International Conference on Human-Computer Interaction
HCII 2021: HCI International 2021 - Posters.
Communications in Computer and Information Science, volume 1420, pp. 400-407
DOI:10.1007/978-3-030-78642-7_54

Sharing Secrets with Agents: Improving Sensitive Disclosures using Chatbots

Oliver Buckley[1], Jason R. C. Nurse[2], Natalie Wyer[1], Helen Dawes[3], Duncan Hodges[4], Sally Earl[1], and Rahime Belen Saglam[2]

[1] University of East Anglia, Norwich, UK
o.buckley@uea.ac.uk
[2] University of Kent, Canterbury, UK
[3] Oxford Brookes University, Oxford, UK
[4] Cranfield University, Defence Academy of the UK, Shrivenham, UK

Abstract. There is an increasing shift towards the use of conversational agents, or chatbots, thanks to their inclusion in consumer hardware (e.g. Alexa, Siri and Google Assistant) and the growing number of essential services moving online. A chatbot allows an organisation to deal with a large volume of user queries with minimal overheads, which in turn allows human operators to deal with more complex issues. In this paper we present our work on maximising responsible, sensitive disclosures to chatbots. The paper focuses on two key studies, the first of which surveyed participants to establish the relative sensitivity of a range of disclosures. From this, we found that participants were equally comfortable making financial disclosures to a chatbot as to a human. The second study looked to support the dynamic personalisation of the chatbot in order to improve the disclosures. This was achieved by exploiting behavioural biometrics (keystroke and mouse dynamics) to identify demographic information about anonymous users. The research highlighted that a fusion approach, combining both keyboard and mouse dynamics, was the most reliable predictor of these biographic characteristics.

Keywords: Chatbot · Conversational Agent · Disclosure · Biometrics · Keystroke dynamics · Mouse dynamics · Information inference

1 Introduction

Conversational agents, also known as chatbots, are applications that interact with users in increasingly empathetic and humanistic ways. These technologies have seen a significant rise in popularity and prevalence, in part thanks to assistants such as Siri and Alexa being embedded in consumer hardware. While these assistants will typically focus on everyday tasks and enquiries there is a growing shift to their use in contexts where users are required to disclose sensitive information, for example, in healthcare.

The ubiquity and uptake of chatbots has continued to grow, despite a number of key questions remaining unanswered. These include areas such as how

a chatbot's characteristics may be used to maximise the depth and accuracy of disclosures that are linked to sensitive information. This is particularly pertinent when considering applications where honest and frank disclosures are prerequisites for an improved service outcome. Additionally, there are a number of privacy and security concerns associated with chatbots, which need to be appropriately managed to enable a better service provision [11].

The Platform for Responsive Conversational Agents to Enhance Engagement and Disclosure (PRoCEED) project aims to address these issues, and provide a framework for delivering agents that are both effective and efficient in their interactions. In this paper we provide a summary of the research to date, along with key findings in relation to agent design and report on two key studies. The remainder of the paper is structured as follows: Section 2 provides an overview of related material, Section 3 outlines the methodology used in the studies, Section 4 reports on the initial results and findings and finally, Section 5 discusses our conclusions and future work.

2 Related Work

As our societal dependence on technology increases, then so does the need for organisations to develop communications that are capable of managing the increasing volume of users. Currently, chatbots are largely seen as a tool for first line customer support, offering a balance between limited organisation resources and providing the users with efficient and direct answers. For example, Juniper Research [9] estimates that annual cost savings derived from the adoption of chatbots could reach $7.3 billion globally by 2023.

The increasing trend towards the use of chatbots, by both government and industry, highlights the need for a greater appreciation of how individuals interact with them and the factors that determine the efficiency of this engagement. This is crucial if they are to increase service effectiveness and efficiency, given the reluctance, historically, to engage with agents [3]. However, a significant portion of research in this domain focuses on creating increasingly human-like agents [7], and developing methods that maximise an individual user's engagement [12]. While these represent important aims, they highlight a lack of research focused on understanding which aspects of an agent encourage user engagement and which of these aspects support deeper and more accurate disclosures.

There are a range of factors that influence how willing individuals are to engage and disclose the correct information to agents, which can relate to the agent itself (e.g. humanity, avatar), the medium (e.g. text, voice) or the individual (e.g. age, need). The literature provides indications about the features that are likely to affect engagement, with the perceived 'humanity' of the agent being vital. Epley et al. [6] suggest that adapting the characteristics of agents makes anthropomorphism more likely. This anthropomorphisation and its nature is something that is determined by the attributes of the perceiver. Specifically, individuals who are more motivated to engage in social interaction (e.g., the chronically lonely) or those who are more motivated to establish control over

their environments (e.g., those wishing to show expertise in their domain) have been shown to be more prone to anthropomorphise agents [5].

Existing research on the use of chatbots for the disclosure of sensitive information has often been limited by the technology. Elmasri and Maeder [4] developed a chatbot to engage with 18-25-year olds to understand alcohol consumption. While the work found that individuals did disclose small amounts of information, and were generally positive about the agent they were largely frustrated by the technical solution. Similarly, in the education sector Bhakta et al. [2] present research that used simple chatbots to cover a range of sensitive topics such as sexual health, drug use and plagiarism. This work found that the use of a chatbot resulted in more words than a standard non-interactive survey, which was considered to be a greater engagement. While more words may constitute increased engagement, the elicitation of more sensitive information requires a more nuanced approach.

The research in this project is informed by broader research into conversational agents and user interactions. Hill et al. [7] concluded that humans communicate differently when they know they are directly interacting with a machine. Their experiments revealed that participants sent twice as many messages to chatbots as they did to humans. There is research (e.g. Jenkins et al. [8]) that suggests that users expect conversational agents to behave and communicate in a very human-like way, whereas other studies (Bailenson et al. [1]) highlight an increase in the depth of disclosures when engaging with a more machine-like agent.

In this project we focus on understanding the depth of disclosure and how the characteristics of both the agent and the user can impact this level of information sharing. The studies in this paper highlight the methods used to quantify the sensitivity of particular disclosures as well as our approach to automatically understand the characteristics of the user.

3 Methodology

The project to date consists of two studies designed to better understand the core characteristics of an agent when considering disclosure, and to inform agent design. Firstly, we discuss a 491-individuals study that aimed to investigate the types of information that people with different characteristics are comfortable sharing with a chatbot vs. a human. The second study presents a 240-person experiment to infer demographic information from mouse and keyboard usage.

3.1 Understanding information disclosure

The perceived sensitivity of personal information is central to the privacy concerns and information disclosure behavior of individuals. Such perceptions motivate how people disclose and share information with others. In this part of the study, we investigated the perspective of citizens in the United Kingdom about three main issues related to chatbot-mediated information disclosure. Firstly, we

identified the factors that lead to certain data items being considered sensitive by the participants. Secondly, we asked them to evaluate the perceived sensitivity of different data items and finally, we analysed how demographic characteristics of the participants (such as age, gender, education), anonymity, context (health vs finance) and interaction means (chatbot-mediated vs human-mediated) influence the perception of sensitivity on an individual level.

In order to understand reasons or factors that lead participants to consider certain personal data as more sensitive, we asked individuals to give their reasons in open-ended questions and conducted a thematic content analysis to identify emerging themes in their responses. Whilst to assess the sensitivity ratings covered in the second part of the study, we used hierarchical cluster analysis and grouped data items based on their perceived sensitivity. Finally, to model the effect of demographic characteristics, anonymity, context and interaction means, we built proportional-odds logistic regression models. Using those models, we explored the effects of these factors on comfort levels of people while disclosing sensitive information.

3.2 Inferring Demographic information using keystroke and mouse dynamics

As identified previously it is clear that encouraging successful disclosures from individuals requires a tailored experience that acknowledges the individual differences between users. Logically, if a chatbot is to provide a tailored experience to a user it must have a way of assessing a set of characteristics about the user in order to select the most appropriate experience. In this study we look to model the ability to infer user characteristics, in this case demographic information such as age and gender, from how the user is interacting with the chatbot.

This study used a custom web-application that facilitated the collection of large volumes of mouse and keystroke dynamic data, in both controlled and uncontrolled scenarios. Participants were required to complete five distinct tasks: three gathering data associated with the use of a mouse and two gathering data from keyboard key presses.

To gather data about the use of a mouse, participants were shown single cross hairs, and were required to click the centre of the target. Once they clicked within 100px of the centre of the target they were presented with a new target, with participants asked to click five targets in total. This used a similar approach to that devised by Van Balen et al. vanbalen. Additionally, more natural mouse data was gathered during the initial phases of the study, when asking users to provide demographic data. In order to collect keystroke data, participants were required to copy a short passage of text from Bram Stoker's Dracula and then to write a brief description of the plot of the last movie that they watched. This approach is designed to both provide data collection which is comparable against all participants and then free collection (e.g. writing creatively rather than copying), to better mimic reality. The study collected data from 239 participants, and used a combination of features resulting in a total of 241 features in total.

4 Initial results

4.1 Understanding information disclosures

The thematic analysis of the responses given for the factors that lead participants to consider a data item to be sensitive identified several common factors. Our findings confirmed the importance of the risk of harm, trust of the interaction means, public availability of data, context of the data, and the re-identification risk in disclosure behavior. Additionally, we found other factors had an impact upon the comfort level of individuals while disclosing information. These factors included concerns about the reactions from others, concerns regarding personal safety or mental health, and the consequences of disclosure on loved ones.

Assessment of the sensitivity ratings of different data items and clustering results demonstrate that from a UK perspective, passwords are the most sensitive data type where 92% of participants gave it the highest rating. Bank account credentials, credit card number are other items appearing in the same high sensitivity category. The next category, sensitive data items, included formally identifiable information such as national ID number and passport number. Accordingly, the least sensitive items were hair colour, gender and height which are unhelpful to identify individuals and are typically observable.

Focusing specifically on chatbots, the proportional-odds logistic regression models that were built in the third part of the study demonstrated that there is a preference for disclosure to humans (over chatbots) especially in a health context. Participants were more comfortable disclosing their health information such as medical diagnosis, chronic diseases and mental health issues directly to a human. This impact was less visible in the finance domain where only the credit score and income level data items showed a significant effect. Therefore, finance may well be a potential context where chatbots can be utilised.

4.2 Inferring demographic information using keystroke and mouse dynamics

As discussed previously the ability to dynamically adapt the chatbot to the user provides the opportunity to present the chatbot that is most likely to elicit deep and accurate disclosures. This section explores the results from the study looking to automatically identify biographic characteristics of the user and hence support the personalisation of the chatbot.

Using a variety of machine learning approaches we explored the ability to predict the characteristics of the users from their interactions with the computer via the keyboard and mouse, the data was randomly split with 95% for training and 5% for testing, utilising random undersampling to account for class imbalance, this process was performed 100 times for each experiment. The research also explored the effect of accurate feature engineering by reducing the number of features used by the classifiers, this used the ANOVA F-values between the label and the features to identify the most discriminatory features. The set of features associated with the mouse dataset was smaller and hence we only present the

results for the full dataset. Initial results are shown in Table 1, which shows the performance for a variety of classification techniques. We would anticipate these scores could be improved by more rigorous feature engineering and tuning of the model parameters but the performance is included as an indication of the classification accuracy.

Table 1. Accuracy scores for each biometric (keystroke, mouse dynamics and a combined measure of the two) and its ability to predict soft biometric features.

		Keystrokes				Mouse				Combination			
	N. features	Gender	Handedness	Age	Electronic Hours	Gender	Handedness	Age	Electronic Hours	Gender	Handedness	Age	Electronic Hours
Random Forest (100)	all	0.57	0.61	0.20	0.06	0.60	0.55	0.23	0.11	0.64	0.58	0.25	0.09
	150	0.58	0.53	0.21	0.12	-	-	-	-	0.66	0.58	0.26	0.13
	100	0.60	0.56	0.20	0.17	-	-	-	-	0.68	0.64	0.27	0.17
Decision Trees (3, 5)	all	0.50	0.58	0.16	0.15	0.58	0.57	0.23	0.10	0.52	0.63	0.20	0.21
	150	0.51	0.58	0.17	0.18	-	-	-	-	0.54	0.67	0.21	0.16
	100	0.52	0.58	0.18	0.19	-	-	-	-	0.58	0.68	0.23	0.20
Decision Trees (10, 3)	all	0.57	0.69	0.15	0.10	0.57	0.52	0.22	0.16	0.54	0.81	0.22	0.05
	150	0.51	0.74	0.16	0.10	-	-	-	-	0.56	0.74	0.21	0.07
	100	0.50	0.66	0.18	0.09	-	-	-	-	0.54	0.81	0.22	0.05
SVM	all	0.44	0.32	0.14	0.02	0.50	0.38	0.14	0	0.44	0.30	0.14	0.01
	150	0.52	0.29	0.15	0.02	-	-	-	-	0.51	0.31	0.15	0.02
	100	0.55	0.29	0.15	0.03	-	-	-	-	0.60	0.31	0.17	0.03
Gaussian Naive Bayes	all	0.50	0.26	0.18	0.10	0.54	0.64	0.26	0.13	0.46	0.32	0.20	0.10
	150	0.50	0.33	0.18	0.08	-	-	-	-	0.53	0.42	0.24	0.15
	100	0.54	0.43	0.24	0.11	-	-	-	-	0.53	0.54	0.28	0.22
KNN	all	0.50	0.6	0.11	0.11	0.58	0.41	0.18	0.14	0.52	0.58	0.14	0.16
	150	0.54	0.54	0.12	0.08	-	-	-	-	0.58	0.56	0.18	0.18
	100	0.52	0.55	0.20	0.13	-	-	-	-	0.59	0.08	0.20	0.12

The results of our study clearly showed that using a combination of both keystroke and mouse features led to a more accurate prediction of a soft biometric characteristic, regardless of the biometric being predicted, the classifier being used, or the number of features the machine learning model was trained on. This was the case even when the biometric was not predicted with a large degree of accuracy (age and the number of hours spent on electronic devices).

In addition to this, our study found that the mouse dynamic features were more discriminatory than the keystroke dynamic features. This is seen in both the accuracy scores when comparing mouse and keyboard data. It is also seen in the combined data set, after the feature engineering stage the data set contains a disproportionate number of features from the mouse dataset.

As can be seen from Table 1 the best results were achieved using a combination of mouse and keyboard dynamics, indicating that it is possible to infer some of user characteristics from their interaction but at a relatively low accuracy.

5 Conclusions and Future Work

In this paper we present initial results for two studies, which aim to understand sensitive information disclosures and identifying soft biometrics using a fusion of keystroke and mouse dynamics. The first study determined participants perceptions with regards to the sensitivity of information across various domains. The key finding was that finance represented the area where participants would be most comfortable disclosing sensitive data to a chatbot.

We also identified from the literature that a dynamic chatbot which identified characteristics of the user and then provided a tailored service is likely to produce deeper and more accurate disclosures. Hence a second study considered the ability to automatically identify demographic information about a user based solely on their interactions with a chatbot (e.g. how they used the mouse and keyboard). The goal being that these assessments could drive the personalisation of the chatbot. Whilst we demonstrate it is possible to infer these user characteristics the accuracy is generally low. This indicates that for chatbots it may not be appropriate to infer these characteristics with a degree of confidence suitable for tailoring the experience.

The two studies in this paper will provide some of the early results from a larger project focusing on the provision of tailored chatbots to improve the sensitive disclosures made by users of these systems. The project is currently undertaking a Wizard-of-Oz (e.g. as seen in Kerly and Bull [10]) style experiment, which aims to understand how the perceived humanity of a chatbot impacts the depth of sensitive disclosures.

These experimental studies will be further contextualised with a number of case studies, exploring the application of chatbots in sensitive domains and providing mechanisms to ensure that the resulting digital services are inclusive and consider the needs of all of those using these services.

Acknowledgements. The research presented in this paper forms part of the *A Platform for Responsive Conversational Agents to Enhance Engagement and Disclosure (PRoCEED)* project funded by the Engineering and Physical Sciences Research Council, UK (EP/S027424/1, EP/S027211/1, EP/S027297/1, EP/S027467/1)

References

1. Bailenson, J.N., Yee, N., Merget, D., Schroeder, R.: The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction. Presence: Teleoperators and Virtual Environments **15**(4), 359–372 (08 2006). https://doi.org/10.1162/pres.15.4.359
2. Bhakta, R., Savin-Baden, M., Tombs, G.: Sharing secrets with robots? In: EdMedia+ Innovate Learning. pp. 2295–2301. Association for the Advancement of Computing in Education (AACE) (2014)
3. Drift: The 2018 state of chatbots report. https://www.drift.com/wp-content/uploads/2018/01/2018-state-of-chatbots-report.pdf (2018)
4. Elmasri, D., Maeder, A.: A conversational agent for an online mental health intervention. In: Ascoli, G.A., Hawrylycz, M., Ali, H., Khazanchi, D., Shi, Y. (eds.) Brain Informatics and Health. pp. 243–251. Springer International Publishing, Cham (2016)
5. Epley, N., Waytz, A., Akalis, S., Cacioppo, J.T.: When we need a human: Motivational determinants of anthropomorphism. Social Cognition **26**(2), 143–155 (2008). https://doi.org/10.1521/soco.2008.26.2.143
6. Epley, N., Waytz, A., Cacioppo, J.T.: On seeing human: a three-factor theory of anthropomorphism. Psychological review **114**(4), 864 (2007)
7. Hill, J., Randolph Ford, W., Farreras, I.G.: Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. Computers in Human Behavior **49**, 245–250 (2015). https://doi.org/10.1016/j.chb.2015.02.026
8. Jenkins, M.C., Churchill, R., Cox, S., Smith, D.: Analysis of user interaction with service oriented chatbot systems. In: Jacko, J.A. (ed.) Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments. pp. 76–83. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
9. Juniper Research: Bank Cost Savings via Chatbots to Reach $7.3 Billion by 2023, as Automated Customer Experience Evolves. https://www.juniperresearch.com/press/press-releases/bank-cost-savings-via-chatbots-reach-7-3bn-2023 (2019)
10. Kerly, A., Bull, S.: The potential for chatbots in negotiated learner modelling: A wizard-of-oz study. In: Ikeda, M., Ashley, K.D., Chan, T.W. (eds.) Intelligent Tutoring Systems. pp. 443–452. Springer Berlin Heidelberg, Berlin, Heidelberg (2006)
11. Sağlam, R.B., Nurse, J.R.C.: Is your chatbot gdpr compliant? open issues in agent design. In: Proceedings of the 2nd Conference on Conversational User Interfaces. CUI '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3405755.3406131
12. Sundar, S.S., Bellur, S., Oh, J., Jia, H., Kim, H.S.: Theoretical importance of contingency in human-computer interaction: Effects of message interactivity on user engagement. Communication Research **43**(5), 595–625 (2016). https://doi.org/10.1177/0093650214534962