Open Research Directions in Micro-task Crowdsourcing

Lecture 7
Gianluca Demartini
University of Sheffield
Summary

• Monday

• Lecture 1 - Introduction to Crowdsourcing
  – An overview of the entire course.
  – Early examples of crowdsourcing (reCAPTCHA, ESP game).
  – Types of incentives: games with a purpose, citizen science, and community based crowdsourcing.

• Lecture 2 - Introduction to Micro-task Crowdsourcing Platforms
  – Key terminology of micro-task crowdsourcing.
  – Popular platforms such as Amazon MTurk and CrowdFlower.
  – How to use such systems as a crowd worker
Summary

• Tuesday
• Lecture 3 – How to Setup a Crowdsourcing Micro-task
  – Dimensions involved in crowdsourcing task design such as pricing, question design, and quality assurance mechanisms (e.g., honeypots).
  – Design and deploy a task during the lecture and see how to collect results back from the crowdsourcing platform.
• Lecture 4 – Micro-task Crowdsourcing Effectiveness
  – Techniques to ensure high quality in crowdsourced tasks (e.g., answer aggregation techniques, push crowdsourcing).
  – Behavior of malicious workers in crowdsourcing platforms.
Summary

• Wednesday
• Lecture 5 - Hybrid Human-machine Systems
  – Advanced example uses of crowdsourcing.
  – Systems that combine both the scalability of machines over large amounts of data as well as the quality of human intelligence
• Lecture 6 - Micro-task Crowdsourcing Scalability
  – In hybrid human-machine systems the latency bottleneck lays on the side of the crowd.
  – Recent research results that proposed techniques to improve the latency of crowdsourcing platforms.
  – Pricing techniques, HIT scheduling
Current Trends in Crowdsourcing

• Hybrid Human-Machine systems (DB/SW)
• User-support (HCI)
  – “Social Physics”, A Pentland
• Innovation (Business)
  – Product design by Customers
• Science (Bio/Physics)
  – “The Fourth paradigm: Data-Intensive Scientific Discovery”
State of Micro-task Crowdsourcing

• Platform side
  – Pull platforms
  – Batch processing

• Worker side
  – Work flexibility
  – Anonymity

• Requester side
  – Web/API
The Future for Requesters

• Push Platforms
• Mobile Access
• Quality and Time guarantees, SLA
• Worker API (enable novel worker UI)
• Know your crowd: Model workers
• Enterprise Crowdsourcing
  – Incentives, priorities, scheduling, profiling
The Future of the Worker side

- Reputation system for workers
- More than financial incentives
- Recognize worker potential (badges)
  - Paid for their expertise
- Train less skilled workers (tutoring system)
The Future of the Worker side

• Promote workers to management roles
  – Create gold labels
  – Manage other workers
  – Make task design suggestions (first-pass validation)

• Career trajectory (based on reputation):
  1. Untrusted worker
  2. Trusted worker
  3. Hourly contractor
  4. Employee

• Platforms logs
  – Which kind of tasks attract skilled workers
Current trends

• Active learning
• Team work in crowdsourcing platforms
  – Building flash teams:
    • collaborative knowledge work
    • Collaborative design
  – More complex tasks
Dagstuhl Seminar 13361
Crowdsourcing: From Theory to Practice and Long-Term Perspectives

Dagstuhl Seminar 14282
Crowdsourcing and the Semantic Web
Who is who in Crowdsourcing (a biased sample)

• Crowdsourcing

• Louis von Ahn, CMU
  – ESP game
  – reCaptcha
  – Duolingo (now)

• Panos Ipeirotis, NYU
  – Mturk
  – mturk-tracker.com
  – Tagasauris
Who is who in Crowdsourcing (a biased sample)

- **Databases**
  - M Stonebraker, MIT
    - Tamr
  - M Franklin, UC Berkeley
    - CrowdDB
- **Companies**: A Marcus at locu/godaddy

- **IR**
  - Matt Lease, U Texas
    - TREC Crowdsourcing
  - S Mizzaro, U Udine
    - Crowd vs TREC
- **Companies**: O Alonso, Microsoft. G Kazai, Lumi
Who is who in Crowdsourcing
(a biased sample)

• **Sem Web**
  
  • Elena Simperl, U Southampthons
    – GWAP, Galaxy Zoo

• **Lora Aroyo, VU Amsterdam**
  – Crowd Truth
  – Cultural Heritage

• **Natasha Noy, Google**
  – ICD Ontology
Who is who in Crowdsourcing (a biased sample)

• **Web Science**
  - Sir Nigel Shadbolt, U Southampton
    - SOCIAM – The Theory and Practice of Social Machines

• **HCI**
  - Michael Bernstein, Stanford
    - Worker side
  - Walter S. Lasecki, U Mich this fall
    - Real-time crowdsourcing
Who is who in Crowdsourcing (a biased sample)

• **Machine Learning**
  - M Jordan, UC Berkeley
    - Active Learning

• M Venanzi, U Southampton
  - Answer Aggregation

• **NLP**

• K. Bontcheva, U Sheffield
  - GATE Crowdsourcing plugin
Where to find crowdsourcing research

• Domain specific conferences
• Special Issues in domain-specific journals

• HCOMP
  – http://www.humancomputation.com/2015/
• HC Journal
  – http://hcjournal.org/
Open research questions

• Which pricing schemes are most appropriate to attract and motivate crowd workers in the long term?

• Can task routing and worker notification improve efficiency of real-time hybrid human-machine systems?

Open research questions

• What is the best method to track worker achievements, port them across platforms, and to develop worker profiles and skills over time?
• Which external information should be provided to workers to positively influence their work?
• How can we automatically identify malicious workers in crowdsourcing platforms?
• How can we define optimal task design guidelines for different task types?
Open research questions

• Can we automatize the design of hybrid human-machine workflows?

• Which are the most appropriate incentive, task designs, and task routing approaches for enterprise crowdsourcing?

• Which information should we to provide to non-expert workers when crowdsourcing domain-specific tasks?
Exam – Individual project

• 1. Research problem and how to apply micro-task crowdsourcing to it
• 2. Design the crowdsourced task/HIT
• 3. Get a data sample and crowdsource it over CrowdFlower
  – Spend no/little/lots money
• 4. Collect results back
• 5. Aggregate, Analyze, Discuss, Draw conclusions on what works and what does not
• 6. Submit 10 pages LLNCS report (to me and S Mizzaro)
Thank you!

Gianluca Demartini

gianluca@sheffield.ac.uk

Slides: gianluca@sheffield.ac.uk