Intro to Machine Learning for Qualitative Research

Corey M. Abramson (University of Arizona)
Zhuofan Li (University of Arizona)
Daniel Dohan (UC San Francisco)

GOALS

- Understand Machine Learning Basics
- Understand how Machine Learning can aid qualitative researchers
- Familiarize with software tools

Funding, Acknowledgements, Team

Partners

MEDICAL CULTURES LAB

Acknowledgements

Alma Hernandez

Melissa Ma

Stephen Z.

Funding (Presented Works):

NIH NIA DP-1 1DP1AG069809-01 (Dohan, PI; Abramson, Co-I)

Research, Development, and Innovation Faculty Seed Grant, University of Arizona (Abramson, PI)

NIH

THE UNIVERSITY OF ARIZONA

DISCLAIMER: Opinions are ours and not those of funders, partners or collaborators.

Outline

I. Overview: Machine Learning and Qualitative Methods

II. Orientation to python (using colab notebook)

III. Hybrid Workflows

IV. Practical Exercises

- Importing Data
- Extending human coding
- Exporting results
- Visualizing
- Other: as time allows

PART I: Core Concepts

Definitions (1)

Qualitative Analysis

- Umbrella term for various data types
  - *includes almost anything that is not closed ended survey*
  - *E.g. interviews, participant observation, focus groups, historical records, digital media*
- Range of epistemic perspectives even within a method, like participant observation (Abramson and Gong 2020)
- Definitions are a point of confusion, contestation, and gatekeeping

SIMPLIFIED HERE: An examination of data that includes text and is not fully reduced to numbers.

Definitions (2)

Coding

- Lots of jargon

SIMPLIFIED HERE: Qualitative "coding" is the process where we tag (or ask a computer to tag) 'qualitative data." …

Coding is akin to the application of researcher #hastags to or keywords to bits of text. Often done through a computer.

Definitions (3)

Machine Learning

- AI makes predictions, by using algorithms to ‘learn’ from data you provide
- Lots of uses : Organize text, identify patterns, visualize information (alongside misuses)

PURPOSE HERE**: **Accurately #tag paragraphs of text by learning from human coders familiar with the data.

But Why?

- Accurately index text with less effort, improve accuracy and save time for higher level analyses (Li, Abramson, Dohan 2021, summary)
  - Can maintain iterative qualitative approach and conventions (e.g. Deterding and Waters 2021)
- Emerging possibilities for generating inductive insights (Nelson 2021)
- Repurposing existing data and opening new possibilities (Abramson et. 2018)

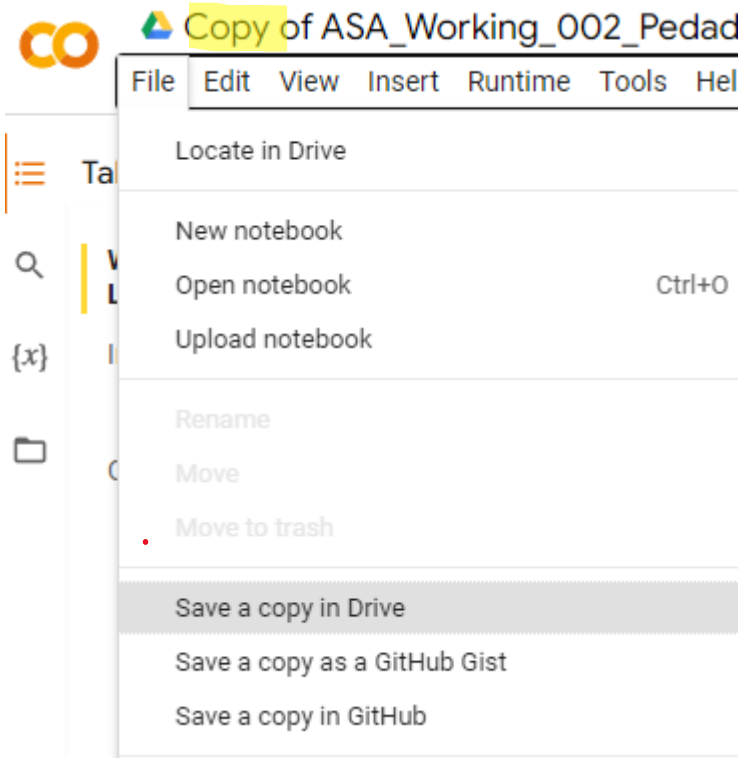
PART II: Using Google Colab

Step 1: Open Colab

- Go to: https://tinyurl.com/4v3mvrvt
- This will require you log in with a google account (sorry)
- All downloads to local machines are disabled in our code.

Step 2: Making a Working Copy

- Save a copy in drive
- Close first browser window
  - *You should be working on a version that says "copy of…"*
- Now, to walk through basics…



PART IV: Workflow

Definitions (4)

Workflow

- Steps to go from start to finish for a task
- Narrower than a process or methodological approach

Workflows

Examples:

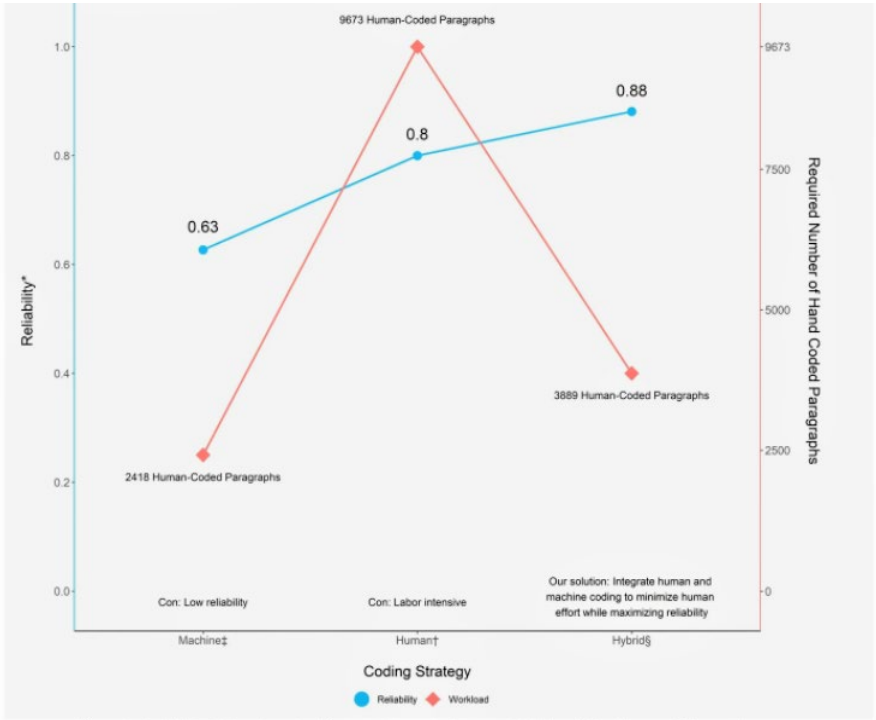
1. Text mining
2. Formal modeling
3. Scale qualitative coding

#3 provides some tools for various qualitative traditions, and is our focus here

SOCIUS
Data Visualization
Socius: Sociological Research for a Dynamic World
Volume 7: 1–4
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23780231211062345
srd.sagepub.com
SAGE

Qualitative Coding in the Computational Era: A Hybrid Approach to Improve Reliability and Reduce Effort for Coding Ethnographic Interviews

Zhuofan Li[1], Daniel Dohan[2], and Corey M. Abramson[1]



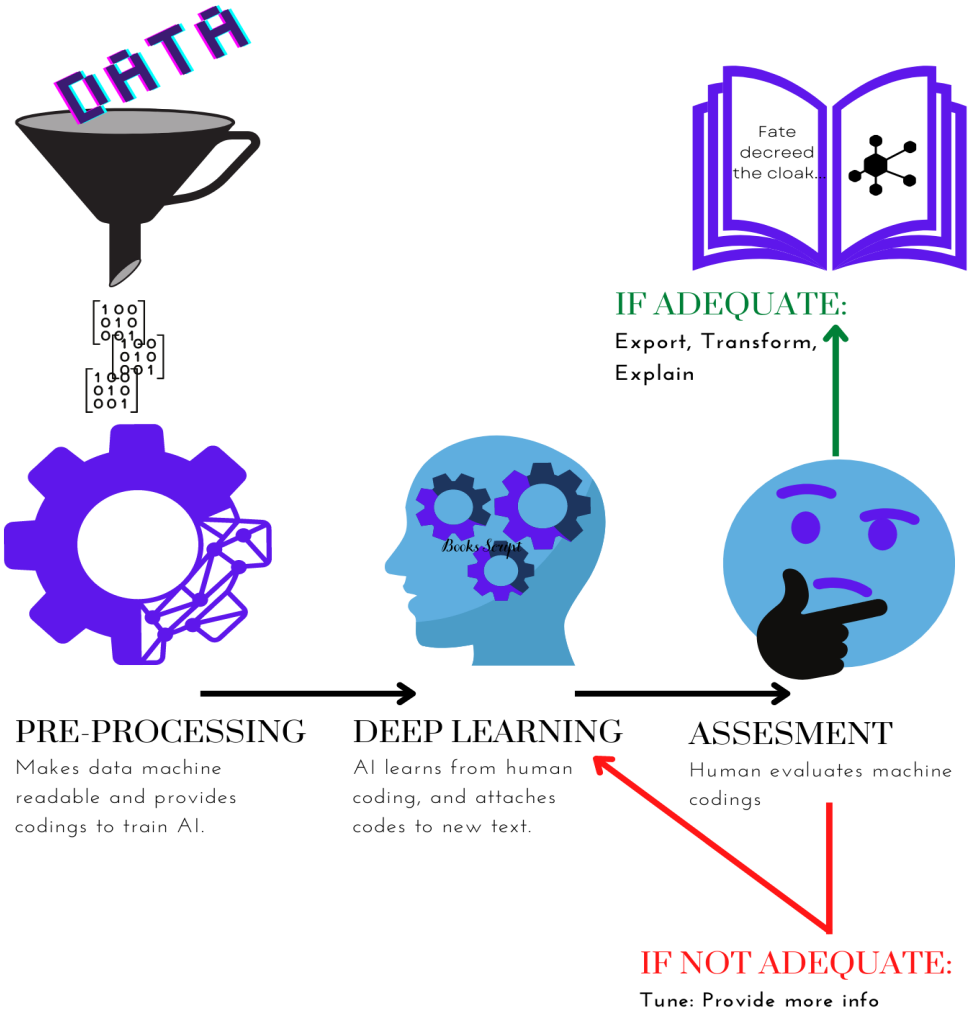
SAMPLE QUAL+MACHINE LEARNING WORKFLOW

Assumes researcher(s) have data and some initial codes that can be applied consistently.

- Code training data
- Preprocess (make python readable):
  - Export from QDA
  - Or match this spreadsheet (p.15)
- Scale coding
  - AI learns from human codings
  - AI applies to new text
- Output/ Export
- *Analyze and explain*

Table 1A: Sample Tabular Output From ATLAS.ti" Before Importation into Python

Interview ID	Paragraph ID	Quotation Content	Codes
518	133	Respondent: [text here]	[Code 1] [Patient Information] [Code 3] ...
518	134	Respondent: [text here]	[Code 1] [Code 2] [Code 3] ...
518	135	Respondent: [text here]	[Code 1] ...
518	136	Respondent: [text here]	[Code 1] ...
518	137	Respondent: [text here]	[Patient Information] ...
518	138	Respondent: [text here]	[Patient Information] [Code 3] ...
518	139	Respondent: [text here]	[Code 1] [Patient Information] ...


DATA
Fate decreed the cloak..
IF ADEQUATE:
Export, Transform, Explain
PRE-PROCESSING
Makes data machine readable and provides codings to train AI.
DEEP LEARNING
AI learns from human coding, and attaches codes to new text.
ASSESMENT
Human evaluates machine codings
IF NOT ADEQUATE:
Tune: Provide more info