Bootstrapping Online Communities Using Decision-theoretic Optimization

Shih-Wen Huang*
UW

Jonathan Bragg*
UW

Isaac Cowhey
AI2

Oren Etzioni
AI2

Daniel S. Weld
UW

*contributed equally

ALLEN INSTITUTE *for* ARTIFICIAL INTELLIGENCE

Overview

- Success of online communities
- Decision-theoretic bootstrapping
- Bootstrapping an AI resources community

Overview

- Success of online communities
- Decision-theoretic bootstrapping
- Bootstrapping an AI resources community

Value of online communities

23 million users, over 4 million English articles*

14 million answers to 8.5 million questions*

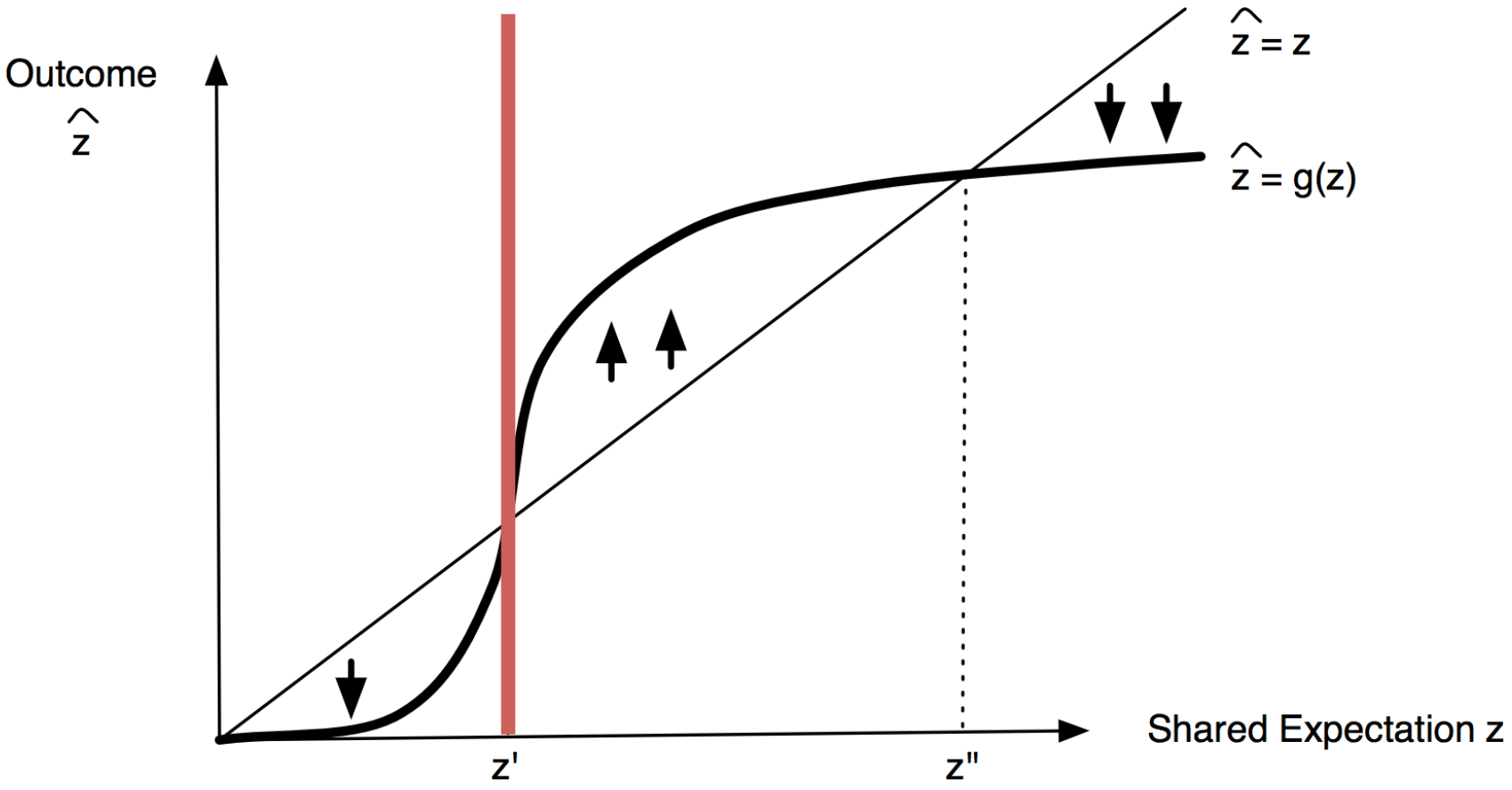
67 million reviews*

*as of December 2014

Most online communities fail

- SourceForge: 10% of projects at least 3 members
- Email groups: 50% received no messages during four-month study

The tipping point

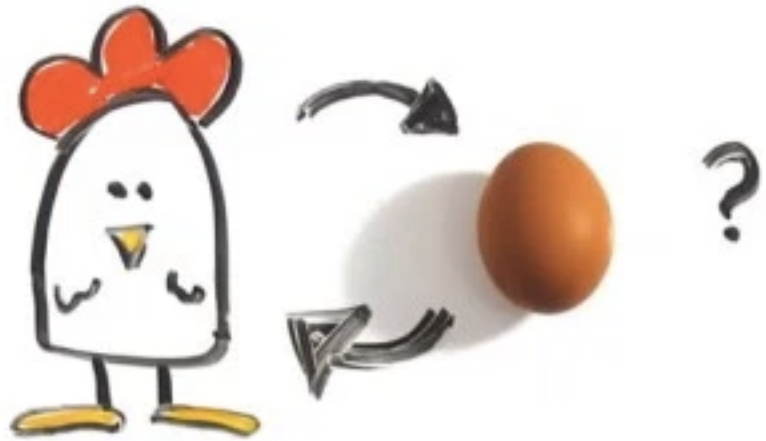


[Figure 17.6, Easley & Kleinberg 2010]

Reaching critical mass

- Recruit through existing members
- Create content that provides value
  - Recruit new members
  - Guide new members to create similar content [Solomon & Wash, CSCW ’12]

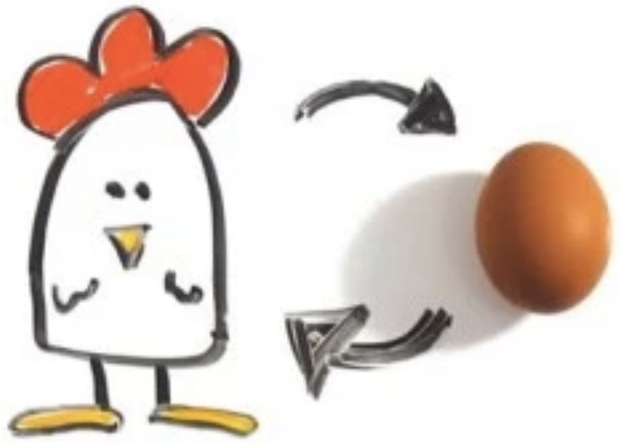
Members Content



Members

Content

movielens

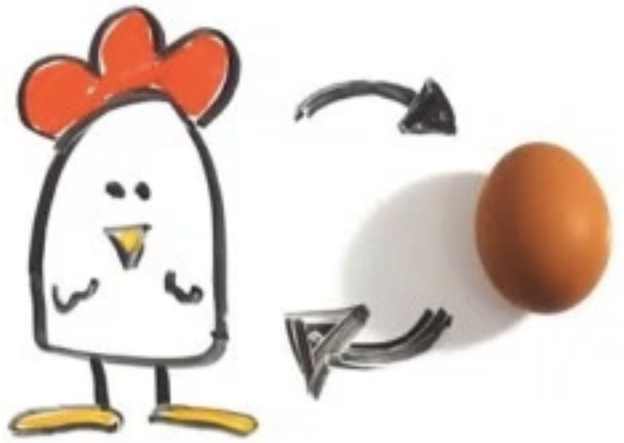


Copy it

Members

Content

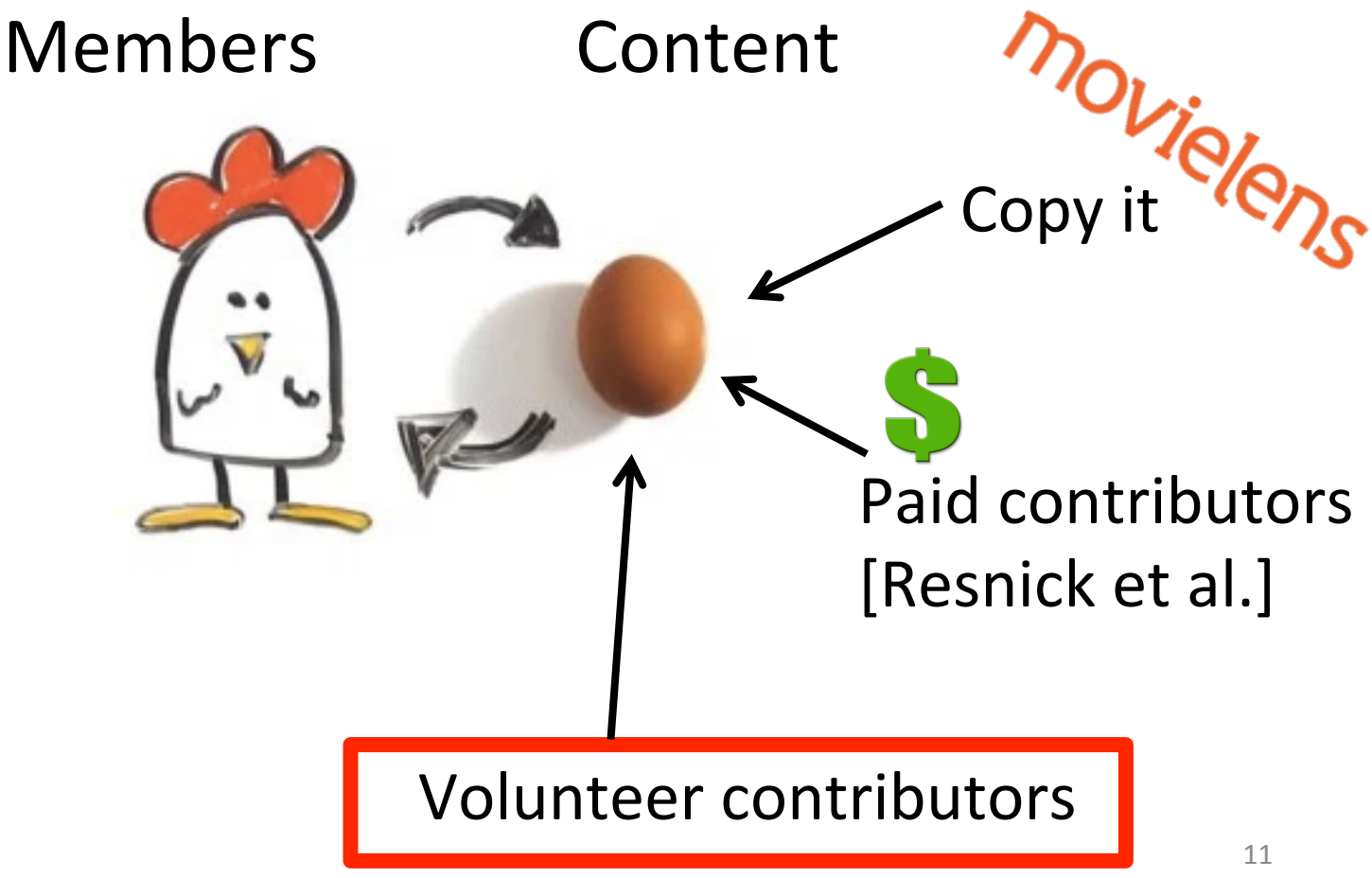
movielens



Copy it

$

Paid contributors
[Resnick et al.]

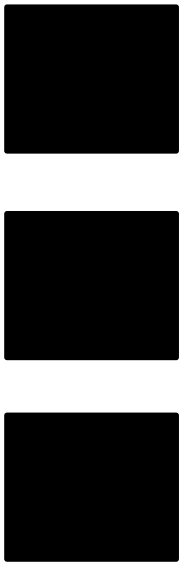

Members
Content
movielens
Copy it
$
Paid contributors
[Resnick et al.]
Volunteer contributors

Overview

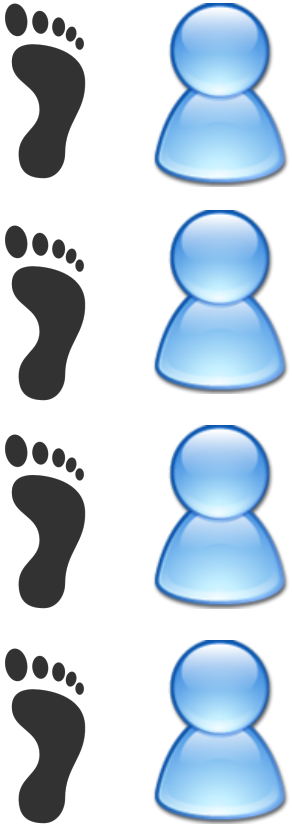
- Success of online communities
- Decision-theoretic bootstrapping
- Bootstrapping an AI resources community

Creating content

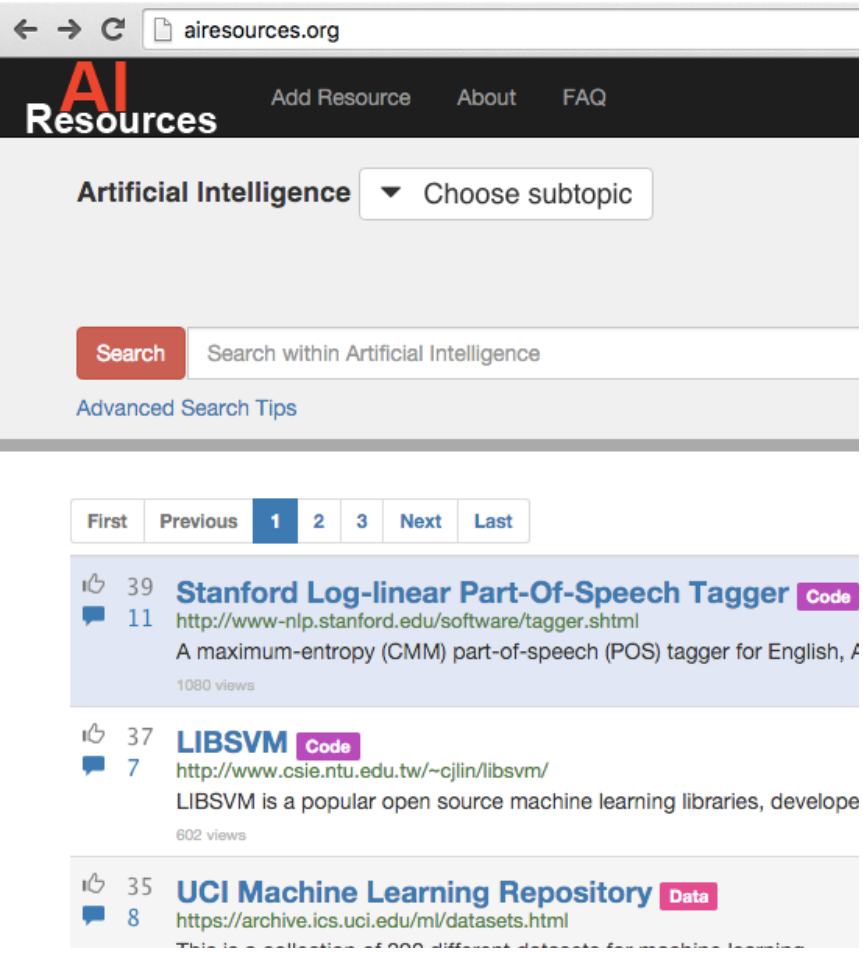
Tasks



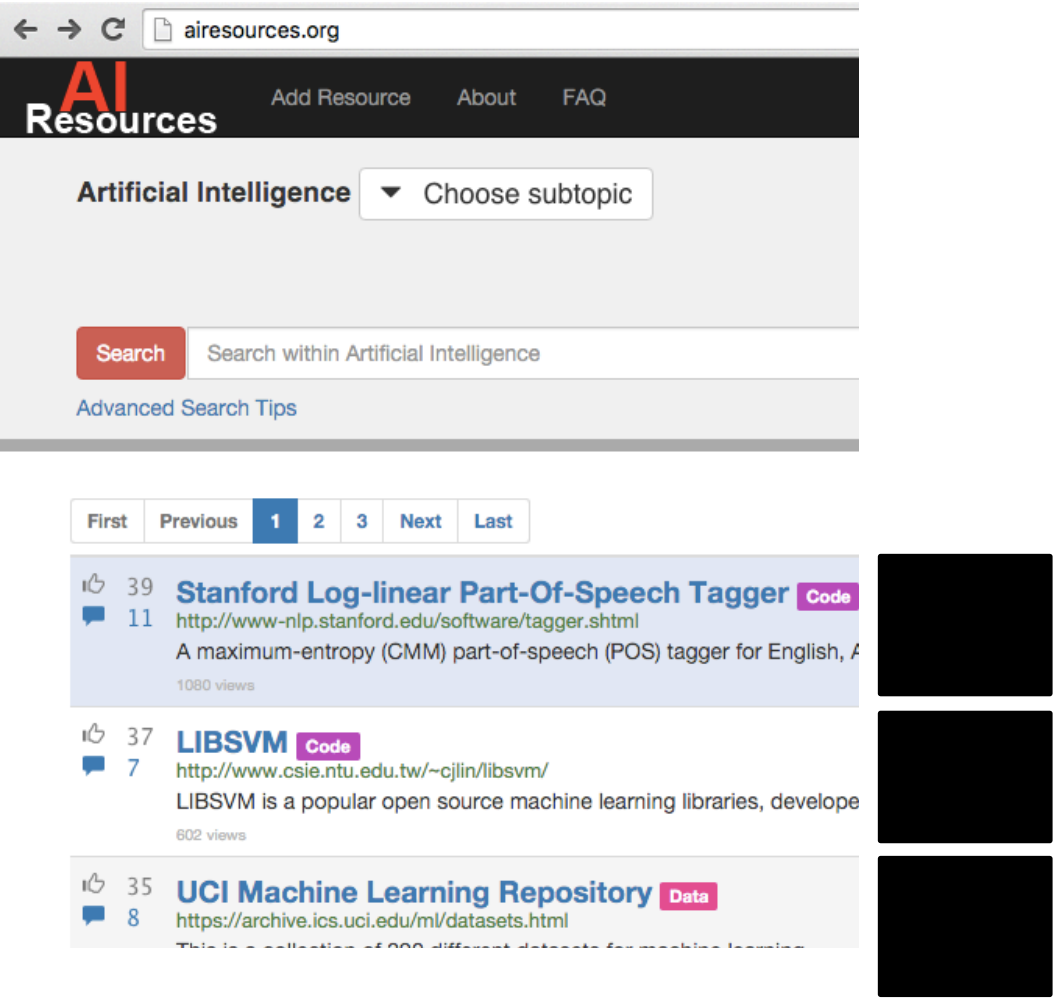
Likely contributors



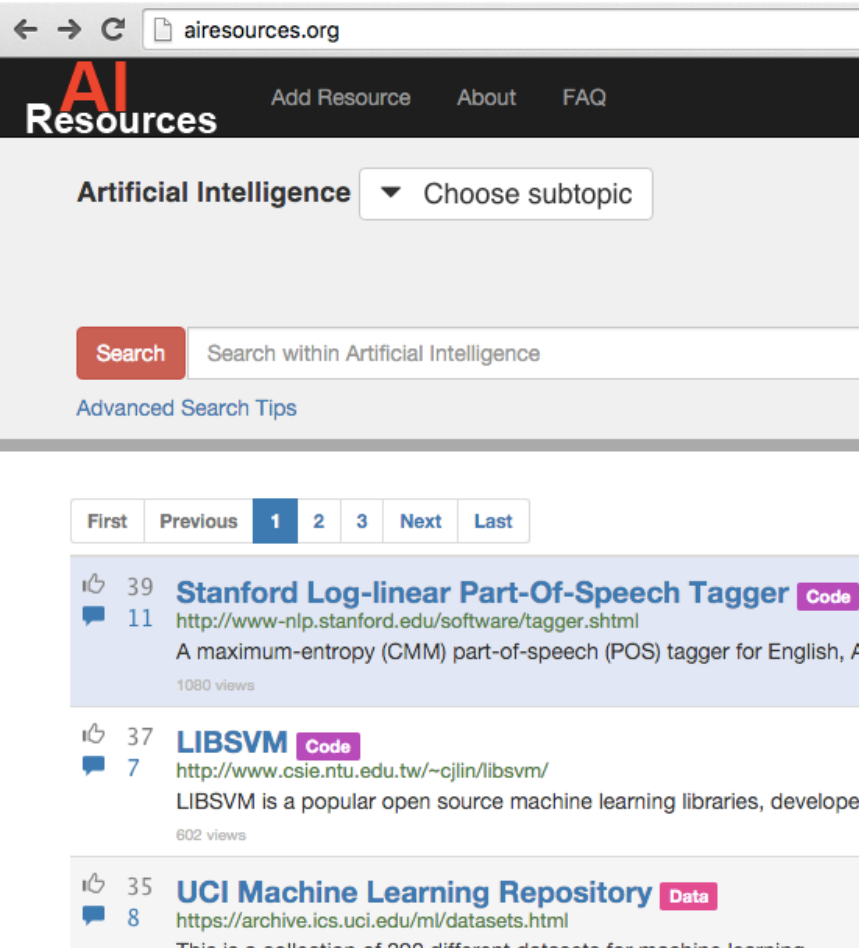
AI resources community

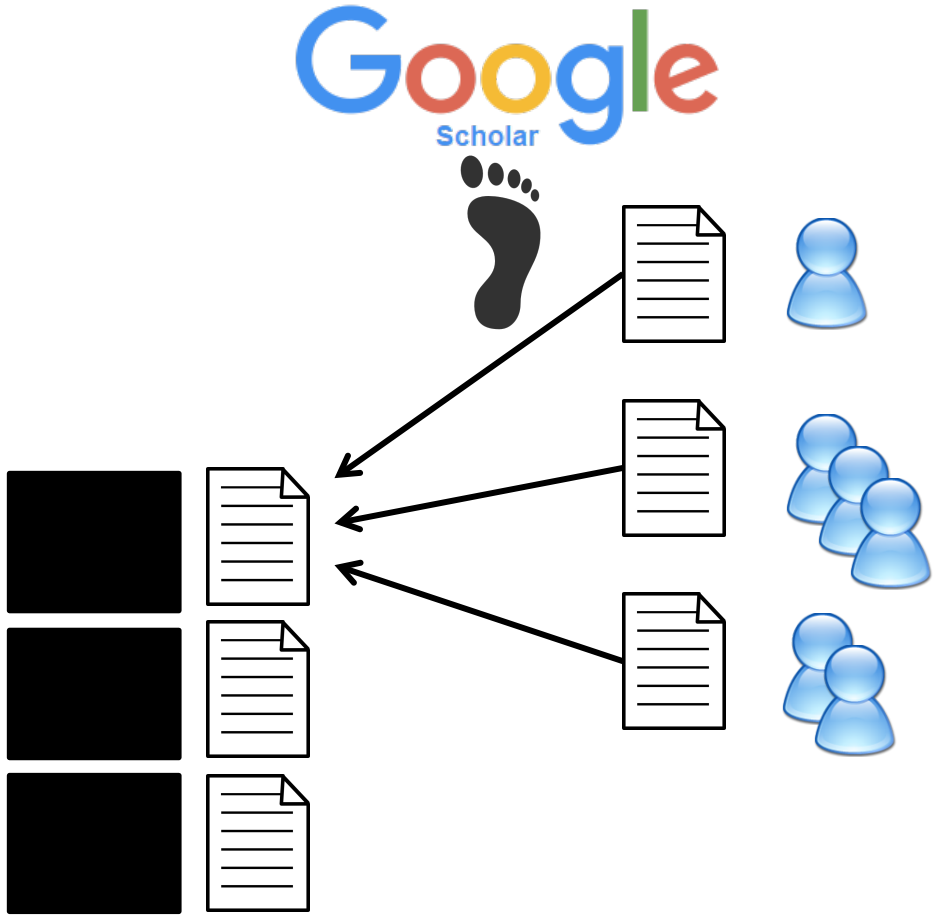


AI resources community



AI resources community





Airbnb community

airbnb

craigslist

<< < prev 1 to 100 of 2500 next >

☆ Feb 29 1 Bedroom 1 Bath with Bonus Den, Balcony and Courtyard View $2800 / 1br - 740ft² - (bayview) pic map

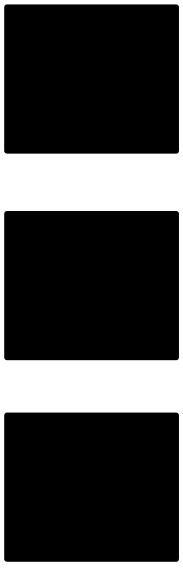
☆ Feb 29 MASTER BEDROOM WITH PRIVATE Bathroom FOR RENT $1175 / 1br - (excelsior / outer mission) map

☆ Feb 29 Private room | Shared Bath | Daily maid service | Caltrain < 5 min $1800 / 1br - 1600ft² - (SOMA / south beach) pic map

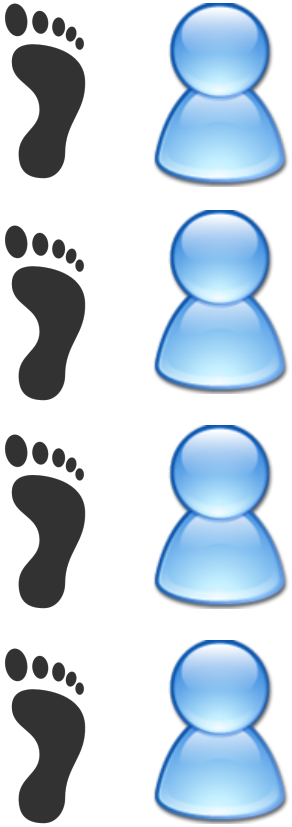
dubAI2

How to get the most (best) content?

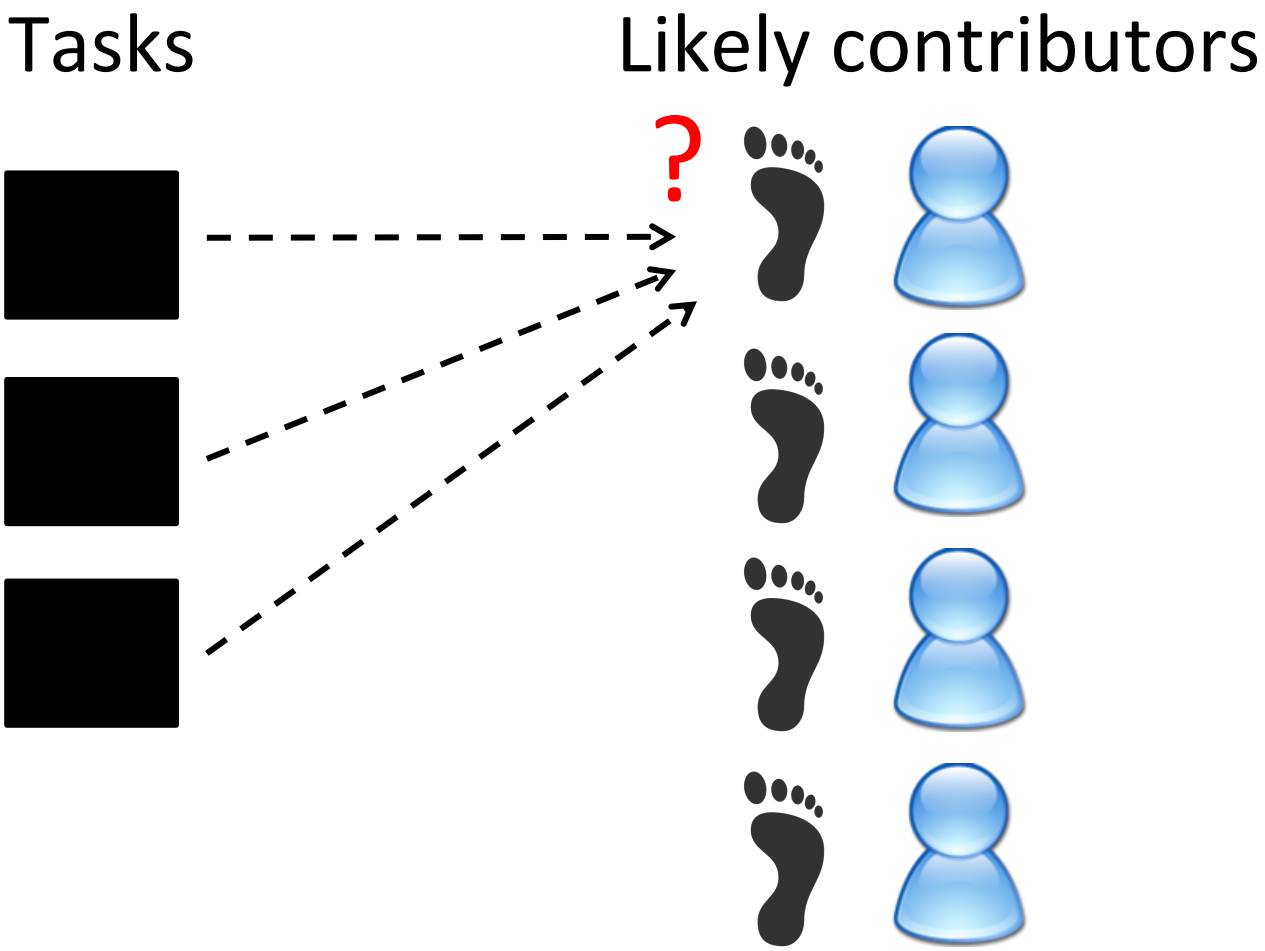
Tasks



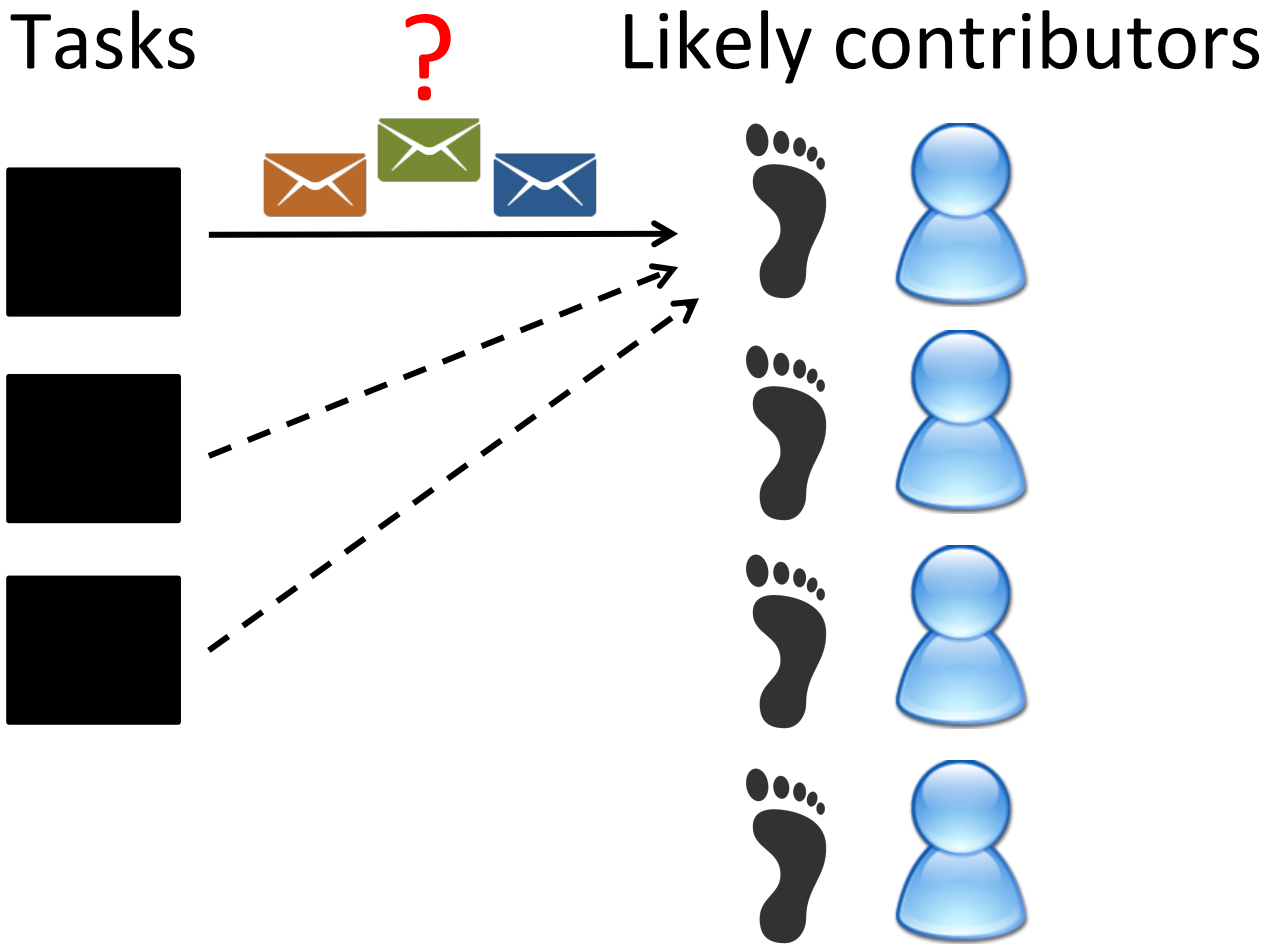
Likely contributors



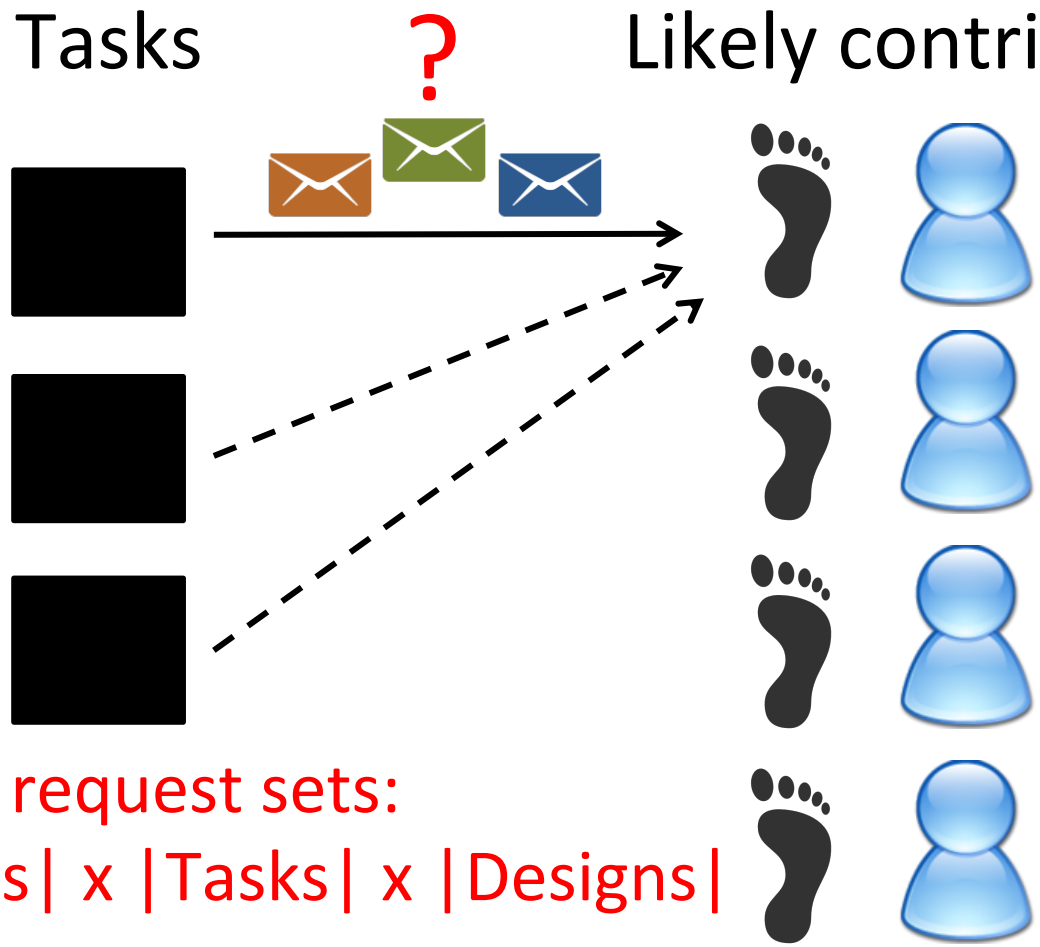
Which task to assign?



Which request design?



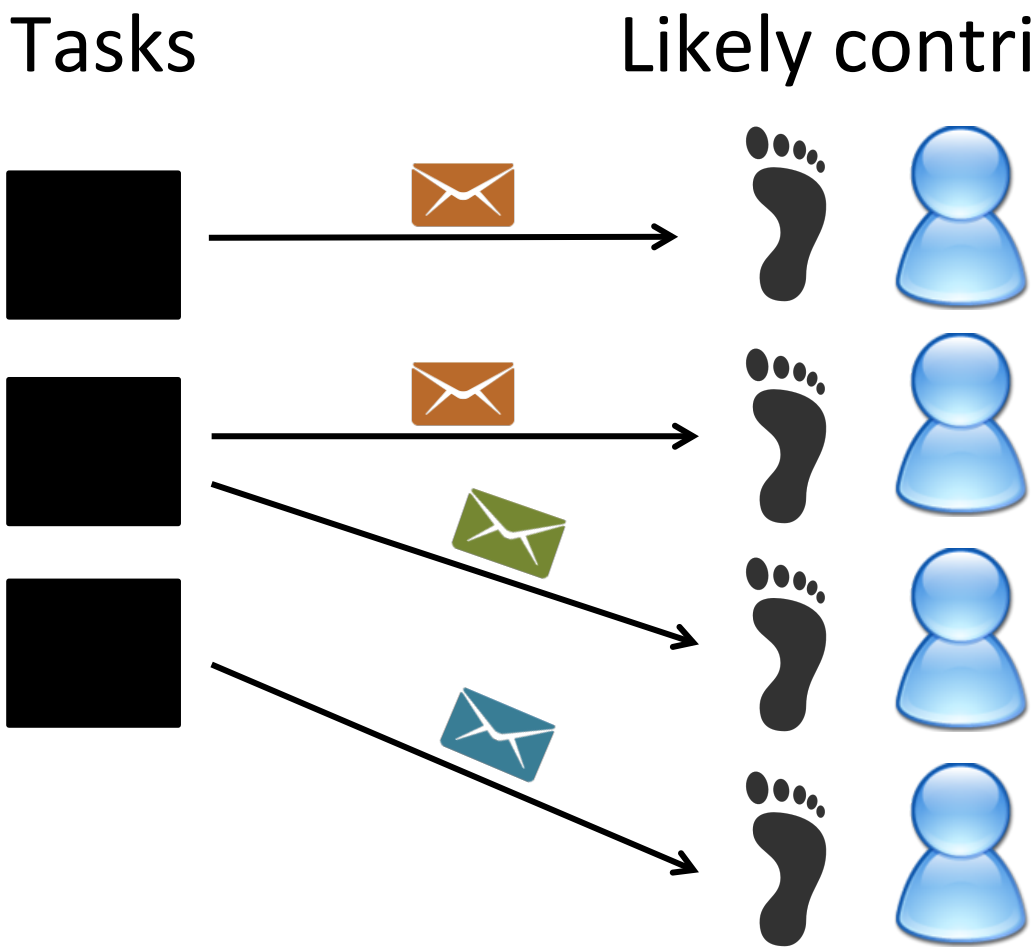
Which request design?



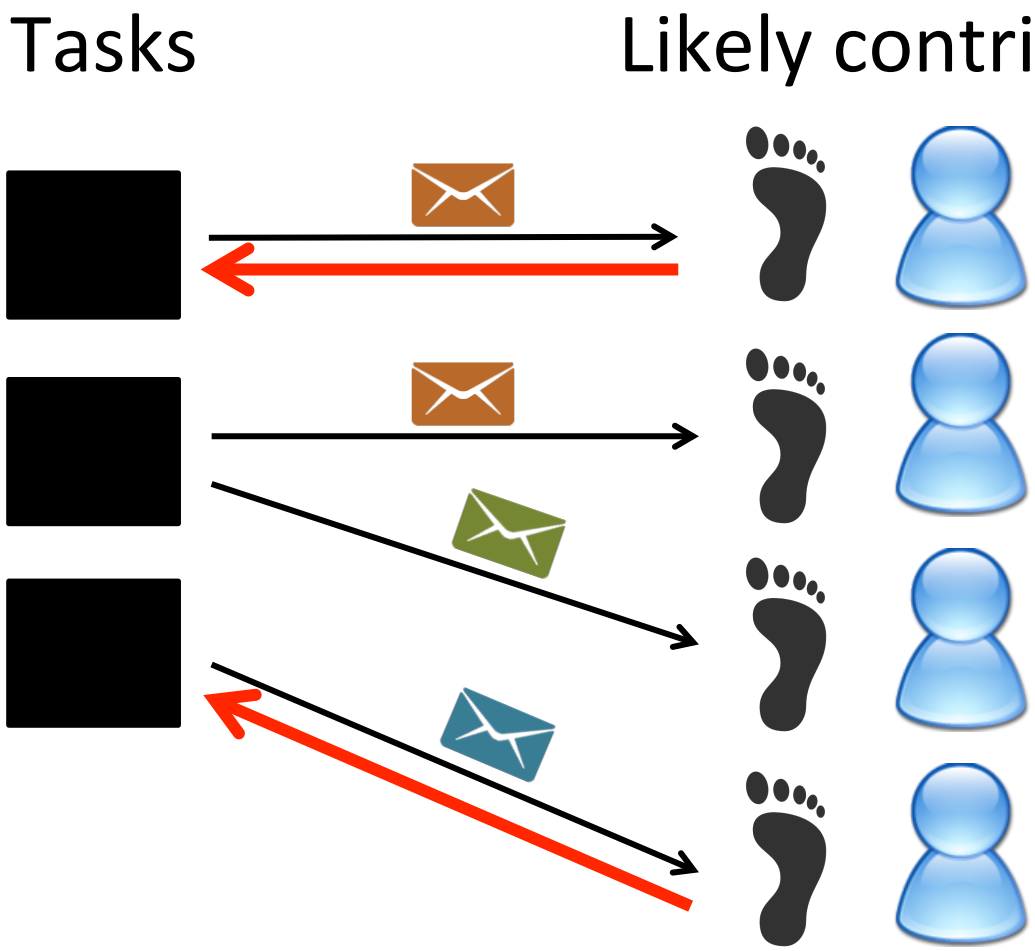
Possible request sets:

|Humans| x |Tasks| x |Designs|

Expected utility of a set of requests



Expected utility of a set of requests



Expected utility of a set of requests

- Subset of requests result in contributions
- Contributions have varying quality
- Possible outcomes:

  (1 + |Qualities|) ^ |Humans|

- Utility of outcome
  - Diminishing returns for each task
  - 1st contribution more valuable than 100th

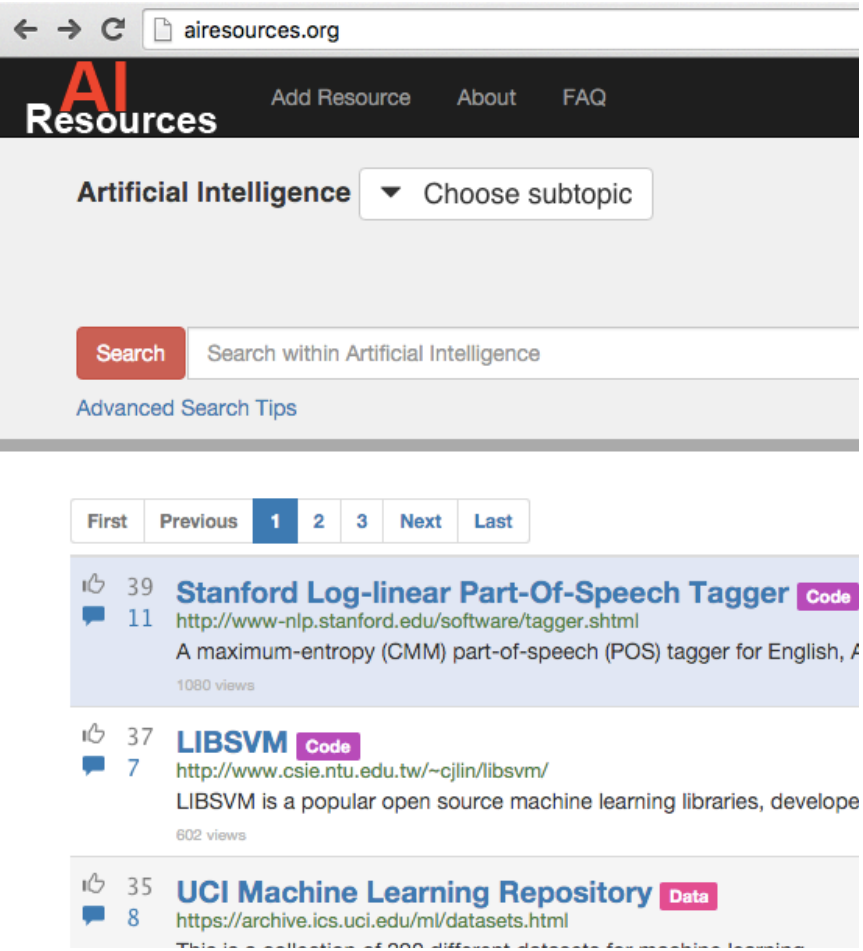
Algorithm to maximize expected utility

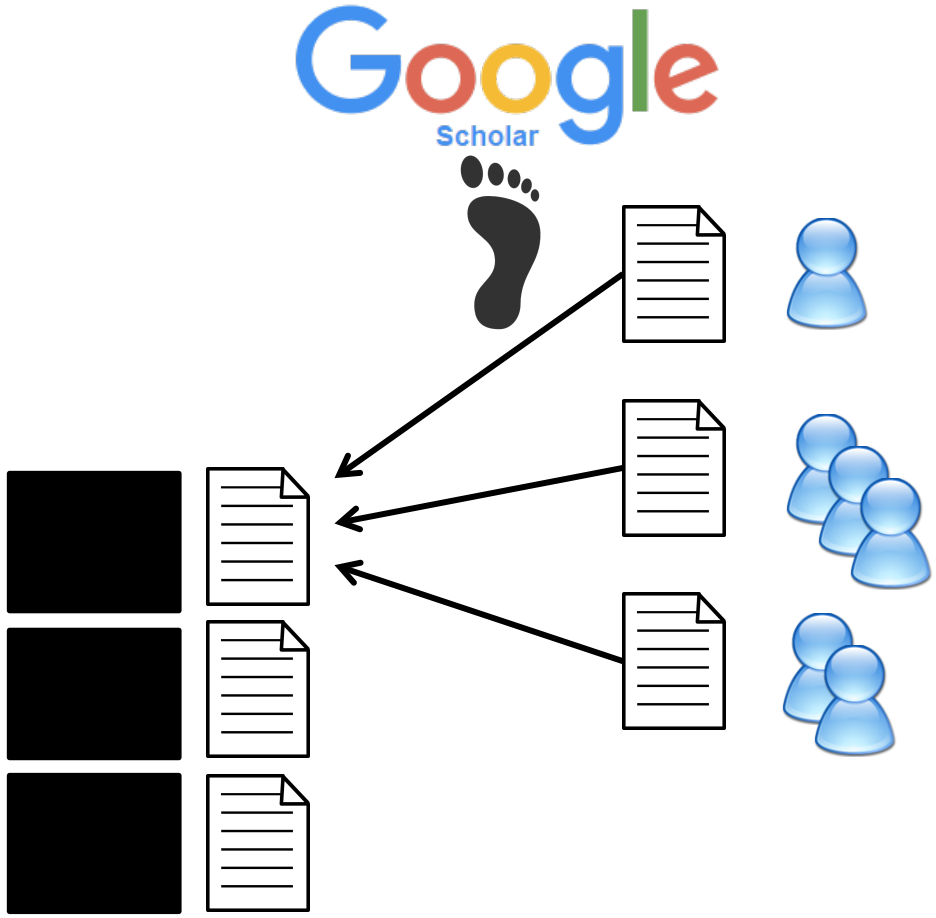
- While someone is unassigned a request:
  - Select request with highest expected utility
  - Assign task and mark person assigned
  - Recalculate expected utilities
- Guaranteed at least ½ of optimal

Overview

- Success of online communities
- Decision-theoretic bootstrapping
- Bootstrapping an AI resources community

AI resources community





Contributing to AI resources

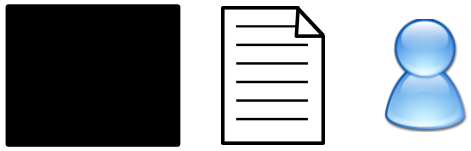
- Submit a new resource
- Update an existing resource
- Comment on a resource

Contributing to AI resources

- Submit a new resource
- Update an existing resource
- Comment on a resource

*"**Scikit** provides a wide variety of Machine Learning and data-processing algorithms, all interfaced through Python. Plus, their website is a great resource for concepts and details about the algorithms."*

Predicting contributions



Text mining features

TEXT-CITE

TEXT-USE

TEXT-SENT

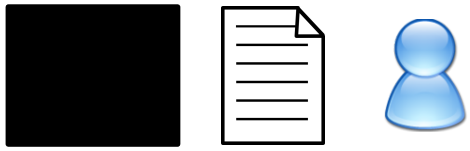
Request design

REQ-FOOT

Email campaign

- 1,339 non-members who cited one of five resources
- Email sent via MailChimp on weekday mornings 10/28/2014 - 11/10/2014

Predicting contributions



Text mining features

TEXT-CITE

TEXT-USE

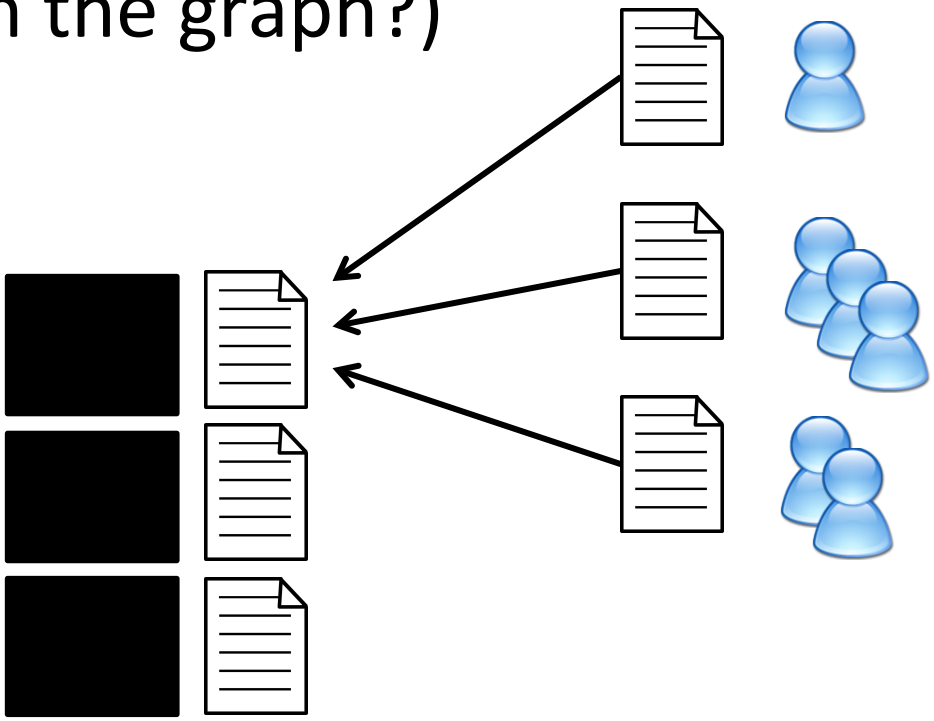
TEXT-SENT

Request design

REQ-FOOT

TEXT-CITE

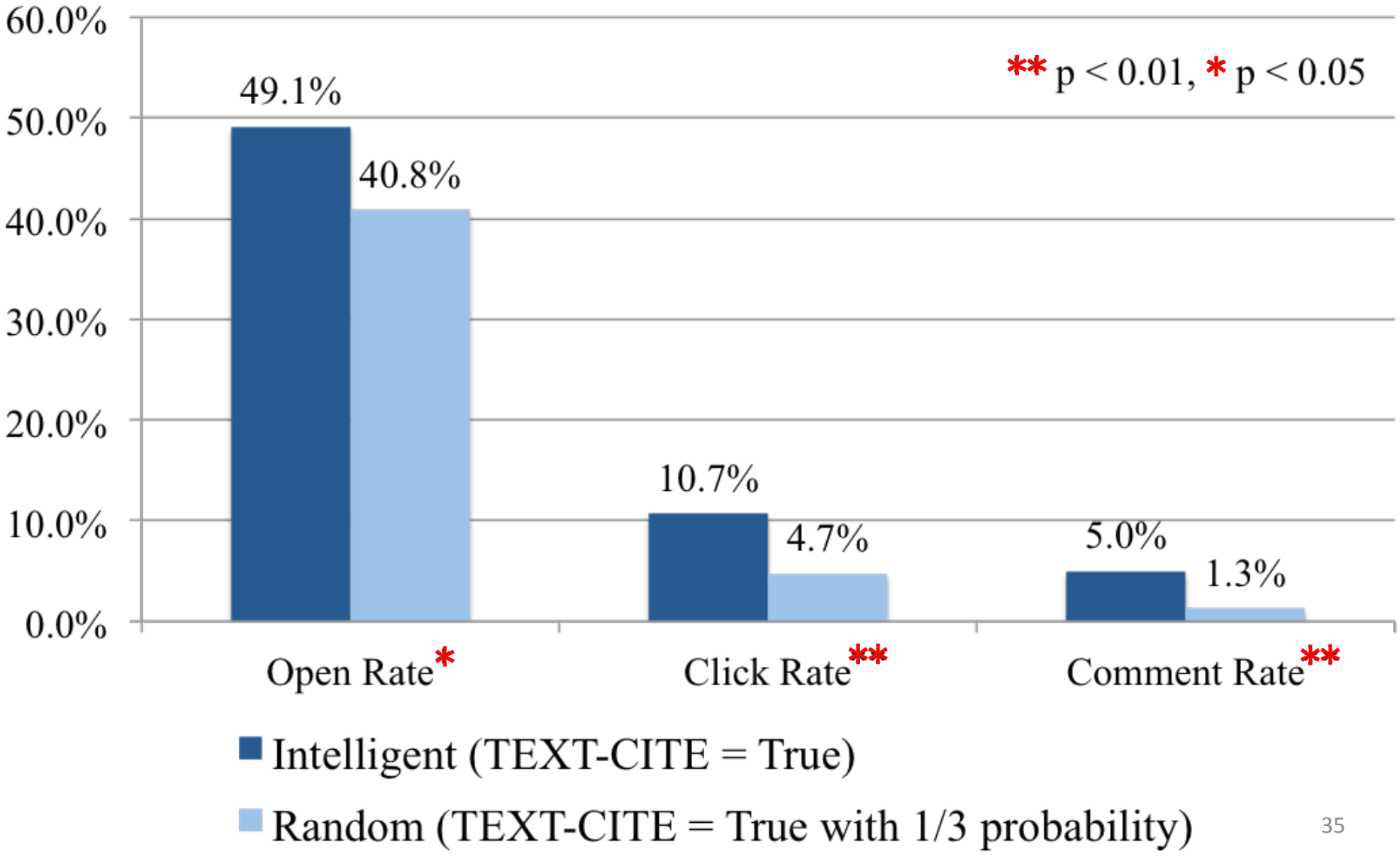
Did the author cite the resource?
(Is there an edge in the graph?)



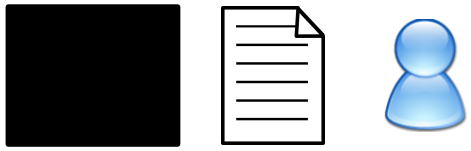
Effects of citing a resource

- Conditions
  - Intelligent (n = 403): TEXT-CITE = True
  - Random (n = 382): P(TEXT-CITE = True) >= 1/3

Effects of citing a resource



Predicting contributions



Text mining features

TEXT-CITE

TEXT-USE

TEXT-SENT

Request design

REQ-FOOT

TEXT-USE

Did the author actually use the resource?

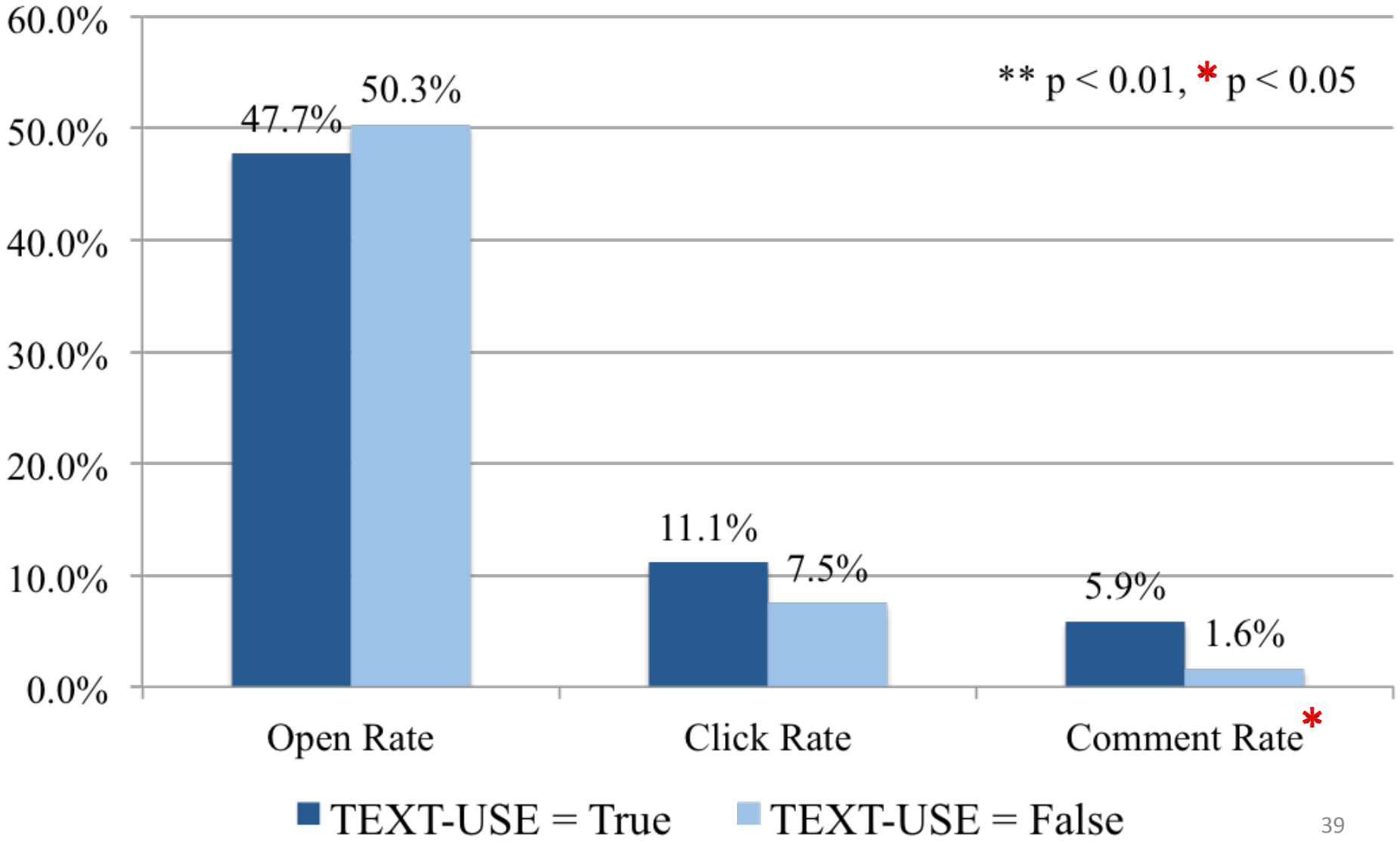
✔ **Souza et al.**: “We apply Stanford NER toolkit to extract named entities from the texts (**Finkel et al., 2005**).”

✖ **Kliegr et al**: “In contrast, NER systems only categorize named entities to several predefined classes (typically ‘organization’, ‘person’, ‘location’, ‘miscellaneous’ (**Finkel et al., 2005**)).”

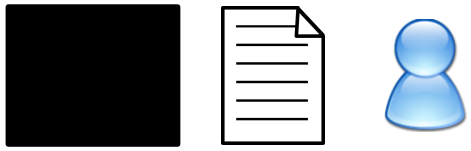
Effects of citation context

- Conditions
  - Used resource (n = 153): TEXT-USE = True
  - Cited but did not use (n = 187): TEXT-USE = False

Effects of citation context



Predicting contributions



Text mining features

TEXT-CITE

TEXT-USE

TEXT-SENT

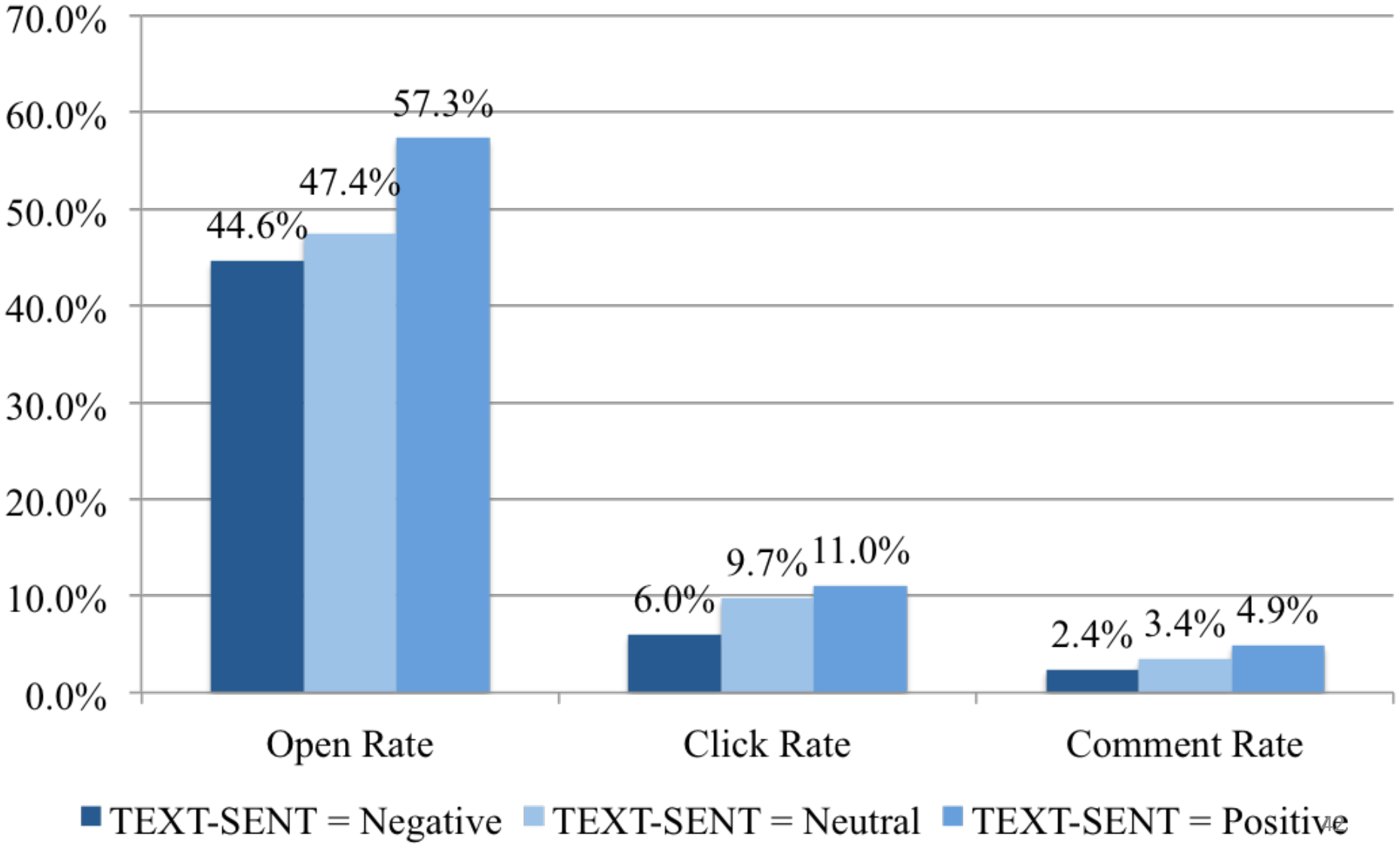
Request design

REQ-FOOT

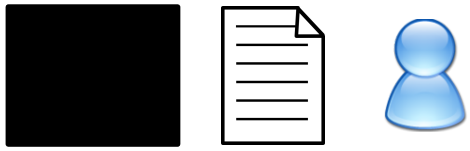
Effects of citation sentiment

- Conditions (sentiment)
  - TEXT-SENT = Negative (n = 83)
  - TEXT-SENT = Neutral (n = 175)
  - TEXT-SENT = Positive (n = 82)

Effects of citation sentiment



Predicting contributions



Text mining features

TEXT-CITE

TEXT-USE

TEXT-SENT

Request design

REQ-FOOT

REQ-FOOT = False

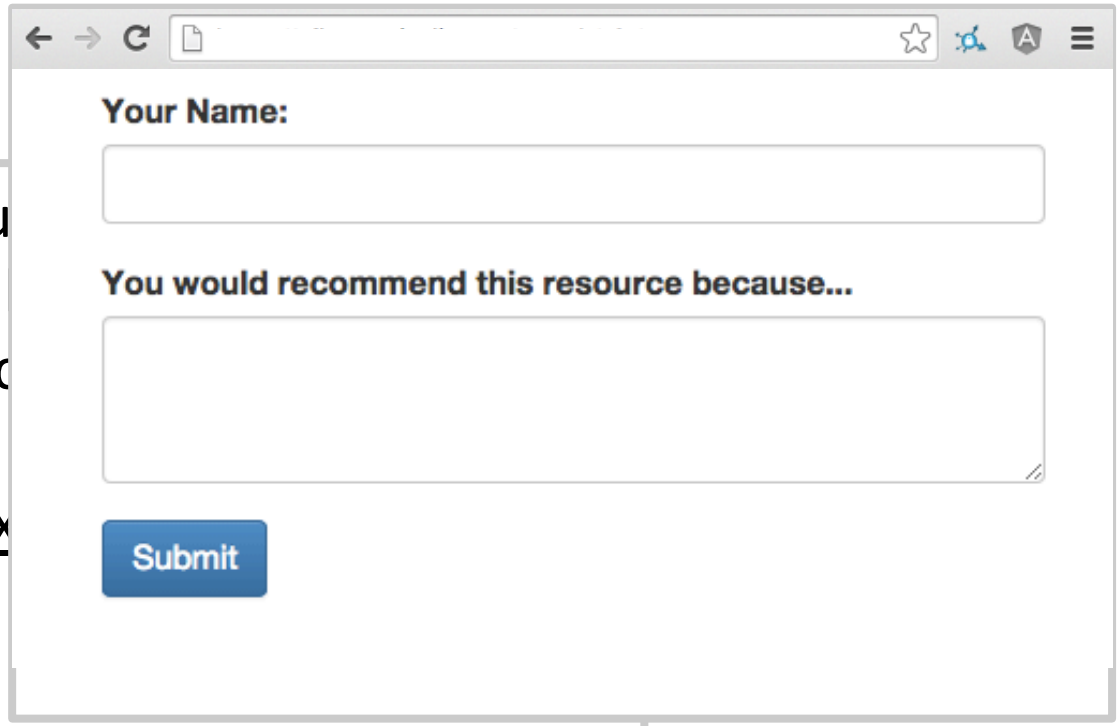
Subject: Did you use [[Resource Name]]?
Body: Hi, We'd love to hear your opinion on [[Resource Name]]!

Tell us about your experience

REQ-FOOT = False

Subject: Did you u
Body: Hi, We'd
opinion on [[Resourc

Tell us about your ex



REQ-FOOT = True

Foot-in-the-door design

Subject: Did you use [[Resource Name]]?
Body: Hi, We'd love to hear your opinion on [[Resource Name]]! Please choose one of the following:

Recommended to other AI researchers
NOT recommended to other AI researchers
Haven't used [[Resource Name]]

REQ-FOOT = True

Foot-in-the-door design

Subject: Did you u

Body: Hi, We'd

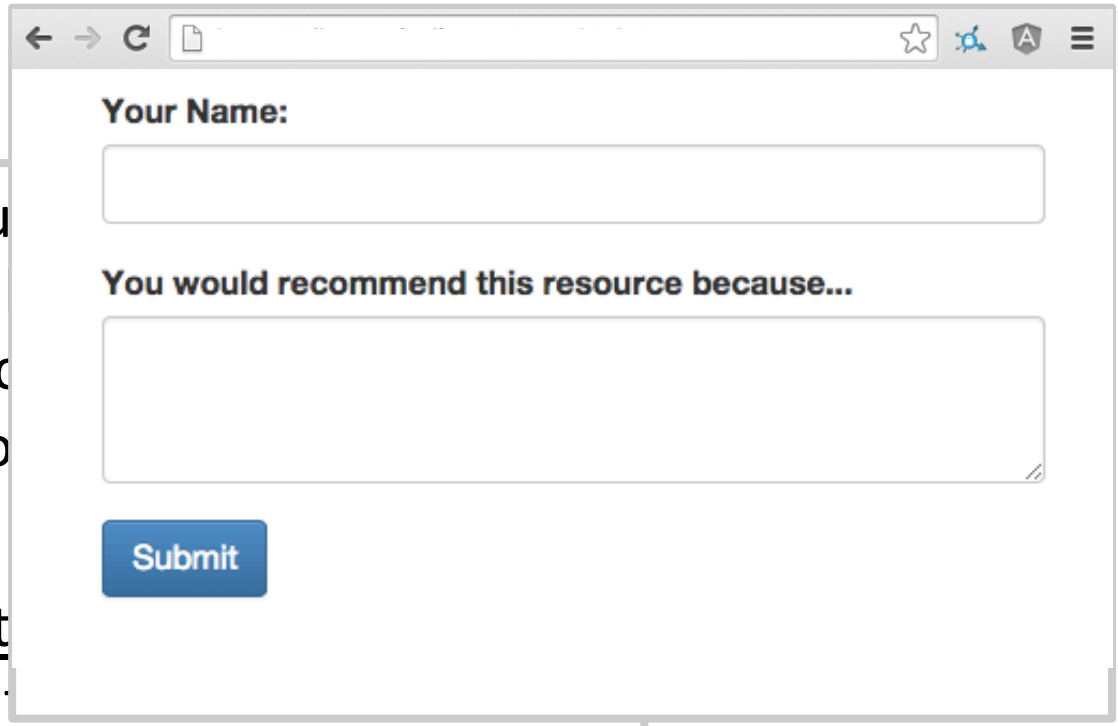
opinion on [[Resourc

choose one of the fo

Recommended to ot

NOT ommended

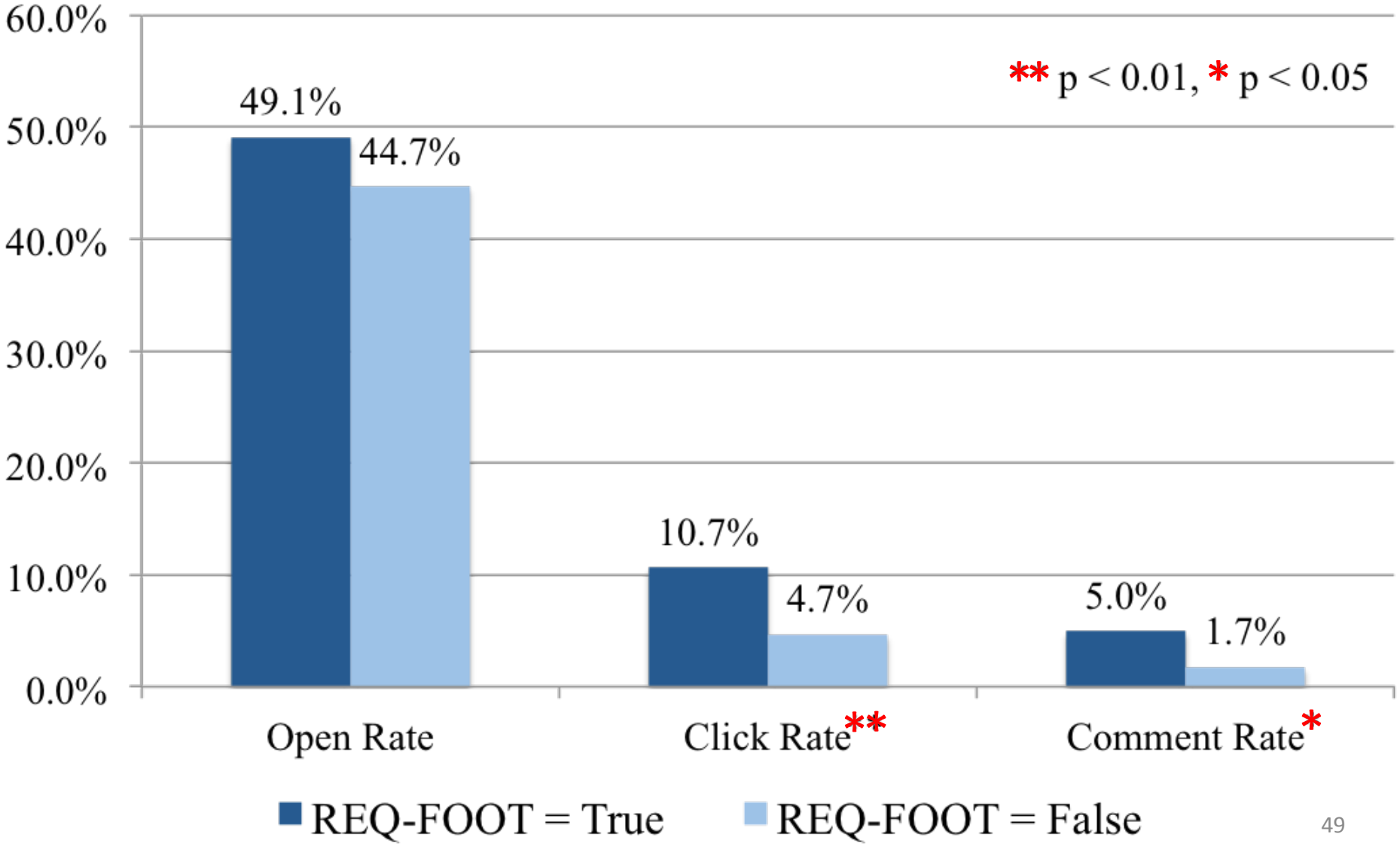
Haven't used [[Resource Name]]



Effects of request design

- Conditions
  - Foot-in-the-door (n = 403): REQ-FOOT = True
  - Basic link (n = 407): REQ-FOOT = False

Effects of request design



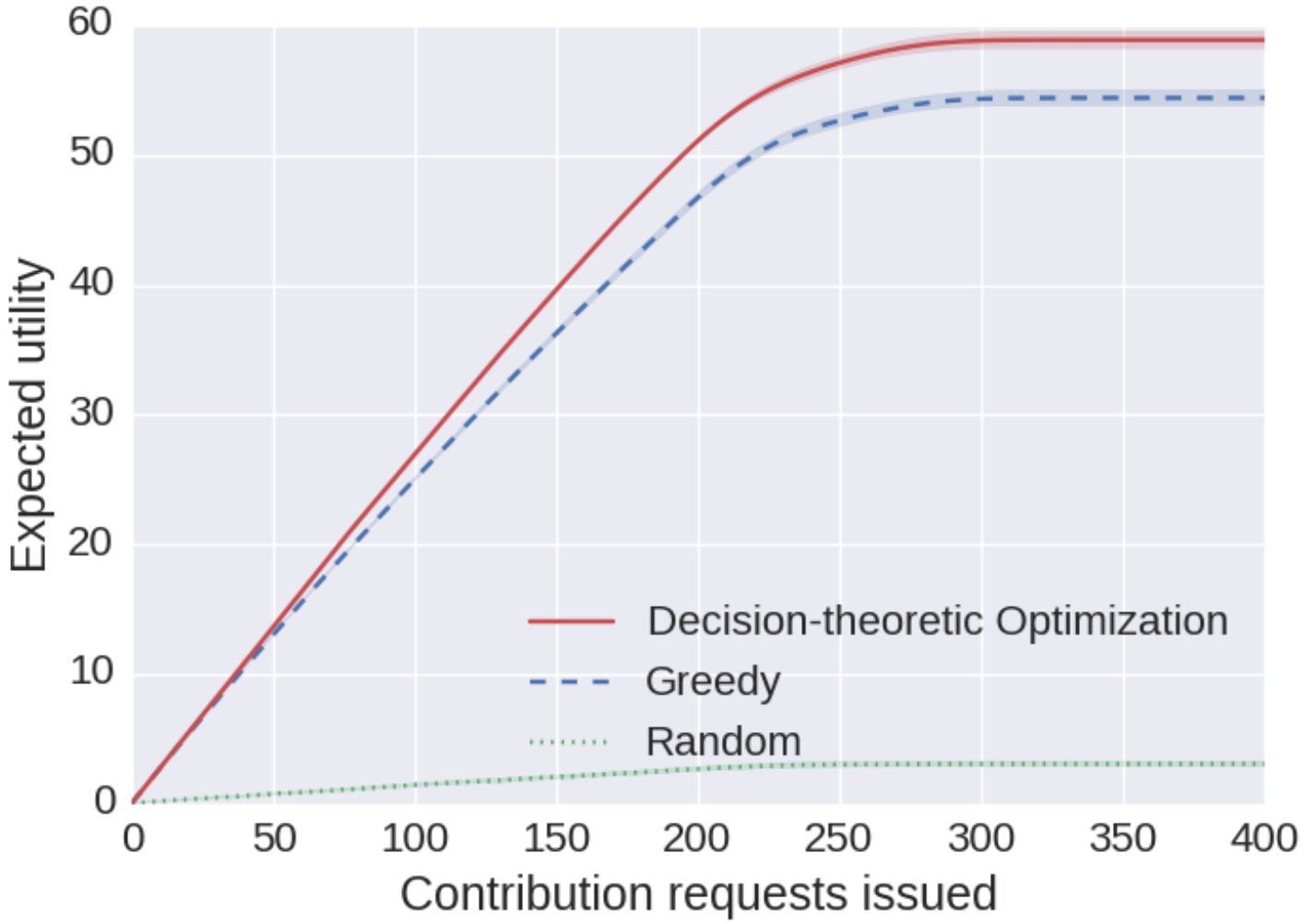
Estimated parameters

Design	Features	P(Contribute)
REQ-FOOT	**TEXT-CITE**	0.050
REQ-FOOT	**TEXT-USE**	0.059
REQ-FOOT	**TEXT-CITE** $\wedge\neg$ **TEXT-USE**	0.016
REQ-FOOT	**TEXT-CITE** $\wedge$ **TEXT-SENT**=*Pos*	0.049
REQ-FOOT	**TEXT-CITE** $\wedge$ **TEXT-SENT**=*Neut*	0.034
REQ-FOOT	**TEXT-CITE** $\wedge$ **TEXT-SENT**=*Neg*	0.024
$\neg$ **REQ-FOOT**	**TEXT-CITE**	0.017

Request issuing experiment

- Generated 100 citation graphs
- Utility function: log of # contributions
- Request strategies:
  - Random: Assign authors to resources randomly
  - Greedy: Assign author to resource most likely to contribute to
  - Decision-theoretic algorithm: Described earlier

Request issuing experiment



Conclusions

- Bootstrapping as decision-theoretic optimization
- Exact solution infeasible, but simple algorithm with guarantees
- Text-mining can predict contribution probabilities
- Effective request design essential
- Learned parameters let our method make fewer requests

Future work

- Bring our community to the tipping point
- Other communities and utility functions
- Richer prediction models
- Request design extensions

Future work

- Bring our community to the tipping point
- Other communities and utility functions
- Richer prediction models
- Request design extensions

Thanks!