SOCIAL METAPHOR DETECTION VIA TOPICAL ANALYSIS
Verb has **selectional preferences** to its arguments.

Can you **eat money**?
Could we capture *metaphors* in social media by selectional preference?

If *not*, why not?
Could *topic model* help?

If *yes*, how?
Is it for *verb only*?
Outline

- Selectional Preference
- 3-Step Framework
  1. Pre-processing
  2. Modeling & Detection
  3. Post-processing
- Topical Analysis
- Experiment & Result
- Conclusion
Selectional Preference

- Selectional Association (SA) (Resnik, 1997)

\[
A_R(p, c) = \frac{1}{S_R(p)} \Pr(c|p) \log \frac{\Pr(c|p)}{\Pr(c)}
\]

\[
S_R(p, c) = D(\Pr(c|p)||\Pr(c))
\]

\[
= \sum_c \Pr(c|p) \log \frac{\Pr(c|p)}{\Pr(c)}
\]

p: predicate
c: noun class
3-step Framework

Pre-processing -
Word Extraction &
Noun Clustering

Modeling & Detection -
SA Outlier Detection

Post-processing -
SA Strength Filter
Step 1: Pre-processing (1)

- Word Extraction
  - Why?
    - Parsing & POS tagging is hard on noisy data
  - How?
    - Using lemma form
    - Set minimal term frequency
    - Set minimal “POS rate”
      - Proportion of occurrence of certain POS
    - Predicates should be more strict than the nouns
      - Noun: TF > 5, POS rate >= 0.7
      - Verb & Adj: TF > 50, POS rate >= 0.8
Step 1: Pre-processing (2)

- Semantic Noun Clustering

**Top 100 Similar Nouns**

**money:**
1. funds
2. cash
3. profits
4. millions
5. monies
6. dollars
7. royalties
...

**Weighted Directed Graph for Nouns**
Step 2: Modeling & Detection

- **Selectional Association**

![Diagram](image)

**Another Candidate Semantic Outlier Word Detection**

- “Semantic Coherence” outlier (Inkpen et al., 2005)
- Based on pair-wise word semantic similarity
- Very High False Positive
  - The influences of “general words”
  - Semantic similarity is not reliable
Step 3: Post-processing

- **SA Strength Filtering** (Shutova, et al., 2010)
  - **SA Strength**
    - Strong (e.g., filmmake)
    - Weak (e.g., “light verb”, put, take, ...)
  - Predicates with weak selectional preference barely “violates” their own preference.

\[
S_R(p, c) = D(\Pr(c|p) \| \Pr(c)) \\
= \sum_c \Pr(c|p) \log \frac{\Pr(c|p)}{\Pr(c)}
\]
Data

- Online breast cancer support community
- All the public posts from Oct 2001 to Jan 2011.
- 90,242 unique users who posted 1,562,459 messages belonging to 68,158 discussion threads. (Wang, et al., 2012; Wen, et al., 2013)
Experiment Setting

Pre-processing -
- Stanford NLP/Parser
- 55k nouns, 3k adjs, and 1.8k verbs

Modeling & Detection -
- 3 deps: nsubj, dobj, amod
- Observe negative pairs

Post-processing -
- Follow (Shutova, et al., 2010)

Topical Model: JGibbLDA, 20 topics (k = 20)
Most outliers are **NOT** metaphors

- **Parsing Error**
  “...yearly breast MRI...”: amod(breast, yearly)

- **Non-metaphor**
  “...cancer cells float around in my blood...”: dobj(float, cancer)

- **Metonymy**
  “If John win tomorrow night, ...”: dobj(win, tomorrow)

Only **very few** metaphors are identified

- “...keep my head occupied ...”: nsubj(occupy, head)
- “... my belly has overtaken the boobs ...”: nsubj(overtake, belly)

- Topic model does **NOT** help much
Could we capture **metaphors** in social media by selectional preference?

**Maybe not by fully-automatic approaches.**

If **not**, why **not**? Could **topic model** help?

**Good parsing** is challenging on social media. Outliers of SA are **not always** metaphors. **Topic modeling does not help much.**

If **yes**, how? Is it for **verb only**?

**Maybe seed-expansion method works better.** **No**, it could also work for **amod** dependency.
Thanks!

- **Acknowledgement**
  
  Zi Yang, Prof. Teruko Mitamura, Prof. Eric Nyberg for academic supports, and Yi-Chia Wang, Dong Nguyen for data collection.

  Supported by the Intelligence Advanced Re-search Projects Activity (IARPA) via Department of Defense US Army Research Laboratory contract number W911NF-12-C-0020.

- **Main References**
  


License

- Except where otherwise noted, content on this slide is licensed under a Creative Commons. Attribution-ShareAlike 3.0 Unported License.
- Material used
  - Background image of slide 2: uxd@Flickr, “Plate of money”, CC BY 2.0
    http://www.flickr.com/photos/uxud/3205640524/
  - Background image of slide 3, 15: jasonahowie@Flickr, “Social Media apps”, CC BY 2.0
    http://www.flickr.com/photos/jasonahowie/8583949219/
  - Background image of slide 12: susangkomenforthecure@Flickr, “Susan G. Komen walkers gear up and take on Day 1 for breast cancer awareness.”, CC BY-NC-ND 2.0
    http://www.flickr.com/photos/susangkomenforthecure/9623480334/
  - Background image of slide 16: jam_project@Flickr, “IMG_0405 [2011-08-23]”, CC BY-NC-ND 2.0
    http://www.flickr.com/photos/jam_project/6075715482/