Classify

Note: Please always classify the galaxy in the centre of the image.

SPIRAL
How many spiral arms are there?

1 2 3 4

More than 4 Can't tell
Killer whales are beautiful animals. I remember seeing these huge, smooth, black and white creatures jumping high into the air at Sea World as a kid.

Crowdsourcing

Hello everybody out there using minix –

I’m doing a (free) operating system (just a hobby, won’t be big and professional like gnu) for 386(486) AT clones. This has been brewing since april, and is starting to get ready. I’d like any feedback on things people like/dislike in minix, as my OS resembles it somewhat (same physical layout of the file-system (due to practical reasons) among other things).

I’ve currently ported bash(1.08) and gcc(1.40), and things seem to work. This implies that I’ll get something practical within a few months, and I’d like to know what features most people would want. Any suggestions are welcome, but I won’t promise I’ll implement them

Linus (torvalds@kruuna.helsinki.fi)

PS. Yes – it’s free of any minix code, and it has a multi-threaded fs. It is NOT portable (uses 386 task switching etc), and it probably never will support anything other than AT-harddisks, as that’s all I have :-(.
Crowdsourcing

“We say that a system is a [crowdsourcing] system if it enlists a crowd of humans to help solve a problem defined by the system owners”

Human Computation

“a paradigm for utilizing human processing power to solve problems that computers cannot yet solve.”

Human Computation

• Human computation:
  • The problems fit the general paradigm of computation, and as such might someday be solvable by computers.
  • The human participation is directed by the computational system or process

“Human Computation: A Survey and Taxonomy of a Growing Field”
Quinn and Bederson CHI 2011
Human Computation

Computer algorithms that call on human effort in order to solve the problems they’re given
Human Computation

Computer algorithms that call on human effort in order to solve the problems they’re given

Programs that call people as subroutines
Collective Intelligence

• “Collective intelligence is a shared or group intelligence that emerges from the collaboration and competition of many individuals and appears in consensus decision making in bacteria, animals, humans and computer networks”

Wikipedia, “Collective Intelligence”
Complementary

• Crowdsourcing: Focuses on bringing people together

• Human computation: Focuses on the algorithms that utilize people

• Collective intelligence: Focuses on the emergent behavior
Can Still Be a Computer Scientist...

... even though you’re building computer programs whose functionality depends on the work of people

- Reliability
- Efficiency
- ....
Common Application Domain: Labeling Data

• “Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers”, Victor S. Sheng, Foster Provost, Panagiotis G. Ipeirotis, *KDD 2008*

• “Cheap and Fast — But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks”, Rion Snow, Brendan O’Connor, Daniel Jurafsky, and Andrew Y. Ng, *EMNLP 2008*

• “Utility Data Annotation with Amazon Mechanical Turk”, Alexander Sorokin and David Forsyth, *IEEE Workshop on Internet Vision, CVPR 2008*
Common Application Domain: Labeling Data

Major workhorse: Majority vote

Give it to $n$ workers
Take the majority vote
# OpenSurfaces

Sean Bell, Paul Upchurch, Noah Snavely, Kaeta Bala
Cornell University

OpenSurfaces is a large database of annotated surfaces created from real-world consumer photographs. Our annotation framework draws on crowdsourcing to segment surfaces from photos, and then annotate them with rich surface properties, including material, texture and contextual information.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Materials</th>
<th>Reflectances</th>
<th>Textures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Good</td>
<td>All</td>
<td></td>
</tr>
<tr>
<td>Labeled</td>
<td>22,218</td>
<td>91,812</td>
<td></td>
</tr>
<tr>
<td>Scenes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whitebalanced Photos</td>
<td>18,199</td>
<td>26,162</td>
<td></td>
</tr>
<tr>
<td>Segmentations</td>
<td>108,082</td>
<td>155,318</td>
<td></td>
</tr>
<tr>
<td>Planar Segmentations</td>
<td>82,196</td>
<td>101,082</td>
<td></td>
</tr>
<tr>
<td>Normal Materials</td>
<td>87,991</td>
<td>104,965</td>
<td></td>
</tr>
<tr>
<td>Normal Objects</td>
<td>71,460</td>
<td>88,868</td>
<td></td>
</tr>
<tr>
<td>Reflective Textures</td>
<td>27,923</td>
<td>38,192</td>
<td></td>
</tr>
<tr>
<td>Reflectance Judgements</td>
<td>NA</td>
<td>3,153,157</td>
<td></td>
</tr>
<tr>
<td>NTLab Assignments</td>
<td>382,042</td>
<td>380,382</td>
<td></td>
</tr>
<tr>
<td>NTLab Users</td>
<td>3,174</td>
<td>3,553</td>
<td></td>
</tr>
<tr>
<td>Hours at Work</td>
<td>NA</td>
<td>10,478</td>
<td></td>
</tr>
<tr>
<td>Submitted Items</td>
<td>NA</td>
<td>12,026,283</td>
<td></td>
</tr>
</tbody>
</table>
Computer Science Question

Given an unlimited supply of unlabeled data, and you want to label each item with probability of error less than epsilon, how many items can you label within a given budget B?
“Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers”
Victor S. Sheng, Foster Provost, Panagiotis G. Ipeirotis
Can Still Be a Computer Scientist...

... even though you’re building computer programs whose functionality depends on the work of people

- Reliability
- Efficiency
- ....
Can Still Be a Computer Scientist...

... and think about people
Financial Incentives and the “Performance of Crowds”
Winter Mason and Duncan J. Watts
2009 KDD Human Computation Workshop
To select a word, first click on the first letter of the word, then click on the last letter of the word. If you are correct, it will turn red and the word will appear to the right of the puzzle.

For each puzzle you will see a set of possible words and their category. Not all of the words listed are in the puzzle! In addition, the number of words in each puzzle changes. The list of possible words follows: ACHIEVE, ATTAIN, BUILDING, CHAIR, COMPETE, GREEN, LAMP, MASTER, MUSIC, PLANT, STAPLE, STEREO, STRIVE, SUCCEED, TURTLE.

For this practice puzzle, you will have to find at least 8 words to continue.

**RANDOM WORDS**

MASTERMOZQKKWLF
RGTRDBUILDINGOT
DSUCUWRJBMQPLC
LGPETEMPOMCAEPF
KZTVCFTBXWQVCAI
SULFJDGWUWICGO
WCEOMPMLRESUZL
XVRQXONTLNNFNW
YHBAINSPLICNEF
EHLDTAVAWISCXU
NKGLTTEFYBEEMCM
OPRPBSALIPJERNX
UVBAZJ MIDHFDGUV
QPIAWTFUNUCHGISX
YKIOOTWLORNQAQWA
Financial Incentives and the “Performance of Crowds”
Winter Mason and Duncan J. Watts
KDD Hcomp Workshop 2009

• More pay for a task
  • Means you get more workers
  • Does not mean you get more accurate answers

• Pay rate sets the human computation algorithm’s “clock rate” – how fast you want things done
Financial Incentives and the “Performance of Crowds”
Winter Mason and Duncan J. Watts
*KDD Hcomp Workshop 2009*

• More pay for a task
  • Means you get more workers
  • Does *not* mean you get more accurate answers

• Pay rate sets the human computation algorithm’s “clock rate” – how fast you want things done

• If you pay on a quota system, performance improves
Honesty in an Online Labor Market
Siddharth Suri, Daniel G. Goldstein, and Winter A. Mason
In *Proceedings of the Third Human Computation Workshop, 2011*
Honesty in an Online Labor Market
Siddharth Suri, Daniel G. Goldstein, and Winter A. Mason
In *Proceedings of the Third Human Computation Workshop*, 2011
Class Mechanics

• Instructor: Prof. Haym Hirsh, cs.cornell.edu/~hirsh
• TA: Eric Wang, ericwang0701@gmail.com, ericwang.info
• Course website: cs.cornell.edu/courses/cs5306
• Discussion: piazza.com/cornell/fall2017/cs5306
• Coursework submission: gradescope.com (entry code 9ZRR3V)
• Exams:
  • Prelim (tentative): 10/17/2017, in class
  • Final: 12/12/2017, 9:00AM - 11:30AM
Required Background

“knowledge of basic computer science principles and skills (such as CS 1110, CS 1114, CS 2110, CS 3110, or equivalent)”

You should feel comfortable doing some programming:
- Project 1: Data analysis
- Project 2: Human computation system or analysis
Coursework

• Prelim: 15-20%
• Final: 15-20%
• Assignments (~5): 10-20%
• Projects (2): 45-55%
• Extra credit: Used if you are near the boundary between grades
• READINGS!
Readings

• *Infotopia: How Many Minds Produce Knowledge*, Cass Sunstein

• Technical papers 2-4 per lecture
Assignments

• Readings for Tuesday:
  • "They saw a game; a case study." Albert H. Hastorf and Hadley Cantril, *The Journal of Abnormal and Social Psychology* 49, no. 1, 1954
    [http://www2.psych.ubc.ca/~schaller/Psyc590Readings/Hastorf1954.pdf](http://www2.psych.ubc.ca/~schaller/Psyc590Readings/Hastorf1954.pdf)
  • "Opinions and social pressure”, Solomon E. Asch, *Scientific American* Nov 1955
Assignment 0

Setting up accounts on Amazon Mechanical Turk and Crowdflower
(to be posted on the course website)