.55%
MoMoBeiDanCi	HUAWEI	50.47%	31.18%	18.35%
	Vivo	30.41%	51.58%	18.01%
	OPPO	48.44%	32.00%	19.56%
	Apple	64.11%	11.20%	24.69%
BuBeiDanCi	HUAWEI	81.94%	3.31%	14.75%
	Vivo	66.70%	1.53%	31.77%
	OPPO	77.91%	2.42%	19.67%
	Apple	90.11%	1.56%	8.33%

4.2.1. Hot High-Frequency Keywords of User Requirements

This section uses Python software to perform the statistics of text similarity heat. The input form is an excel table. This section only contains the comments on three APPs. The specific implementation steps are as follows:

(1) Pre-processing data

Use the Pandas data analysis package in Python to obtain the input information, and use the Jieba Chinese word segmentation database in Python to segment the input long comment sentences to form a two-dimensional array.

(2) Processing dictionary

Gensim is a Python library for automatically extracting semantic topics from documents, which can be used to process unstructured numerical text, that is, plain text. The `corpora.Dictionary()` method in the library generates a dictionary from the two-dimensional array formed by word segmentation, thereby constructing a dictionary based on the input comment text information, and uniquely identifying a word with a digital number.

(3) Processing corpus

The bag-of-words model means that all words are packed into a bag, regardless of the morphology and word order, that is, each word is independent. Supposing you create a dictionary [Jane, wants, to, go, Shenzhen, Bob, Shanghai], the sentence "Jane wants to go to Shenzhen" can be represented by [0,0,1,1,1,1,2], whose value is the number of times the word at the corresponding position in the dictionary. Based on this, the two-dimensional array is transformed into a sparse vector by the `doc2bow()` method to form a corpus.

(4) Calculating text similarity

The Latent Semantic Indexing (LSI) model uses Singular Value Decomposition (SVD) to decompose the word-document matrix. SVD can be seen as finding irrelevant index variables from the word-document matrix and mapping the original data into the semantic space. Documents that are not similar in the word-document matrix may be relatively similar in the semantic space. The text topic matrix obtained through LSI can be used for text similarity calculation.

Term Frequency-inverse Document Frequency (TF-IDF) is a statistical method used to evaluate the importance of a word to a document set or a document in a corpus. The importance of a word increases in proportion to the number of times it appears in the document, but at the same time, it decreases in inverse proportion to the frequency of its appearance in the corpus.

In actual operation, the `LsiModel()` method of the `models` module is used to calculate the TF-IDF of the words in the corpus, and the `keys()` keyword acquisition method in the

dictionary is used to obtain the number of features in the dictionary. Finally, the TF-IDF of the words in the corpus and the number of features in the dictionary are substituted into the SparseMatrixSimilarity() similarity calculation method of the similar module to establish the sparse matrix similarity to obtain an index.

(5) Calculating similarity between test and sample data

By reading each user comment text in the input excel document, segmenting word by Jieba, calculating sparse vector of the test data through doc2bow, and finally calculating the similarity between the test and sample data, this paper classifies the data with the similarity greater than 0.6 as one type.

(6) Calculating popular concerns

Calculating the amount of data contained in each type of user's question, using this as the concerns. Sorting the questions according to their popularity and getting the hot high-frequency keywords. The results are shown in Tables 3–5.

Table 3. BaiCiZhan comments on content hot high-frequency keywords.

Select Comments	Related Words
Users overall comment content (Contains positive, neutral and negative comment content)	Pictures and texts, login, repetition, novelty, pronunciation, feedback, root, reward, example sentences, spoken language, clear, compatible, version, phonetic transcription, diversified, convenient, listening, paid, classroom scene, concise and clear, support, screen, annotation, affix, Advertising, association, annotation, interface design, boring, personalized, rote, screen, fun
Neutral and negative comments from HUAWEI, Vivo and OPPO	Upgrade, scallop, animation, context, forget, convenience, example, refinement, customization, Ebinho curve, dependency, defect, sensitivity, accuracy, audio, compound words, vocabulary, word page
Neutral, negative reviews from Apple	Adapt to ipad, horizontal screen, customer service, non-adaptation, version, screen, pronunciation, advertisement, example sentence, login, interface, update, mall, pictogram, font, coupon, revision, black screen, bug, extension, derivative word, complete, Hd version

Table 4. MoMoBeiDanCi comments on content hot high-frequency keywords.

Select Comments	Related Words
Users overall comment content (Contains positive, neutral and negative comment content)	Forgetting curve, example sentence, sign-in, repeat, convenience, thesaurus, picture, design, pronunciation, upper limit, dictionary, homophonic, paraphrase, purchase, humanization, expanded vocabulary, conciseness, payment, sense of accomplishment, pronunciation, fast, video, clarity, Context, circulation, rote memorization, concise atmosphere, Chinese and English, grammar, color, fun, spoken language, bells and whistles, intelligence, color matching, screen, comfort, fatigue
Neutral and negative comments from HUAWEI, Vivo and OPPO	Word limit, vocabulary, vocabulary test, get word limit, page, earn words, dictation mode, easy to use, phonetic transcription, limited number of times, page conciseness, synonyms, boring, proficiency, etymology, fancy, multiple choice, convenient, Black screen, accuracy, follow-up, pictogram, reward
Neutral, negative reviews from Apple	Reimbursement, night mode, high-frequency vocabulary, false memory

Table 5. BuBeiDanCi comments on content hot high-frequency keywords.

Select Comments	Related Words
Users overall comment content (Contains positive, neutral and negative comment content)	Root affix, spelling, interface, conciseness, vocabulary, picture, dictation, phrase, free, repeat, pattern, context, check-in, flashback, advertisement, derivative word, paraphrase, dictation, payment, phrase, algorithm, Purchase, thesaurus, humanization, concise atmosphere, expansion, interest, grammar, phonetic transcription, loop, beauty, color control, fun, sense of accomplishment, audio, analysis, login, practicality

From Tables 3–5, the features that user comments mentioned frequently and expressed more concerns about can be summarized. Because most comments of BuBeiDanCi are positive, and the hot high-frequency keywords in negative and neutral comments are the same as total comments, Table 5 only shows the hot high-frequency keywords of the user’s total comments. In order to dig deeper into the feature attributes mentioned by users through high-frequency keywords, the new word discovery module of the NLPiR platform is combined with the new word discovery module to expand the above-mentioned hot high-frequency keywords, and the user’s requirements and concerns for English vocabulary APPs are obtained. The results are as follows shown in Table 6.

Table 6. BaiCiZhan, MoMoBeiDanCi, BuBeiDanCi new word mining.

APP	New Termlist
BaiCiZhan	Customer service, English pronunciation, expanded vocabulary, word details, teaching version, dark mode, simple and clear, quick to recite words, question type, lock screen, horizontal screen, screenshot, memory rule, convenient and fast, pause, vertical screen, No ads, unsuitable, follow-up, dark mode, real-person pronunciation, root affixes, combination of pictures and text, night mode, planning, word interpretation, associative memory, relying on pictures, circular memory, filling in the blank spelling, memory rules, convenient and fast, quantitative learning, easy to be confused, English essays, practicality, dictation, analysis, screen, vocabulary improvement, video, novel, humanized, word spelling, expanded vocabulary, word dictation, suitable for beginners, picture guessing, lock screen, follow-up, Playback failure, network abnormality, freeze, unable to log in, pause, authorization failure, compatibility, resolution, black border, card replacement, vertical screen, service attitude, paid version, offline package
MoMoBeiDanCi	Associative memory, root affixes, real questions, number of words, customer service, anti-forgetting, enhanced memory, auxiliary memory, artificial customer service, homophonic stalk, night mode, supplementary sign, British pronunciation, single quota, real pronunciation, rolling review, synonymous words, forgotten critical points, derivative words, page simplification, late-night mode, ad insertion, ease of use, clever memorization, VIP required, upper limit of words, different from person to person, very intelligent, memory curve, cognitive level, scientific system, word order
BuBeiDanCi	Real context, original example sentences, top-up, audio example sentences, historical questions, high-frequency vocabulary, word upper limit, convenient and quick, correct pronunciation, colorful content, unlimited repetition, personalized settings, machine pronunciation, fun, word skills, derived vocabulary, Flashback, uninstall, face value, come from movies, review planning, note-taking function, expanded vocabulary, follow-up function, sense of experience, British pronunciation, don’t be fancy

4.2.2. Classification of User Requirements

According to the above-mentioned hot high-frequency keywords and expanded new words mentioned in the user comments, K-means text clustering is adopted for user requirements. Based on user requirements element hierarchical model established by literature [30] and semantic space vocabulary construction method of the bicycle modeling demand questionnaire by reference [31], clusters are set into five categories. Based on label results obtained by the clustering, user requirements are subdivided into 26 sub-requirements according to the label results obtained by the clustering, and table of user requirement elements is constructed. The results are shown in Table 7.

Table 7. Table of user requirement elements.

User Need	Concrete Subclassification	K-Means Clustering Involves Keywords
Appearance requirements	Interface design	Clear, concise and clear, screen, interface, design, simple and clear, interface design, concise, atmosphere, screen, word page, font
	Match colors	Fancy, color, beauty, color matching, color control, beauty, value, fancy
Functional requirements	Memorizing words	Pictures and texts, associations, diversification, cyclic memory, memory rules, context, phonetic symbols, repetition, listening, classroom scenes, annotations, rote memorization, annotations, word details, combination of pictures and texts, cyclic memory, follow-up reading, question types, language Context, example, associative memory, fill-in spelling, dictation, analysis, pictogram, sign-in, repetition, dictionary, homophonic, paraphrase, video, loop, grammar, dictation mode, multiple choice, spelling, phrase
	Whether there are roots or affixes	Root, affix
	Quantity of vocabulary	Expand vocabulary, vocabulary, upper vocabulary, compound words, expand vocabulary, extension, derived words, thesaurus, upper limit, number of words, derivative words, synonymous words, upper limit of words, earn words, etymology, expand vocabulary
	Whether to personalize the words memorized daily according to the Ebbinghaus Forgetting Curve	Memory law, forgetting, Ebinho curve, forgetting curve, anti-forgetting, rolling review, forgetting critical point, memory curve graph, algorithm, review planning
	Is there an offline package	Offline package
	Whether there are Chinese and English homophonic	Chinese and English, homophonic stalk
	Whether there is an authentic English test over the years	High-frequency vocabulary, historical real questions
	Effect of pronunciation	British pronunciation, pronunciation, spoken, pause, audio, pronunciation
Emotional requirements	Convenience	Convenient, log-in, quick to recite words, convenient and fast, easy to use, proficient
	Practicability	Humanization, follow-up, suitable for beginners, practicality, improve vocabulary, sense of accomplishment, rote memorization, strengthen memory, auxiliary memory, clever memorization, scientific system, analysis, practicality, note-taking function
	Sensitivity	<i>Sensitivity, accuracy</i>
	Enjoyment	Novelty, animation, boring, diverse, interesting, interesting, false memory, interest, fun
	Image aided	Guess the picture, rely on the picture
	Intelligentize	Original example sentences, sound examples, machine pronunciation, scene diversification, lock screen, example sentences, personalization, customization, planning, quantitative learning, intelligent, different from person to person, accuracy, intelligence, cognitive level, correct pronunciation
	Comfort level	Fatigue, resolution, clarity, comfort, streamlined and colorful pages
Adaptability requirements	Software pattern	Dark mode, night mode, late night mode, mode
	Software version (compatible with multiple mobile versions)	hd version, compatible with ipad, compatible, version, teaching version, not compatible, upgrade, update, revision
	Whether the screen is adjustable	Horizontal screen, vertical screen, screen, lock screen, screenshot

Table 7. Cont.

User Need	Concrete Subclassification	K-Means Clustering Involves Keywords
Network technology Environment requirements	Whether there is a lag	Playback failure, network abnormality, pause, login, blemish, easy to be confused, video, black screen, bug, stuck, unable to log in, authorization failure, black border, re-card, re-sign, flash back, uninstall
	Whether there is a black screen	
	Whether there is a logon failure	
Service requirements	Whether there are any advertisements	Ads, no ads, ad insertion
	Whether there is a customer service reply	Feedback, customer service, service attitude, manual customer service
	Whether there are paid items	Paid version, VIP, payment, rewards, single limit, mall, coupons, purchase, cool coins, free, top-up

4.2.3. Analyzing Key User Requirements

In order to dig out the key user requirements, the user requirements established in Table 7 are quantified and the index vocabulary for each type of need is set, and the index vocabulary lexicon is constructed. The index lexicon is traversed and counted, and the frequency of the occurrence of words is used as the evaluation method of the index. The results are shown in Tables 8–10.

Table 8. BaiCiZhan user requirement elements sort.

User Requirements Subclassification	Key Feature Word Frequency	BaiCiZhan APP User Focus Sorting	User Requirement Type
Memorizing words	365	1	Functional requirements
Convenience	333	2	Emotional requirements
Image aided	228	3	Emotional requirements
Enjoyment	187	4	Emotional requirements
Whether there are paid items	136	5	Service requirements
Quantity of vocabulary	126	6	Functional requirements
Software version (compatible with multiple mobile versions)	96	7	Adaptability requirements
Effect of pronunciation	96	7	Functional requirements
Interface design	80	8	Appearance requirements
Practicability	77	9	Emotional requirements
Whether there are any advertisements	53	10	Service requirements
Intelligentize	50	11	Emotional requirements
Network technical environment (problems such as lag, black screen, login, etc.)	50	11	Network technology Environment requirements
Software pattern	45	12	Adaptability requirements
Whether the screen is adjustable	22	13	Adaptability requirements
Whether there is a customer service reply	20	14	Service requirements
Whether there are roots or affixes	16	15	Functional requirements
Match colors	8	16	Functional requirements
Whether to personalize the words memorized daily according to the Ebbinghaus Forgetting Curve	8	16	Functional requirements
Comfort level	7	17	Emotional requirements
Whether there is an authentic English test over the years	1	18	Functional requirements
Sensitivity	1	18	Emotional requirements
Whether there are Chinese and English homophonic	0	19	—
Whether there is an offline package	0	19	—

Table 9. MoMoBeiDanCi user requirement elements sort.

User Requirements Subclassification	Key Feature Word Frequency	MoMoBeiDanCi APP User Focus Sorting	User Requirement Type
Whether to personalize the words memorized daily according to the Ebbinghaus Forgetting Curve	3265	1	Functional requirements
Memorizing words	2769	2	Functional requirements
Quantity of vocabulary	2461	3	Functional requirements
Interface design	1750	4	Appearance requirements
Whether there are paid items	813	5	Service requirements
Intelligentize	811	6	Emotional requirements
Practicability	747	7	Emotional requirements
Convenience	666	8	Emotional requirements
Whether there are roots or affixes	328	9	Functional requirements
Whether there are any advertisements	319	10	Service requirements
Software pattern	317	11	Adaptability requirements
Effect of pronunciation	289	12	Functional requirements
Image aided	278	13	Emotional requirements
Match colors	272	14	Appearance requirements
Enjoyment	226	15	Emotional requirements
Network technical environment (problems such as lag, black screen, login, etc.)	148	16	Network technology Environment requirements
Comfort level	139	17	Emotional requirements
Software version (compatible with multiple mobile versions)	121	18	Adaptability requirements
Whether there is a customer service reply	98	19	Service requirements
Whether there is an authentic English test over the years	84	20	Functional requirements
Whether there are Chinese and English homophonic	34	21	Functional requirements
Whether the screen is adjustable	12	22	Adaptability requirements
Sensitivity	2	23	Emotional requirements
Whether there is an offline package	0	24	—

It can be seen from Tables 8–10 that when users use BaiCiZhan, they often mention the way of memorizing words in this software, such as the combination of pictures and vocabularies, circular memory, associative memory, etc. These methods are full of fun, reducing the boring feeling in the process of word memory. For the way in which Baizizhan connects words with actual situations, “rely on pictures” appears in the new word mining in Table 6, which shows that users believe that memorizing words in this way will increase their dependence on pictures, and does not deepen memory from the roots and affixes. In addition, whether the software is fast to use, whether there is enough thesaurus to choose from during use, and whether there are paid items are all users’ concerns when choosing an English vocabulary APP. For the MoMoBeiDanCi, users appreciate the feature of using Ebbinghaus Forgetting Curve to personalize the words memorized every day. Most users feel very intelligent when using this APP, and the interface design is simple and clear. In the process of memorizing words, there are Chinese and English homophones, roots, affixes and other auxiliary memory. However, its obvious shortcoming is that the number of vocabularies is insufficient, and there are very few vocabularies available for free use. After the vocabulary reaches the upper limit, you need to pay for it. What users appreciate most in the BuBeiDanCi APP is the interface design. This software focuses on the in-depth analysis of the structure and grammar of the word. From the ranking of the user’s attention points in BuBeiDanCi APP, it can be known that the function requirements “whether to personalize the words memorized daily according to the Ebbinghaus Forgetting Curve” ranks higher, indicating that most users want to add the function of personalizing the forgetting curve to this software. In addition, the requirements of the network technology environment are rarely mentioned in user comments, because these three APPs rarely have

black screens, freezes, and inability to log in, and their use environments are relatively stable. Compared with other user requirements elements, the need to improve the network technology environment is less necessary, and enterprises do not need to spend too much investment on this element.

Table 10. BuBeiDanCi user requirement elements sort.

User Requirements Subclassification	Key Feature Word Frequency	BuBeiDanCi APP User Focus Sorting	User Requirement Type
Interface design	1920	1	Appearance requirements
Memorizing words	1650	2	Functional requirements
Intelligentize	1158	3	Emotional requirements
Whether there are roots or affixes	565	4	Functional requirements
Whether there are paid items	485	5	Service requirements
Convenience	376	6	Emotional requirements
Match colors	279	7	Appearance requirements
Whether to personalize the words memorized daily according to the Ebbinghaus Forgetting Curve	261	8	Functional requirements
Whether there are any advertisements	232	9	Service requirements
Practicability	214	10	Emotional requirements
Quantity of vocabulary	201	11	Functional requirements
Effect of pronunciation	188	12	Functional requirements
Software pattern	143	13	Adaptability requirements
Enjoyment	142	14	Emotional requirements
Software version (compatible with multiple mobile versions)	136	15	Adaptability requirements
Image aided	136	16	Emotional requirements
Network technical environment (problems such as lag, black screen, login, etc.)	84	17	Network technology Environment requirements
Whether there is an authentic English test over the years	67	18	Functional requirements
Comfort level	58	19	Emotional requirements
Whether there is a customer service reply	34	20	Service requirements
Whether the screen is adjustable	20	21	Adaptability requirements
Whether there are Chinese and English homophonic	7	22	Functional requirements
Sensitivity	1	23	—
Whether there is an offline package	0	24	—

Through the analysis of the requirements elements in the user comments of the above three APPs, users are more concerned about the functional requirements of English vocabulary APPs, followed by appearance requirements > emotional requirements > service requirements > adaptability requirements > network technical environment requirements. Therefore, if companies want to improve user satisfaction, they need to further optimize the software's functions and appearance.

5. Analysis of the Impact of User Online Comments on Product Downloads

Generally, new users will choose whether to download a product based on user comments, historical downloads, APP ratings, etc. At the same time, companies will also adjust the software functions according to user comments, historical downloads, APP ratings, advantages of competitive product, etc., to update the software. In fact, when users choose to download an English vocabulary APP, they mainly pay attention to whether the software meets their own needs. Due to longer use and different users' preferences, different users' requirements for APPs gradually increase and APP downloads increase. At this time, English vocabulary APPs should not only meet basic functional requirements, but also improve the intelligence, such as cooperating with personalized setting of learning content, etc. Therefore, where is user's focus on APPs with different downloads? How

does the company adjust the product based on users’ downloads behavior? What factors can affect the number of users’ downloads behavior? What are the requirements that affect user downloads?

5.1. Model Construction

In order to solve the above problems, the downloads of English vocabulary APPs are selected as an indicator of the popularity of APPs, and a correlation model between the English vocabulary APP downloads and user comments is constructed. Due to the large range of downloads, user ratings and other indicators, the conditional mean regression model results are difficult to extend to non-central positions. Therefore, this paper selects a quantile regression model to describe the relationship between user comments and user downloads at different quantile points. The variables involved in the model are shown in Table 11.

Table 11. Variable description of quantile regression model for English vocabulary APP downloads.

Variable Name	Variable Symbol	Explanation
Dependent variable: downloads (take the logarithm)	Down (lgdown)	The number of times users downloaded English vocabulary APPs
Rates of positive emotions in user reviews	Pos	User’s positive and negative emotional value to the product
Negative emotion rates of user reviews	Neg	
The proportion of very satisfied users	Lik6	Users’ satisfaction with English vocabulary APPs according to their own usage (By referring to the emotionality dictionary of Taiwan University and import it into ROST CM6.0, the Likert score of each comment is obtained, which mainly includes seven kinds of satisfaction values: −3, −2, −1, 0, 1, 2, 3)
The proportion of satisfied users	Lik5	
Part of the number of satisfied users	Lik4	
The proportion of the number of users with neutral satisfaction	Lik0	
The proportion of partially dissatisfied users	Lik3	
The proportion of unsatisfied users	Lik2	
Percentage of highly dissatisfied users	Lik1	
Percentage of mobile APP5 star ratings	5S	Users score English vocabulary APPs according to their own usage
Percentage of mobile APP4 star ratings	4S	
Percentage of mobile APP3 star ratings	3S	
Percentage of mobile APP2 star ratings	2S	
Percentage of mobile APP1 star ratings	1S	
Appearance requirements	waiguan	This requirement = 1 is mentioned in user comments, and this requirement = 0 is not mentioned in user comments
Functional requirements	gongneng	
Emotional requirements	qinggan	
Adaptability requirements	shipeidu	
Network technology Environment requirements	wangluo	
Service requirements	fuwu	

Quantile regression can be seen as an extension of median regression, and its parameters are estimated to minimize the objective function:

$$\hat{\beta}(\tau) = \operatorname{argmin}_{\beta \in \mathbb{R}^k} \sum_{i=1}^n \rho_{\tau}(y_i - x_i' \beta) \tag{1}$$

The loss function of the general τ quantile regression is:

$$\rho_{\tau}(u) = u(\tau - I(u < 0)) \tag{2}$$

Among them, $0 < \tau < 1$ and $I(u)$ is the indicator function, which is called the regression coefficient estimation under the τ th quantile.

In order to study the difference in quantile points of different downloads of English vocabulary APPs by influencing factors, quantiles are taken every 0.05, and

{0.05,0.1,0.15,0.2,0.25,0.3,0.35,0.4,0.45,0.5,0.55,0.6,0.65,0.7,0.75,0.8,0.85,0.9,0.95} quantile regression are selected to establish the following quantile regression model:

$$\begin{aligned}
 \text{lgdown} = & C + \beta_1 \text{Pos} + \beta_2 \text{Neg} + \beta_3 \text{Lik6} + \beta_4 \text{Lik5} + \beta_5 \text{Lik4} + \beta_6 \text{Lik0} \\
 & + \beta_7 \text{Lik3} + \beta_8 \text{Lik2} + \beta_9 \text{Lik1} + \beta_{10} \text{5S} + \beta_{11} \text{4S} + \beta_{12} \text{3S} + \beta_{13} \text{2S} + \beta_{14} \text{1S} \\
 & + \beta_{15} \text{waiguan} + \beta_{16} \text{gongneng} + \beta_{17} \text{qinggan} + \beta_{18} \text{shipeidu} + \beta_{19} \text{wangluo} + \beta_{20} \text{fuwu}
 \end{aligned}
 \tag{3}$$

5.2. Quantile Regression Model Results

In order to make downloads of new user more time-sensitive, the user downloads of three APPs from 22 February 2021 to 22 March 2021 are selected as the dependent variable. The independent variable value in each comment is obtained according to the quantification method of user requirements. The regression analysis is performed on the Eviews measurement software: first, the regression model is tested by ordinary least squares (OLS) and stepwise regression test to find out which variables are significant in the regression. Secondly, quantile regression is performed at different quantile points. Finally, the significance between OLS regression and quantile regression is performed variables to illustrate the validity is compared of the digit regression model. The results are shown in Tables 12 and 13.

Table 12. Ordinary least square (OLS) regression results.

Variable	Coefficient	Std. Error (Standard Error)	Statistic (T-Statistic Value)	Prob (Significance Test Value)
C(constant)	4.87	1.063	4.59	0.0000
Neg	−0.17	1.04	−0.17	0.8675
Pos	−0.51	1.02	−0.49	0.6205
Lik3	−0.14	0.49	−0.28	0.7780
Lik2	0.17	0.56	0.31	0.7589
Lik1	0.6	0.75	0.8	0.4252
Lik0	0.11	0.34	0.31	0.7594
Lik4	−0.45	0.45	−1.02	0.3122
Lik5	−0.06	0.41	−0.16	0.8751
Lik6	−0.67	0.41	−1.68	0.0985
waiguan	−0.01	0.003	−2.98	0.0039
gongneng	0.004	0.00	4.78	0.0000
qinggan	−0.002	0.003	−0.86	0.3905
shipeidu	−0.02	0.009	−1.73	0.0889
fuwu	0.00	0.005	0.05	0.9626
wangluo	−0.001	0.02	−0.08	0.9399
1S	−0.006	0.01	−0.38	0.7026
2S	0.01	0.03	0.39	0.6980
3S	0.05	0.02	2.46	0.0167
4S	−0.00	0.008	−0.23	0.8166
5S	−0.00	0.001	−0.88	0.3821

R-Squared = 0.659657; Adjusted R-squared = 0.5565; Prob = 0.0000.

Table 13. Step by step backward regression results.

Variable	Coefficient	Std. Error	T-Statistic	Prob
C(constant)	4.71	0.13	36.23	0.000
POS	−0.37	0.17	−2.25	0.027
Lik4	−0.45	0.18	−2.44	0.0169
Lik1	0.65	0.6	1.08	0.2828
gongneng	0.004	0.00	5.65	0.0000
waiguan	−0.007	0.002	−3.21	0.0020
3S	0.05	0.018	2.72	0.0080
shipeidu	−0.017	0.008	−1.99	0.0491
Lik6	−0.62	0.15	−4.01	0.0001
5S	−0.001	0.00	−1.16	0.2492
qinggan	−0.002	0.002	−1.03	0.3056

R-Squared = 0.654187; Adjusted R-squared = 0.608685; Prob = 0.0000.

Figure 7 shows OLS regression fitting residuals. The adjusted fit is 0.5565. Table 12 shows the significant relationship between the variables. P is selected as the value of 0.05 and it demonstrates that significant relationship between downloads and appearance requirements, functional requirements, and 3S. At the same time, the correlation coefficient between appearance requirements and downloads is −0.01, that is, the more appearance requirements are mentioned in user comments. Also, it has a negative impact on other users’ downloads behavior; In addition, the correlation coefficient between functional requirements and downloads is 0.004, indicating that the more functional requirements involved in the user’s comment information, the more other users download this APP. Table 13 shows the regression using the backward stepwise method. Comparing the results of OLS regression, it can be seen that the fit of the model improves when the variables related to downloads increase.

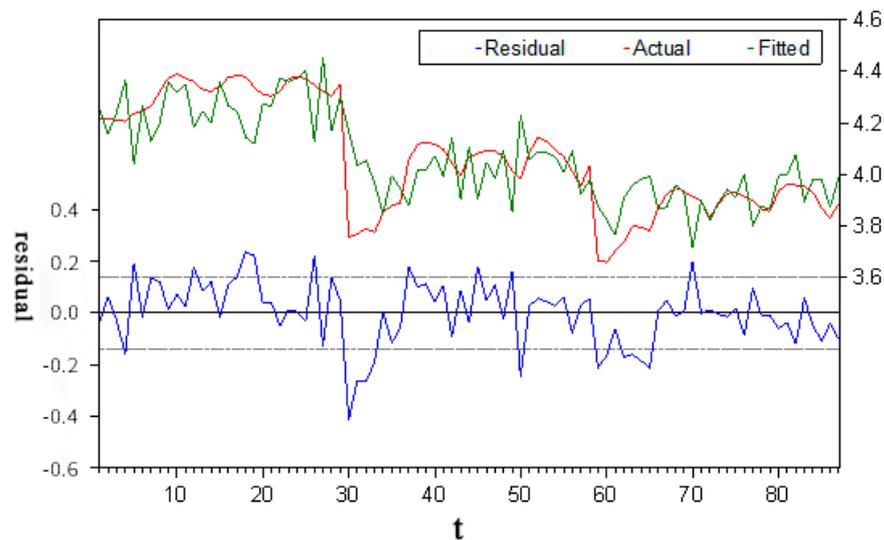


Figure 7. OLS regression fitting residuals.

We further analyze the variables significant under the quantile of {0.05,0.1,0.15,0.2,0.25,0.3,0.35,0.4,0.45,0.5,0.55,0.6,0.65,0.7,0.75,0.8,0.85,0.9,0.95}, the results are shown in Tables 14–16.

Table 14. Variables significant at 0.05 to 0.3 points.

Variable	(0.05) Prob	(0.1) Prob	(0.15) Prob	(0.2) Prob	(0.25) Prob	(0.3) Prob
Pos	0.0256 (−2.54)	—	0.0198 (−2.76)	—	—	—
Neg	—	—	—	—	—	—
fuwu	—	—	—	—	—	—
gongneng	0.0237 (0.003)	—	0.0094 (0.0035)	0.0169 (0.0035)	0.0032 (0.004)	0.0048 (0.0038)
qinggan	—	—	—	—	—	—
shipeidu	—	—	—	—	—	—
waiguan	—	—	—	0.0296 (−0.01)	—	—
wangluo	—	—	—	—	—	—
Lik0	—	—	—	—	—	—
Lik1	—	—	—	—	—	—
Lik2	—	—	—	—	—	—
Lik3	—	0.0368 (−1.23)	—	—	—	—
Lik4	0.0335 (−1.08)	0.0361 (−1.52)	—	—	—	—
Lik5	—	—	—	—	—	—
Lik6	0.0309 (−1.06)	0.0327 (−1.31)	—	0.0192 (−1.19)	—	—
1S	—	—	—	—	—	—
2S	—	—	—	—	—	—
3S	0.0001 (0.089)	0.0006 (0.078)	0.0063 (0.07)	—	—	—
4S	—	—	—	—	—	—
5S	—	0.0487 (0.003)	—	—	—	—

Note: Blank units indicate that they have not passed the significance test. Units with given data indicate the relationship between the *p*-value of the significance test and the regression value respectively.

Table 15. Variables significant at 0.35 to 0.6 points.

Variable	(0.35) Prob	(0.4) Prob	(0.45) Prob	(0.5) Prob	(0.55) Prob	(0.6) Prob
Pos	—	—	—	—	—	—
Neg	—	—	—	—	—	—
fuwu	—	—	—	—	—	—
gongneng	0.0343 (0.0035)	0.0321 (0.0033)	0.004 (0.0045)	0.0044 (0.0048)	0.002 (0.005)	0.0009 (0.005)
qinggan	—	—	—	—	—	—
shipeidu	—	—	—	—	0.0496 (−0.02)	0.0119 (−0.025)
waiguan	0.0335 (−0.0073)	0.0302 (−0.0073)	0.0036 (0.0092)	0.0052 (−0.0092)	0.0018 (−0.01)	0.0009 (−0.0091)
wangluo	—	—	—	—	—	—
Lik0	—	—	—	—	—	—
Lik1	—	—	—	—	—	—
Lik2	—	—	—	—	—	—
Lik3	—	—	—	—	—	—
Lik4	—	—	—	—	—	—
Lik5	—	—	—	—	—	—
Lik6	—	—	—	—	—	—
1S	—	—	—	—	—	—
2S	—	—	—	—	—	—
3S	—	—	—	—	—	—
4S	—	—	—	—	—	—
5S	—	—	—	—	—	—

Note: Blank units indicate that they have not passed the significance test. Units with given data indicate the relationship between the *p*-value of the significance test and the regression value respectively.

Table 16. Variables significant at 0.65 to 0.95 points.

Variable	(0.65) Prob	(0.7) Prob	(0.75) Prob	(0.8) Prob	(0.85) Prob	(0.9) Prob	(0.95) Prob
Pos	—	—	—	—	—	—	—
Neg	—	—	—	—	—	—	—
fuwu	—	—	—	—	—	—	0.0116 (0.014)
gongneng	0.0005 (0.005)	0.0009 (0.005)	0.0008 (0.004)	0.0003 (0.005)	0.001 (0.004)	0.01 (0.004)	0.0000 (0.004)
qinggan	—	—	—	—	—	—	0.0017 (−0.00932)
shipeidu	0.0044 (−0.03)	0.0051 (−0.025)	0.007 (−0.0225)	0.0059 (−0.022)	—	—	—
waiguan	0.0005 (−0.0092)	0.0006 (−0.0089)	0.0005 (−0.009)	0.0002 (−0.0092)	0.0005 (−0.01)	0.0031 (−0.0071)	0.0001 (−0.0077)
wangluo	—	—	—	—	—	—	—
Lik0	—	—	—	—	—	—	—
Lik1	—	—	—	—	—	—	0.0077 (1.67)
Lik2	—	—	—	—	—	—	—
Lik3	—	—	—	—	—	—	—
Lik4	—	—	—	—	—	—	0.025 (−0.8213)
Lik5	—	—	—	—	—	—	—
Lik6	—	—	—	—	—	—	—
1S	—	—	—	—	—	—	—
2S	—	—	—	—	—	—	—
3S	—	—	0.0425 (0.05)	0.0361 (0.048)	—	—	—
4S	—	—	—	—	—	—	—
5S	—	—	—	—	—	—	0.01 (−0.002)

Note: Blank units indicate that they have not passed the significance test. Units with given data indicate the relationship between the *p*-value of the significance test and the regression value respectively.

Due to limited space, only the slope equality test and symmetry test at point 0.25 are carried out, and the results are as shown in Tables 17 and 18. At this time, the Prob values are all less than 0.05, indicating that each quantile has passed the slope equality test and the symmetry test.

Table 17. Slope equality test at point 0.25.

Test Summary	Chi-Sq.Statistic (Chi-Squaretest)	Chi-Sq.d.f. (Degree of Freedom)	Prob. (Significance Value)
Wald Test	38.5	40	0.005

Table 18. Symmetry test at point 0.25.

Test Summary	Chi-Sq.Statistic (Chi-Squaretest)	Chi-Sq.d.f. (Degree of Freedom)	Prob. (Significance Value)
Wald Test	8.75	21	0.009

5.3. Analysis of Quantile Regression Results

From the results in Tables 14–16, it can be seen that the independent variables Neg, wangluo, Lik0, Lik2, Lik5, 1S, 2S, 4S are not significant at any quantile, indicating that these eight variables are not significant for users’ downloading behavior. From the explanation of the independent variables in Table 11, it can be seen that the negative comments represented by the independent variable Neg and the positive comments represented

by Pos are both values derived from the emotional tendencies in the user comments, so they are classified as the same type during the analysis. The wangluo variable represents network technology environment requirements; Lik0, Lik1, Lik2, Lik3, Lik4, Lik5, Lik6 variables all represent user satisfaction; 1S, 2S, 3S, 4S, 5S variables all represent user ratings. The following analysis of all the independent variables mentioned in Tables 14–16, where downloads with a quantile between 0.05 and 0.45 are defined as APPs with fewer downloads. Downloads with a quantile between 0.5 and 0.7 are defined as APPs with average downloads, and downloads with a quantile between 0.75 and 0.95 are defined as APPs with higher downloads. The specific results are as follows.

Emotional tendency of user comments: Compared with negative comments, positive comments have a more significant impact on users' downloading behavior. The independent variable Pos is significant at both the quantile points 0.05 and 0.15, and the coefficients are -2.54 and -2.76 , which indicates that for those APPs with fewer downloads, the user's positive comments have a counterproductive effect on the increase in downloads, while for those APPs with higher downloads, the user's positive comments have no effect on the increase in downloads. Comparing the results of the OLS test in Table 12, the independent variable Pos has no significant impact on the downloads, and the independent variable Pos in the backward stepwise regression in Table 13 has a significant impact on the downloads. It can be seen that OLS and backward stepwise regression impact on downloads compared to quantile regression. The significance test of the independent variable is not accurately portrayed according to the different distribution of the dependent variable, but only gives a total test result.

Network technology environment requirements: User comments rarely mention APP-related network technical environment such as lag, black screens, and logins. On the one hand, it shows that such problems are less common in the three APPs. On the other hand, it also shows that companies do not have to spend more on the network technical environment when upgrading products in the later stage. The existing technology is sufficient to support users' requirements.

User satisfaction: Lik0, Lik2, and Lik5 have no significant impact on the increase of user downloads, while Lik1, Lik3, Lik4, and Lik6 have a significant impact, and Lik1 only has positive impact in the 0.95 quantile. Lik3 only has negative impact in the 0.1 quantile, Lik4 has negative impact in the 0.05, 0.1, and 0.95 quantiles. Lik6 has negative impact on the 0.05, 0.1, and 0.2 quantiles. This shows that the comments of some dissatisfied users and very satisfied users only have a negative inhibitory effect on the low quantile samples. At the same time, some satisfied users' comments have a negative inhibitory effect on the low quantile and the highest quantile. Satisfied users' comments are only helpful for the highest quintile. Comparing the regression test results in Tables 12 and 13, Lik0~Lik6 are not significant in the OLS test, and Lik1 is not significant in the stepwise regression. It can be seen that OLS and backward stepwise regression are not accurate comparing to the quantile regression for the significance of independent variables.

User ratings: 1S, 2S, 4S have no significant impact on the increase in user downloads; 3S has a positive and significant impact on the 0.05, 0.1, 0.15, 0.75, and 0.8 quantiles; 5S has a significant impact on the 0.1 and 0.95 quantiles, and the regression coefficient changes from 0.003 to -0.002 . This shows that as the quantile increases, comments with a user score of 5S have a negative impact on downloads. This shows that for APPs with fewer downloads, 5S comments have a positive effect on the increase in downloads. For APPs with higher downloads, 5S has a negative effect on the increase in downloads.

Service requirement and emotional requirement: From the quantile regression results, it can be seen that both service requirement and emotional requirement have a significant impact on the 0.95 quantile. Among them, service requirement promotes APPs with high downloads, and emotional requirement has the opposite effect on downloads of APPs with high downloads. This means that companies can further improve product and service requirements, such as reducing the number of advertisements in English vocabulary APPs, and improving the role of customer service, without spending too much money on the

convenience, sensitivity, intelligence, fun, and comfort of the APP and other emotional requirements. In OLS and stepwise regression tests, service requirement and emotional requirement did not pass the significance test on APP downloads. In fact, these two types of independent variables have a significant impact on APP downloads at high quantile points. There is no significant correlation between the middle and low quantile points, which shows that OLS and backward stepwise regression are not accurate in characterizing the significance test of independent variables compared with quantile regression.

Functional requirements: According to the quantile regression results, functional requirements have a positive and significant impact on the low, middle, and high quantiles, which shows that functional requirements have a significant role in promoting the increase in downloads. When the downloads of English vocabulary APPs gradually increase, companies should strengthen the improvement of product functions, such as enriching the way of memorizing words, expanding the vocabulary database, and personalizing the words memorized every day. With the spread of APPs, companies need to continuously enrich functions to attract more users to download. This is also the same as the analysis results of key elements of user requirements in Section 4.2.3.

Adaptability requirements: Adaptability requirements have a significant negative impact on the mid-to-high quantiles of 0.55 to 0.8, which means that in the process of increasing downloads, the company's modification of software versions and modes will reduce user downloads. However, users are unwilling to try new usage models.

Appearance requirements: appearance requirements have a significant negative impact on the 0.2, 0.35~0.95 points. As the quantile increases, the absolute value of the regression coefficient first increases and then decreases, indicating that for the general English vocabulary APPs, users value the appearance design of the APPs. If the appearance design of English vocabulary APPs cannot attract users, the download increment for new users will be greatly reduced. For APPs with high downloads, improving the appearance design will not increase downloads.

6. Conclusions

In the previous literature studies, less consideration is given to the impact of user needs on consumer behavior. Therefore, this paper collects user experience data on three typical APPs (BaiCiZhan, MoMoBeiDanCi, BuBeiDanCi). By analyzing user comments to dig out user requirements, emotion and satisfaction, this paper constructs quantile regression equation affecting users' downloads behavior so as to give suggestions for further improvement, and proposes user demand factors that need to be improved for different downloads of APPs. The results show that:

- (1) Positive comments have a negative effect on the increase in downloads of APPs with fewer downloads, while having no effect on the increase in downloads of APPs with higher downloads. In addition, negative comments have no significant impact on downloads. When optimizing products in the future, companies should not only pay attention to negative comments but also pay more attention to the points mentioned in user positive comments.
- (2) Since users will refer more to the content of 5S user comments when downloading English vocabulary APPs, companies should focus on the content of 5S user comments when positioning user requirements later.
- (3) The comments of some dissatisfied users, very satisfied users, and some satisfied users have a negative effect on the promotion of English vocabulary APPs with lower downloads. Companies should focus on the needs and concerns of these three types of users.
- (4) The promotion and improvement of network technical environment requirements, emotional requirements, adaptability requirements, and appearance requirements will not have much effect on English vocabulary APPs with high downloads.
- (5) Companies can further optimize their functional requirements and service requirements to increase English vocabulary APPs downloads. For example, they can design

some interactive activities in the learning process, which can enhance the learning effect while enhancing the learning interest, so as to ensure the effectiveness of the learning content of the English vocabulary APP.

As can be seen from the above, for enterprises, when the number of APP users is relatively small, they should dig out the product problems mentioned in users' negative online comments, timely improve and enrich the product functions of APP to meet user needs. When APP has a large number of users, mining 5-star user comment information, improving product service demand (such as reducing advertising, improving customer service attitude, etc.), and functional demand (increasing ways to memorize words, expanding vocabulary, planning memory curve, etc.) will help increase the usage of new users. However, this paper still has the following limitations that need to be further improved:

- (1) The sample data can be further enriched. This paper mainly collects the public data of English vocabulary APPs such as MoMoBeiDanCi, Biaicizhan, and BuBeiDanCi. Later, data of other types of APPs can be obtained, and the competition analysis of same types can be carried out. In the later stage, a questionnaire can be set up to show the feelings of the users after using such software, in order to obtain more opinions and opinions on product improvement.
- (2) Features extracted by test mining can be enriched [32]. While constructing a user behavior variable system, the number of independent variables can be increased to obtain a more complete quantile regression equation.
- (3) For extraction of the user demand points that need to be improved in the APPs [33], quantile regression is adopted in this paper. Although it has passed the significance test, there are some non-strong significant correlations, which can be optimized by the non-parametric model in the later stage.
- (4) In the future research, we will further supplement the influence of graphical arrangement or explanation of functions of different types of mobile APPs on user experience.

Author Contributions: T.C. described the proposed framework, L.P. wrote the whole manuscript; G.C. collected data; J.Y. revised the English. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Social Science Foundation of China (Grant No. 20BTQ059), the China (Hangzhou) cross-border electricity business school, Contemporary Business and Trade Research Center and Center for Collaborative Innovation Studies of Modern Business of Zhejiang Gongshang University of China (Grant No. 14SMXY05YB), as well as the Characteristic & Preponderant Discipline of Key Construction Universities in Zhejiang Province (Zhejiang Gongshang University-Statistics).

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Shuib, L.; Shamshirband, S.; Ismail, M. A review of mobile pervasive learning: Applications and issues. *Comput. Hum. Behav.* **2015**, *46*, 239–244. [[CrossRef](#)]
2. Chang, C.; Liang, C.; Yan, C. The Impact of College Students' Intrinsic and Extrinsic Motivation on Continuance Intention to Use English Mobile Learning Systems. *Asia Pac. Educ. Res.* **2013**, *22*, 181–192. [[CrossRef](#)]
3. Jiang, P.; Lv, K. Research on customer experience evaluation of b2c e-commerce based on comprehensive fuzzy empirical method. *Bol. Tec. Tech. Bull.* **2017**, *55*, 267–274.
4. Shi, L.; Ming, Y. Mining frequent and infrequent features from Chinese customer reviews. *J. Theor. Appl. Inf. Technol.* **2013**, *48*, 193–199.

5. Bassett, G.; Koenker, R. Asymptotic Theory of Least Absolute Error Regression. *Publ. Am. Stat. Assoc.* **1978**, *73*, 618–622. [[CrossRef](#)]
6. Chen, T.; Peng, L.; Yin, X.; Rong, J.; Yang, J.; Cong, G. Analysis of User Satisfaction with Online Education Platforms in China during the COVID-19 Pandemic. *Healthcare* **2020**, *8*, 200. [[CrossRef](#)] [[PubMed](#)]
7. Chen, T.; Yin, X.; Peng, L.; Rong, J.; Yang, J.; Cong, G. Monitoring and Recognizing Enterprise Public Opinion from High-Risk Users Based on User Portrait and Random Forest Algorithm. *Axioms* **2021**, *10*, 106. [[CrossRef](#)]
8. Shan, X.; Zhang, X.; Liu, X. Research on User Portrait Based on Online Reviews. *Inf. Theory Pract.* **2018**, *41*, 99–105. (In Chinese)
9. Hussain, S.; Wang, G.; Jafar, R. Consumers' online information adoption behavior: Motives and antecedents of electronic word of mouth communications. *Comput. Hum. Behav.* **2018**, *80*, 22–32. [[CrossRef](#)]
10. Kang, Y.; Zhou, L. Rule-based Methods for Extracting Product Features from Online Consumer Reviews. *Inf. Manag.* **2017**, *54*, 166–276. [[CrossRef](#)]
11. Wong, C.; Qi, S. Tracking the evolution of a destination's image by text-mining online reviews—The case of Macau. *Tour. Manag. Perspect.* **2017**, *23*, 19–29. [[CrossRef](#)]
12. Yang, D. *User Needs Analysis and Research in Online Product Communities*; Tianjin University: Tianjin, China, 2017. (In Chinese)
13. Tirunillai, S.; Tellis, G.J. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *J. Mark. Res.* **2014**, *51*, 463–479. [[CrossRef](#)]
14. Yu, C. Mining Perspectives from Product Reviews Principle and Algorithmic Analysis. *Inf. Theory Pract.* **2009**, *32*, 124–128. (In Chinese)
15. Kasiviswanathan, S.; Ramalingam, D. Development and Application of User Review Quality Model for Embedded System. *Microprocess. Microsyst.* **2020**, *74*, 103029. [[CrossRef](#)]
16. Xu, X.; Wang, X.; Li, Y.; Mohammad, H. Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. *Int. J. Inf. Manag.* **2017**, *37*, 673–683. [[CrossRef](#)]
17. Wang, Y.; Lu, X.; Tan, Y. Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines. *Electron. Commer. Res. Appl.* **2018**, *29*, 1–11. [[CrossRef](#)]
18. Chen, T.; Peng, L.; Yin, X.; Jing, B.; Yang, J.; Cong, G.; Li, G. A Policy Category Analysis Model for Tourism Promotion in China During the COVID-19 Pandemic Based on Data Mining and Binary Regression. *Risk Manag. Healthc. Policy* **2021**, *13*, 3211–3233. [[CrossRef](#)] [[PubMed](#)]
19. Kim, S.; Baek, T.H.; Kim, Y.K. Factors affecting stickiness and word of mouth in mobile Applications. *J. Res. Interact. Mark.* **2016**, *10*, 177–192. [[CrossRef](#)]
20. Gonnmans, M.; Krishnan, K.; Scheffold, K. From Brand Loyalty to E-Loyalty: A Conceptual Framework. *J. Econ. Soc. Res.* **2001**, *3*, 43–58.
21. Cai, S.; Wang, W.; Zhang, W.; Cui, X. An Empirical Analysis on the Influencing Factors of Negative Word-of-Mouth Communication Willingness on the Internet. *Stat. Decis. Mak.* **2016**, *31*, 116–119. (In Chinese)
22. Shan, C.; Lu, Y. An Empirical Study on the Impact of Positive and Negative Online Reviews on the Initial Trust of C2C Merchants. *Libr. Inf. Work* **2010**, *54*, 136–140. (In Chinese)
23. Cai, K. *Mobile Store Research Based on User Adoption*; Huazhong University of Science and Technology: Hubei, China, 2010. (In Chinese)
24. Duan, W.; Gu, B.; Whinston, A. Do Online Reviews Matter an Empirical Investigation of Panel Data. *Decis. Support Syst.* **2008**, *45*, 1007–1016. [[CrossRef](#)]
25. Chen, Y. Auction Fever: Exploring Information Social Influences on Bidder Choices Cyber psychology. *Behav. Soc. Netw.* **2011**, *32*, 437–446.
26. Moon, S.; Bergey, P.K.; Iacobucci, D. Dynamic Effects among Movie Ratings, Movie Revenues, and Viewer Satisfaction. *J. Mark.* **2010**, *74*, 108–121. [[CrossRef](#)]
27. Li, H. *The Impact of Negative Online Reviews and Their Remedial Measures on Customers' Purchase Intention*; Donghua University: Shanghai, China, 2012. (In Chinese)
28. Changjo, Y.; Jonghee, P.; Deborah, J.M. Effects of Store Characteristics and In-Store Emotional Experiences on Store Attitude-Science Direct. *J. Bus. Res.* **1998**, *26*, 83–100.
29. Lu, P.; Hu, Y. Analysis of the Impact of Developer and User Behavior on Mobile APP Downloads. *Pract. Underst. Math.* **2019**, *49*, 108–116. (In Chinese)
30. Zhao, G.; Liu, W.; Liu, D. Variation Indexes Used to Determine the Influence of Dynamic User Demand on Product Redesign. *J. Chongqing Univ.* **2003**, *26*, 56–59. (In Chinese)
31. Li, Y.; Zhu, L. Research on Product Image Modeling Design Based on Associative Analysis. *J. Graph.* **2012**, *33*, 121–128. (In Chinese)
32. Chen, T.; Rong, J.; Yang, J.; Cong, G.; Li, G. Combining Public Opinion Dissemination with Polarization Process Considering Individual Heterogeneity. *Healthcare* **2021**, *9*, 176. [[CrossRef](#)]
33. Chen, T.; Wang, Y.; Yang, J.; Cong, G. Modeling multidimensional public opinion polarization process under the context of derived topics. *Int. J. Environ. Res. Public Health* **2021**, *18*, 472. [[CrossRef](#)]