

# Boosting Crowdsourcing

Yugian Li

Broader Crowdsourcing | Journalism | Project Detail

# Broader Crowdsourcing [1]

## Bradham's 4 types

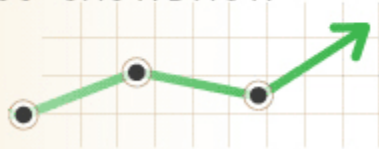
- Knowledge discovery and management  
e.g. Peer-to-Patent Community Patent Review
- Broadcast search  
e.g. Innocentive: <http://www.innocentive.com/>
- Peer-vetted creative production  
e.g. threadless: <http://www.threadless.com>
- Distributed human intelligence tasking (HIT)  
e.g. Amazon's Mechanical Turk, Microtask: <http://www.microtask.com>

# Welcome To InnoCentive

Where the World Innovates

Need solutions to real problems?  
Click below or call 1-855-CROWDNOW


Learn More »



Want to solve important problems  
and make a positive impact?

Become A Solver »





Are you suffocating under piles of paper forms?  
We convert paper forms into Excel spreadsheets

UNITED SHIPPERS LTD Job application form

LAST NAME	FIRST NAME
ROTASQUE	Mick
BIRTH DATE	DRIVER LICENSE NO.
16.02.1977	F255-921-50-094-0
EMAIL ADDRESS	
mick_rotasque@gmail.com	

An illustration from a first-person perspective of a person with pink skin and a black collar sitting at a desk. They are looking at a laptop screen that displays the email address trial@microtask.com. A mouse cursor is hovering over the email address. The person's hand is on the laptop keyboard. In the background, there are some office supplies like a printer and a box.

We give you an email address where you can send your forms.

# Broader Crowdsourcing [1] (cont.)

## Howe's taxonomy

- Crowd wisdom  
similar to broadcast search
- Crowd creation  
similar to peer-vetted creative production
- Crowd voting  
e.g. prediction markets
- Crowd funding  
e.g. crowd-based microlending institutions,  
spot.us <http://spot.us>

# Journalism Aspect of Crowdsourcing

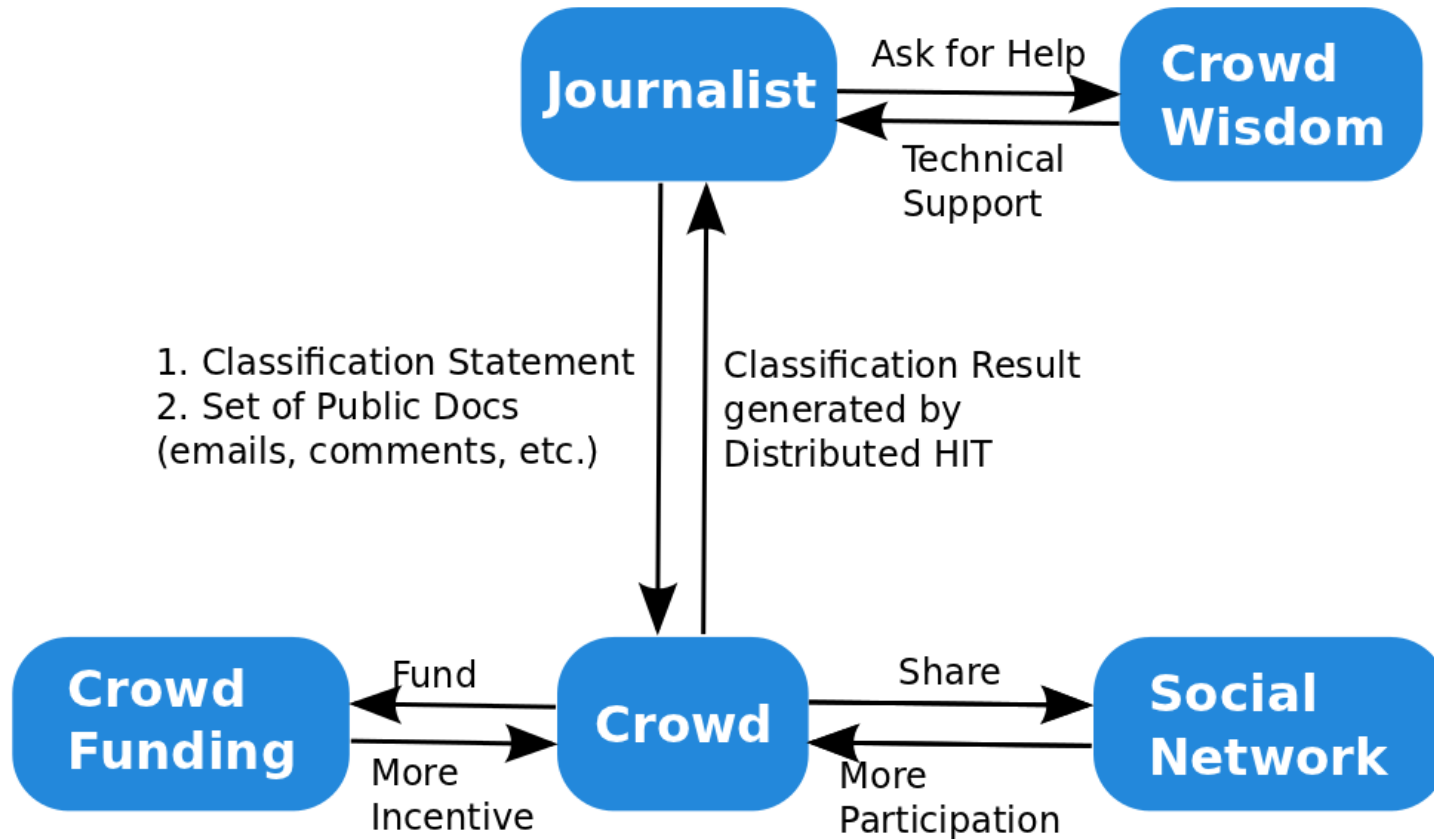
- Co-creation, crowdsourcing as journalism
  - meaningful **process** for crowd, unlike CAPCHA, e.g. audience will know more related information about a topic when they help classifying articles/comments
- Crowd wisdom, ask for help
  - It's too hard to get a universal, powerful while easy-to-use tool for reporters. But for a specific task of an reporter, it's easy to find an expert to help him. e.g. TopCoder <http://community.topcoder.com/tc>, Craigslist <http://raleigh.craigslist.org/cpg/>
- Crowd funding + Human Intelligence Tasking
  - We're interested in your story. If we have tim, we can do HITs. If we don't, we donate money to encourage who don't have interests to do so.

# A Small Task: Classification

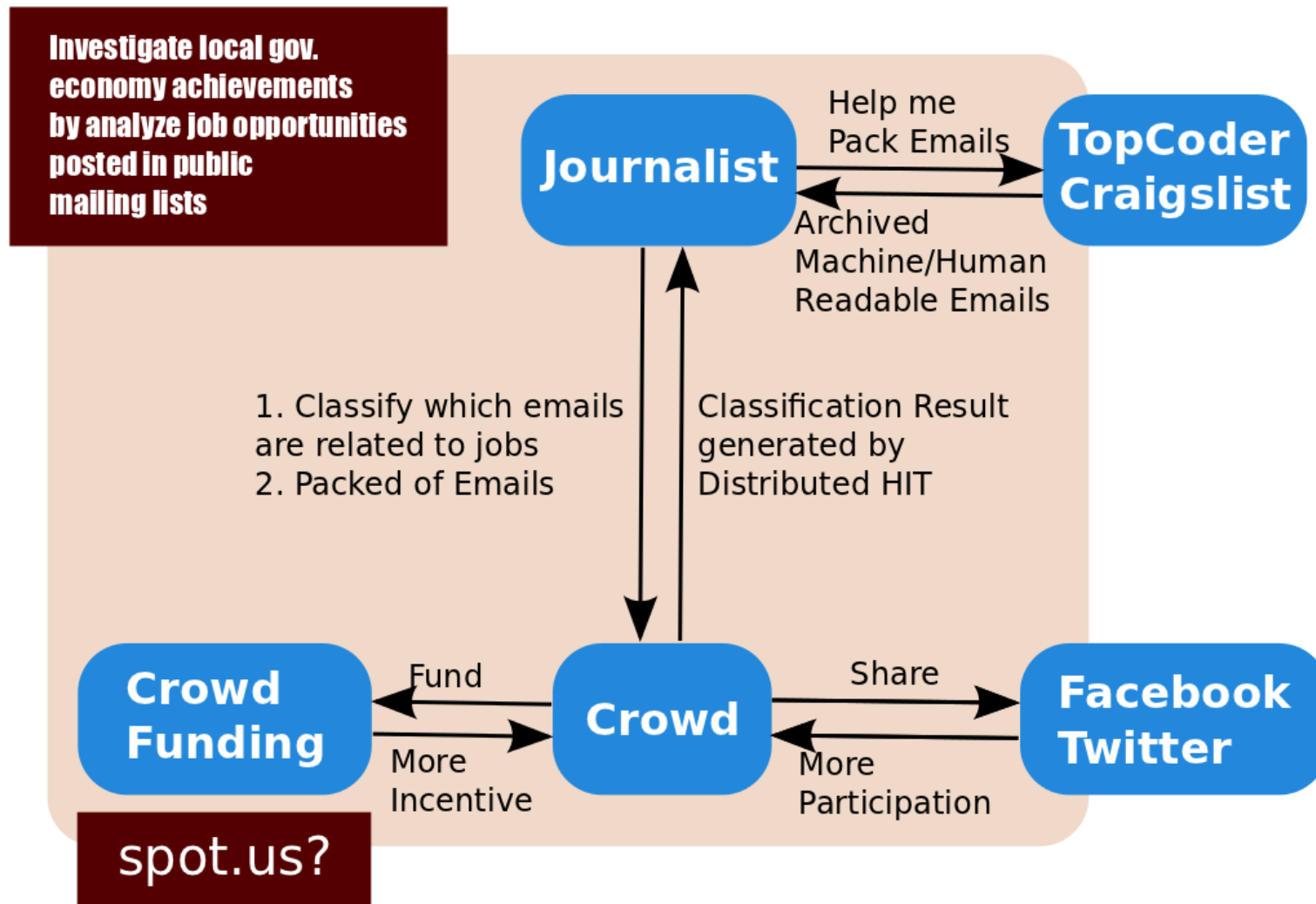
Why choose this? Easy input/output, hard for computer alone, important for journalism

- Filtering by "related or not" among:
  - tens of thousands of comments/tweets
  - thousands of emails in a public mailing-list
  - hundreds of public documents
- Clustering documents by "relation"
  - Clustering by topics, DOC x TOPICS
  - Clustering by pairwise relation, DOC x DOC

# Human Boosting

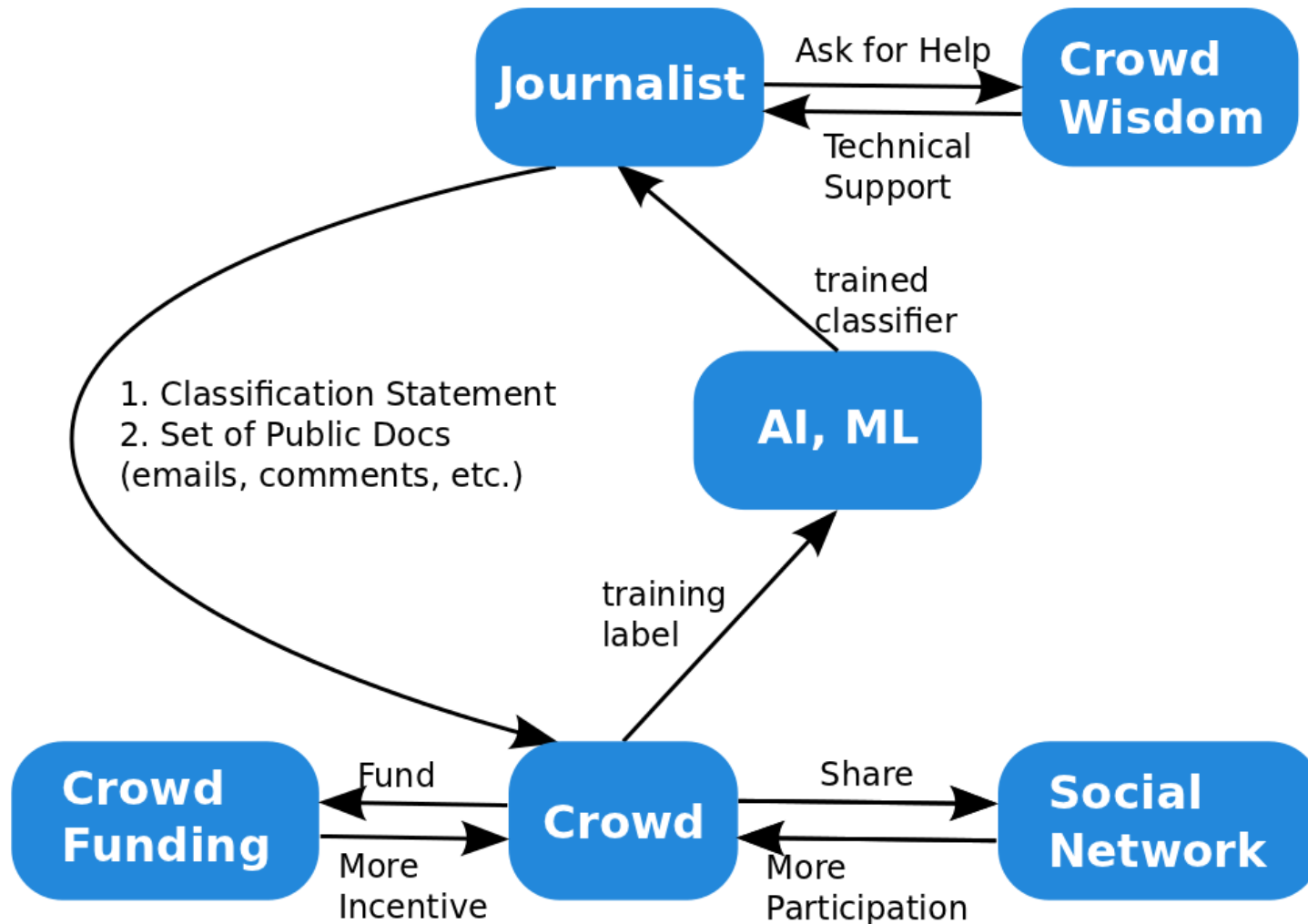


# Human Boosting Example





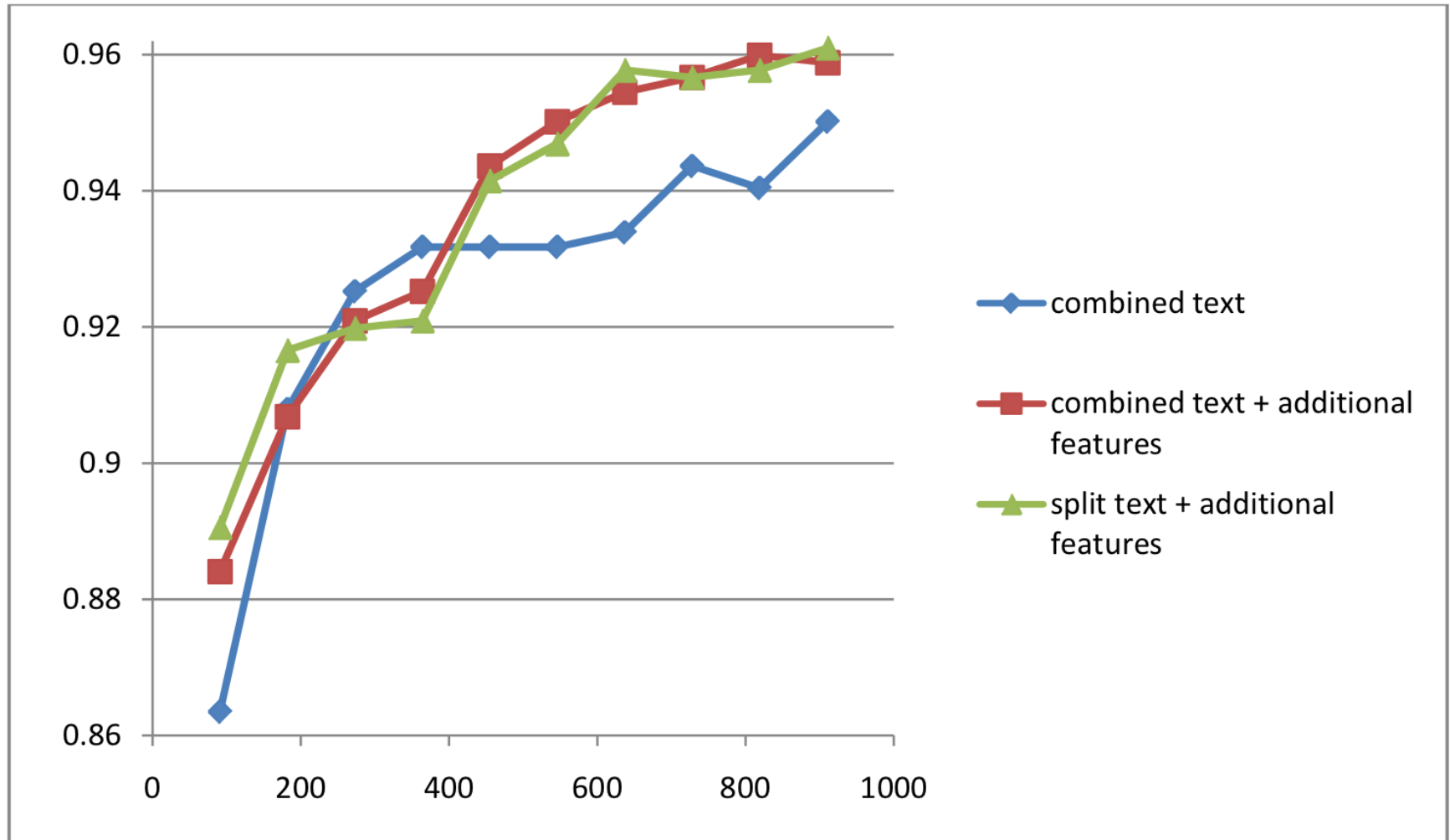
# Machine Boosting



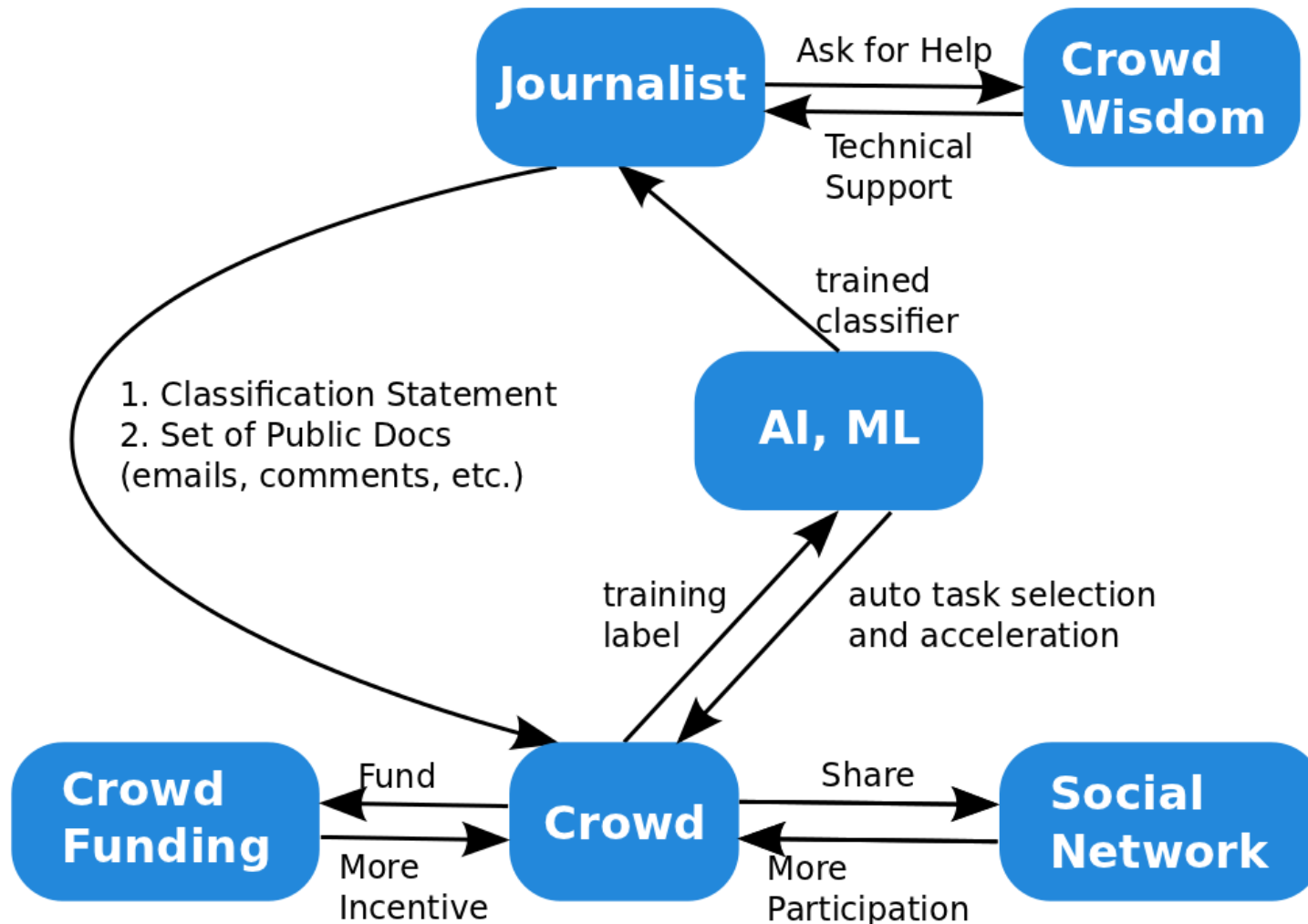
# Classify sales/rent/supply in intl@duke.edu mailing list

Fletcher Maynard	Duke/intlhouse/others	[intlhouse] NO!! BRIDGES LUNCH TOMORROW ( 11/28/11) - Dear D
dipys macho	Duke/intlhouse/others	[intlhouse] Durham to DC on December 17th or 18th - I would be driv
Teng-Yee Ng	Duke/intlhouse/supply	[intlhouse] MOVING SALE--- UPDATED - Dear all, I updated the mo
Yui	Duke/intlhouse/supply	[intlhouse] Bedroom set - Hello, We have a nice full size bedroom set
Tiago Mata (2)	Duke/intlhouse/supply	[intlhouse] TOYOTA RAV4 1998 for sale - 3,600\$ - To UNSUBSCRIBE
Francisco Duque	Duke/intlhouse/supply	[intlhouse] Bedroom in Chapel Hill: One month sublease - To UNSUB
Tiago mata	Duke/intlhouse/others	[intlhouse] - To UNSUBSCRIBE from International House List Serv gr
Jiaoni Zhou (2)	Duke/China Duke/intlhouse/others	[intlhouse] Ride from Durham to DC today (Nov 27) - Hi,
Sambuddha Banerjee	Duke/intlhouse/others	[intlhouse] Ride from DC to Durham today - Hi, Is any one traveling fr
Luis Vergara	Duke/intlhouse/others	[intlhouse] Hola....Spanish Hour on Monday at 6:30 PM - In Spanish t
Mira Han	Duke/intlhouse/supply	[intlhouse] flying pigeon men's bike \$90 - To UNSUBSCRIBE from Int
Shuqun Yang	Duke/China Duke/intlhouse/supply	[intlhouse] Brand new uncharted 3 for \$50 - To UNSUBS

# Classify sales/rent/supply in intl@duke.edu mailing list

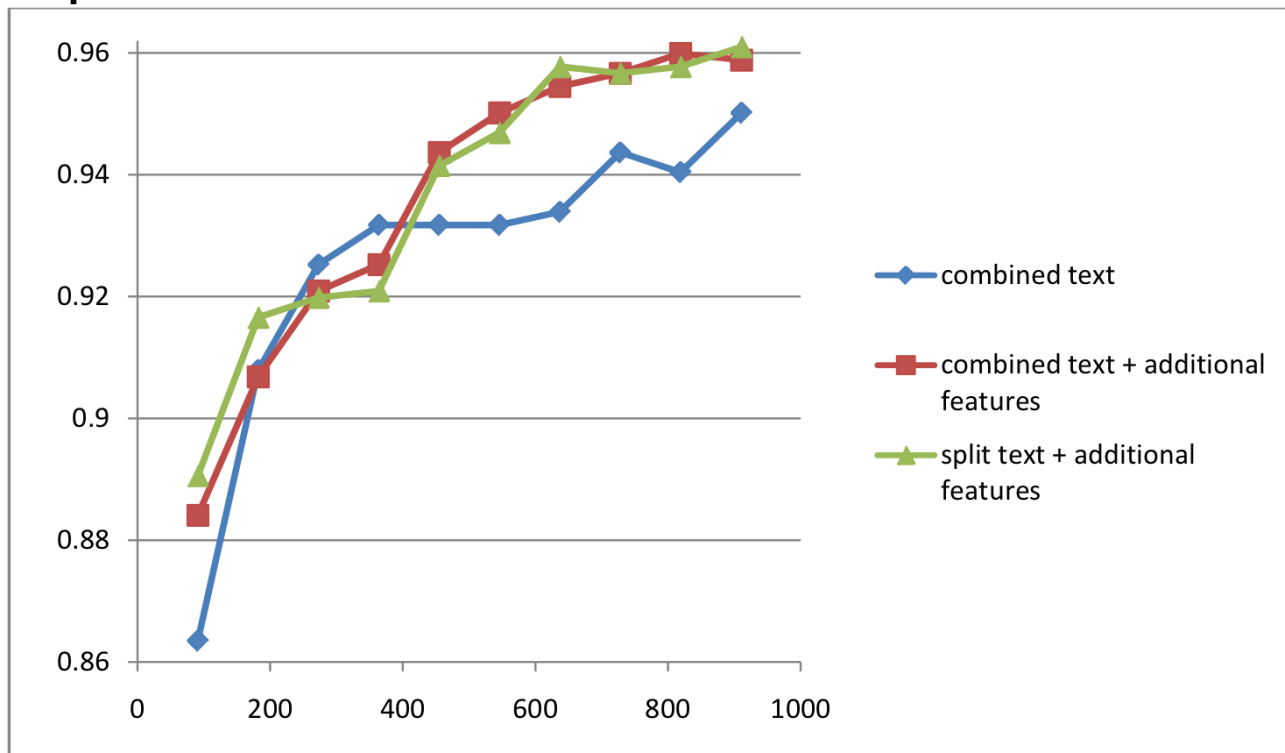


# Interact Boosting



# Why Interact?

- Human make errors
  - In previous example, 95% human accuracy, no improvement from 400 to 1000 without interaction



# Why Interact?

- **Select Critical Documents**
  - Use classifier to predict which documents are more important to label (human errors, characteristic)
- **Accelerate**
  - Use classifier to predict a new document, use that result as a hint to human
- **Evolve Iteratively**
  - As human give larger amount and more accurate labels, classifiers evolves, and it'll be used to refine/assisit/accelerate human labels, and human give even larger amount and more accurate labels, classifiers evolves again...

# Boosting Summary

- Crowd wisdom
  - ask an expert for preparing a specific problem
- Crowd funding
  - Incentive to crowd and from crowd
- Social Network
  - More participation
- Interactive AI/ML module
  - Better quality
  - Less human work (with hint)
  - Classify more even we don't have so many crowd