

Factors Impacting the Quality of User Answers on Smartphones

Ivano Bison, Haonan Zhao*

DISI, University of Trento, Italy.

Abstract

So far, most research investigating the predictability of human behavior, such as mobility and social interactions, has focused mainly on the exploitation of sensor data. However, sensor data can be difficult to capture the subjective motivations behind the individuals' behavior. Understanding personal context (e.g., where one is and what they are doing) can greatly increase predictability. The main limitation is that human input is often missing or inaccurate. The goal of this paper is to identify factors that influence the quality of responses when users are asked about their current context. We find that two key factors influence the quality of responses: user reaction time and completion time. These factors correlate with various exogenous causes (e.g., situational context, time of day) and endogenous causes (e.g., procrastination attitude, mood). In turn, we study how these two factors impact the quality of responses.

Keywords

Human-Machine interaction, Context, Quality of user answers, Diversity awareness

1. Introduction

Smartphones facilitate pervasive and continuous interaction between humans and machines by enabling timely notifications and data collection, while simultaneously gathering sensor data such as Bluetooth, GPS, and accelerometer. This creates opportunities for smartphone applications to learn how to ask the right questions at the right moment and in the right context, while also verifying the accuracy and consistency of user input. Previous research, as demonstrated in [1, 2], has explored these capabilities and their potential to establish a symbiotic relationship between humans and machines, wherein machines can learn about every aspect of daily life. This work holds high-impact applications in Artificial Intelligence, Psychology (e.g., the Experience Sampling Method), and the Social Sciences. Additional research on this topic can be found in [3, 4, 5].

However, an outstanding issue remains with user-provided answers as they often fail to meet the necessary quality standards. There are many possible reasons for this, as described in Furnham's work [6], including recall bias, which occurs when participants cannot recall previous events [7]. These issues can lead to various problems, as highlighted by Schneier [8].

HHAI-WS 2023: Workshops at the Second International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 26–27, 2023, Munich, Germany

*Corresponding author.

✉ ivano.bison@unitn.it (I. Bison); haonan.zhao@unitn.it (H. Zhao)

🆔 0000-0002-9645-8627 (I. Bison); 0000-0003-0825-7188 (H. Zhao)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

In ESM, fixed time intervals are typically used for asking questions, which may result in sending questions at inopportune moments, disrupting people's daily lives and leading to low-quality data collection and, consequently, a limited number of responses. Much research has focused on understanding this phenomenon and improving the response rate by minimizing missed answers, as shown by the works in [9, 10], but they did not find which factors can impact the quality of answers.

The goal of this paper is to provide an in-depth study of the parameters that influence the quality of answers. By quality, we refer to a low number of mistakes. Different individuals possess diverse understandings of questions and context, which leads to variations in the quality of their answers. In this study, we discovered that the correctness of an answer is directly negatively affected by the reaction time, which, in turn, depends on the characteristics of the respondent and the situational context. Thus, controlling the reaction time becomes crucial for establishing meaningful hybrid machine-artificial intelligence.

2. Methodology

The reasons for respondents making mistakes when providing answers can be categorized into four main causes: (i) the context in which the answer is provided (e.g., being alone, at university, using social media, on Monday) [11, 12]; (ii) the cognitive task involved in the response process [13, 14]; (iii) a range of psycho-physical attitudes, such as mood, procrastination, response behaviors, and habits [14, 15]; and (iv) technical problems related to the functioning of the smartphone. These causes are independent of each other but interact over time and across respondents. Based on this understanding, we propose a hypothesis that the quality of answers can be influenced by the reaction time (time elapsed between receiving a question and starting to answer) and completion time (time taken to complete an answer). Both the reaction time and completion time can be affected by the four categories of mistakes described above.

Initially, we hypothesized that longer reaction times could negatively affect answer quality due to memory recall issues. The longer the time elapsed since the request, the greater the risk of memory-related problems and, consequently, making mistakes in the answer. Additionally, we hypothesized that longer completion times could also have a negative influence on answer quality. Furthermore, the combined effect of reaction and completion times could further amplify their negative impact on answer quality.

According to our hypothesis, the primary challenge of our study is to determine how variations in reaction time and completion time contribute to errors in answers. However, solving this problem is complex due to the causal relationship that exists between reaction and completion time in a temporal sequence. To address this issue, we have chosen a causal model [16, 17], which establishes a hierarchy that considers the temporal order and logical derivation of variables along the pathway connecting explanatory features and objectives. In this paper, we assess answer correctness by examining the temporal consistency between the respondent's indicated location at home and the GPS position of the phone at the time of notification.

3. Results and Descriptions

To achieve the goal of this paper, we analyzed the SmartUnitn 2 database [18], which designed an experiment that lasted four weeks and involved students from the University of Trento. Throughout the experiment, the iLog application [19] was used to ask a set of questions diachronically, enabling the selective asking of multiple questions for situational context, which defined in [20, 21]. And then collecting sensor data from mobile phones.

Table 1

Chain of errors: Multilevel structural equation model and a Structural Equation model.

Features	ML-SEM Model (β parameters)	SEM Model (β parameters)
<i>M1: Home phone distance (meter)</i>		
Completion time	0.2401***	0.0863***
Reaction time	0.4725**	0.0238**
GPS accuracy	0.3439**	0.0237**
Constant	7.3587	0.0457
<i>M2: Unanswered Questions (count)</i>		
Reaction time	0.0252***	0.8809***
Constant	-0.1165*	-0.0691
<i>M3: Reaction time (minutes)</i>		
User Characteristics: Procrastination	1.0325***	0.1216***
User Characteristics: Mood	1.4515***	0.0168*
Event context: Activity	0.5991***	0.0530***
Study time	1.6248***	0.1127***
Study time2	-0.3105***	-0.0716***
Social context: Alone vs not alone	-13.5724***	-0.1170***
Constant	18.7450***	0.3596***
<i>M4: Completion time (second)</i>		
Pending notification (count)	-0.4689***	-0.0946***
Question delivery delay time (sec.)	-0.0012**	-0.0212*
Event context: Activity	-0.2697***	-0.1784***
Social context: Alone vs not alone	-2.521***	-2.521***
Study time	-0.1818***	-0.0844***
Study time2	0.0203***	0.0357***
User Characteristics: Procrastination	0.0766***	0.0644***
Constant	13.7792***	1.7036***

Note: "P-value", providing a measure of the statistical significance of that feature compared to the reference feature. Here "***" means $p < 0.1$, namely, relatively weak (statistical) significance; "**" is $p < 0.05$, a strong significance; "*" is $p < 0.01$, a very strong significance.

We study the correctness of the user answers by building a causal model able to capture both the direct and indirect effects of the various features. We compute the correctness of answers by comparing the distance from home, as computed by the device via the GPS information, any time the respondent declares she is at home. Notice that each participant was asked to provide their home address, we then used google maps were compute their home GPS coordinates; Google maps were used to calculate the GPS coordinates of the house.

The results are reported in Table 1. Horizontally we have a path model composed of four models (*M1-M4*). The variable written in italics near the name of the model is the dependent variable, the variables listed below are the input variables. In both models, we report the correlation (i.e., "Coeff β ") between input and dependent variable, for each model in *M1-M4*. Out of the two models, the first is a Multilevel Structural Equation Model (ML-SEM) [22], and a Structural Equation Model (SEM) [23]. Both models (MSEM, SEM) support the idea of a chain of events on answer quality. In fact, the input variables are (mostly strongly) relevant and the fit indices show that the model fits the observed data very well (RMSEA= 0.022; SRMR=0.013; $\chi^2=95.41$ (20); $R^2 = 0.127$). Moreover, the model explains 13.0% of fluctuation in answer quality.

In *M1*, completion and reaction time correlate positively with "Home \Leftrightarrow phone distance" controlled by the smartphone accuracy. Their interaction has a series of negative consequences on answer quality. In fact, a longer reaction time has two effects. The first is a direct effect on increasing the distance between "Home \Leftrightarrow phone". The second is indirect, through the increase of the "pending notifications" (*M2*), which, in turn, induce a shorter completion time which in turn decreases the error. This is not a contradictory result, but the twofold way in which an error can appear.

Our results prove that subjective annotations present a certain degree of error due to both exogenous and endogenous factors (*M3, M4*) affecting the quality of responses. Context history, cognitive ability, attention, effort, motivation, burden, procrastination, mood, and technical problems can play a role in terms of raising the probability of stopping the interaction with the machine, of not complying with the interaction protocol, of a decrease in the level of attention and, consequently, they can cause a decrease of the accuracy of answers.

4. Conclusion

This paper has investigated the effects of various factors on the quality of answers, in terms of missed answers as well as their correctness. The key results of this paper are of two kinds. The first kind is that in the future the researcher's attention should be placed on several factors related to: (a) controlling the situational and temporal context to find the best moment for administering a notification; (b) focusing on the human-machine interaction not only on the layout of the apps, but on the structure and order of the response alternatives, the ease of filling in, and finally on the support of the machine to help respond so as to reduce the response time and improve its quality. The second kind of results is related to the cognitive and more generally psychosocial traits of the respondents. It is clear that not all subjects are cooperative and follow the research protocols carefully. In the future, it will be a matter of finding what and how cognitive factors act differently and how to extrapolate their data and replace missing data from the few and fragmented data provided. The future work will focus on how to minimize the reaction time by establishing the best moment for asking a question.

Acknowledgments

The research by Ivano is funded by the European Union's Horizon 2020 FET Proactive project "WeNet – The Internet of us", grant agreement No 823783. The work by Haonan receives funding from the China Scholarships Council (No.202107820038).

References

- [1] A. Bontempelli, S. Teso, F. Giunchiglia, A. Passerini, Learning in the wild with incremental skeptical gaussian processes, in: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI), 2020.
- [2] A. Bontempelli, M. R. Britez, X. Li, H. Zhao, L. Erculiani, S. Teso, A. Passerini, F. Giunchiglia, Lifelong personal context recognition, 2022. URL: <https://arxiv.org/abs/2205.10123>. doi:10.48550/ARXIV.2205.10123.
- [3] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, A. T. Campbell, Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones, in: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing, 2014, pp. 3–14.
- [4] F. Giunchiglia, X. Li, M. Busso, M. Rodas-Britez, A context model for personal data streams, in: Web and Big Data: 6th International Joint Conference, APWeb-WAIM 2022, Nanjing, China, November 25–27, 2022, Proceedings, Part I, Springer, 2023, pp. 37–44.
- [5] X. Li, M. Rodas-Britez, M. Busso, F. Giunchiglia, Representing habits as streams of situational contexts, in: Advanced Information Systems Engineering Workshops: CAiSE 2022 International Workshops, Leuven, Belgium, June 6–10, 2022, Proceedings, Springer, 2022, pp. 86–92.
- [6] A. Furnham, Response bias, social desirability and dissimulation, *Personality and individual differences* 7 (1986) 385–400.
- [7] M. Porta, *A dictionary of epidemiology*, Oxford university press, 2014.
- [8] B. Schneier, *Secrets and lies: digital security in a networked world*, John Wiley & Sons, 2015.
- [9] N. van Berkel, J. Goncalves, S. Hosio, Z. Sarsenbayeva, E. Velloso, V. Kostakos, Overcoming compliance bias in self-report studies: A cross-study analysis, *International Journal of Human-Computer Studies* 134 (2020) 1–12.
- [10] V. Mishra, B. Lowens, S. Lord, K. Caine, D. Kotz, Investigating contextual cues as indicators for ema delivery, in: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, 2017, pp. 935–940.
- [11] P. J. Lavrakas, T. N. Tompson, R. Benford, C. Fleury, Investigating data quality in cell phone surveying, in: annual American Association for Public Opinion Research conference, Chicago, Illinois, 2010, pp. 13–16.
- [12] A. Wenz, Do distractions during web survey completion affect data quality? findings from a laboratory experiment, *Social Science Computer Review* 39 (2021) 148–161.

- [13] J. A. Krosnick, Response strategies for coping with the cognitive demands of attitude measures in surveys, *Applied cognitive psychology* 5 (1991) 213–236.
- [14] P. Lynn, O. Kaminska, The impact of mobile phones on survey measurement error, *Public Opinion Quarterly* 77 (2013) 586–605.
- [15] B. Read, Respondent burden in a mobile app: Evidence from a shopping receipt scanning study, in: *Survey Research Methods*, volume 13, European Survey Research Association, 2019, pp. 45–71.
- [16] P. W. Holland, Statistics and causal inference, *Journal of the American statistical Association* 81 (1986) 945–960.
- [17] J. J. Hox, et al., Multilevel regression and multilevel structural equation modeling, *The Oxford handbook of quantitative methods 2* (2013) 281–294.
- [18] I. Bison, F. Giunchiglia, M. Zeni, E. Bignotti, M. Busso, R. Chenu-Abente, Trento 2018 - an extended pilot on the daily routines of university students, DataSet soon to be available at <https://ri.internetofus.eu>, 2021. DataScientia dataset descriptors.
- [19] M. Zeni, I. Zaihrayeu, F. Giunchiglia, Multi-device activity logging, in: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication.*, ????
- [20] F. Giunchiglia, Contextual reasoning, *Epistemologia*, special issue: I Linguaggi e le Macchine 16 (1993) 345–364.
- [21] F. Giunchiglia, E. Bignotti, M. Zeni, Personal context modelling and annotation, in: *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, IEEE, 2017, pp. 117–122.
- [22] S. Rabe-Hesketh, A. Skrondal, A. Pickles, Generalized multilevel structural equation modeling, *Psychometrika* 69 (2004) 167–190.
- [23] S. Rabe-Hesketh, A. Skrondal, X. Zheng, Multilevel structural equation modeling, in: *Handbook of latent variable and related models*, Elsevier, 2007, pp. 209–227.