Collective Intelligence and Crowdsourcing

Maria Maistro
mm@di.ku.dk

Web Science Lecture
15 February 2021

UNIVERSITY OF COPENHAGEN

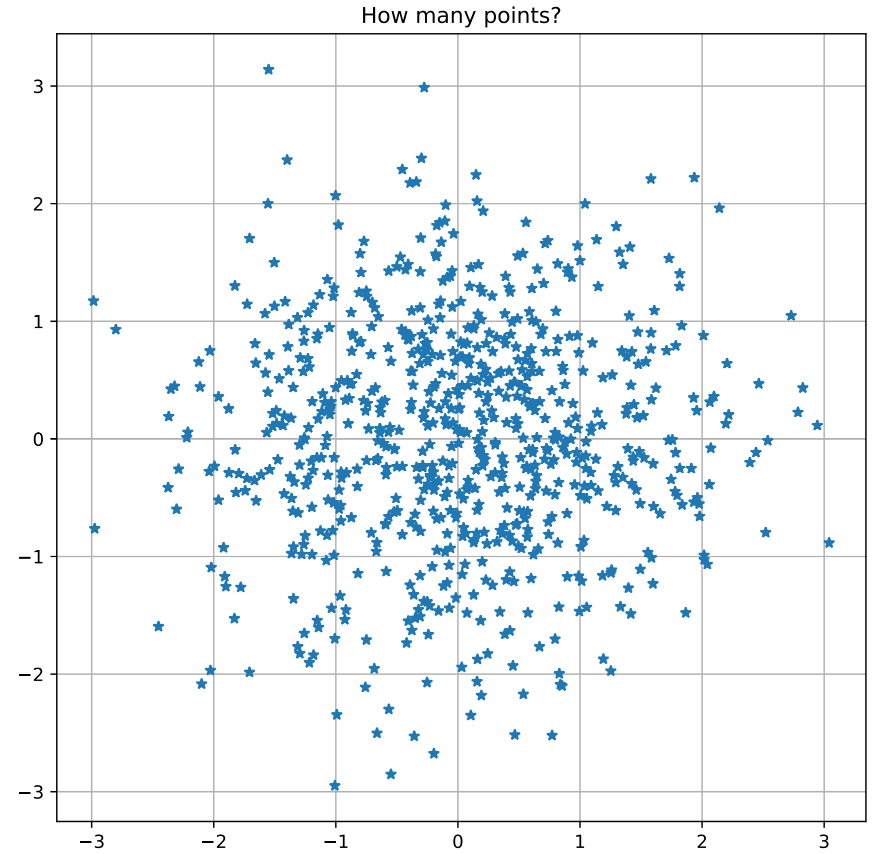
Credit for many of the slides: Christina Lioma

Outline

- The Basic of Collective Intelligence
- Collective Intelligence in Nature and in AI
- Crowdsourcing
- Advantages and Challenges of Crowdsourcing
- Expectation Maximization (EM)
- EM Extensions

COLLECTIVE INTELLIGENCE

Let’s Start with an Experiment



Guess the number of points in the plot!

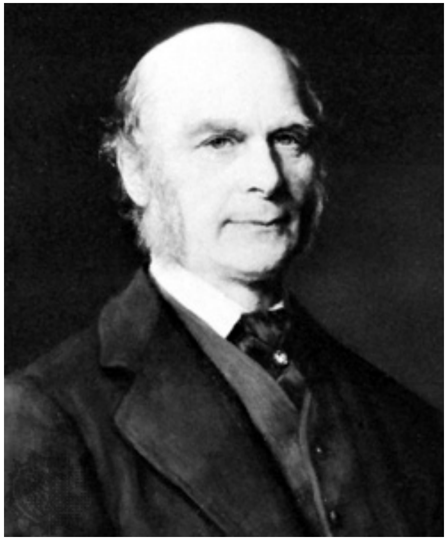
Exercise 1: https://tinyurl.com/stwoone

Exercise 2: https://tinyurl.com/wnbsds7

Answer

- Correct answer: 748;
- Exercise 1 (Individual): Average answer 460;
- Exercise 2 (Pair): Average answer 511;
- Best guess: 700 (Yuze);
- The average was closer to the correct answer than >50% of your individual guesses;
- The average was closer to the correct answer than >63% of your group guesses.

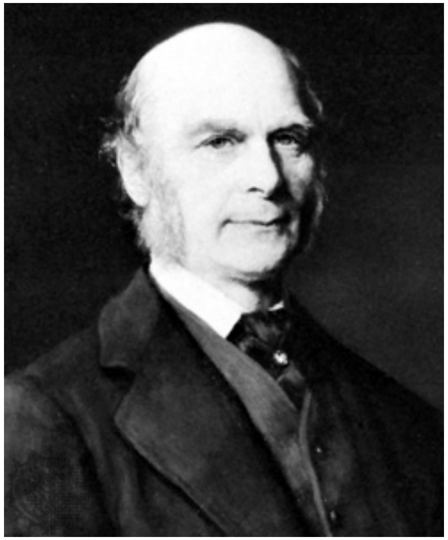
Sir Francis Galton, Cambridge University 1906



Victorian Polymath
Charles Darwin's half-cousin

- Created statistical concept of correlation;
- Introduced questionnaires & surveys;
- Coined terms "eugenics", "nature versus nurture";
- Founded psychometrics & differential psychology;
- Invented fingerprint classification in forensics;
- Co-founded scientific meteorology, first weather map.

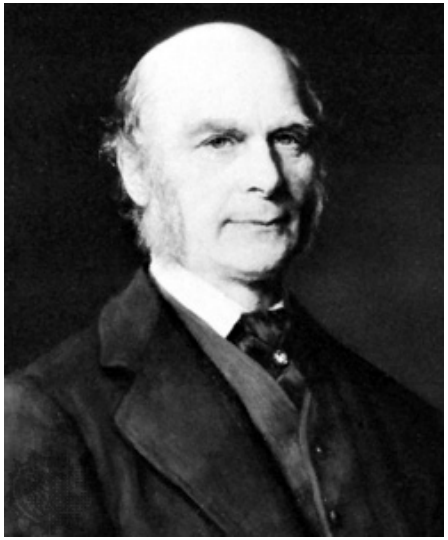
Sir Francis Galton, Cambridge University 1906



Guess the weight of an ox – prize for the best guess:

- People placed wagers on the ox’s weight;
- Galton analysed 800 guesses;
- Crowd’s median guess: 542kg;
- No one found the exact weight: 543kg.

Sir Francis Galton, Cambridge University 1906



If you put together a big enough and diverse enough group of people and ask them to make a decision, that group's decision will, over time, be intellectually superior to the isolated individual .

"Collective Intelligence" (CI)

"The group can be smart in a way that none of its members is."

- Levy (philosopher) 1994

Intelligence definition: the ability to solve problems

A system is more intelligent than another system if, in a given time interval, it can:

- solve more problems, or
- find better solutions to the same problems.

Collective Intelligence: a group can find more or better solutions than the whole of all solutions that would be found by its members working individually.

Social Insects

Complex & highly intelligent how:

- ants map out their environment;
- termites build their mounds.

Individually they are not intelligent (limited information processing capacities). But they are **many**.

Self-organising system: very large numbers of simple components interacting locally to produce global organization and adaptation.

Termites Building Mounds:

- At first, different termites drop mud randomly;
- The presence of a mud heap incites other termites to add mud to that heap, rather than start a new heap;
- The larger the heap, the more attractive it is to other termites;
- Small heaps will be abandoned;
- Large heaps will grow into tall columns;
- Columns will grow towards each other until they touch → arch;
- Arch will grow until it touches other arches → intricate structure of interlocking arches.

Swarm Intelligence (in AI)

Emergent collective intelligence of groups of simple agents.

The key is that complex behaviour (problem solving) arises out of a composition of simple behaviours.

Ant colonies, bird flocking, animal herding, microbial intelligence, etc.

Collective Intelligence and Robotics

https://www.ted.com/talks/radhika_nagpal_what_intelligent_machines_can_learn_from_a_school_of_fish

Science fiction visions of the future show us AI built to replicate our way of thinking, but what if we modeled it instead on the other kinds of intelligence found in nature?

https://www.ted.com/talks/vijay_kumar_the_future_of_flying_robots

Precision farming: swarms of robots map, reconstruct and analyze every plant and piece of fruit in an orchard, providing vital information to farmers that can help improve yields and make water management smarter.

Six Characteristics of Collective Intelligence:

1. **Group** of individuals (e.g., people, robots, software agents, or a mixture of these);
2. **Big** enough group, **diverse** enough group
3. **Local** and often **random behaviours;**
4. **Simple interactions**, often following simple rules;
5. **Self-organising** over time;
6. **De-centralised** (no dictating how to behave).

→ Globally intelligent: more or better solutions than if working individually.

The World Wide Web

CROWDSOURCING

Collective Intelligence and Crowdsourcing

Collective intelligence

- Possibility to seamlessly integrate machine and human intelligence at large scale;
- One of the holy grails of AI, known as **mixed-initiative systems** (Horvitz 2007):
  - Intelligent systems that can collaborate effectively & naturally with humans (reasoning, communication, planning, execution, learning).

Crowdsourcing (type of Collective Intelligence)

- Taking a job traditionally done by a designated person and outsourcing it to an undefined, large group of people.

Crowdsourcing

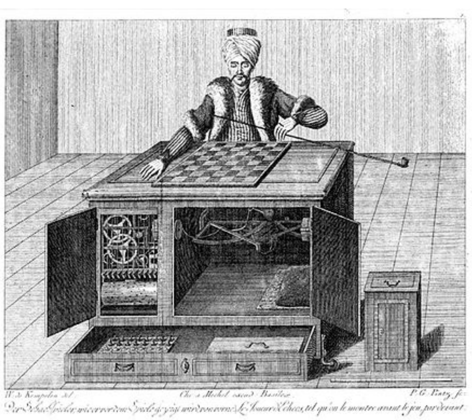
Crowdsourcing: human-based computation

Use humans as processors in a distributed system.

- Wikipedia, Captcha.

Amazon Mechanical Turk (mturk.com):

- Crowdsourcing platform;
- Named after Mechanical Turk, fake chess playing machine in 18C;
- Remote human labour hidden behind a computer interface.



What is Crowdsourcing?

Crowdsourcing:

- Process for obtaining work or knowledge from large crowds through a system of invitations and incentives;
- A requestor (with some **internal objective**) solicits a group of workers to perform tasks in service of that objective.

Requestor's objective → utility function to be maximized.

Example: upper bound on the amount of money that the requestor can spend.

Amazon Mechanical Turk

- Requesters create human intelligence tasks (HITs) via web services API or dashboards;
- Workers log in, choose HITs, perform them:
  - Initially mainly in US, educated, female, money not the key consideration (mothers at home with their children);
  - Increasingly diverse and international, workers in some countries often more interested in money.
- Requesters assess results, pay per HIT satisfactorily completed;
- Currently >200,000 workers from >100 countries; millions of HITs completed.

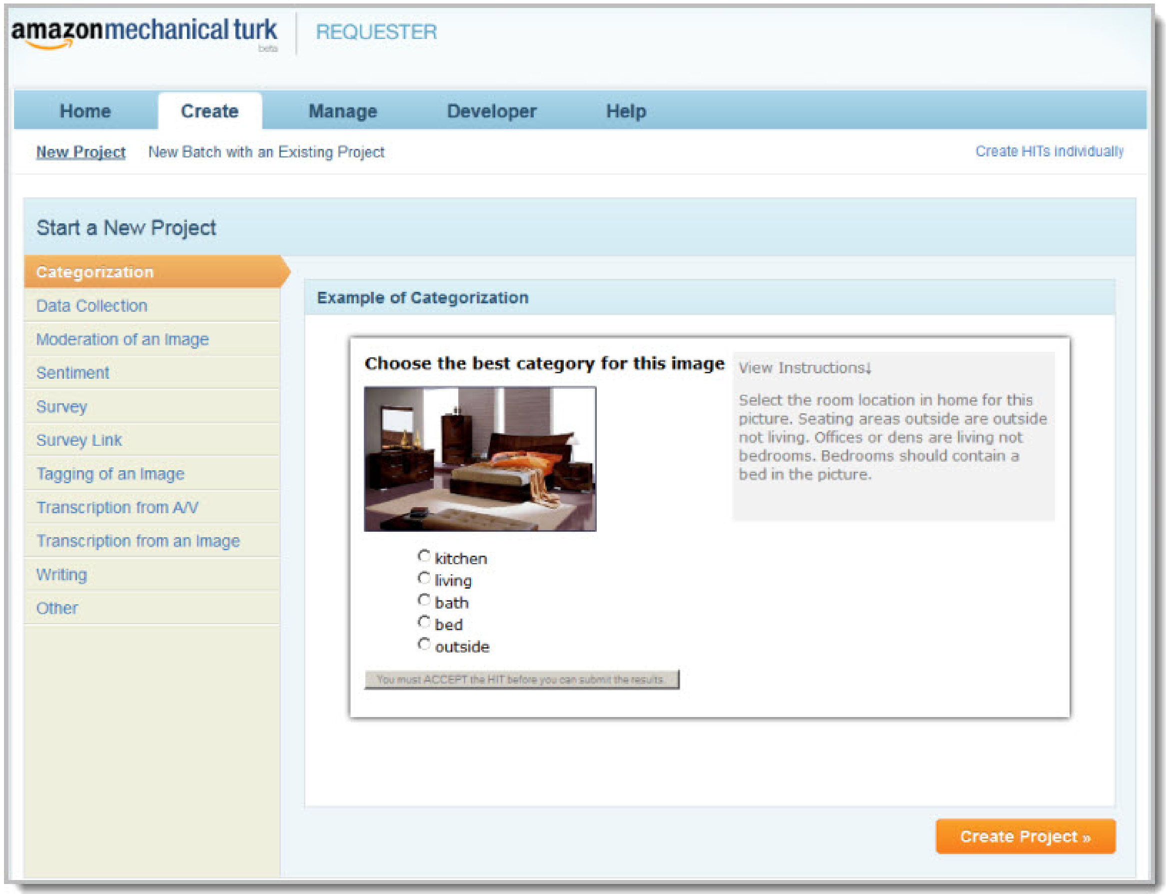
Crowdsourcing Example

Example: obtain labels for a set of images.

Requestor’s objective includes:
- Maximising average quality of labels;
- Minimising the number of tasks given to workers (to reduce own cost and/or worker burden);
- Minimising latency (get answers quickly).

Requestor must choose:
- How many workers should be given a labelling task;
- How their responses should be aggregated to estimate the best answer.

Example: Amazon Mechanical Turk



amazonmechanical turk

Crowdsourcing Assumptions

Assumptions:

- Workers act independently, interacting only through shared tasks;
- Each worker has own utility function, which is often different from the collective's utility function.

Crowdsourcing for AI: use crowdsourcing to label training sets as input for data-hungry supervised algorithms.

Crowdsourcing Advantages and Challenges

Advantages:

- Wisdom of crowds;
- Round the clock availability of people;
- Low cost;
- Allows rapid early stage experimentation;

Challenges:

- Highly varying skills, abilities, motivation of workers;
- Different tasks may be individually easier or more difficult, requiring less or more work (iterations) on them.

The same HIT will be assigned to multiple workers!

ADDRESSING CROWDSOURCING CHALLENGES

Gaming Behaviour

Common problem: **gaming behavior**.

- Workers answer the first few questions and then decrease effort and accuracy on remaining questions.

Solution:

- Intermix questions with known (gold questions) and unknown answers;
- Whenever a worker fail a gold question, the worker will be discarded.

Reputation

- Workers can build a reputation within a crowdsourcing system;
- Details of previous tasks undertaken, etc can be available to requesters to help them to decide which offers to accept;
- This can act as an incentive for a requestor to accept their offer to undertake a task, and to trust the likely quality of their work;
- Requester can also select a worker based on their own previous experience of the worker.

Addressing Issues with Crowdsourcing

Methods to address these challenges:

1. Machine learning algorithms that model the accuracy of crowd workers;
2. Aggregation methods for predicting true answers from error-prone and disagreeing workers;
3. Control algorithms that choose which tasks to request and which individuals should work on them.

Crowdsourcing quality: highly dependent on 1-3

Crowdsourcing Strategies

- Post hoc response aggregation (after responses have been received);
- Dynamically fit tasks to workers (while responses are being received).

Input and Goals

Given:

- Set of questions, workers, and their responses (some questions are answered by more than one worker);
- Questions are objective (each has a unique answer);
- Workers may not answer correctly;
- Majority of workers are more likely to answer correctly than not.

Goal:

1. Exploit redundancy;
2. Learn and track worker skill.

Goals

1. Exploit redundancy by comparing different workers responses to the same question;

- Snow et al 2008: with simple majority voting a crowd of novices outperformed an expert on sentiment analysis and semantic word similarity.

2. Learn and track the skills of workers

- Instead of majority voting, weigh worker responses by using models of workers' abilities. E.g. if a worker is excellent at translating French to English, assume that their English to French translations are of high quality.

Majority Vote

Majority vote:

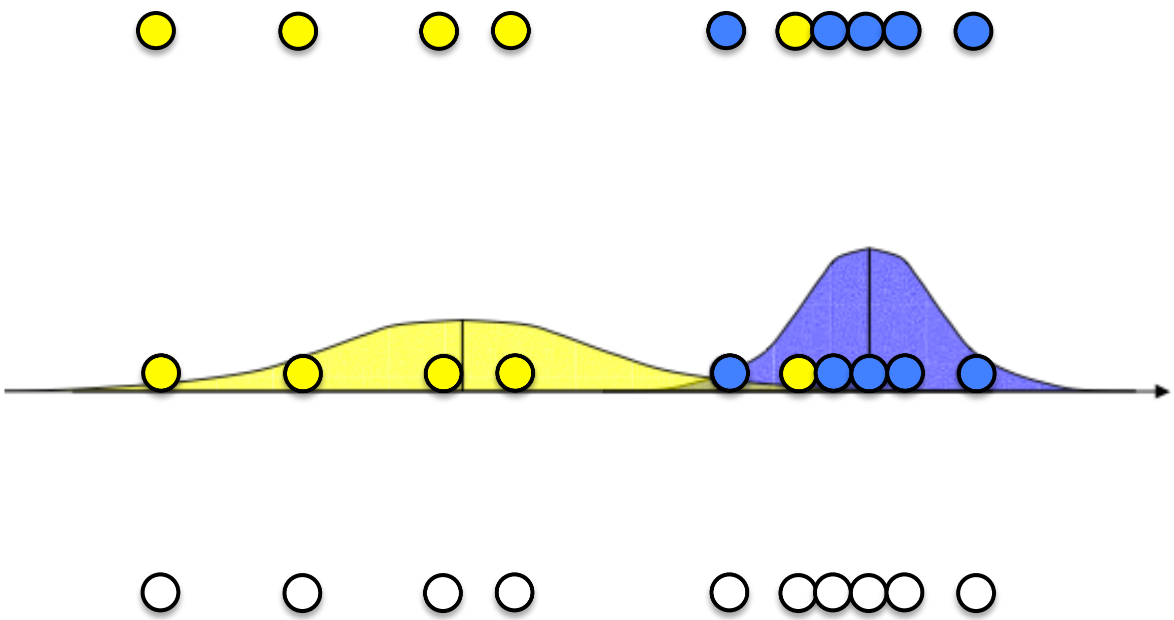
Each worker is a voter, we assign the answer which gets the majority of the votes. In case of ties, we toss a coin or consider the average (if possible).

Improve on majority voting by modelling worker skill

- Supervised learning: give workers questions for which gold standard answers are already known;
- If worker fails to correctly answer: dismiss or have their weights lowered.

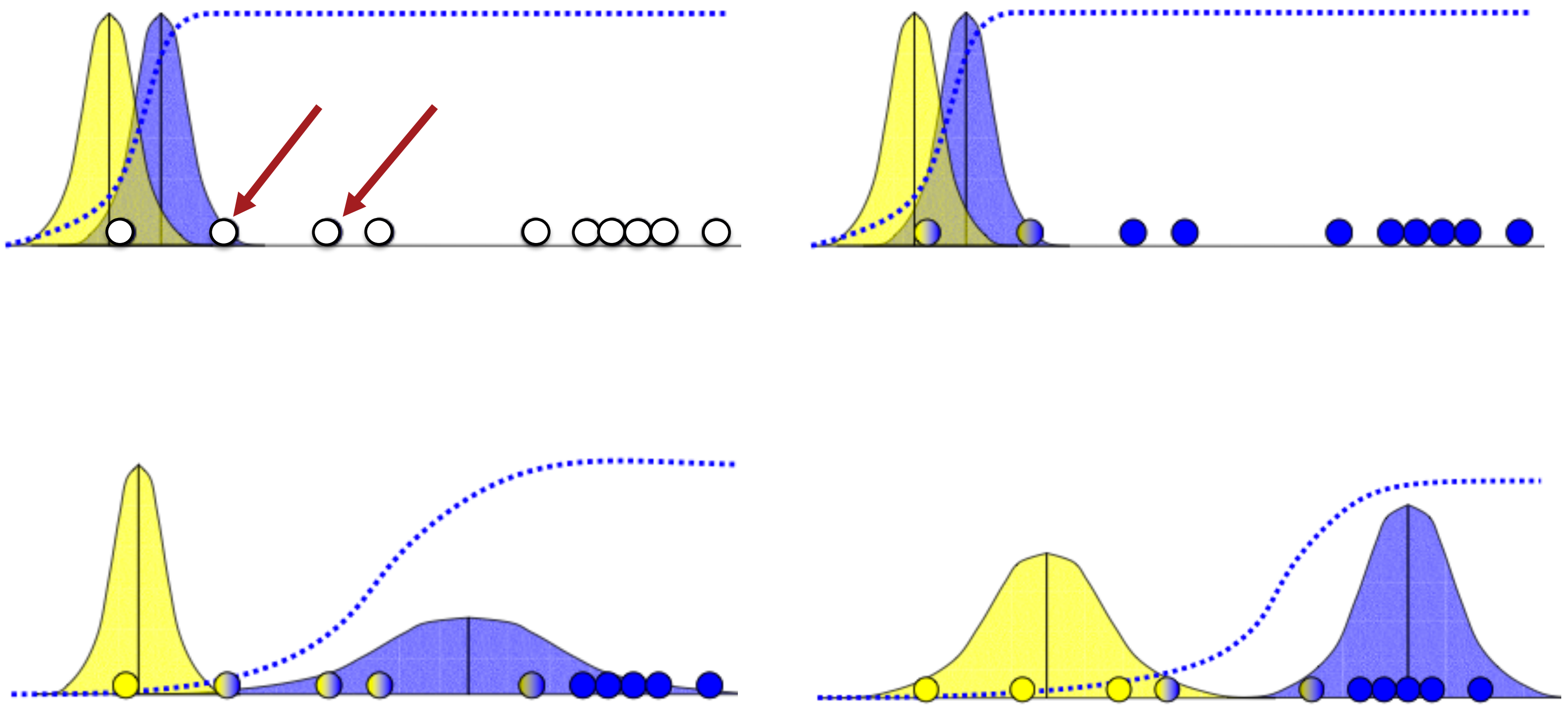
EXPECTATION MAXIMIZATION

1-Dimensional Example



Slides from Victor Lavrenko:
https://www.youtube.com/watch?v=iQoXFmbXRJA&ab_channel=VictorLavrenko

Expectation Maximization (EM)



Slides from Victor Lavrenko:
https://www.youtube.com/watch?v=iQoXFmbXRJA&ab_channel=VictorLavrenko

Early use of Expectation Maximisation (EM)

David & Skene 1979 (medical diagnosis)

- There is a single question with an unknown correct answer;
- $P_w(r|a)$: probability that worker w will give response r when true answer is a;
- If $r \neq a$, an expert worker would have $P_w(r|a) = 0$;
- Worker responses are conditionally independent of each other, given the true answer;
- Expectation Maximisation (EM): iterative algorithm to estimate which answers are correct at the same time that the algorithm learns the model of worker accuracies.

EM Algorithm

- Start by taking a majority vote and use that to determine an initial guess of the correct answer to each question;
- Score each worker based on how many answers he/she got right;
- Weigh each worker's votes based on their score (better workers count more);
- Since weighted votes likely produce a different set of correct answers, recompute each worker's score;
- Repeat until convergence.

Expectation and Maximisation Steps

Expectation: given estimates of all the probabilities, compute the probability of each possible answer using Bayes rules and assuming conditional independence.

Maximisation: given the posterior probability of each possible answer, compute new parameters than maximise the likelihood of each voter's response.

- Assigns higher weights to good workers;
- Allows for single strong worker to overrule multiple weak workers;
- Predicted answer may not always be the majority vote.

EM Extension (Whitehill et al. 2009)

Let on average students have an 80% chance of correctly answering questions → Jane has 80% chance of answering correctly Q13.

If everyone else answers Q13 wrong → Q13 is especially hard and we must lower Jane's chances.

Idea: use worker errors to update estimate of task difficulty and worker accuracy.

Same idea as in EM, but probability computations include worker expertise and task difficulty.

EM extension (Welinder et al. 2010)

- Questions have many features → Workers modelled as linear classifiers who respond by weighing those features;
- Account not only for question difficulty and worker skills, but also for arbitrary question and worker features.

E.g. a worker can be good at identifying types of birds, but only when viewed from the back.

Question features do not need to be specified a priori: the algorithm learns the features.

Excellent performance but low interpretability.

Further extensions (I)

Prelec and Seung 2007

Find correct answers missed by the majority by asking workers to predict coworkers mistakes.

Worker is asked to:

- Endorse the response mostly likely to be true, and
- Predict the proportion of the sample workers that will endorse each possible response.

Lin et al. 2012

When requesters cannot enumerate all possible answers for the worker or when the solution space is infinitely large

Further extensions (II)

Raykar et al. 2010

Fully automatic:

- Learn worker abilities;
- Infers correct responses;
- Learns a logistic regression classifier to predict future crowd responses or the answer (no need to consult human workers in the future).

Instead of only reconciling multiple responses, get the answers themselves.

Further extensions (III)

Dekel and Shamir 2009

Limit influence of bad workers:

- Use worker responses to train a SVM;
- Add constrains to the loss function such that no bad worker can overly influence the learned weights;
- Prune bad workers.

Wauthier and Jordan 2011

Relax the idea that tasks must contain a correct answer.

- Predict each worker's response to a future subjective question.

Thank you!