Article n.47

Managing Web Experiments for Psycholinguistics: An Example from Experimental Semantics/Pragmatics

Article by J. Degen and J. Tonhauser Presentation by Sungwoo Kwon

Linguistics and Data Science 2023F

October 29, 2023

Outline

- Introduction
- Data Management
- 3 Data Collection
- 4 Data Analysis
- Conclusion
- 6 Quiz

Introduction

- Presenting the workflow of a study conducted in the past by the authors, rather than data itself.
- Experimental Data from HITs(Human Intelligence Tasks).
- Research Topic
 - Investigating the extent to which variability in a content's projectivity is predicted by that content's at-issueness.
 - Does Felipe regret that Mike visited Alcatraz?
 - Pelipe doesn't regret that Mike visited Alcatraz.
 - Ooes Felipe think that Mike visited Alcatraz?
 - Felipe doesn't think that Mike visited Alcatraz.

Data Management

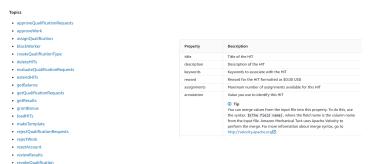
- The Organizational Framework for Collaboration
 - ► Github, Git, Slack
- Project Repository Structure
 - Generally contains
 - ★ README.md
 - * experiments
 - ★ results (or analysis)
 - paper
 - Additionally contains
 - ★ data
 - * models
 - ★ talks

Data Collection

- Amazon Mechanical Turk
 - A crowdsourcing platform for HITs
 - many advantages over in-lab studies
 - ★ Larger and diverse participants
 - ★ Rapid completion times
- Workflow in AMT
 - ▶ Web-programmed experiments using JS,HTML,CSS etc.
 - Supersubmiterator for interfacing with AMT
 - Directory management

Supersubmiterator

- A tool for managing external HITS on Amazon Mechanical Turk
- Makes it easy to retrieve the data once the experiment is completed, with participant IDs anonymized and results formatted into a .csv file.
- Up-to-date with current AMT configurations.
- Provides automatic batching feature.



https://docs.aws.amazon.com/AWSMechTurk/latest/AWSMturkCL/Welcome.html

Supersubmiterator

```
1 {
2  "liveHIT":"yes",
3  "title":"Language study",
4  "description":"You will rate short dialogs.",
5  "experimentURL":"https://web.stanford.edu/~jdegen/exp1a/",
6  "keywords":"language research fun cognitive science linguistics university",
7  "USonly?":"yes",
8  "minPercentPreviousHITsApproved":"95",
9  "frameheight":"650",
10  "reward":"1.00",
11  "numberofassignments":"250",
12  "assignmentduration":"1800",
13  "hitlifetime":"2592000",
14  "autoapprovaldelay":"60000"
1  "autoapprovaldelay":"60000"
1  "
```

Web-Based Experiment Pipeline

- 1. Program experiment as external website and copy it to web space.
- 2. Create mturk directory inside the experiments directory where all Mechanical Turk-related files are stored and from where the Submitterator command will be called. 6. Run from command line to reformat data for worker
- 3. Create experiment.config file (see example in figure 47.1) in mturk.
- 4. Run from command line to post experiment to 7. Copy generated data file experiment.csv to its corre-Mechanical Turk:
 - submiterator posthit experiment

5. Run from command line to retrieve data from Mechanical Turk:

submiterator getresults experiment

- ID anonymization and easy analysis in R: submiterator reformat experiment
- sponding data directory in results directory.

- Open Source Statistical Software.
- Ease in managing active analysis scripts and plots.
- Visualisation.
- Syntax highlighting and prediction.
- Integration with R Markdown and knitr.

Data Analysis

- Ime4, brms, tidyverse, ggplot2, Ismeans
- Results Directory
 - data
 - graphs
 - rscripts

Conclusion

- Organizational framework is useful.
 - Reproducibility
 - Generalization
- Preregistration
 - Documenting one's research plan prior to its realization in a public repository, such as OSF.
 - Separating the confirmatory aspects of the research from the exploratory aspects.
 - Allowing for input from colleagues.
 - Allowing journal to accept researchers' papers conditionally.

Quiz

 What are the main reasons Organizational Framework is useful for studies involving human-generated data, especially in Humanities and Social Sciences? (Two answers)