Crowdsourcing

CS329: Computational Linguistics
02/15/2016
Crowdsourcing

What is “Crowdsourcing”?  
- Process of obtaining needed services, ideas, or content by soliciting contributions from a large group of people.

Why is it important?  
- Similar concept to parallel processing.  
- Variety of opinions, cross validation, etc.

How does it relate to Computational Linguistics?  
- Provide annotated data for building models.  
- Intuitive way of obtaining natural language data.
Annotation

Provides semantic/contextual meaning of the raw data

<table>
<thead>
<tr>
<th>ID</th>
<th>Word-form</th>
<th>POS tag</th>
<th>Extra features</th>
<th>Head ID (Malt)</th>
<th>Dependency label (Malt)</th>
<th>Head ID (Stanford)</th>
<th>Dependency label (Stanford)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I</td>
<td>PRP</td>
<td></td>
<td>nsubj</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>would</td>
<td>MD</td>
<td></td>
<td>ROOT</td>
<td>3</td>
<td>aux</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>like</td>
<td>VB</td>
<td></td>
<td></td>
<td>2</td>
<td>VC</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>to</td>
<td>TO</td>
<td></td>
<td></td>
<td>5</td>
<td>VMOD</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>issue</td>
<td>VB</td>
<td></td>
<td></td>
<td>3</td>
<td>VMOD</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>a</td>
<td>DT</td>
<td></td>
<td></td>
<td>7</td>
<td>NMOD</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>challenge</td>
<td>NN</td>
<td></td>
<td>OBJ</td>
<td>5</td>
<td>dobj</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>to</td>
<td>TO</td>
<td></td>
<td></td>
<td>5</td>
<td>VMOD</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>corporate</td>
<td>JJ</td>
<td></td>
<td></td>
<td>10</td>
<td>NMOD</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>America</td>
<td>NNP</td>
<td></td>
<td>PMOD</td>
<td>8</td>
<td>pobj</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>.</td>
<td>P</td>
<td></td>
<td></td>
<td>2</td>
<td>PUNCT</td>
<td></td>
</tr>
</tbody>
</table>
Active Learning

Human validation on semi-annotated data

- Teach machine to learn from mistakes.
- Mark correct or incorrect predictions to adjust model weights.
Character Mining

1. Identify Speaker(Character) in multi-party conversations

2. Knowledge graph construction

3. Character analysis & predictions

Data: Scripts from TV shows (Friends, Seinfeld, Big Bang Theory, etc)

Amazon Mechanical Turk
Amazon Mechanical Turk

A market place for trading “human intelligence” and “time”

- Categorization
- Data Collection
- Moderation of an Image
- Sentiment

- Survey
- Survey Link
- Tagging of an Image
- Transcription from A/V

- Transcription from an Image
- Writing
- Other

---

**Make Money by working on HITs**

HITs - Human Intelligence Tasks - are individual tasks that you work on. [Find HITs now.](#)

**As a Mechanical Turk Worker you:**
- Can work from home
- Choose your own work hours
- Get paid for doing good work

**Find an interesting task**

**Work**

**Earn money**

---

**Get Results from Mechanical Turk Workers**

Ask workers to complete HITs - Human Intelligence Tasks - and get results using Mechanical Turk. [Get Started.](#)

**As a Mechanical Turk Requester you:**
- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you’re satisfied with the results

**Fund your account**

**Load your tasks**

**Get results**

---

Henry Chen
Work Flow

- Task Description
- Template Design
- Data Preparation
- Task Submission
- Result Analysis
- Data Formulation

Batch Iteration
Task Description

- What is the end goal of desired annotation?
  - Identify the corresponding speaker of referents.

- What does the input/output is going to look like?
  - Referent identified => Selected speaker from a given list.

- How much time will the task take for individuals?
  - 8-10min (Should pay $6-7 to the turkers).

- Does the data need to be validated?
  - Cross validation from 2-3 turkers depends on the agreement.
Instructions

The following are the instructions for this task:

- Potential referent to specific character(s) is highlighted with a superscript to the top-right of it indicating the number of the question.
- Please select the name of the character(s) the referent is referring to for the questions.
- For questions that says "DO NOT ANSWER", simply ignore the question since it serves as a placeholder.
- Please select "Collective(You)", if the referent "you" refers to multiple characters in the scene.
- Please select "Unknown", if you are not sure or cannot find the character(s) option the referent(s) is referring to.
- Please select "Unknown" for "you", if "you" refers to nobody or address general cases.
  *i.e. Still, "you" - "you" say Minnie, "you" hear Mouse14.*
- Please select "Error", if the referent does not refer to a person or people.

Friends Season 1 Episode 3 Scene 1

Phoebe: Hi guys!
All: Hey, Pheeb1! Hi!
Ross: Hey. Oh, oh, how'd it go?
Phoebe: Uh, not so good. He2 walked me to the subway and said 'We should do this again!'
All: Ohh. Ouch.
Rachel: What? He3 said 'we should do it again', that's good, right?
Monica: Uh, no. Loosely translated 'We should do this again' means 'You4 will never see me naked'.
Rachel: Since when?
Joey: Since always. It's like dating language. Y'know, like 'It's not you5', means 'It is you6'.
Chandler: Or 'You7' 're such a nice guy8 means 'I'm gonna be dating leather-wearing alcoholics and complaining about them to you9'.
Phoebe: Or, or, y'know, um, 'I think we should see other people' means 'Ha, ha, I already am'.
Batch Iterations

Result Analysis

- Clarify instructions
- Optimize task duration
- Optimize reward
- Redesign template

Does the result make sense?
Was the task flow efficient?

Testing with Smaller Batches

20 hits → 25 hits → 50 hits → All hits
Evaluation

“Simple” Agreement Score

\[ p_0 = \frac{\text{# of agreement}}{\text{# of total assignment}} \]

Cohen’s Kappa Coefficient

- A more robust measure.
- Take in account of agreement by chance.

\[ \kappa = \frac{p_0 - p_e}{1 - p_e} \]

\[ p_e = \sum_{i=1}^{n} \prod_{j=1}^{k} \frac{a_{i,j}}{\sum a_{i,j}} \]

A

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>5</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>15</td>
</tr>
</tbody>
</table>

\[ p_0 = \frac{35}{50} = 0.7 \]

\[ P(A, Yes) = \frac{30}{50} = 0.6, P(A, No) = \frac{20}{50} = 0.4 \]

\[ P(B, Yes) = P(B, No) = \frac{25}{50} = 0.5 \]

\[ p_e = (.6 \times .5) + (.4 \times .5) = 0.5 \]

\[ \kappa = \frac{0.7 - 0.5}{1 - 0.5} = 0.4 \]

Henry Chen
Rachel (to Monica):

... I know that you\textsuperscript{54} and I have kinda drifted apart, but you\textsuperscript{55} 're the only person\textsuperscript{56} I knew who lived here in the city.

... that you and
... but you 're
... only person
... O O O O

... how much Barry\textsuperscript{52} looks like Mr. Potato Head\textsuperscript{53}.

... much Barry looks
... Potato Head
... B-Barry / Mr. Potato Head
... B-Barry L-Barry

BILOU
B: Begin
I: Inside
L: Last
O: Outside
U: Unit