Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/101/20231661

Effect of Different Chatbot Features on Customer Satisfaction in E-commerce

Jiangmeng Ding[1,a,*], Xing Wang[2,b], Xiang Chen[3,c], Mingze Li[4,d]

[1]Revelle College, University of California, San Diego, San Diego, 92023, United States
[2]College of Arts and Sciences, Boston University, Boston, 02215, America
[3]College of Electronic Information Engineering, South China Agricultural University, Guangzhou, 510642, China
[4]Victoria Shanghai Academy, Hong Kong, China
a. j3ding@ucsd.edu, b. tigerw@bu.edu, c. cx1378011297@gmail.com,
d. ps20153048@student.vsa.edu.hk
**corresponding author*

Abstract: By incorporating additional questions related to chatbots, we construct a questionnaire inspired by the work of previous researchers. This questionnaire aims to assess the relationship between various chatbot attributes, including usability, language style, and response time, and customer satisfaction within the e-commerce sector. To mitigate issues of multicollinearity and optimize the number of variables within a multiple linear regression model, we employ LASSO regression. Upon conducting hypothesis testing, our findings manifest a positive correlation between the overall chatbot features and customer satisfaction index. Specifically, an increase in the levels of language style and usability is consistent with increased levels of customer satisfaction, with language style exerting a larger effect.

Keywords: Chatbot, E-commerce, Customer Satisfaction

1. Introduction

The recently concluded COVID-19 pandemic has expedited the digital transformation of numerous industries. With the rapid advancements in e-commerce and artificial intelligence (AI), the applications of AI in e-commerce have become ever so prominent. From AI-powered predictive search and recommendations to chatbot customer service that is available 24/7, AI technology has become a crucial factor in improving customer's online shopping experience. As the digital world offers more and more consumer choices, businesses are constantly pushed by competition to attract and retain customers.

One essential factor in improving consumer satisfaction is online communication [1]. The instrument adopted by businesses to achieve this goal is AI chatbots. The adoption of AI chatbots in e-commerce has been shown by multiple studies to have positive impacts on consumer satisfaction in numerous ways*(Studies)*. Misischia's research [2], for instance, centers on discerning disparities in service quality between AI and human-assisted services. Their findings indicate that chatbots enhance service quality, shedding light on the rationale behind the increasing trend of enterprises opting for AI-powered customer service. Furthermore, J.-S. Chen has conducted a study on how the usability and responsiveness of a chatbot influence customers' extrinsic and intrinsic values [3]. Their findings reveal that usability positively impacts extrinsic values, and responsiveness has a favorable effect on intrinsic values. This paper provides valuable guidance for chatbot developers seeking avenues for enhancement. Nonetheless, the paper does not extensively explore the connection between customer satisfaction and chatbot features, thus limiting the intuitiveness of future efforts aimed at enhancing e-commerce service quality through chatbot utilization.

Within this context, several researchers are keen on elucidating strategies for enhancing chatbot performance in the realm of e-commerce. However, prior to such improvements, gaining insights into the intricate relationship between chatbot features and customer preferences is imperative. After examining multiple previous studies done on the topic of how consumer satisfaction can be improved by AI chatbots, we find that the many features of a chatbot affect consumer experience differently. M. Li's research further contributes to this discourse by revealing that customers prefer chatbots equipped with an informal language style over a formal one [4]. This underscores the influential role of chatbot language style in shaping customer satisfaction. Jenneboer [5] suggested that customers prefer chatbots to respond in a manner similar to real humans, and altering the language style of chatbots is a crucial step towards achieving this. Additionally, Nichifor also conducted a study on various aspects of chatbots [6], revealing that quality, response time, and the relevance of chatbots all exert varying degrees of influence on customer satisfaction. However, none of them made a comparison between different chatbot features. Our research will offer more insight into the finer details of how each chatbot feature affects consumer satisfaction and find the most influential feature of chatbots.

To gain a deeper understanding of which specific features wield the most significant influence on customer satisfaction, we have selected three pivotal chatbot attributes: Usability, responsiveness, and language style. Building upon prior research, we posit that these variables are the most crucial factors in shaping chatbot performance. In our study, we will employ multiple linear regression as the analytical framework to meticulously examine the influence of different chatbot features on customer satisfaction.

2. Methodology

2.1. Instrument Design

The primary method of data collection utilized in this study was a survey questionnaire. The questionnaire consisted of Likert-type scale questions, with response options ranging from 1 (strongly disagree) to 5 (strongly agree). However, during the data processing phase, a slight modification was made, and the scale was changed to -2 to 2. The survey questionnaire comprised questions that focused on the three features that we will be examining, namely language style, response time, and usability. Specifically, there were 9 questions related to usability, 4 questions targeting response time, 6 questions concerning language style, and 4 questions related to satisfaction. All the questions included in our survey were adapted from a previous paper (*Chen et al. (2021))*, which was one of the sources that we examined during our literature review. By incorporating these adjustments and drawing on established research in the field, we are aiming to gather comprehensive data on the various dimensions being investigated in our study.

2.2. Sample

We distributed the survey to a total of 87 participants, and remarkably, we received a 100% response rate with all 87 responses deemed valid after checking for missing or invalid values. Table 1 provides an overview of the sample profile, highlighting key demographic characteristics. Notably, the sample demonstrates a higher representation of male participants(58.62%) as compared to female participants (34.48%). In terms of age groups, the majority of respondents fell within the young-adults category

DOI: 10.54254/2754-1169/101/20231661

of 18-20 years old (43.68%), followed by teenagers below the age of 17 (26.44%). Regarding the level of education, the largest proportion of respondents are currently enrolled in college (52.87%), while the second largest group consists of high school students (19.54%). Lastly, in terms of nationality, the majority of respondents were from China (94.25%), while the other respondents come from the United States and Australia (5.75%). Here, we used snowball sampling. Despite possibly leading to biased samples, it was the most efficient way for us to receive valid data in a short span of time, while maintaining a sufficient sample space.

Table 1: Sample Profile

Characteristics	***No. of people***	***%***
Gender		
Male	51	58.62
Female	30	34.48
Others	6	6.90
Age (in years)		
Younger than 17	23	26.44
18-20	38	43.67
21-29	22	25.29
30-39	2	2.30
40-49	0	0.00
50-59	0	0.00
Older than 60	2	2.30
Highest Level of Education		
Lower than high school	17	19.54
High school degree or equivalent	15	17.24
Some college but no degree	46	52.87
Associate degree	0	0.00
Bachelor degree	6	6.90
Graduate degree	3	3.45
Nationality		
Chinese	82	94.25
Other	5	5.75

2.3. Process of Data Analysis

In our research, we employed multiple linear regression analysis to investigate and assess our hypothesis. We selected multiple linear regression as it was the most appropriate statistical method to suit our research design. The model examines three independent variables, being language style (L_i), response time (R_i), and usability (U_i), respectively, whereas the dependent variable being examined is customer satisfaction (S_i), in which subscript i represents an individual respondent who completed the survey with i = [1, …, 87]. Since we partitioned our questionnaire into distinct sections corresponding to each independent variable, each variable would be written in one indicator equation through a calculation of the average. In which we have:

DOI: 10.54254/2754-1169/101/20231661

$$SS_i = \frac{\sum_{i=1}^{4} S_i}{4} \tag{1}$$

$$LL_i = \frac{\sum_{i=1}^{6} L_i}{6} \tag{2}$$

$$UU_i = \frac{\sum_{i=1}^{9} U_i}{9} \tag{3}$$

$$RR_i = \frac{\sum_{i=1}^{4} R_i}{4} \tag{4}$$

where the LHS is the final indicator equation; the denominator referring to the total number of question that examines that particular feature; LL_i, RR_i, UU_i, SS_i each represent their respective variable for $i = [1, \dots, n]$, where n refers to the total number of questions in each sector. As a result, we gain the following final multiple linear regression model:

$$S_i = \beta_0 + \beta_1 \cdot L_i + \beta_2 \cdot U_i + \beta_3 \cdot R_i + \varepsilon_i \tag{5}$$

where ε_i represents the residual for individual i.

2.4. Data Cleaning

To ensure the validity of our findings, it is crucial to address the potential multicollinearity amongst our independent variables. Notably, our independent variables, from a general sense, demonstrate minimal intercorrelation. Nonetheless, we still conducted deeper investigation, in which we divided the independent variables into its corresponding questions from the survey, in which we are ensuring that a question examining one independent variable is not interdependent with another question that examines another variables, e.g. a question that examines response time should not be intercorrelated with a question that examines language style. Consequently, we established a threshold and selectively emitted the questions that exceeded this value.

To establish a threshold for selecting questions, we considered the desired level of intercorrelation that would be acceptable for our research topic and field. Taking into account the specific objectives of our research and scrutinizing relevant literature, we established a threshold value that would indicate a level of correlation beyond which intercorrelation between questions would be considered problematic to our research. Ultimately, we decided to set the threshold as 0.6, which was a value that was able to sift out problematic intercorrelations, while maintaining the quantity of the data set. Furthermore, we implemented a LASSO regression, which we will discuss in detail in the latter parts, in order to solve this issue. As a result of our data cleaning, we omitted a total of 10 questions: consisting of 5 questions on usability, 1 question on language style, all 4 questions on response time. Practically, we found response time to be a problematic variable because of its high p-value of 0.9 and intercorrelation; therefore, we ultimately decided to withdraw response time from our dataset. Through removing these highly correlated questions, we aimed to minimize the potential bias and inaccuracy or confounding effects caused by interdependence among the independent variables.

3. Results

We first examine the relationship between three chatbot features and customer experience with the specification of equation (5). As shown in Figure 1, the adjusted R-squared reveals that approximately 79% of the variation of the customer satisfaction can be explained by the independent variables in this linear regression model. Moreover, the overall p-value for the model is close to 0, which further shows that the result is significant. However, the p-value for responsiveness nearly approaches one, which manifests that the -0.01 coefficient for responsiveness is insignificant and should be rejected.

It is possible that the result is generated by multicollinearity between the independent variables. Hence, it is necessary to remove some independent variables for model precision. Despite that multicollinearity cannot be eliminated, the problem could be mitigated through The Least Absolute Shrinkage and Selection Operator (LASSO) regression.

```
Call:
lm(formula = sat ~ use + res + lan, data = sdata)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.88663 -0.18790 -0.05293  0.19430  1.09793 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.17302    0.06443   2.685  0.00875 ** 
use          0.30318    0.09667   3.136  0.00237 ** 
res         -0.01064    0.08631  -0.123  0.90218    
lan          0.67656    0.07317   9.246  2.1e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3782 on 83 degrees of freedom
Multiple R-squared:  0.7956,    Adjusted R-squared:  0.7882 
F-statistic: 107.7 on 3 and 83 DF,  p-value: < 2.2e-16
```

Figure 1: R code output of multiple linear regression model (5)

LASSO regression is a regularization technique that reduces the complexity and selects features of the linear regression model to alleviate multicollinearity. It adds an L1 penalty norm to encourage the sparse coefficient estimates in the model. The similar model using L2 as a norm to deal with multicollinearity, Ridge Regression, is not being used in this case, as it is unable to automatically delete features. On the other hand, through forcing the less important variables, LASSO shrinks their coefficients to zero and eliminates these variables. As the complexity of the model reduces, LASSO guarantees a more stable coefficient estimation and prevents the problem of overfitting by lowering the fitting noise. The model requires a threshold (alpha) to control the sparsity of the coefficients in the model. In this case, we set the default value of alpha equals to 1 and add all of the features (specified as different question notations in the questionnaire) as independent variables. In Figure 1, lambda represents a coefficient of the penalty strength. The plot indicates a tendency for the independent variable features to change as lambda increases. Lambda.min, illustrated as the left gray line in Figure 1, is the optimal number of independent variables to be contained in the model. Based on the results presented in Figure 2, it is evident that the LASSO regression, with a lambda.min value of 10, retains ten independent variables, namely being: (intercept), U5, U6, U7, U9, L2, L3, L4, L5, L6; while the LASSO regression simultaneously removes all insignificant indices to zero. As seen, all responsiveness questions, {R1, R2, R3, R4}, are all eliminated from the model, which further evinces our conclusion that responsiveness is uncorrelated to satisfaction.

DOI: 10.54254/2754-1169/101/20231661

```
20 x 1 sparse Matrix of class "dgCMatrix"
                     s1
(Intercept) 1.07980676
U1          .
U2          .
U3          .
U4          .
U5          0.04973718
U6          0.31245203
U7          0.12913070
U8          .
U9          0.47481800
L1          .
L2          0.63735468
L3          0.10064587
L4          0.09546154
L5          0.50626874
L6          0.95716966
R1          .
R2          .
R3          .
R4          .
```

Figure 2: LASSO Regression analysis

Based upon this, the new index for usability is calculated by the mean value of four variables, and the new index for language style is calculated by the mean of five variables. The updated specification includes two independent variables:

$$S_i = \beta_0 + \beta_1 \cdot L_i + \beta_2 \cdot U_i + \varepsilon_i \tag{6}$$

where the dependent variable is still composed of four features categorized by questions. As demonstrated in Figure 3, the adjusted R-squared suggests that approximately 82% of the dependent variables can be explained by the new model, which is slightly more accurate than that in equation (5). The p-value is still close to zero, which testifies the significance of the model. Similar to the previous results, both usability and language style are positively correlated to satisfaction. The low p-value still indicates the high significance for the coefficient of language style that one unit increase in the index of language style rating is associated with approximately 69% increase in customer satisfaction. Notably, the significance of the coefficient of usability has improved compared to the previous one, as the p-value reduces to a lower level. Besides, one unit increase in usability is associated with approximately 30% increase in customer satisfaction. This further demonstrates that language style positively correlates with customer satisfaction than usability.

DOI: 10.54254/2754-1169/101/20231661

```
Call:
lm(formula = sat ~ use + lan, data = ndata)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.66690 -0.13581 -0.05187  0.16825  1.12607 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.11627    0.04851   2.397   0.0188 *  
use          0.29600    0.06058   4.886 4.87e-06 ***
lan          0.68829    0.06599  10.430  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3507 on 84 degrees of freedom
Multiple R-squared:  0.822,	Adjusted R-squared:  0.8178 
F-statistic:   194 on 2 and 84 DF,  p-value: < 2.2e-16
```

Figure 3: R code output of multiple linear regression model (6)

4. Conclusion

Examining a customer satisfaction survey regarding e-commerce on 87 participants on their ratings upon chatbot features online, we find that improvements in chatbot features correlate an enhancement in customer satisfaction. In particular, we delve into features including usability, responsiveness, and language style of the chatbots. With the help of LASSO regression to settle the multicollinearity problem, we conclude that responsiveness is insignificantly correlated with customer satisfaction, while an increase in language style and usability is associated with an increase in customer satisfaction index. Additionally, language style may possess a more positive correlation compared to usability.

5. Limitations and Future Study

The first limitation lies within the data collection process of this study. This study uses a snowball sampling and yielded a rather small sample size which, to some extent, hinders the representativeness and generalizability of the sample to the broader population. The lack of representativeness of this study is also reflected in the skewed demographic with almost half of all respondents being young adults of ages 18-20. Future studies may utilize a systematic sampling technique to ensure an even spread in the demographics of respondents — a random sampling technique was done in previous studies and also resulted in a skewed demographic with overly many young adults — while expanding on a greater sample space, ensuring the universality of the study. Secondly, the language style segment of the survey is adapted from a post experiment questionnaire, which may hinder its reliability under our method of data collection. Additionally, since our respondents were primarily Chinese, we included a translated version of the adapted questionnaire to ensure that there was no confusion for our respondents; however, this calls for the need of triangular validation, which was unable to be done within the restricted time frame of our study. Finally, the results of this study are limited to correlation, as the data collected did not provide the conditions to perform valid causality analysis. Therefore, the conclusions that can be drawn regarding the relationships between the chatbot features and consumer satisfaction is limited to an extent. Future studies may focus on analyzing the causal relationships between the chatbot features and consumer satisfaction. Furthermore, although the multicollinearity problem was mitigated in our study using the LASSO regression, future studies could avoid this problem in the first by conducting pilot tests with the survey questions. Lastly, the type of chatbot interaction in our study was not measured, so future studies could also focus on the chatbot model and conduct analysis using that as a starting point.

Acknowledgment

Jiangmeng Ding, Xing Wang, Xiang Chen, and Mingze Li contributed equally to this work and should be considered co-first authors.

References

[1] *Nichifor, E., Trifan, A. and Nechifor, E.M., 2021. Artificial Intelligence in Electronic Commerce: Basic Chatbots and the Consumer Journey. Amfiteatru Economic, 23(56), pp. 87-101.*

[2] *J.-S. Chen, T.-T. Le, D. Florence. Usability and responsiveness of artificial intelligence chatbot on online customer experience in e-retailing. Int. J. Retail Distrib. Manag., 49(11)(2021), pp.1512-1531, 10.1108/IJRDM-08-2020-0312.*

[3] *Nguyen, H.V.; Vu, T.D.; Nguyen, B.K; Nguyen, T.M.N.; Do, B.; Nguyen, N. Evaluating the Impact of E-Service Quality on Customer Intention to Use Video Teller Machine Services. J. Open Innov. Technol. Mark. Complex. 2022, 8, 167.*

[4] *Misischia, C. V., Poecze, F., & Strauss, C. (2022). Chatbots in customer service: Their relevance and impact on service quality. Procedia Computer Science, 201, 421-428.*

[5] *Jenneboer, L.; Herrando, C.; Constantinides, E. The Impact of Chatbots on Customer Loyalty: A Systematic Literature Review. J. Theor. Appl. Electron. Commer. Res. 2022, 17, 212-229.*

[6] *M. Li, R. Wang, 2023. Chatbots in e-commerce: the effect of chatbot language style on customers' continuance usage intention and attitude toward brand. J. Retailing Consum. Serv., 71 (2023), Article 103209.*