

Title

BRUNO GAZONI, ANDRE LIMA, and MARCELO MANZATO

[Insert abstract here]

Additional Key Words and Phrases: recommender systems, music, surprise, explanation, transparency, user study

ACM Reference Format:

Bruno Gazoni, Andre Lima, and Marcelo Manzato. 2018. Title. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/1122445.1122456>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Association for Computing Machinery.

Manuscript submitted to ACM

O objetivo desse *logbook* é registrar o histórico dos entendimentos e decisões sobre o projeto, até para lembrar depois como a gente conduziu isso tudo.

1 DISCUSSÃO SOBRE AS QUESTÕES DE PESQUISA

Com base na discussão que tivemos no encontro de 19 de abril, entendi que essas são as questões de pesquisa que queremos investigar e porquê:

RQ1 Existe associação entre o nível de surpresa reportado pelo participante e sua motivação para buscar mais informações sobre a recomendação apresentada?

Nossa motivação: Se essa associação existir e for detectável, então podemos personalizar o grau de detalhamento das informações oferecidas a respeito de uma recomendação (e isso nos permite contribuir com a pesquisa do Luan). Também ajudaria o Bruno a avançar na pesquisa com serendipidade e me ajudaria a retomar um projeto sobre variação de surpresa no tempo.

RQ2 Quando o participante busca por mais informações sobre a recomendação apresentada, existe uma categoria de informação (i.e., grupo de *features*) pela qual o participante demonstra mais interesse?

Nossa motivação: Se existe uma categoria de informação pela qual o participante se interessa mais, então podemos personalizar os conteúdos da explicação que é oferecida para a recomendação apresentada (e isso nos permite contribuir com a pesquisa do Luan, caso ele decida incluir o domínio de recomendação de músicas no escopo da pesquisa dele).

RQ3 Oferecer informações adicionais exerce efeito sobre a intenção do participante em consumir o item recomendado?

Nossa motivação: Se a intervenção de oferecer informações adicionais no formato “explicações à la carte” (Tarefa 3b) aumenta a intenção do participante em seguir a recomendação, então essa intervenção tem poder de persuasão (e isso nos permite contribuir com a pesquisa do Luan e também acrescentar evidências mais refinadas à literatura sobre o efeito persuasivo de oferecer “explicações à la carte”).

De acordo com esse entendimento, descrevo a seguir nossas previsões iniciais (i.e., as hipóteses que queremos avaliar) e também como avaliá-las cada uma. Para situar as variáveis de pesquisa adotadas, a Figura 1 ilustra o fluxo do participante no estudo que estamos desenhando, bem como o momento em que cada variável é coletada.

H1 Sim, existe associação entre o nível de surpresa reportado e o grau de motivação em buscar mais informações. Em outras palavras, conhecido o nível de surpresa (S), e tomando a quantidade de informações consultadas (E) como medida indireta do grau de motivação para buscar mais informações, então é possível prever o grau de motivação para buscar mais informações (E), com alguma precisão. Formalmente,

$$E = \alpha S + \epsilon, \quad (1)$$

onde α é uma constante inferida sobre os dados, S é o nível de surpresa reportado, ϵ é o erro de predição, e E corresponde ao número de *clicks* em *links* que oferecem mais informação sobre a recomendação, incluindo *links* de primeiro e segundo níveis.

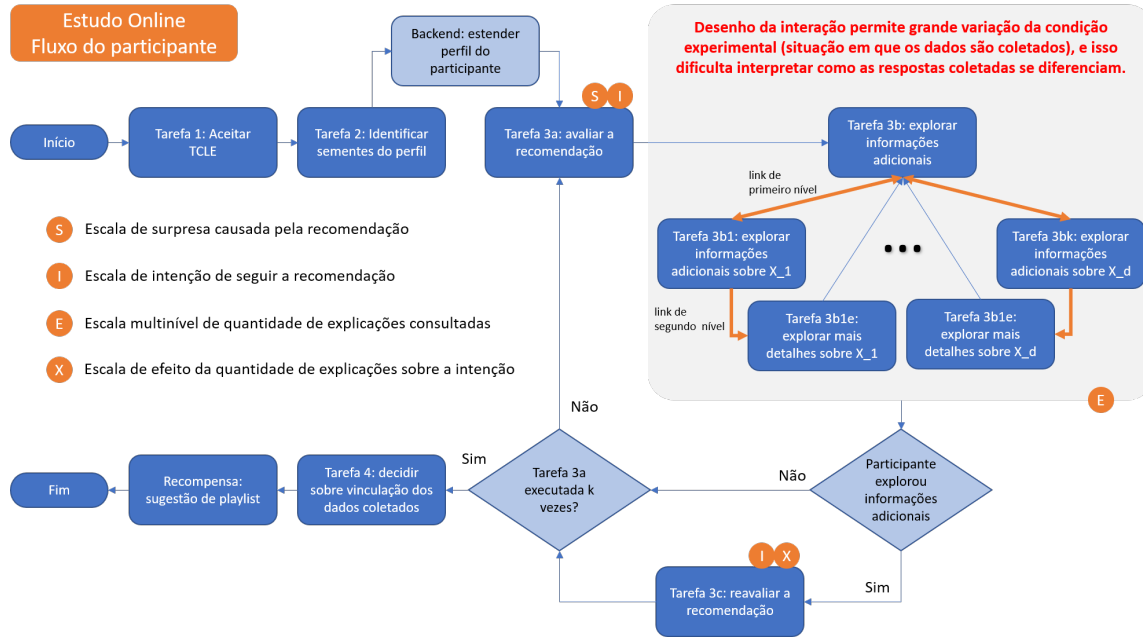


Fig. 1. Fluxo do participante no estudo online.

H2 Sim, o participante busca preferencialmente por uma categoria específica de informação quando se sente motivado a fazê-lo. Em outras palavras, quando o participante se sente motivado a buscar mais informações sobre o item recomendado, uma das categorias é predominantemente mais consultada do que as demais. Formalmente:

$$\exists i : E^i > E^j, \forall j \neq i, \quad (2)$$

onde E^i corresponde ao número de *clicks* em *links* que direcionam o participante para uma página que oferece informações adicionais da categoria i .

H3 Sim, oferecer informações adicionais exerce efeito sobre a intenção do participante em consumir o item recomendado. Formalmente:

$$E > 0 \implies |I(t = 1) - I(t = 0)| > 0, \quad (3)$$

onde $I(t = 0)$ corresponde ao nível de intenção em seguir a recomendação reportado na Tarefa 3a, e $I(t = 1)$ é o nível de intenção reportado na Tarefa 3c. **Como essa tarefa introduz muita variação nas condições em que E é coletada, receio que os dados não suportem a Inequação 3.** Para reduzir esse risco, sugiro incluirmos na Tarefa 3c uma escala para avaliar o efeito das informações consultadas sobre a intenção (X):

$$X \notin [\bar{X} - \theta, \bar{X} + \theta] \iff |I(t = 1) - I(t = 0)| > 0, \quad (4)$$

onde θ é uma constante definida empiricamente e que divide a amostra de respostas em duas classes. Um outro benefício dessa inclusão é que teremos **evidência direta** de que a variação na intenção é causada pela oferta de informações adicionais.

Vale dizer que, nas Inequações 2, 3 e 4, as variáveis aleatórias serão interpretadas como intervalos de confiança (IC) para a média para fins de teste de hipótese¹. Particularmente, a Inequação 4 será avaliada pela acurácia obtida na separação das duas classes. Já a Equação 1 será avaliada da forma usual para modelos de regressão linear (IC do parâmetro α).

¹Este procedimento segue o princípio de que “não existem pontos, somente bolas”, ou, mais formalmente, devido a limitações profundas e incontornáveis em se avaliar o verdadeiro valor de qualquer propriedade (e de qualquer entidade), o único resultado possível de qualquer ato de mensuração é a obtenção de um intervalo que inclui o verdadeiro valor da propriedade de interesse com uma probabilidade aceitável.

A Tabela 1 registra algumas características importantes das variáveis de pesquisa adotadas.

Table 1. Variáveis de pesquisa. Coluna p indica o número de níveis da escala base, q representa o número de itens da escala específica da variável, min e max são respectivamente o menor ou maior valor que pode ser assumido pela variável, “modelo de mensuração” descreve a interpretação do relacionamento ente itens e a variável do experimento (vista como uma variável latente), e “composição por” registra a forma como os escores obtidos de cada item são agregados para compor o valor assumido pela variável.

Var.	Descrição	p	q	min	max	modelo de mensuração	composição por
S	Nível de surpresa causada pela recomendação	5	3	3	15	reflectivo	soma ou soma ponderada
I	Nível de intenção em seguir a recomendação (H3: $\Delta I = \beta E + \epsilon = \gamma X + \epsilon$)	5	2	2	10	reflectivo	soma
E	Número de <i>clicks</i> em <i>links</i> para mais informações. É assumida como uma medida indireta para o grau de motivação para buscar mais informações. (H1: $E = \alpha S + \epsilon$)	2	4	0	9	não tradicional	soma
X	Efeito da quantidade de informações adicionais consultadas na intenção de seguir a recomendação	5	3	3	15	reflectivo	soma

Finalmente, tendo feito as considerações acima, assumo que o dataset que coletaremos deve ter um formato como o descrito na Tabela 2.

Table 2. Tabulação das respostas coletadas. S corresponde ao nível de surpresa reportado; I_0 é o grau de intenção em seguir a recomendação coletado na Tarefa 3a; E é o número de *clicks* sobre *links* para mais informações; I_1 é o grau de intenção em seguir a recomendação coletado na Tarefa 3c; X é o nível de efeito da quantidade de explicações sobre a intenção de seguir a recomendação. H1, H2 e H3 indicam se a amostra contribui na avaliação de qualquer dessas hipóteses. Adotados $k = 4$ e $d = 4$

Participante	Item	S	I_0	E				I_1	X	H1	H2	H3
Amostras ilustrativas que suportam as hipóteses												
Alice	xcI3QA6OaW	15	2	1 1	1 1	1 1	1 1	10	15	✓	✓	✓
Alice	9uPkyYEcCv	3	2	1 0	0 0	0 0	0 0	2	3	✓	✓	✓
Amostras ilustrativas que enfraquecem as hipóteses												
Alice	3BksjNYQ5y	15	10	1 1	0 0	0 0	0 0	10	3	✓	✓	✓
Alice	khKZz8Bml1	3	10	0 0	1 1	0 0	0 0	10	15	✓	✓	✓
Outras amostras ilustrativas para casos relevantes												
Beatriz	LIKGHhRFiP	3	10	1 0	1 0	0 0	0 0	5	15	✓	✓	✓
...												
Zuleika	NG4mXllqRC	3	2	0 0	0 0	0 0	0 0	-	-	✓	✓	-
Zuleika	NG4mXllqRC	3	2	0 0	0 0	0 0	0 0	2	-	✓	✓	✓

2 DISCUSSÃO SOBRE AS ESCALAS USADAS NO ESTUDO ONLINE

Assumo que todos os itens das escalas descritas a seguir estão associados a uma escala Likert de 5 níveis, variando de “descordo totalmente” com “concordo totalmente”.

2.1 Escala para avaliar surpresa (variável *S*)

Itens candidatos para a escala de surpresa foram extraídos de Kotkov et al. 2018. Originalmente, esses itens foram desenhados para avaliar o construto *unexpectedness*. Entretanto, por conta de uma análise do coletado usado no estudo, acredito que esses itens possam ser usados para avaliar surpresa (ao receber uma recomendação, em contraste com avaliação após consumir um item), como concebida por Kaminskis and Bridge 2014:

- s5 This movie is different (e.g., in style, genre, topic) from the movies I usually watch.
- s8 Watching this movie broadened my preferences.
- s4 This is the type of movie I would not normally discover on my own.
- s6 I was surprised that Movielens picked this movie to recommend to me.

Adaptação sugerida: traduzir e adaptar para o contexto de aplicação.

- s5 Essa música é diferente (em estilo, em ...²) das músicas que eu costumo ouvir.
- s8 Eliminar porque não condiz com o estudo que estamos desenhando.
- s4 Eu raramente encontro músicas como essa quando busco novas músicas para ouvir.
- s6 Eu fiquei surpreso pelo fato dessa música ter sido recomendada para mim.

2.2 Escala para avaliar intenção (variável *I*)

Itens candidatos para a escala de intenção extraídos de Van Der Heijden 2004. Originalmente, esses itens foram desenhados para avaliar o construto *behavioural intention (of using the system)*, como proposto no modelo TAM2 [Venkatesh and Davis 2000].

- b1 I intend to revisit [the system] shortly.
- b2 I predict that I will revisit [the system] in the short term.

²Essa lista deve referir apenas detalhes que estejam expostos na interface da Tarefa 3a.

Adaptação sugerida: como o nosso propósito para essa escala é medir a intenção do participante em seguir a recomendação, acho que é melhor adaptar os itens substituindo o comportamento avaliado de *use the system* para *consume the item*.

b1 Eu pretendo ouvir a música que me foi recomendada assim que possível.

b2 É bem provável que eu escute essa música nos próximos dias.

2.3 Escala para avaliar efeito da intervenção (variável X)

Itens candidatos para essa escala extraídos de Yi and Davis 2003. Originalmente, esses itens foram desenhados para avaliar o construto *motivation (to use the system)*.

m1 The training provided information that motivated me to use Excel.

m2 The training helped me see the usefulness of Excel.

m3 The training increased my intention to master Excel.

m4 The training showed me the value of using Excel in solving problems.

Adaptação sugerida: como o nosso propósito para essa escala é medir o efeito que “prover explicações” (vista como uma intervenção) exerce sobre a intenção de seguir uma recomendação, acho que é melhor adaptar os itens substituindo a intervenção de *provided training* para *provided explanations*³.

m1 As explicações que eu li me motivaram a ouvir a música que me foi recomendada.

m3 Depois de ler as explicações, fiquei com vontade de ouvir essa música.

m2 As explicações que consultei me ajudaram a ver algo interessante sobre a música.

m4 **Eliminar porque este item é centrado na funcionalidade do sistema e não nos resultados que se obtém dessa funcionalidade.**

É importante notar que essa escala, diferentemente das anteriores, possui dois polos: valores altos para a variável X indicam que as informações consultadas aumentaram o interesse do respondente em ouvir a música, ao passo que valores baixos indicam que as informações consultadas tiveram o efeito contrário, o de reduzir o interesse em ouvir a música recomendada.

³Não estou certo de que temos boas justificativas para classificar esse objeto como uma explicação.

REFERENCES

- Marius Kaminskas and Derek Bridge. 2014. Measuring surprise in recommender systems. In *Proceedings of the workshop on recommender systems evaluation: dimensions and design (Workshop programme of the 8th ACM conference on recommender systems)*. Citeseer.
- Denis Kotkov, Joseph A. Konstan, Qian Zhao, and Jari Veijalainen. 2018. Investigating Serendipity in Recommender Systems Based on Real User Feedback. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing (Pau, France) (SAC '18)*. Association for Computing Machinery, New York, NY, USA, 1341–1350. <https://doi.org/10.1145/3167132.3167276>
- H. Van Der Heijden. 2004. User acceptance of hedonic information systems. *MIS Quarterly: Management Information Systems* 28, 4 (2004), 695–704. <https://doi.org/10.2307/25148660> cited By 2057.
- V. Venkatesh and F.D. Davis. 2000. Theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science* 46, 2 (2000), 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926> cited By 8960.
- M.Y. Yi and F.D. Davis. 2003. Developing and validating an observational learning model of computer software training and skill acquisition. *Information Systems Research* 14, 2 (2003), 146–169+219. <https://doi.org/10.1287/isre.14.2.146.16016> cited By 366.