Unit 7: Data Annotation

Human coding

Lesson 50

Derek Ruths

Overview of unit

Objectives:
- Understand how to approach data annotation (whether automated or manual)
- Know how to run a small human annotation task

1. Typologies
2. Building a typology
3. Applying a typology – human coding
4. Confirming an annotation
5. Applying a typology – classifier

Lesson overview

Objectives

- Understand how to obtain human-provided annotations
- Understand design pitfalls when using human coding

Outline

- What is human coding?
- Human coding approaches
- Ethics approval
- Training

What is human coding?

Approaches to human coding

Expert coding

- You and your collaborators
- A small group of trained participants (paid or unpaid)

Crowd sourcing ← *mostly strangers, little quality control*

- Amazon Mechanical Turk

cheaper

higher fees for better coding

how to use design coding to get high quality results

Number of annotators?

• One annotator… can make mistakes

↳ one person ⇒ no one can check their work !!

• Two or more annotators … will inevitably disagree. What to do with disagreements?
  • Code them yourself?
  • Have more coders annotate them?
  • Have another category?

- score by proportion
- score by majority

category/task issue

- review and decide for yourself (most labor intensive)

- throw data point out (cheap & undesirable)

- throw data out & note it but not super compelling

Ethics approval

* do no harm
* If there is risk, the benefits must outweigh them

what is the impact
  ↳ on the people doing the annotating
  ↳ people who see the end result
  ↳ the researchers

content moderation is a form of annotation

- after project decided, taxonomy defined and # coders picked

Training

Help coders understand the concepts and how they apply at a deep level.

- Prepare a **code book** (an extension to the typology definition document)
- Prepare challenging examples
- Run group sessions with ample time to code content and discuss codings

Often training can help refine the typology.

The human coding process

① Preparation phase
- Decide on human coding group (experts, small team, crowd)
- Develop code book/training materials
- Develop annotation software (if needed)

② Annotation phase
- Recruit coders
- Run training session
- Run coding exercise (often spans days, sometimes weeks)

③ Validation phase
- Throughout, assess the quality and level of agreement among coders

Lesson wrap-up

Takeaways

- Human coding is a lot of work
- It's also often the only way to get good annotations

Up next

- Confirmation annotation quality