Supersense and Sensibility
Proxy Tasks for Semantic Annotation of Prepositions

Luke Gessler, Shira Wein, and Nathan Schneider
Georgetown University

LAW XIV at COLING’2020
Prepositional Supersense

A coarse-grained label describing a preposition’s meaning

(1) I rented an apartment in Locus Boston
(2) I hope to see you in Time the future
(3) It’s gone downhill since a change in Characteristic ownership
Annotating prepositional supersenses

- ...is currently **hard** and **expensive**
Annotating prepositional supersenses

- ...is currently **hard** and **expensive**

- **SNACS v2.5:**
  - 50 supersenses
  - 100-page manual
Crowdsourcing prepositional supersenses

- Direct supersense annotation is too hard for crowdworkers
Crowdsourcing prepositional supersenses

- Direct supersense annotation is **too hard** for crowdworkers

- But: could we have crowdworkers solve a **simpler task**, and **infer** a full supersense label?
Crowdsourcing prepositional supersenses

- 2 proxy task designs
  - Intended for implementation on Amazon Mechanical Turk
  - Amazon MT already explored for crowdsourcing senses for content-words

- 4 pilot studies with a handful (5-7) of graduate student crowdworkers

- Goal: assess whether designs can work under ideal conditions
Design 1: Preposition Substitution

Step 1: crowdworkers generate substitute phrases

Great spot to kick back for a cup of joe and a snack

with to get having
Design 1: Preposition Substitution

Step 1: crowdworkers *generate* substitute phrases

**Clean rooms, great for the price and cheapest on the exit**

- consider
- given
- with respect to
Design 1: Preposition Substitution

Step 2: crowdworkers select substitute phrases

Clean rooms, great for the price and cheapest on the exit

But otherwise, it can feel pricey for what you get.

- with
- to get
- having
- considering
- given
- with respect to
- Other:
Design 1: Preposition Substitution

Step 2: Crowdworkers select substitute phrases
Design 1: Preposition Substitution

Result: a distribution across substitutes, given a preposition as used in a sentence

But otherwise, it can feel pricey for what you get
Design 1: Preposition Substitution

Result: a distribution across substitutes, given a preposition as used in a sentence which we can use for supersense inference.

But otherwise, it can feel pricey for what you get.
Design 1: Preposition Substitution

Distributions for in aggregated by gold supersenses
Design 1: Preposition Substitution

Distributions for **in** aggregated by gold supersenses

**in** in **Locus** paraphrased as **at** 8 times, **to** 4 times, ...
Design 1: Preposition Substitution

Distributions for in aggregated by gold supersenses

\(\text{in}_{\text{Time}}\) paraphrased as \textit{during} 8 times, \textit{around} 2 times, ...
Design 2: Neighbor Selection

Step 1: given a **target instance**

I am in love with a giant plate of nachos!
Design 2: Neighbor Selection

Step 1: given a target instance, retrieve similar instances from a gold corpus

I am in love with the giant plate of nachos!

- I am now more at peace and my food craving is about 99% gone
- They are very rude over the phone and in person
- They worked around the clock to ensure that my puppy life was saved.
Design 2: Neighbor Selection

Step 1: given a **target instance**, retrieve **similar instances** from a **gold corpus** (by taking **supersense tag probability vectors** for the target instance and gold-tagged instances and ranking them)

I am in love with the giant plate of nachos!
- I am now more at peace and my food craving is about 99% gone
- They are very rude over the phone and in person
- They worked around the clock to ensure that my puppy life was saved.

Cosine Similarity Ranking

\[
\frac{A \cdot B}{\|A\| \|B\|}
\]

[...; p(Time|s); p(Locus|s); ...]
Design 2: Neighbor Selection

Step 2: Crowdworkers select an appropriate neighbor, the majority winner’s gold tag is used for the target instance
Design 2: Neighbor Selection

Step 2: Crowdworkers select an appropriate neighbor, the majority winner’s gold tag is used for the target instance

- I am in love with the giant plate of nachos!
I am in love with the giant plate of nachos!

- I am in love with the giant plate of nachos!

**Design 2: Neighbor Selection**

Step 2: Crowdworkers select an appropriate neighbor, the majority winner’s gold tag is used for the target instance.
Design 2: Neighbor Selection

Two questions
- What if no neighbor has a supersense that matches the target?

| I am in ?? (Characteristic|Locus) love with the giant plate of nachos! |
|---------------------------------------------------------------|
| - Again, a great outing for the kids, a frustration for an out\_Locus of town climber |
| - They are very rude over the phone and in\_Circumstance|Locus person |
| - They worked around\_Time the clock to ensure that my puppy life was saved. |
| - None |
Design 2: Neighbor Selection

Two questions
- What if no neighbor has a supersense that matches the target?
  → crowd should select none

I am in love with the giant plate of nachos!

- Again, a great outing for the kids, a frustration for an out of town climber
- They are very rude over the phone and in person
- They worked around the clock to ensure that my puppy life was saved.
- None
Design 2: Neighbor Selection

Two questions
- What if no neighbor has a supersense that matches the target?
  → crowd should select none

- Is the quality of the neighbors seriously affected if the predicted tag for the target is wrong?

I am in Circumstance|Locus love with the giant plate of nachos!
- I am now more at Characteristic|Locus peace and my food craving is about 99% gone
- They are very rude over the phone and in Circumstance|Locus person
- They worked around Time the clock to ensure that my puppy life was saved.
- None
Design 2: Neighbor Selection

Two questions

- What if no neighbor has a supersense that matches the target?
  → crowd should select none

- Is the quality of the neighbors seriously affected if the predicted tag for the target is wrong?
  → hopefully, a neighbor with the appropriate tag will still be retrieved

I am in Circumstance|Locus love with the giant plate of nachos!

- I am now more at Characteristic|Locus peace and my food craving is about 99% gone
- They are very rude over the phone and in Circumstance|Locus person
- They worked around Time the clock to ensure that my puppy life was saved.
- None
Design 2: Neighbor Selection

<table>
<thead>
<tr>
<th>Case</th>
<th>Tagger</th>
<th>Crowd</th>
<th>“None”</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Tagger correct, gold present)</td>
<td>17/17</td>
<td>17/17</td>
<td>0/17</td>
</tr>
<tr>
<td>2 (Tagger incorrect, gold present)</td>
<td>0/12</td>
<td>6/12</td>
<td>5/12</td>
</tr>
<tr>
<td>3 (Tagger correct, gold absent)</td>
<td>3/3</td>
<td>0/3</td>
<td>2/3</td>
</tr>
<tr>
<td>4 (Tagger incorrect, gold absent)</td>
<td>0/8</td>
<td>0/8</td>
<td>5/8</td>
</tr>
</tbody>
</table>

Pilot study

- 40 instances, 5 workers
- Divide instances for analysis by (1) tag correctness, (2) presence of the target instance’s gold tag among neighbors
Design 2: Neighbor Selection

<table>
<thead>
<tr>
<th>Case</th>
<th>Tagger</th>
<th>Crowd</th>
<th>“None”</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Tagger correct, gold present)</td>
<td>17/17</td>
<td>17/17</td>
<td>0/17</td>
</tr>
<tr>
<td>2 (Tagger incorrect, gold present)</td>
<td>0/12</td>
<td>6/12</td>
<td>5/12</td>
</tr>
<tr>
<td>3 (Tagger correct, gold absent)</td>
<td>3/3</td>
<td>0/3</td>
<td>2/3</td>
</tr>
<tr>
<td>4 (Tagger incorrect, gold absent)</td>
<td>0/8</td>
<td>0/8</td>
<td>5/8</td>
</tr>
</tbody>
</table>

Pilot study

- What if no neighbor has a supersense that matches the target?
  → crowd often selects none

- Is the quality of the neighbors seriously affected if the predicted tag for the target is wrong?
  → appropriate neighbors are still retrieved sometimes
Conclusion

- Under **ideal conditions**, can we obtain **labels** from **proxy tasks**, by having crowdworkers solve an easy task, and then inferring the labels from their output?
  - Yes!

- Future work
  - **Implementation** on a platform like Amazon Mechanical Turk
  - Different **ranking strategies** for the neighbor selection design
  - Explore **tradeoffs** between the two designs
  - Application to **different annotation tasks**