Using Data and Hypothesis Testing

CSCI 16: Intro to the CS Research Process
6th Meeting, Thursday, January 12, 2023

Kelly Shaw

Attribution

Except where otherwise noted, ERSP and all associated resources were developed by Christine Alvarado, UC San Diego, Department of Computer Science and Engineering.

These slides were originally jointly developed by Christine Alvarado and Mai ElSherief. They have been adapted by Kelly Shaw for CSCI 16 at Williams College.

Today's Plan

- From Experiments to Data Interpretation
- Discussion of paper topics

CHI '19 Paper: *Emotional Utility and Recall of the Facebook News Feed* by Nontasil and Payne

We report a laboratory study (N=53) in which participants browsed their own Facebook news feeds for 10-15 minutes, choosing exactly when to quit, and later rated the overall emotional utility of the episode before attempting to recall threads. Finally, the emotional utility of each encountered thread was rated while looking over a recording of the interaction. We report that Facebook browsing was, overall, an emotionally positive experience; that recall of threads exhibited classic primacy and recency serial order effects; that recalled threads were both more positive and more valenced (less neutral) on average, than forgotten threads; and that overall emotional valence judgments were predicted, statistically, by the peak and end thread judgments. We find no evidence that local quit decisions were driven by the emotional utility of threads. In the light of these findings, we discuss the suggestion that emotional utility might partly explain the attractiveness of reading the news feed, and that an emotional memory bias might further increase the attractiveness of the newsfeed in prospect.

What does this mean?
What are the implications?

Overarching motivation

- The researchers aimed to understand "more of the structure of [users'] emotional responses [to the Facebook News Feed] [in order to] understand why, beyond its broad function, the news feed is attractive."

But how?

Specific Research Questions

- Read section 3 in the paper. If you need to understand terms, you can refer back to section 2 where they are defined.
- **RQ1: What is the remembered emotional experience of a brief period of Facebook news feed reading, and how is it related to the emotional response to each encountered thread?**
- **RQ2: Does memory for threads on the news feed show classic serial order effects, despite the variety of thread content?**
- **RQ3: Is memory for news feed threads predicted by their emotional valence?**
- **RQ4: When do people give up reading the news feed?**

Laboratory Study

- 53 participants
- 5 phases, in the same order for each participant:
  1. "Interacting with Facebook by reading the news feed"
  2. "responding to a [demographic and Facebook usage] questionnaire"
  3. " judging the overall emotional utility of the Facebook–browsing episode"
  4. "recalling Facebook threads"
  5. "judging the emotional utility of each encountered Facebook thread"

Match each phase to the relevant research question(s). RQs may match more than one phase.

Laboratory Study

- 53 participants
- 5 phases, in the same order for each participant:
  1. "Interacting with Facebook by reading the news feed" RQ1, RQ2, RQ3, RQ4
  2. "responding to a [demographic and Facebook usage] questionnaire"
  3. " judging the overall emotional utility of the Facebook–browsing episode" RQ1
  4. "recalling Facebook threads" RQ2, RQ3
  5. "judging the emotional utility of each encountered Facebook thread" RQ1, RQ3, RQ4

Descriptive Data

- Read Section 5: Descriptive Data + Table 1
- When presenting results, it is common to start with "descriptive data".
  - What descriptive data is presented here?
  - What purpose does it serve? Does it directly answer any research questions?

Answering RQ1: **What is the remembered emotional experience of a brief period of Facebook news feed reading, and how is it related to the emotional response to each encountered thread?**

- The authors will use *correlation* to answer this question. So first, let's get an intuition for what correlation is.
  - https://www.youtube.com/watch?v=ugd4k3dC_8Y

Answering RQ1: **What is the remembered emotional experience of a brief period of Facebook news feed reading, and how is it related to the emotional response to each encountered thread?**

- Read section 5: RQ1
- What are the variables on the X and Y axes in each correlation?

- What are the data points on each plot?

- Bonus: What is the p value in each result? (You can look this up if you need to)

What are p-values?

- A p-value is the probability of the *null hypothesis* being true. The *null hypothesis* states that the results you are seeing are due to chance instead of a true phenomenon.
  - In this case, the null hypothesis is that there is actually no relationship between the variables, and the p-value gives us a probability that the apparent relationship we are seeing is just a normal chance occurrence.
- The lower the p-value, the less likely it is that the null hypothesis is true and the more you can trust the results you are seeing.
- Typically, in CS research, $p < 0.05$ indicates a sufficiently high chance that the underlying relationship the model shows can be believed. The result is said to be *statistically significant* (or just *significant*).

Correlation summary

- Correlation is used to determine the strength of the relationship between two variables.
- The absolute value of the correlation coefficient (r) is higher when a change in one variable more reliably corresponds to a change in the other.
- The p-value indicates how likely it is that the relationship between the variables is just due to random chance. The lower the p value, the more you can trust the correlation coefficient.

Regression: Like correlation, but a little more flexible, with "direction"

"To further test the peak-end rule we performed a multiple regression with peak and end as the predictor variables and overall rating as the dependent variable. This regression was significant ($F(3,49) = 4.70$, $p < .05$, $r^2 = .22$)."

https://www.youtube.com/watch?v=zPG4NjIkCjc

Answering RQ3: **Is memory for news feed threads predicted by their emotional valence?**

"RQ3: Emotional utility and recall

We predicted that people would better recall emotionally more positive threads. The data in Table 1 show that, on average, the recalled threads were rated more positively than the forgotten threads."

Why is this not enough to answer the research question?

Answering RQ3: **Is memory for news feed threads predicted by their emotional valence?**

Hypothesis: Recalled threads were rated more positively (or negatively) than the forgotten threads (and this was not due to chance!)

Null Hypothesis:

Answering RQ3: **Is memory for news feed threads predicted by their emotional valence?**

Hypothesis: Recalled threads were rated more positively (or negatively) than the forgotten threads (and this was not due to chance!)

Null Hypothesis: On average, recalled threads were rated no more positively or negatively than the forgotten threads.

Answering RQ3: **Is memory for news feed threads predicted by their emotional valence?**

Hypothesis: Recalled threads were rated more positively (or negatively) than the forgotten threads (and this was not due to chance!)

Null Hypothesis: On average, recalled threads were rated no more positively or negatively than the forgotten threads.

Statistical technique: Paired t-test

https://www.youtube.com/watch?v=pTmLQvMM-1M

Read section Results: RQ3

Answering RQ3: **Is memory for news feed threads predicted by their emotional valence?**

Hypothesis: Recalled threads were rated more positively (or negatively) than the forgotten threads (and this was not due to chance!)

Null Hypothesis: On average, recalled threads were rated no more positively or negatively than the forgotten threads.

Statistical technique: Paired t-test

Data points:

How to tell if the difference is non-trivial

"The effect was significant (t (52) = 5.858, p<.01, **d = 0.805**)."

- A t-test can reject the null hypothesis, but the difference between the groups might be trivial
- Cohen's d is a measure of *effect size* which tells you how non-trivial the difference is.
  - Small effect = 0.2
  - Medium Effect = 0.5
  - Large Effect = 0.8

t-test summary

- A t-test can be used to test whether or not two groups have the same means. A paired t-test requires points in the two groups to be linked.
- Data in each group should be normally distributed.
- A low p value means it is unlikely the difference in means is due to chance and tells you how strongly you can believe your results ($p < 0.5$ is typically used for "statistical significance").
- The effect size gives a measure of how non-trivial the difference is.

Stats we've seen, and other tests to consider

- Descriptive Statistics – to get an overview of your data
- Correlation – relationship between two variables
- Regression – relationship between independent variable(s) and dependent variable
- T-test –to test whether two groups have different means
- ANOVA – like a t-test but with more than two groups
- Non-parametric tests (Chi-squared, Wilcoxon Sign-Rank, etc.) – Compare groups when you don't have normally distributed data.
- How to choose? https://quizlet.com/gb/256872448/bio-stats-flow-chart-diagram/

Paper Discussion