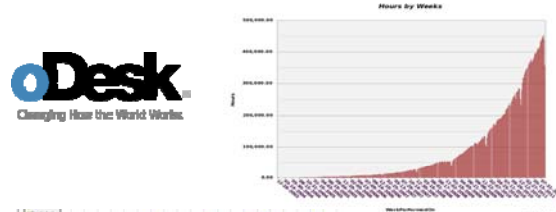
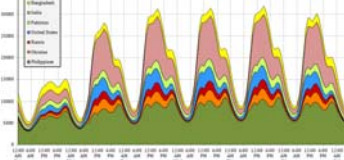



Crowdsourcing


thanks to

Mausam Peng Dai Chris Lin
NSF, ONR, Google

63,000 FTE Daily Workers (at 8h/day)

Built in 1770 by Wolfgang von Kempelen



Powerset

Find Substitutable Words

In the sentence below, what words or phrases could replace the **bolded** word without changing the meaning? F

Example:
In most countries **children** are required by law to attend school.

You might enter:
kid
youngster
pupil
young person

Try to enter single words or short phrases like "water bottle" or "post office." You are encouraged to use the "a station".
Avoid descriptive phrases, e.g. "a container you drink out of," or "a place you mail things from" unless you abs
Further, tell us how easy or difficult it is to assign one of several possible meanings for the **bolded** word in the

Your sentence is: The term silver dollar is often used for any large white metal coin issued by the United States with a face value of one dollar ; although purists insist that a dollar is not silver unless it contains some of that metal .

Enter *one term* per box. \$0.05

Fast & Cheap, but is it Good? [Snow et al. EMNLP-08]

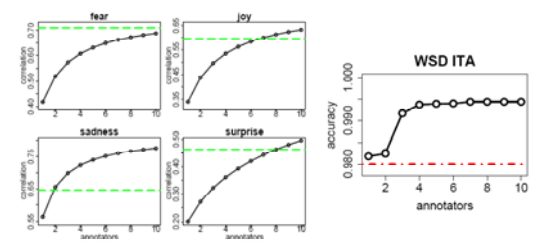


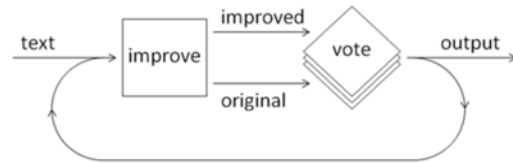
Figure 1: Non-expert correlation for affect recognition

Need for Workflows

- **Challenges**
 - Reliability & skill of individual workers vary
 - Small work units
- **Therefore**
 - Use workflow to aggregate results & ensure quality
 - Manage workers with (unreliable) workers
 - Accomplish large tasks from small contributions
 - *Eg, TurkIt: programming language for workflows*

Complex Jobs

- **Casting Words**
- **Turkit**



Iterative Improvement



A partial view of a pocket calculator together with some coins and a pen.

Iterative Improvement



"A close-up photograph of the following items:

A CASIO multi-function, solar powered scientific calculator.

A blue ball point pen with a blue rubber grip and the tip extended.

Six British coins; two of £1 value, three of 20p value and one of 1p value.

Seems to be a theme illustration for a brochure or document cover treating finance - probably personal finance."

Clowder



AI, ML & Decision Theory ... for Crowdsourcing

- **Declarative language to specify workflows**
 - HTNs
- **Shared models for common tasks**
 - Eg, voting, discrete choices, content improvement
- **Integrated modeling of workers**
- **Comprehensive decision-theoretic control**

Quality Control & Reputation

- **Probably the most important deterrent**
 - to wide adoption of mechanical turk
- **Recently: more spammers than usual**
- **Necessitates**
 - automatic detection of spammers
 - automatic rewarding of diligent workers
 - automatically achieving quality goals

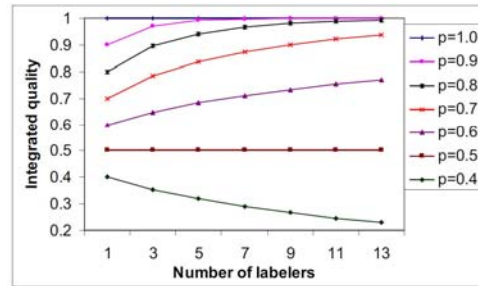
Quality Control

- **Simple tasks**
 - Majority vote
 - Quality-corrected vote based on worker parameters
 - assumptions on worker independence
 - Learning worker parameters using gold standard
 - Joint estimation of votes and worker parameters
- **Complex tasks: workflows**
 - Decision-theoretic control of workflows

17

Majority Voting

[Sheng et al, 2008; Snow et al, 2008]



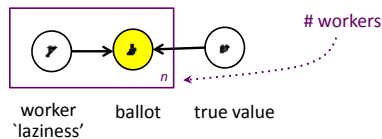
Majority vote of 8 Turkers better than expert labeling

18

Quality-Corrected Voting

[Clemen and Winkler, 1990]

Assumption: workers independent of each other



$$P(v | b_1, \dots, b_n, \gamma_1, \dots, \gamma_n) \sim P(b_1, \dots, b_n | v, \gamma_1, \dots, \gamma_n) P(v) = P(v) \prod_i P(b_i | v, \gamma_i)$$

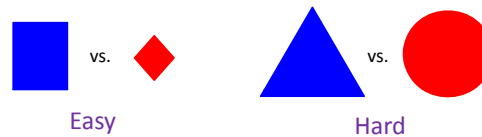
Outperforms majority vote

19

Quality-Corrected Voting 2

[Whitehill et al, 2009; Dai et al, 2010]

Are workers really independent?



Intrinsic difficulty (d) measures how hard is problem

Conditional Independence

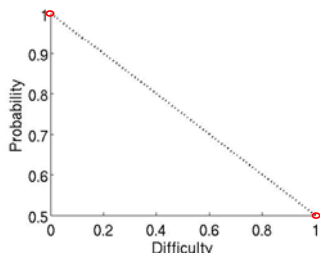
- workers independent given intrinsic difficulty

20

Probability of a Correct Answer

$$\text{accuracy}_w(d) = \frac{1}{2}[1+(1-d)^w]$$

Assume: no malevolence



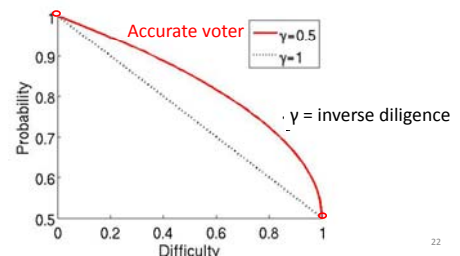
11/28/2012

21

Probability of a Correct Answer

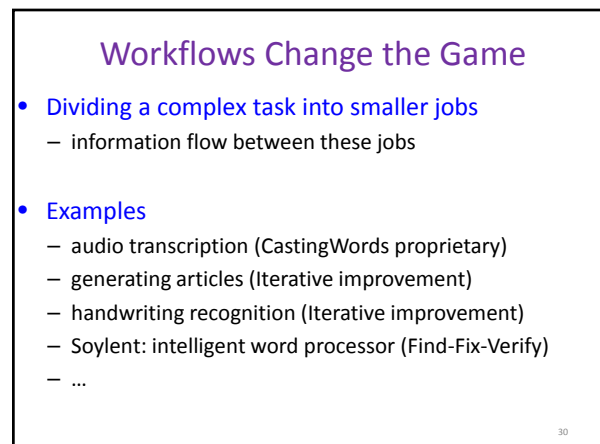
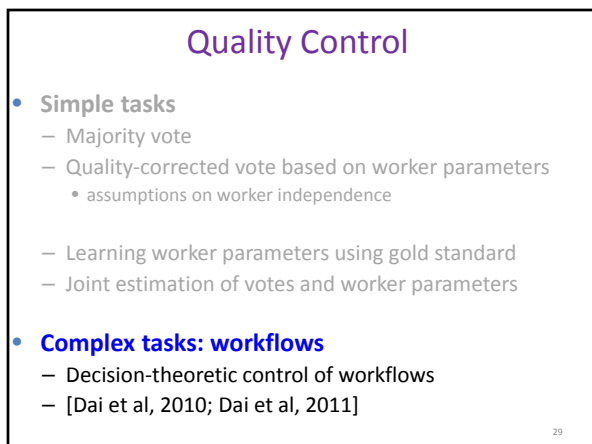
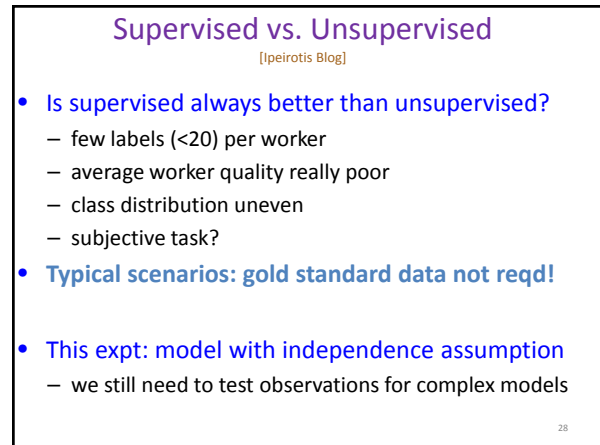
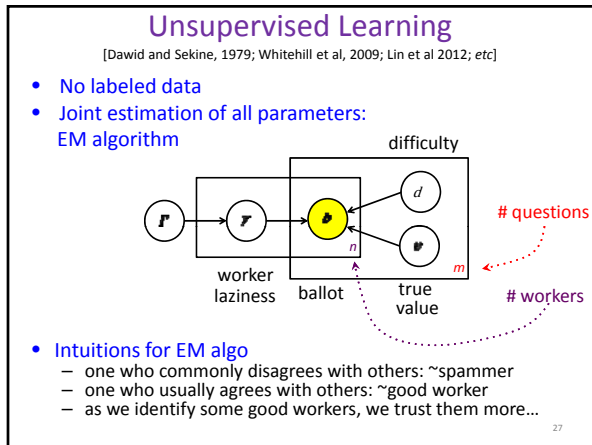
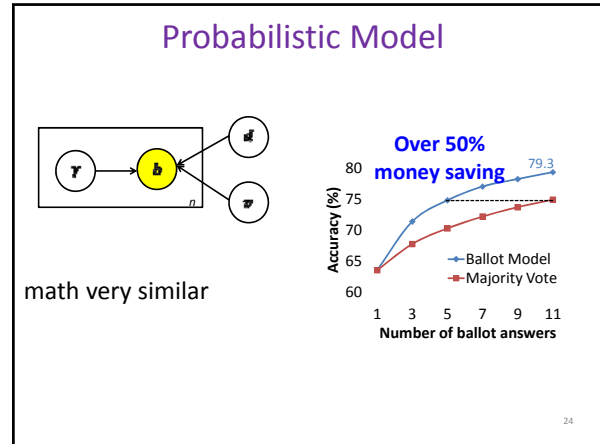
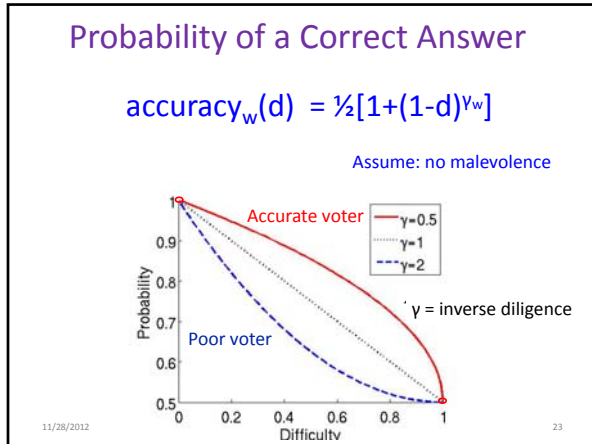
$$\text{accuracy}_w(d) = \frac{1}{2}[1+(1-d)^w]$$

Assume: no malevolence



11/28/2012

22



Iterative Improvement

[Little et al, 2010]

The diagram illustrates an iterative improvement process. It starts with 'text' which goes into an 'improve' block. The output of 'improve' is 'improved', which then goes to a 'vote' block. The 'vote' block outputs 'output'. A feedback loop goes from 'output' back to 'improve'. A smaller diagram above shows 'text' going to 'improved' and 'original', which then go to 'vote'.

Clowder

The Clowder architecture diagram shows a central 'rendered job' (purple oval) connected to several components: 'HTN library' (blue cloud), 'DT planner' (blue box), 'renderer' (blue box), 'worker marketplace' (purple cloud), 'executor' (purple box), 'learner' (blue box), 'task models' (blue cloud), and 'user models' (blue cloud). Arrows indicate the flow of information between these components.

- Belief states = distribution over world states
- Actions = probabilistic transitions

Decision-Theoretic Execution Control

- **POMDP:** Partially-Observable Markov Decision Process
- **Belief State:** Probability distribution over world states
- **Action Transitions:** Submission & observation of HITS
 - Eg, Improvement: Prob distribution of new artifact
 - Eg, Ballot Observation: Bayesian updates on qualities
EM update worker accuracies
- **Objective:** Maximize $E[R(q) - \Sigma c]$

TurKontrol of Iterative Improvement

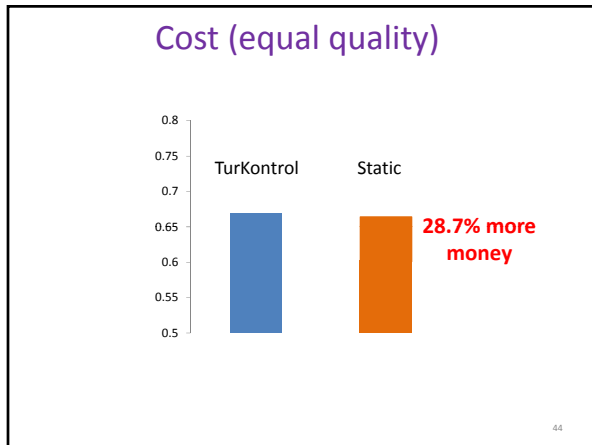
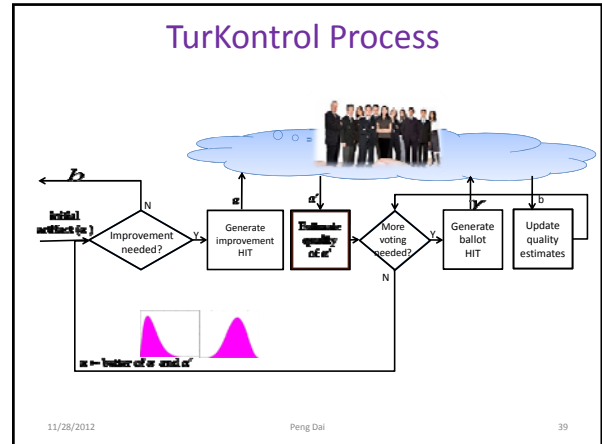
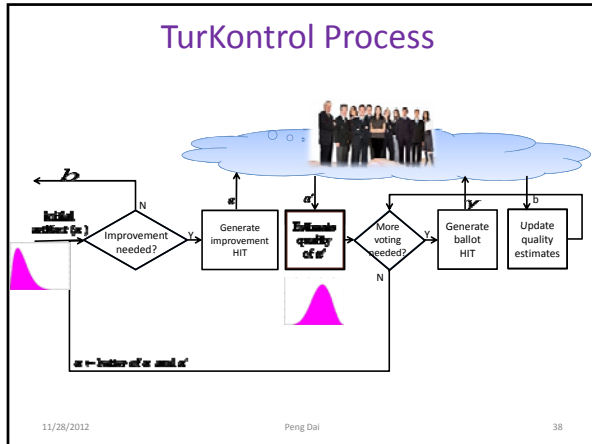
The TurKontrol process flowchart starts with 'Initial artifact (a_0)' leading to a decision 'Improvement needed?'. If 'Yes', it goes to 'Generate improvement HIT', then 'Estimate quality of a^1 ', then 'More voting needed?'. If 'Yes', it goes to 'Generate ballot HIT' and 'Update quality estimates'. If 'No', it loops back to 'Improvement needed?'. If 'No' at the first decision, it loops back to 'Initial artifact (a_0)'.

TurKontrol Process

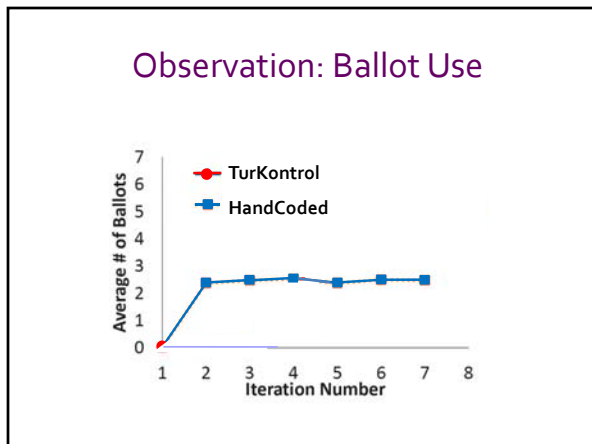
This slide shows the same TurKontrol process flowchart as slide 35, detailing the steps from artifact generation to quality estimation and voting.

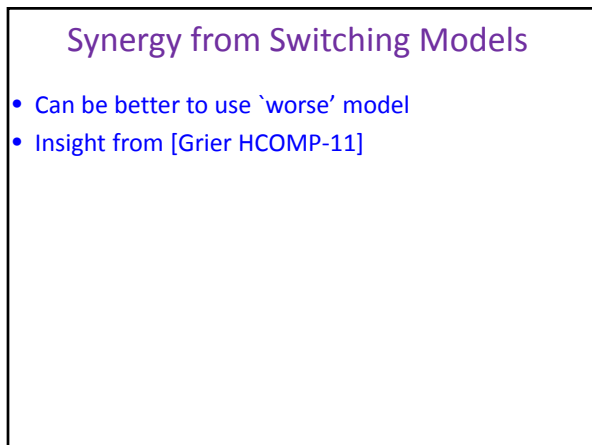
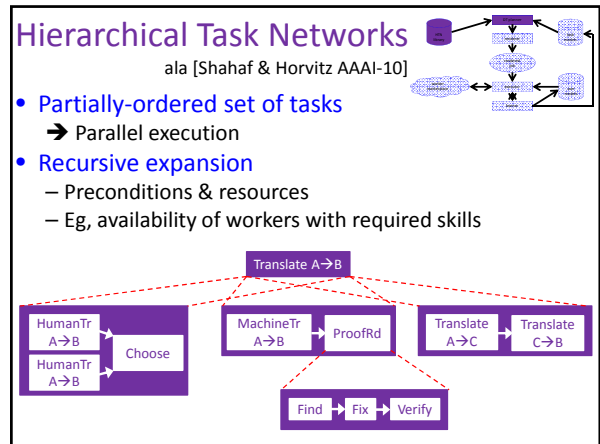
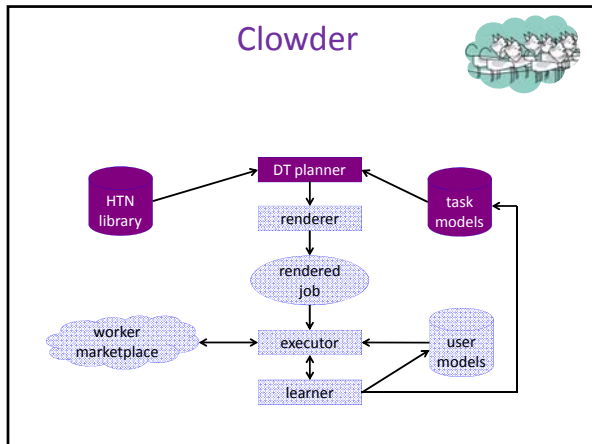
TurKontrol Process

This slide shows the same TurKontrol process flowchart as slide 35, detailing the steps from artifact generation to quality estimation and voting.



- ### Anecdotal Observations
- 7 images: TurKontrol fewer iterations than static
 - 6 of those resulted in higher quality!!
 - once: TurKontrol trusted the first vote
 - the worker was known to be higher quality
 - intelligent ballot use
- 45





Example Task: Named-Entity Recognition

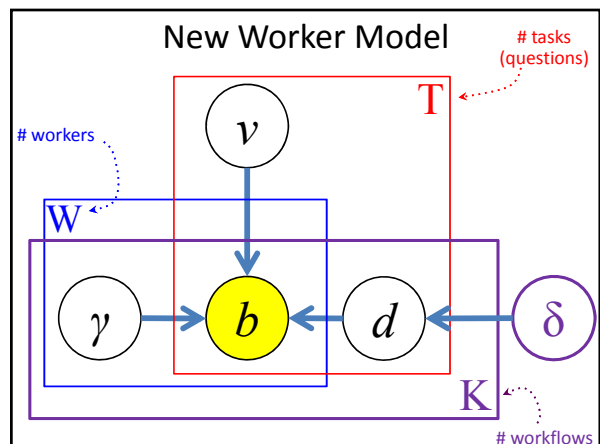
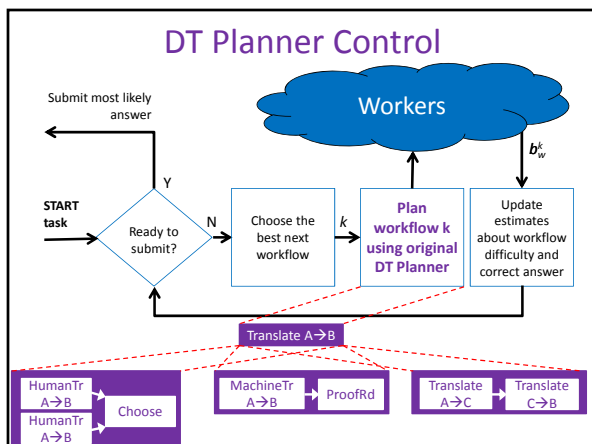
Only two states -- Vermont and **Washington** -- this year joined five others requiring private employers to grant leaves of absence to employees with newborn or adopted infants

Which of the following Wikipedia articles defines the word "**Washington**" in exactly the way it is used in the above sentence?

- [Washington](http://en.wikipedia.org/wiki/Washington)
Washington, D.C., formally the District of Columbia and commonly referred to as Washington, "the District", or simply D.C., is the capital of the United States....
- [Washington \(state\)](http://en.wikipedia.org/wiki/Washington_(state))
Washington () is a state in the Pacific Northwest region of the United States located north of Oregon, west of Idaho and south of the Canadian province of British Columbia, on the coast of the Pacific Ocean....

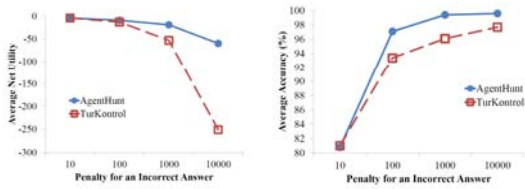
Which of the following sets of tags best describes the word "**Washington**" in the way it is used in the above sentence?

- location
- us_county location citytown



Experiments

- Training Data:
 - 50 NER Tasks
 - 40 Wikipedia jobs and 40 direct tagging jobs
- 1000 simulations



- 106 NER Tasks using Mechanical Turk

	AGENTHUNT	TURKONTROL	TURKONTROL ₁₀₀	AGENTHUNT _{RL}
Avg Accuracy (%)	92.45	85.85	84.91	93.40
Avg Cost	5.81	4.21	6.26	7.25
Avg Net Utility	-13.36	-18.35	-21.35	-13.85

Research Agenda

- every workflow needs AI support
 - optimal pricing
 - optimal parameter estimation
 - optimal control
 - comparison between multiple workflows for a task
- designing a generalized workflow optimizer
 - HTN language: express a workflow in the language
 - automatically optimize parameters and control

60