Surveying User Reactions to Recommendations Based on Inferences Made by Face Detection Technology

Jennifer Marlow
FXPAL
Palo Alto, California, United States
marlow@fxpal.com

Jason Wiese
University of Utah
Salt Lake City, Utah, United States
wiese@cs.utah.edu

ABSTRACT

It is increasingly possible to use cameras and sensors to detect and analyze human appearance for the purposes of personalizing user experiences. Such systems are already deployed in some public places to personalize advertisements and recommend items. However, since these technologies are not yet widespread, we do not have a good sense of users' perceptions of the benefits and drawbacks of public display systems that use face detection as an input for personalized recommendations. We conducted a user study with a system that inferred participants' gender and age from a facial detection and analysis algorithm and used this to present recommendations in two scenarios (finding stores to visit in a mall and finding a pair of sunglasses to buy). This work provides an initial step towards understanding user reactions to a new and emerging form of implicit recommendation based on physical appearance.

CCS CONCEPTS

H.5 Information interfaces and presentation

KEYWORDS

Face recognition; Personalization; Public displays

1 INTRODUCTION

Recommendation systems typically focus on developing algorithms and solutions to present tailored options to people online based on their online behavior (e.g. Amazon product reviews, movie recommendations, etc.). However, advances in technology are now moving the possibilities for personalization and individualized recommendations off the computer or mobile screen and out into the real world.

There is a new trend for "real-world personalization" using cameras combined with face detection and recognition to adapt and personalize information that people view in public display settings. While not yet widespread, there are examples of camera-based adaptive and personalized billboards [18], vending machines [5, 15], and even meal recommendation at fast food restaurants [7]. Currently these systems detect easily observable characteristics, (e.g. gender and age), although future sensing technologies could make more sophisticated inferences (e.g. emotional state, heart rate, or skin condition).

Given the relative recency with which camera-based public detection and personalization/recommendation systems have begun to emerge, we currently do not have a good idea of how users will react when encountering such technology. On one hand, people might see these sensing systems as beneficial (or at least as not harmful). In contrast to the online world, in a public place there is often little information about a person interacting with a public system. In these situations, even using coarse features such as age and gender to personalize an interface could help a system determine how to present more tailored information. On the other hand, people might see these types of inferences either as too invasive or as too generic to be helpful.

This work seeks to develop an understanding of user reactions to facial-recognition-based recommendations in the context of a shopping mall public display. Specifically, we explore two questions: First, how will people react to recommendations provided based on inferred gender/age? Second, what are people's attitudes towards systems that recommend items based on facial recognition in a public setting?

To explore these questions, we conducted a test of a prototype system. In this study, U.S.-based participants viewed and evaluated a personalized list of stores to visit and objects to buy. This study offers a first insight into user reactions to face-detection-driven recommender systems. Our findings suggest that in this context, participants find age to be less useful than gender as a recommendation input (with neither being considered particularly sensitive). Additionally, their comfort decreases when inaccurate detections occur, particularly of gender. These findings can help inform the design of future public displays that use facial recognition technology.

2 RELATED WORK

Providing recommendations based on an individual's preferences or characteristics can help users find relevant items in a larger item set, improve decision making, and assist a user in exploring a set of options [9]. Historically, recommendation systems have gathered information about a user's characteristics and tailored the items displayed using explicit and/or implicit feedback [21].

RecSys '17, August 27-31, 2017, Como, Italy
 ACM ISBN 978-1-4503-4652-8/17/08…$15.00
http://dx.doi.org/10.1145/3109859.3109875

Explicit feedback requires an action from a user (e.g. by providing ratings on products, restaurants, or movies). Implicit feedback, on the other hand, involves inferring interest based on online actions (such as what people are clicking on in a rating list). Each approach has pros and cons: While implicit feedback is easy to gather because it does not disrupt a user's browsing patterns, it can result in "overspecialization" where the system only recommends items that the user likes [12].

Recently, research in recommender systems has sought to move beyond making recommendations based on online behavior like clicks and browsing behavior, and has considered the role of using other (non-invasive) means of tailoring results using types of personal traits or behaviors. These have included using personality [3, 10, 20] and gaze prediction [21] as new forms of feedback to inform recommender systems. Because feedback based on physiological factors such as gaze activity is not yet widespread, it is unclear how the public will react to systems that incorporate it. For example, Zhao et al. [21] mention that gaze prediction could become commonplace, or it could face resistance from users due to privacy concerns. A similar dynamic could occur for systems that incorporate personality details – such a method requires collecting sensitive data about the user, raising the concern of privacy invasion [1].

This work focuses on assessing user reactions to recommendations based on high-level information that is detectable in public by camera: specifically, age and gender. Prior work suggests that people may be amenable to providing information about their gender and age, at least in an online context. For example, Knijnenburg and Kobsa [11] found in a survey about disclosing different types of information online that 94% of respondents did not mind disclosing their gender and 93% were willing to disclose their age. Other work has found several reasons why users of recommendation systems do not volunteer personal information like gender and age; In some cases it is intentional (to protect privacy), while other times it is unintentional (out of laziness or disinterest) [19].

It is increasingly possible to infer a user's gender and age without them agreeing to it, simply based on online behavior. Gender and other demographics can be predicted just from traces of online behavior, such as web browsing patterns [8] or posts and activity on social media [4, 13, 22] and are often used for targeting ads. Other work has shown that using gender and age outside of online ad placement might be considered less sensitive or more acceptable if users believe it actually serves a constructive purpose or will be of some benefit to them [17].

While inferring age or gender from a user's online behavior without their explicit consent is controversial, detecting this information without requiring any prior information or observations from a user (which is possible using a camera) has the potential to be even more concerning. This work explores participant reactions to exactly that type of sensing.

3 ONLINE STUDY OF USER REACTIONS

To understand user reactions to a public display that used a camera and face detection to display different kinds of information based on inferred age and gender, we first had to select a scenario to provide context to the system. We chose the scenario of a store directory in a shopping mall because shopping is a general and easily-relatable topic. We also expected that age and gender-based personalized recommendations in such a context could be useful to help shoppers explore and find relevant items in a larger item set [9].

Next, we created a simulated store directory system to investigate users' reactions after they gained a more realistic first-hand experience with the technology and its recommendations. We integrated Microsoft's Face Detection API (via webcam) into an online survey and restricted participation in the study to people who had working webcams.

We conducted the study online as a HIT on Amazon Mechanical Turk. We chose Mechanical Turk as a platform to reach a more diverse set of participants than we could using local or convenience samples [2]. Participants based within the United States who had completed over 500 tasks with a greater than 95% acceptance rate were eligible to view the HIT. They received $3.00 for participating in the 15 to 20 minute task.

Participants received the following instructions: *We are interested in getting people's reactions to a new type of technology for helping people in malls and department stores. We will show you two shopping scenarios and ask for your opinions and preferences. In order to participate in the task, we will be asking you to share your webcam (only on Page 3). This is to help simulate and give you a sense of the technology scenario.* ***Your image will not be stored or saved as part of your responses to the survey.***

The rest of the task description set up the sample scenario: Participants were asked to imagine they were on vacation in Singapore and that the airline had lost their luggage. They needed to visit an unfamiliar mall to buy (1) some clothes to wear while walking around the city and (2) a pair of sunglasses. Participants were then taken to a page representing the store directory for the mall. Instructions here were as follows: *Now we will use the webcam to recommend some items specifically for you. Make sure your face is visible below, then click the button to "verify your face appears above" - and hit "Next" in the survey. If you get an error, the webcam may not be detecting your face well so please try a few times.*

When the user clicked the button as directed, the webcam took a picture of their face and used Microsoft's Cognitive Services Face Detection API [16] to infer their gender and age from the picture (this information was used to drive the personalization of the results presented in the personalization condition). Because we did not want participants to know what the system concluded about them, (to simulate the use case where detection happens automatically and unobtrusively), we did not explicitly mention or present the system's age and gender inferences to the participant until the very end.

Participants then performed the store and sunglasses selection tasks (beginning with the store selection task). In each task, participants saw a set of 15 options (stores or sunglasses) presented together in a vertical column on one page. In the

Personalized condition, the first five items in the column that corresponded to a participant's inferred gender/age category (with a yellow "recommended" banner), followed by five items labeled as being for both genders, then five items that corresponded to the opposite gender/age category. In the Unpersonalized condition, the order of the items was reversed (so there was still organization based on gender, as one would find in an ordinary mall directory,) but the most relevant items were not called out nor located at the top of the list. Approximately half of participants saw personalized stores and unpersonalized sunglasses; the other half had the reverse order of personalization.

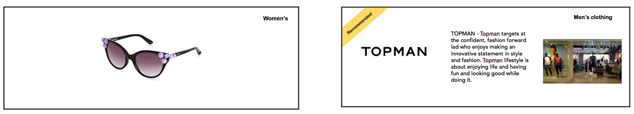

Figure 1: Sample of "unpersonalized" women's glasses and "personalized" men's clothing store

For both tasks, the study asked participants to look at the webpage listing all 15 items and then to check or select three items from the list that they were interested in trying. We recorded the position in the list of each item, and also recorded the amount of time taken to make the decision. At the end, we asked participants to rate how satisfied they were with the process of choosing stores and items within a store.

After completing the two selection tasks, we explained the purpose of the study and then showed participants what the system thought their age and gender were. Age was presented as a single number based on the Microsoft Cognitive Services API's estimate. We asked participants to rate whether or not the age was accurate within five years.

Next, the survey asked: *Now that you've experienced a sample of what it's like, how comfortable would you feel with a store directory detecting your age and gender and using this information to recommend stores/products to you? (You can assume that your picture isn't saved or recorded).* We also solicited participants' opinions about the relevance and appropriateness of using face detection for recommendations in an open-ended question. We concluded with two Likert Scale questions about the perceived benefit of the system, perceived privacy intrusion, and two questions about privacy attitudes more broadly.

4 RESULTS

In total, 44 participants took part in the study. First, we wanted to examine the role of providing recommendations on choice behavior. Figure 2 shows the percentages of participants who chose items defined as being for their "same gender" (e.g. women's items if the person was a woman), "neutral" items (things defined as being for men and women), and "opposite gender" items.

Overall, there were no significant differences of "recommending" an item for males in terms of the items they chose. Men chose "men's" items 67% of the time with no recommendations, and 61% of the time with recommendations. However, female participants were significantly more likely to choose a "female" item when it was recommended (79%) than when it was not recommended (61%) (X^2 (2,114)=7.24, p=.02). Further analysis of participants' choices showed that people in the "unpersonalized" condition frequently scrolled down to the bottom of the list to choose their own gender items. There were no significant differences in time taken to complete either task with or without the recommendations. There were also no significant differences in how easy the task was reported to be with or without recommendations.

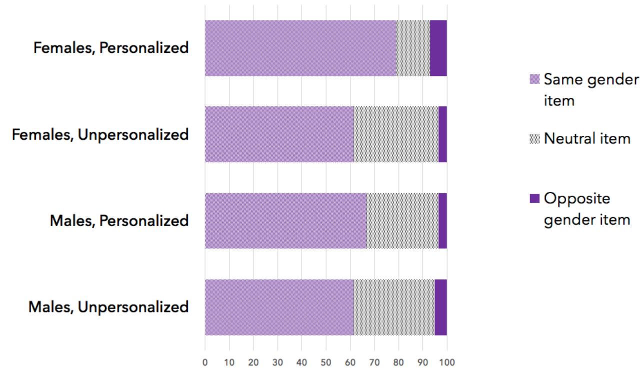

Figure 2: Percentage of participants choosing same gender, neutral, and opposite gender items

4.1 Reactions to system

Based on participants' responses, the system accurately detected the correct gender and age (+/- five years) for 24 participants. The system incorrectly detected age for an additional 14 participants, of whom the system also incorrectly classified six females as male. Table 1 presents reported comfort with gender detection and age detection by prediction accuracy (0=completely uncomfortable, 100=completely comfortable).

Table 1: Comfort with gender & age detection (Mean, SD)

Results of face detection algorithm	N	Comfort, gender detection	Comfort, age detection
Correct gender + age	24	85.3 (21.6)	67.8 (30.6)
Correct gender, Incorrect age	14	88.5 (14.9)	74.8 (21.8)
Incorrect gender, Correct age	2	49.0 (33.9)	50.0 (35.4)
Incorrect gender + age	4	25.0 (38.2)	22.50 (33.4)

Overall, participants were significantly more comfortable with a camera-based store directory system detecting information about gender (M=79.2) than they were with it detecting information about age (M=65.1) (F(1,86)=4.92, p=.03). Interestingly, there was a significant drop-off in comfort levels for people whose gender was misclassified. In these cases, they were not only less comfortable with gender detection but they were also less comfortable with age detection. Accuracy did not influence other measures such as satisfaction or ease of the task; however, this could be due to the fairly small set of choices.

We examined the open-ended responses to understand more about why people reacted as they did to the system and its recommendations. We used an open-coding process to categorize the responses as positive or negative. Then we used affinity diagramming to cluster the responses around common themes, which are presented in order of frequency in Table 2.

Table 2: Categorization of open-ended responses

Positive reaction	Negative reaction
PR1. Convenience, help narrow down options, speed up process [n=9]	NR1. Recommendations were based on stereotypes; not helpful [n=4]
PR2. Not a problem, not concerning [n=7]	NR2. Age not helpful for recommendations [n=4]
PR3. Age and gender are not sensitive info [n=5]	NR3. Prefer to make own choices when shopping [n=3]
PR4. Machine inferences are the same as what a person or salesperson could infer [n=3]	NR4. General dislike for cameras [n=2]
PR5. Recommendations were good [n=3]	NR5. Time consuming [n=1]

24 participants had only positive reactions, 10 provided a mixture of positive and negative reactions, and 9 had purely negative reactions. PR1 was the most commonly cited positive reaction, where people thought a camera-based recommendation system for shopping could be convenient in helping narrow down options, speed up the process, or filter results: *I think it's kind of cool too. I feel it could really help give me stores that would be relevant to me, and exclude stores that are more focused on the other gender or for different age groups* (P13).

The following quote exemplifies PR3 (a feeling that age and gender are not sensitive): *I don't mind it at all. age and gender aren't a secret. So I actually consider this to be an advantage. I like this idea and can see it catching on as long as it doesn't dig deeper than those two categories* (P6). Similarly, PR4 equated the inferences made by the system with a human's capabilities: *I feel fine with that information being used because any salesman in the mall can take a quick look at me and make the same recommendations based on age and gender.*

On the negative side, NR1 expressed a dislike for stereotype-based recommendations: *The system is using stereotypes to recommend things to me* (P25). NR2 had to do with the fact that age was seen as less relevant for recommendations than gender: *why would I like something just because I am a certain age. I still shop at "teen" shops even though I am almost 30*! (P35). NR3 had to do with a general preference to look around and discover things on one's own: *I think it's fine except I don't really want a machine recommending me things to buy. I have my own tastes and I doubt a machine can determine that* (P20).

Seven comments addressed the role of accuracy. People were happy to have their age underestimated (no one mentioned that their age was overestimated): *First of all, I like that it underestimated my age. I'm going to be 77 and it guessed 59.2!* (P3). In terms of gender, one woman who was misclassified as a male wrote: *I feel uncomfortable because it thought I was male. I'm actually female* (P16). People who did not experience such detrimental inaccuracies were more pleased with the results: *As it stands it cut off 13 years and got my gender correct. I liked it and it gave me some good suggestions* (P12).

5 DISCUSSION

These results seem to indicate different reactions to detecting age and gender. Gender was seen as less problematic and more helpful for recommendations when it was assessed correctly, but was problematic when assessed incorrectly. Users' reactions to misclassified age were positive when it was underestimated. As a result, in cases of uncertainty it may be beneficial to underestimate rather than overestimate. This fits with the typical cultural assumption that it is preferable to be mistaken as younger than one actually is. However, we also did not have any reactions where participant ages were overestimated, so we cannot say from this data how harmful it would be to overestimate somebody's age.

Participants' reasons for disliking the system in this study point to a potentially problematic future for these systems. These concerns seem unlikely to go away (i.e. improving the technology accuracy will not relieve the concerns). Perceived benefits may increase if privacy is ensured (e.g. systems are prevented by regulations from storing face images, recommendations are not viewable by passers-by) or the use case is convincing (using gender/age inferences for health or fitness recommendations may be more useful than for shopping).

This exploratory study has some limitations: it was done using Mechanical Turk with participants who agreed to share their webcams, and was not in a publicly-deployed context. In the future, we will try a physical instantiation of the store directory and also explore other types of scenarios (for example, health/gym kiosk, museum kiosk, movie ticket recommender) to see how participants' reactions vary in these different settings. We might encounter different reactions in different contexts, as users' information disclosure decisions are highly dependent on context and presumed benefit [14, 17]. While participants saw some value in clothing recommendations based on gender (less so on age), trying other contexts will help us explore the design space in which age and gender are seen as more or less useful and helpful for making recommendations.

REFERENCES

[1] Amos Azaria and Jason Hong. 2016. Recommender Systems with Personality. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 207–210.
[2] Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. 2012. Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political Analysis* 20, 3: 351–368.
[3] Shlomo Berkovsky, Ronnie Taib, and Dan Conway. 2017. How to Recommend?: User Trust Factors in Movie Recommender Systems. *Proceedings of the International Conference on Intelligent User Interfaces.* ACM, 287-300.

[4] Smriti Bhagat, Udi Weinsberg, Stratis Ioannidis, and Nina Taft. 2014. Recommending with an agenda: Active learning of private attributes using matrix factorization. In *Proceedings of the 8th ACM Conference on Recommender systems*, 65–72.

[5] Lauren Davidson. 2014. The vending machine of the future is here, and it knows who you are. http://www.telegraph.co.uk/finance/newsbysector/retailandconsumer/11274179/The-vending-machine-of-the-future-is-here-and-it-knows-who-you-are.html

[6] F. Maxwell Harper, Funing Xu, Harmanpreet Kaur, Kyle Condiff, Shuo Chang, and Loren Terveen. 2015. Putting users in control of their recommendations. In *Proceedings of the 9th ACM Conference on Recommender Systems*, 3–10.

[7] Amy Hawkins. 2017. KFC China is using facial recognition tech to serve customers – but are they buying it? https://www.theguardian.com/technology/2017/jan/11/china-beijing-first-smart-restaurant-kfc-facial-recognition

[8] Jian Hu, Hua-Jun Zeng, Hua Li, Cheng Niu, and Zheng Chen. 2007. Demographic prediction based on user's browsing behavior. In *Proceedings of the 16th international conference on World Wide Web*, 151–160.

[9] Dietmar Jannach and Gediminas Adomavicius. 2016. Recommendations with a Purpose. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 7–10.

[10] Raghav Pavan Karumur, Tien T. Nguyen, and Joseph A. Konstan. 2016. Exploring the value of personality in predicting rating behaviors: a study of category preferences on movielens. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 139–142.

[11] Bart P. Knijnenburg and Alfred Kobsa. 2013. Making decisions about privacy: information disclosure in context-aware recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 3, 3: 20.

[12] Bart P. Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. 2016. Recommender systems for self-actualization. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 11–14.

[13] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110, 15: 5802–5805.

[14] Scott Lederer, Jennifer Mankoff, and Anind K. Dey. 2003. Who wants to know what when? privacy preference determinants in ubiquitous computing. In *CHI'03 extended abstracts on Human factors in computing systems*, 724–725.

[15] Elaine Lies. 2010. Japan vending machine recommends drinks to buyers. http://www.reuters.com/article/us-japan-machines-idUSTRE6AE0G720101115

[16] Microsoft Cognitive Services Face Detection API https://www.microsoft.com/cognitive-services/en-us/face-api

[17] Helen Nissenbaum. 2010. Privacy in context: Technology, policy, and the integrity of social life. Stanford University Press.

[18] Andrew Tarantola. 2015. Hey ladies, this German billboard wants to sell you a beer. https://www.engadget.com/2015/05/21/hey-ladies-this-german-billboard-wants-to-sell-you-a-beer/

[19] Udi Weinsberg, Smriti Bhagat, Stratis Ioannidis, and Nina Taft. 2012. BlurMe: Inferring and obfuscating user gender based on ratings. In *Proceedings of the sixth ACM conference on Recommender systems*, 195–202.

[20] Wen Wu, Li Chen, and Liang He. 2013. Using personality to adjust diversity in recommender systems. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, 225–229.

[21] Qian Zhao, Shuo Chang, F. Maxwell Harper, and Joseph A. Konstan. 2016. Gaze Prediction for Recommender Systems. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 131–138.

[22] Xin Wayne Zhao, Yanwei Guo, Yulan He, Han Jiang, Yuexin Wu, and Xiaoming Li. 2014. We know what you want to buy: a demographic-based system for product recommendation on microblogs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1935–1944.