Creative Language Processing: Metaphors

Casey R. Kennington

June 8, 2010

Paper

- Comprehending and Generating Apt Metaphors: A Web-driven, Case-based Approach to Figurative Language
- Tony Veale, Yanfen Hao
- 2007, Association for the Advancement of Artificial Intelligence

Metaphors and Similies

- Similes: T is as P as [*a*|*an*] V
- Example: John is as tall as a tree.
- P is shared by T and V, but also P is a salient property of V
- Explicit similes are the low hanging fruit of figurative language, and are easily identifiable
- Similes use bridge words like "as" or "like"

Metaphors and Similies

- Similes: T is as P as [*a*|*an*] V
- Example: John is as tall as a tree.
- P is shared by T and V, but also P is a salient property of V
- Explicit similes are the low hanging fruit of figurative language, and are easily identifiable
- Similes use bridge words like "as" or "like"
- Metaphors tend to be more subtle (no "as is")

Similes

- ...as hard as nails
- ...as pure as snow
- ...as silly as a goose
- ...as straight as an arrow
- ...time flies like an arrow

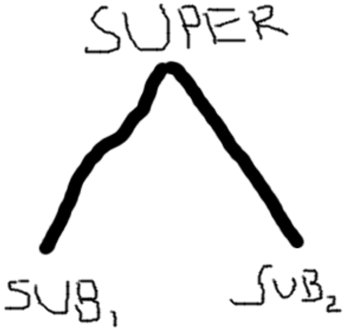
Similes

- ...as hard as nails
- ...as pure as snow
- ...as silly as a goose
- ...as straight as an arrow
- ...time flies like an arrow
- Question: what are some approaches to finding similes?

Taxonomical Approach

- Taxonomy: a way of classification (typically using a supertype)
- Example: cigarettes are like time bombs
- Problem: symmetry
- time bombs are like cigarettes?
- if something has the same supertype, then they should work in either order

Type Hierarchy



Structural Approach

- Stucture-Mapping Theory (Falkenhainer et al 1989)
- uses semantic structures as a process of graph alignment
- map between systematic elements; mapping across domains
- ignore surface features and and find matches based on the structure of representation
- Example: a pen is like a sponge (both can dispense liquid) (Wikipedia)
- use connected knowledge over independant facts

Sardonicus

- Neither fully taxonomic or structural, but is compatible with both
- Similar to the MIDAS approach, but looks more at common similes
- Goal: automatically find a simile
- later use these data to generate other similes or metaphors

Using Google

- Sardonicus uses Google to retrieve similes from the web
- Use wildcards *
- Example: * is as a *
- Keep ones with form: as ADJ as *a*|*an* N
- Gather a large database of similes (representative sample)

Using Google Continued

- Use a list of ADJ from WordNet
- Example: “cold” or “hot”
- Query Google for: * as ADJ as *
- get top 200 results
- Ascertain which noun values around the ADJ
- Further search: as * as a N
- Ascertain common adjectives around N
- Idea: obtain many examples for each ADJ and N

Results

- Set of 74,704 simile instances
- with 42,618 unique similes
- 3769 different adjectives
- 9286 different nouns

Cleaning the Data

- Some similes had NP values
- Checked against WordNet as lexical unit
- Example: “gang of thieves” is a lexical unit
- Throw out the others

Annotation

- Some similes were ironic
- Example: as hairy as a bowling ball
- difficult to automate (as they are creative)
- A human judge annotated 30,991 similes
- 12,259 as non-ironic
- 4,685 as ironic
- can further extend knowledgebase using antonyms, hyponyms, and synonyms with WordNet
- can now be used to help Sardonicus determine ironic or bona-fide similes

Comprehension

- With the data, Sardonicus can determine salient properties
- Example: funeral
- sad, orderly, unfortunate, dignified, solemn, serious
- Example: wedding
- joyous, joyful, decisive, glorious, expensive, emotional

Comprehension Continued

- Similes are not categorizations, but comparisons
- Consider the metaphor: weddings are funerals
- Consider also: funerals are weddings
- Sardonicus determined that the former was legitimate (funeral-like wedding), while the latter (wedding-like funeral) was either not valid or wholly original
- See previous slide to see why
- Checked against Google

Generation

- The number of possibilities of N and ADJ is very large
- huge search space, unwanted metaphors
- goal-driven where user picks tenor and a property of the tenor

Generation

- The number of possibilities of N and ADJ is very large
- huge search space, unwanted metaphors
- goal-driven where user picks tenor and a property of the tenor
- Example: novel (to Sardonicus) noun: Paris Hilton with tenor "skinny"
- results: post, pole, stick, miser, stick insect
- "Paris Hilton is a pole"
- pole: straight, skinny, thin, slim, stiff, scrawny

Limitations (and upsides)

- Limit: cannot abstract more than what Google can find
- Upside: resulting interpretations are well adapted to their targets
- Sardonicus can employ abstraction using WordNet
- As long as web expands, so can Sardonicus

Dynamic Learning

- Unique nouns are no big deal because it can look on the web
- Example: Atlantis is a myth
- Query for: Atlantis is a *
- Query for: * is a myth (if not already known)
- Find properties for myth:
- religious(3), famous(3), strong(3), heroic(2), improbable(2), timeless(1), historical(1), innacurate(1)
- Adapt to tenor Atlantis (Atlantis is a myth):
- famous(1283), strong(178), historical(93), religious(10), inaccurate(6), timeless(5), heroic(5), improbable(3)
- Whereas “Herucles is a myth” shows prominance for strong(295) and heroic(140), etc

Evaluation Metric

- Use a metric that associates certain positive or negative feelings, values, or ideas
- Whissel (1989) produced a “dictionary of affect”
- 8,000 words were given a numeric value between 1.0 and 3.0 (most pleasant)
- Use Whissel score for ADJs, find weighted average, then predict the N score, compare to the Whissel score
- tall as a tree (trees are tall, green, leafy, strong, old, etc)

Data Sets

- A. Only bona-fide similes
- B. All similes
- C. Only ironic similes
- D. All ADJ used for a specific N (from corpus)
- E. All ADJ used for a specific N (from WordNet)

Results

- A (bona-fide only). highest correlation (+0.514)
- C (ironic only). lowest correlation (-0.243)
- B (together). middling (0.347) which shows 4 to 1 non-ironic/ironic ratio
- D (corpus ADJ). 0.15
- E (WordNet ADJ). 0.278

Concluding Remarks

- Web is a vast resource for Sardonicus
- Sardonicus has limits, but can grow as long as it can use the web
- Only 3.6 percent of WordNet glosses with ADJ N associations (as strong as espresso) had examples on the web
- WordNet may not have the properties of how people actually think of, and use certain words and categories
- Could have other uses (MT, parsing)?

Metaphor Modeling

- *A Fluid Knowledge Representation for Understanding and Generating Creative Metaphors*
- Tony Veale, Yanfen Hao (2008)
- Metapor modeling requires semantically accomodating representation
- They present *Talking Points*, a flexible knolwedge representation

Slipnet

- Metaphors can be viewed as a stretching of linguistics convention to cover new conceptual ground
- Hofstadter and Mitchell (1994) introduces *slipnet*
- slipnet: a probabilistic network in which concepts are linked to others into which they can slip or be substituted with

Slipnet Example

- Governor of California
- = governor of 12 percent of U.S.
- = leader of 12 percent of U.S.
- = president of 12 percent of U.S.
- = president of 100 percent of U.S.
- president of 100 percent of U.S.

Talking Points

- Use WordNet and Google
- Use specific patterns:
- ADJ+ N
- talking point becomes: isADJ:N
- example: isTall:tree, composes:music(composer)
- change right or left side of colon, build statistical slip stream

Empirical Evaluation

- Use clustering
- Find simile examples on web, check against slip stream
- as * as the....
- yields 90.2 percent accuracy (ex1, 214 nouns)
- yields 69.85 percent accuracy (ex2, 402 nouns)

Thanks

- Thank you!