Privacy Expectations and Preferences in an IoT World (SOUPS2017)

2017/10/10

担当　森

- Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Cranor, and Norman Sadeh,
- "Privacy Expectations and Preferences in an IoT World",
- SOUPS 2017, Jul 2017

- https://privacyassistant.org/
  の中の人たち

本日の進め方

- 著者らの発表スライドを使って、ざっくり何をやったかを説明（省エネ）。

- 口頭発表では割愛されていた技術詳細や、論文の書き方としてうまいと感じた点、その他細かい気づきなどを本スライドで説明。

- 所感

データ収集

- 1,014 の Mturk workers
  - 米国在住
  - HIT (Human Intelligence Task) approval rate > 95%
- 平均して16分かかるタスク⇒5分以内に完了したユーザを除外
- ひとりあたり14個のシナリオを提示

質問紙 (appendix)

D. DEMOGRAPHIC QUESTIONS

この選択肢もつける

Q1. How old are you?

Q2. What is your gender? (Choices: Female, Male, Other, Prefer not to answer)

Q3. What is the highest degree you have earned? (Choices: No high school degree, High school degree, College degree, Professional degree (masters/PhD), Associates degree, Medical degree, Prefer not to answer)

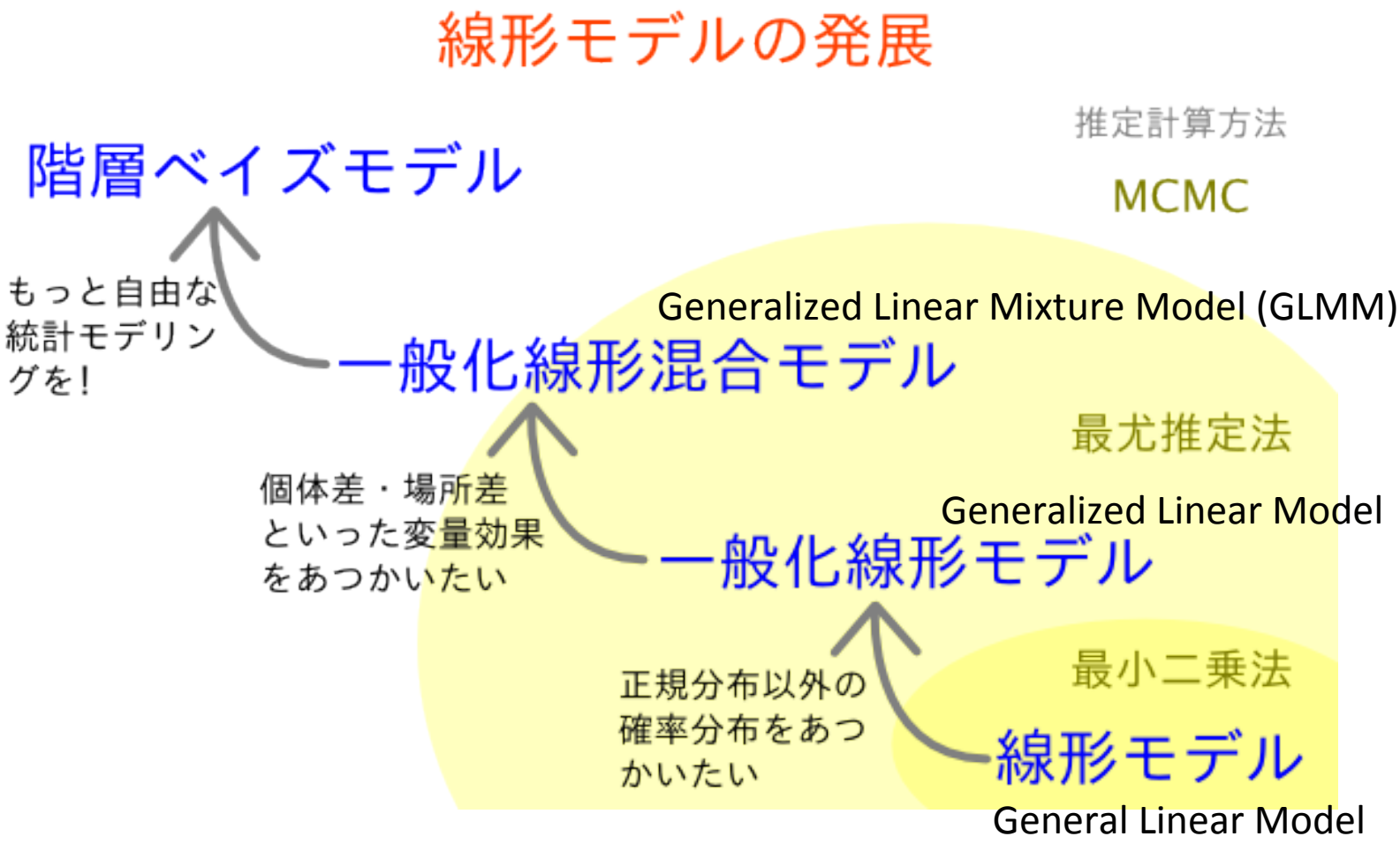
Q4. What is your income range? (Choices: Less than $15,000/ year, $15,000/ year - $24,999/year, $25,000/ year - $34,999/ year, $35,000/ year - $49,999/ year, $50,000/ year - $74,999/ year, $75,000/ year - $99,999/ year, $100,000/ year - $149,999/year, $150,000/year - $199,999/ year, $200,000/ year and above, Prefer not to answer)

Vignette Study （ビネット調査）

- 架空の人物や状況を評価対象として設定し、それに対する回答者の評価や判断を測定する方法

- [en.wikipedia]
  A vignette in psychological and sociological experiments presents a hypothetical situation, to which research participants respond thereby revealing their perceptions, values, social norms or impressions of events. Peter Rossi and colleagues[1] developed a framework for creating vignettes by systematically combining predictor variables in order to dissect the effects of the variables on dependent variables. For example, to study normative judgments of family status, "there might be 10 levels of income; 50 head-of-household occupations, and 50 occupations for spouses; two races, white and black; and ten levels of family size". [2] Since this approach can lead to huge universes of stimuli – half a million in the example – Rossi proposed drawing small random samples from the universe of stimuli for presentation to individual respondents, and pooling judgments by multiple respondents in order to sample the universe adequately. Main effects of predictor variables then can be assessed, though not all interactive effects.[3]
- [1] Rossi, Peter H., and Steven L. Nock, Eds. Measuring Social Judgments: The Factorial Survey Approach (Sage, 1982); Rossi, P. H., and Richard A. Berk (1985). "Varieties of normative consensus" American Sociological Review 50: 333-347

GLMM



http://hosho.ees.hokudai.ac.jp/~kubo/ce/LinksGlmm.html

General Linear Model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \epsilon_i$$

目的変数

説明変数

comfort level $\in \{0, 1\}$

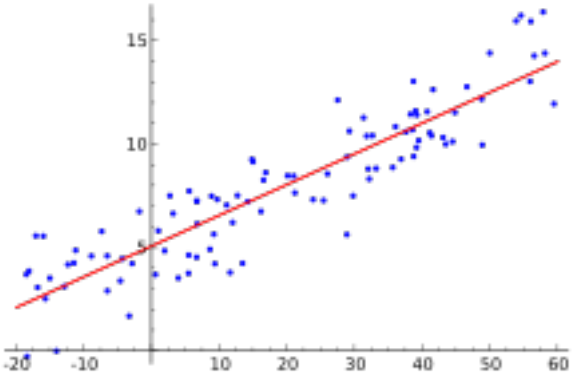
deny/allow decision $\in \{0, 1\}$

data-type

location

being-shared

⋮



GLMMの結果の解釈

Table 4: Generalized linear mixed model regression output for the comfort level model. A positive estimate (effect size) indicates inclination toward comfort and a negative estimate shows inclination toward discomfort. Factors are ordered by their contribution: the factor with the lowest BIC contributes most to explaining participants' comfort level.

p値が小さければ有意な説明変数である

つまり、その説明変数が目的変数に影響を与えない（効果がない）と仮定したときに、実際に観測した値が出現する確率

Table 4 (Comfort level)

Factor	Estimate	Std Err	Z-value	p-value	BIC
data type:happening today					14633
baseline=friend's house:not happening today					
video:happening today	1.39	0.20	6.83	**0.00**	
biometric:happening today	0.89	0.15	5.80	**0.00**	
presence:happening today	0.91	0.18	12.57	**0.01**	
temperature:happening today	0.95	0.22	4.26	**0.00**	
data (baseline=specific position)					15843
biometric	-1.45	0.13	-11.12	**0.03**	
presence	1.42	0.16	8.99	**0.00**	
temperature	2.50	0.20	12.57	**0.00**	
video	-0.30	0.19	-1.62	0.11	
user perceive benefit:location					15866
baseline=beneficial:friend's house					
not beneficial:department store	0.00	0.32	0.00	0.99	
purpose unspecified:department store	-0.07	0.24	-0.30	0.76	
not beneficial:house	-0.15	0.48	-0.30	0.76	
purpose unspecified:house	0.05	0.28	0.19	0.85	
not beneficial:library	-0.45	0.33	-1.38	**0.00**	
purpose unspecified:library	-0.17	0.24	-0.70	0.48	
not beneficial:public restroom	-0.40	0.36	-1.10	0.27	
purpose unspecified:public restroom	-0.48	0.26	-1.85	**0.01**	
not beneficial:work	-0.49	0.36	-1.38	0.17	
purpose unspecified:work	-0.11	0.24	-0.47	0.63	

Table 4 (Comfort level) 下の方

happen today (baseline=disagree)					16491
agree	10.98	333.4	0.03	0.97	
location (baseline=friend's house)					17987
library	1.00	0.18	5.54	**0.00**	
work	0.87	0.18	4.82	**0.01**	
house	-0.88	0.20	-4.34	**0.00**	
department store	0.76	0.18	4.24	**0.00**	
public restroom	0.29	0.19	1.48	0.14	
being shared (baseline=not being shared)					18079
being shared	-0.68	0.09	-7.86	**0.00**	
IUIPC					
collection	-0.59	0.05	-11.47	**0.04**	18081
retention (baseline=not specified)					18103
week	0.25	0.11	2.25	**0.00**	
year	0.16	0.11	1.45	0.14	
purpose specific	0.0.56	0.15	4.85	**0.02**	
not deleted	0.10	0.10	0.99	0.32	

Introduction の構成: 6パラグラフ

- IoT の利便性とプライバシーリスク
- 研究の動機（後述）＋実例
- この研究でやったこと
- この研究の結果＋価値
- この研究の貢献（2つ）
  - 過去の研究との差異も述べる
- 論文の構成

これらがうまくつながっている
特に「動機」⇒「やったこと」⇒「結果＋価値」のフローが良い

うまいと感じた表現

- 研究の動機:
  - To fully realize the potential of IoT, individuals need to be sufficiently knowledgeable and aware to make informed decisions. Thus, IoT devices need to inform their users about their data collection practices and offer privacy choices that respect individual privacy preferences.

あるべきIoTデバイスのデザインを示す
（ユーザがIoTデバイスのデータ収集をよく理解していることが重要）

うまいと感じた表現

- やったこと:
  - We conducted a large-scale online vignette study to identify the contribution of different factors (such as the type of data, retention time, purpose of data collection, and location of data collection) in promoting or inhibiting individuals' self-professed comfort levels.

Vignette study は such as … のような要因を明らかにする手法である

その手法により、個々人が自認する"comfort level" を促進、阻害する要因を特定できる

うまいと感じた表現

- 研究の価値:
  - The results of our study informs the design of more transparent IoT-connected systems—we envision our results can be used to improve privacy notices for IoT devices, and develop more advanced personal privacy assistants [25].

実際のシステム設計に有用であることを主張

うまいと感じた表現

- 関連研究との比較:
  - Many of our findings are consistent with observations made in prior work, but our quantitative methodology and the scale of our experiment allows us to understand the effect of individual factors and their relative importance more precisely

Related work にも十分なページ数が割かれている

関連研究の構成

- 2.1 IoT Privacy Challenges
  - こういう問題が出始めているという観点
- 2.2 Privacy Interfaces for IoT Systems
  - 色々な提案がなされてきているが、本研究がこれらと違うのは、ビネット調査により複数のシナリオを同時に扱った点であるという主張
- 2.3 Factors Impacting Privacy Preferences
  - この話題は過去に色々とやられてきたので、それらを列挙。差分は下記
  - We aim to expand on these findings by evaluating these factors in a larger scale study, and in combination with additional factors capturing more contextual nuances that are specific to IoT environments.
- 2.4 Predicting Privacy Preferences
  - これも過去には違う文脈で色々やられている（モバイルアプリ等）
  - In our work, we incorporate additional factors into a larger scale study, using similar techniques to make predictions with the goal of achieving improved prediction accuracy relative to prior work.

Funding

- SBE: Medium: Towards Personalized Privacy Assistants
- https://www.nsf.gov/awardsearch/showAward?AWD_ID=1513957
- Sep 2015 – Aug 2018

- Specifically, this research harnesses recent advances in privacy preference modeling, machine learning and dialogue technologies to develop personalized privacy assistants capable of learning people's privacy preferences and of semi-automatically configuring many privacy settings on their behalf.

所感

- 書き物としてのクオリティがとても高い。
  - 特にintro と関連研究との差異を示すくだり
  - その一方で、数字に関して詰めが甘いところもあるが。
- IoTならでは、という点を vignette study で浮き彫りにしたところに価値がある（シナリオが色々考えうるような対象）
- GLMMでモデルを作った結果を用いた要因分析を行っているわりに、予測は非線形な識別器を使っているのはOK？（つまり、予測するには要因間の自明ではない掛け合わせが必要）