DATA COLLECTION : LABELING

Motivation : often, labels not readily available in sampled data

Examples :
- Search - The results a customer wants to see
- Music categorization - genre
- Sentiment Analysis - overall attitude of the writer
- Digitization - the transcription of the handwriting
- Object detection - the location of objects in the image

Sometimes labels can be inferred (e.g. from click - through data)

Labeling Components

Guidelines
- Instructions to labelers
- Critical to get right
- Minimize ambiguity

Labeling Tools

Technology
- Excel spreadsheets
- Amazon mechanical turk
- Custom tools

Questions
- Human intelligences tasks (HITS) should be
  - Simple
  - Unambiguous

Poor design of either can
1: impact labeler productivity & quality
2: introduce bias

Labeling Tools

Amazon Mechanical Turk
- Obtain human intelligence on demand
- Access a global, on-demand 24x7 workforce
- Pay for what you use
- Use for labeling

Managing Labelers

Motivation
Plurality
Gold Standard Hits
Auditors
Labeler Incentives
Quality & Productivity Metrics

Managing Labelers

Plurality

Assign each HIT to multiple labelers to identify difficult or ambiguous cases, or problematic labelers.

Gold Standard HITs

HITs with known labels mixed in to identify problematic labelers

Auditors

- Experts to adjudicate labeler disagreements and/or sample results for quality

Incentives

Compensation, rewards, voluntary, gamification

Quality and Productivity

- e.g. consistency, accuracy can detect problems with labelers

Sampling And Treatment Assignment

	Random Assignment	No Random Assignment	Generalizat
Random Sampling	Ideal experiments: causal conclusion and can be generalized.	Typical survey or observation studies. Cannot establish causation but can find correlation and be generalized	
No Random Sampling	Most experiments: causal conclusion only for the sample (more work to generalize)	Badly designed survey or pooled studies. Cannot establish causation, cannot be generalized	No Gen