thanks to

Mausam Peng Dai Chris Lin
NSF, ONR, Google

Built in 1770 by Wolfgang von Kempelen

Find Substitutable Words
In the sentence below, what words or phrases could realize the bolded word without changing the meaning of the sentence?

In most countries education is required by law to attend school.

You might enter:
- kids
- education
- school

Try to enter single words or short phrases like "water bottle" or "cook office." You are encouraged to use the to avoid descriptive phrases, e.g., "a container you drink water from." If you think of something you don't think fits, tell us how easy or difficult it is to assign one of several possible meanings for the bolded word in the sentence.

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]
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 50p 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

---

**Quality-Corrected Voting**

[Clemen and Winkler, 1990]

Assumption: workers independent of each other

\[
P(v | b_1, ..., b_n, \gamma_1, ..., \gamma_n) = \frac{P(v)}{P(b_1, ..., b_n | v, \gamma_1, ..., \gamma_n)}P(v)
\]

Outperforms majority vote

---

**Majority Voting**

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

Majority vote of 8 Turkers better than expert labeling

---

**Probability of a Correct Answer**

- **Accuracy**:\[\text{accuracy}_w(d) = \frac{1}{2}[1+(1-d)^\gamma]\]
  - Assume: no malevolence

- **Accurate voter**:\[\text{accuracy}_w(d) = \frac{1}{2}[1+(1-d)^\gamma]\]
  - \(\gamma = \text{inverse diligence}\)
  - Assume: no malevolence
Probability of a Correct Answer

\[ \text{accuracy}_w(d) = \frac{1}{2} \left[ 1 + (1-d)^y \right] \]

Assume: no malevolence

Unsupervised Learning

[Dawid and Skene, 1979; Whitehill et al, 2009; Lin et al 2012, etc]

- No labeled data
- Joint estimation of all parameters: EM algorithm

Supervised vs. Unsupervised

[Ipeirotis Blog]

- Is supervised always better than unsupervised?
  - few labels (<20) per worker
  - average worker quality really poor
  - class distribution uneven
  - subjective task?
- Typical scenarios: gold standard data not reqd!
  - This expt: model with independence assumption
    - we still need to test observations for complex models

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
  - [Dai et al, 2010; Dai et al, 2011]

Workflows Change the Game

- Dividing a complex task into smaller jobs
  - information flow between these jobs
- Examples
  - audio transcription (CastingWords proprietary)
  - generating articles (Iterative improvement)
  - handwriting recognition (Iterative improvement)
  - Soylent: intelligent word processor (Find-Fix-Verify)
  - ...
Iterative Improvement

[Little et al., 2010]

Clowder

POMDPs at the core

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

\[ \text{Belief states} = \text{distribution over world states} \]
\[ \text{Actions} = \text{probabilistic transitions} \]

Decision-Theoretic Execution Control

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

TurKontrol of Iterative Improvement

TurKontrol Process
Improvement needed?

HIT Generate ballot
HIT Update quality estimates

More voting needed?

Y Y

11/28/2012 Peng Dai

Cost (equal quality)

<table>
<thead>
<tr>
<th></th>
<th>TurKontrol</th>
<th>Static</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

28.7% more money

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

Observation: Ballot Use

![Graph showing average # of ballots over iteration number]

TurKontrol
HandCoded
Clowder

Hierarchical Task Networks
ala [Shahaf & Horvitz AAAI-10]
- Partially-ordered set of tasks
  ➔ Parallel execution
- Recursive expansion
  - Preconditions & resources
  - Eg, availability of workers with required skills

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

Synergy from Switching Models
• Can be better to use ‘worse’ model
• Insight from [Grier HCOMP-11]

New Worker Model

Example Task: Named-Entity Recognition
Which of the following Wikipedia articles defines the word “Washington” in exactly the way it is used in the above sentence?
- 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)
  Washington (state) 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
city
town

DT Planner Control

Wikipedia articles define the word "Washington" in exactly the way it is used in the above sentence?
Experiments

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

• 106 NER Tasks using Mechanical Turk

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