Interpretation

Advanced Machine Learning for NLP
Jordan Boyd-Graber
HUMAN IN THE LOOP EVALUATION

- User study
- 40 minutes
- Sort documents into categories
- What information / interface helps best

- User study
- 40 minutes
- Sort documents into categories
- What information / interface helps best

- User study
- 40 minutes
- Sort documents into categories
- What information / interface helps best
  - Train a classifier on human examples
  - Compare classifier labels to expert judgements

- User study
- 40 minutes
- Sort documents into categories
- What information / interface helps best
  - Train a classifier on human examples (don't tell them how many labels)
  - Compare classifier labels to expert judgements

- User study
- 40 minutes
- Sort documents into categories
- What information / interface helps best
    - Train a classifier on human examples (don't tell them how many labels)
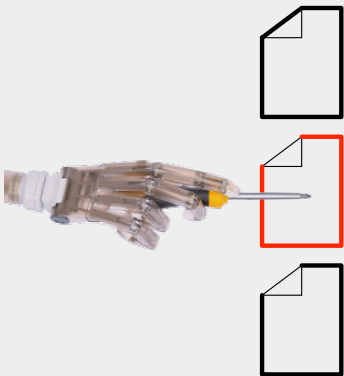   - Compare classifier labels to expert judgements (purity)

$$\text{purity}(\mathbf{U}, \mathbf{G}) = \frac{1}{N} \sum_l \max_j |U_l \cap G_j|, \tag{1}$$

Who should drive?

Which is more Useful?

Active Learning

Topic Models

TOPIC 1 — computer, technology, system, service, site, phone, internet, machine

TOPIC 2 — sell, sale, store, product, business, advertising, market, consumer

TOPIC 3 — play, film, movie, theater, production, star, director, stage

- ○ evacuation
- ○ safety
- ● shelter

[new label name]

add label

rename label **delete label**

Covered Themes Progress:

flood coast marine
restoration coastal vessel
fish gulf wildlife species
pollution council great
fishery fishing waters
ecosystem monitoring
fisheries mitigation

remain
expended heading
disaster september
appropriation transferred
obligation division unit
capital acquisition
inspector purchase funded
procurement units corps
repair salaries

To provide for payments to certain natural resource trustees to assist in re...

A bill to authorize the Secretary of the Army to carry out activities to man...

A bill to prevent forfeited fishing vessels from being transferred to private ...

To reauthorize various Acts relating to Atlantic Ocean marine fisheries.

To amend the Magnuson-Stevens Fishery Conservation and Managemen...

To prevent forfeited fishing vessels from being transferred to private parti...

To amend the Magnuson-Stevens Fishery Conservation and Managemen...

A bill to require the Secretary of the Army to study the feasibility of the hy...

Making appropriations for disaster relief requirements for the fiscal year e...

To rescind any unobligated discretionary appropriations returned to the F...

To amend the Robert T. Stafford Disaster Relief and Emergency Assistanc...

Making appropriations for energy and water development and related age...

To provide for payments to certain natural resource trustees to assist in re...

A bill to authorize the Secretary of the Army to carry out activities to man...

A bill to prevent forfeited fishing vessels from being transferred to private ...

To reauthorize various Acts relating to Atlantic Ocean marine fisheries.

To amend the Magnuson-Stevens Fishery Conservation and Managemen...

To prevent forfeited fishing vessels from being transferred to private parti...

To amend the Magnuson-Stevens Fishery Conservation and Managemen...

A bill to require the Secretary of the Army to study the feasibility of the hy...
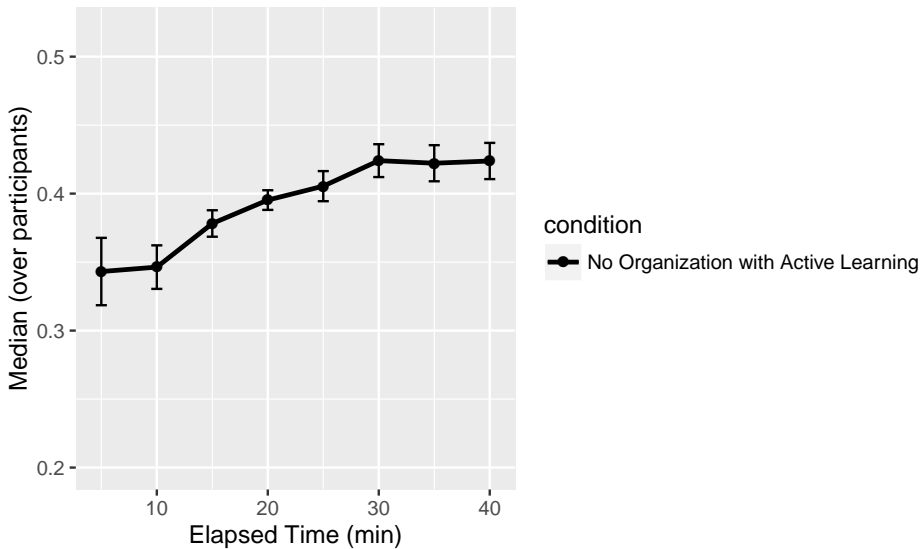
Making appropriations for disaster relief requirements for the fiscal year e...

To rescind any unobligated discretionary appropriations returned to the F...
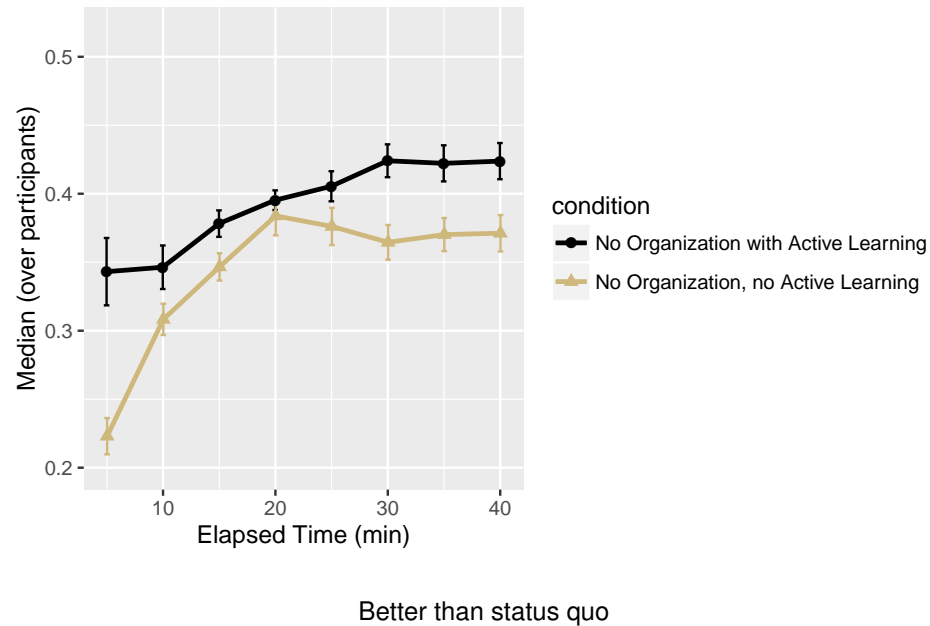
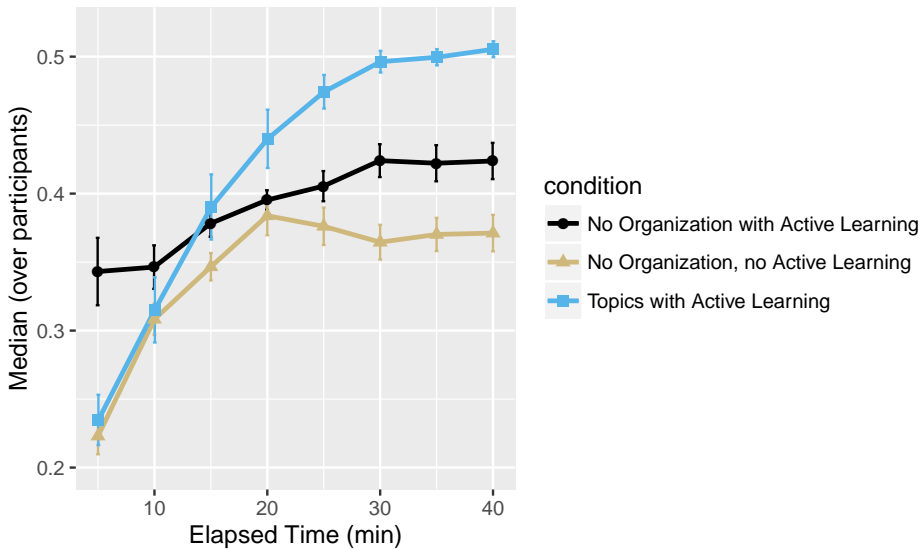To amend the Robert T. Stafford Disaster Relief and Emergency Assistanc...

Making appropriations for energy and water development and related age...

- evacuation
- safety
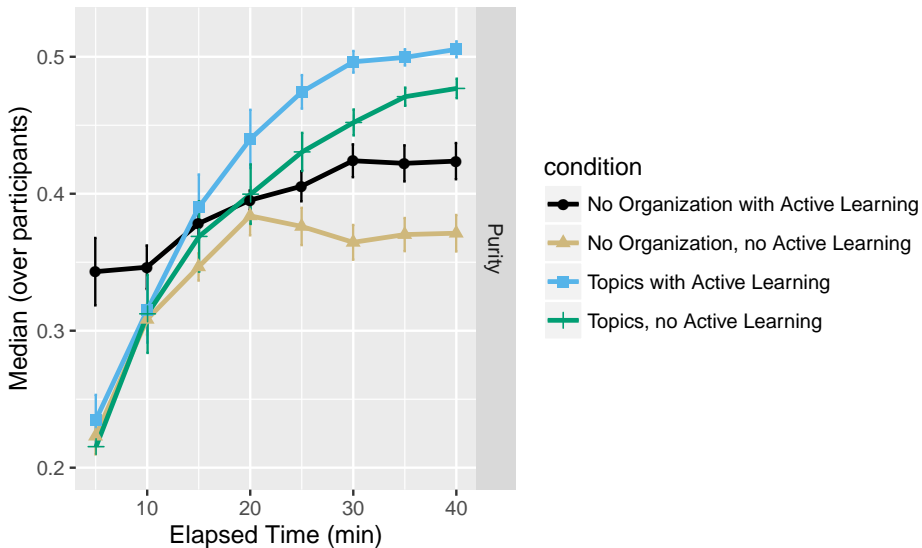- **shelter**

new label name

add label

rename label   delete label

Covered Themes Progress:

flood coast marine
restoration coastal vessel
fish gulf wildlife species
pollution council great
fishery fishing waters
ecosystem monitoring
fisheries mitigation

remain
expended heading
disaster september
appropriation transferred
obligation division unit
capital acquisition
inspector purchase funded
procurement units corps
repair salaries

Direct users to document

Active learning if time is short

Better than status quo

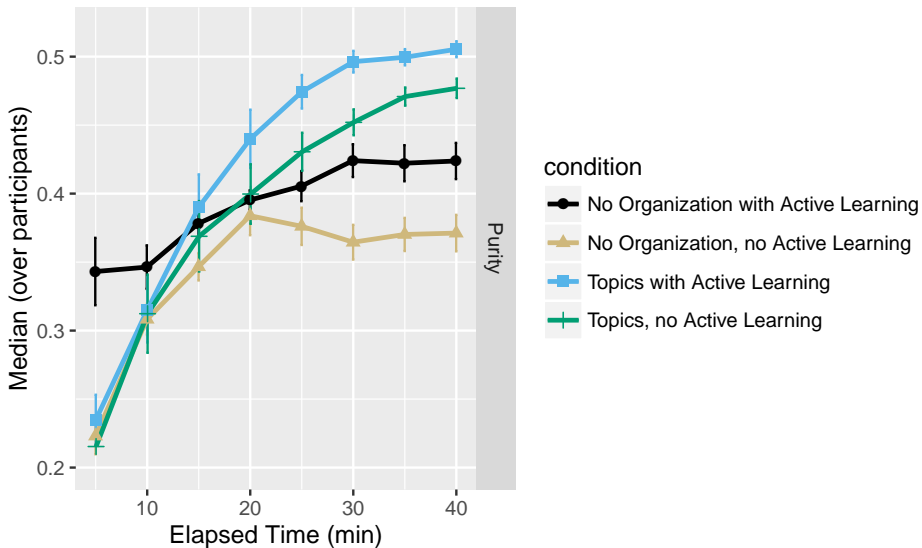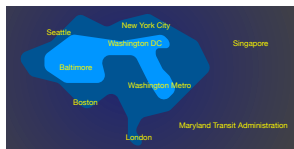Active learning can help topic models

Topic models help users understand the collection
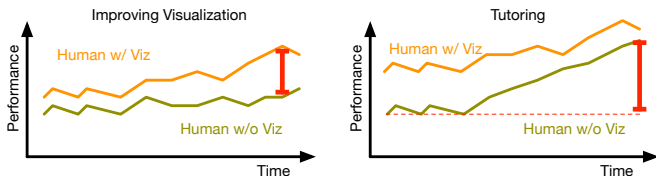
Moral: machines and humans together (if you let them)

This city's mass transit system's director, Paul Wiedefeld, undertook SafeTrack to restore reliability of the system. For ten points each …
[10] Name this American city with stations designed by Harry Weese.

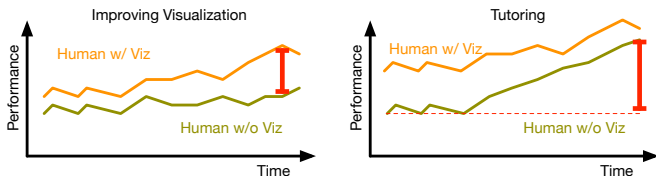| Prediction	Evidence	
34% Washington Metro	stations designed by Harry Weese	undertook SafeTrack
31% Baltimore	Paul Wiedefeld	This city's mass transit
25% Washington, D.C.	this American city	SafeTrack

- Have a task humans are doing
- Show them predicitons of good system
- Measure how much human performance improves (degrades)

- If you have feedback on performance, can use RL
- Change parameters of visualizations
  - How many/which hypotheses are shown
  - How much/which evidence is shown
  - Is anything shown at all?
- Can optimize for team performance

Improving Visualizations

- If you have feedback on performance, can use RL
- Change parameters of visualizations
  - How many/which hypotheses are shown
  - How much/which evidence is shown
  - Is anything shown at all?
  - Which questions are shown
- Can optimize for team performance or just human performance

- Measure what you care about
- If you care about prediction, likelihood is good
- If you care about a particular task, measure that