Brighter than Gold: Figurative Language in User Generated Comparisons

Vlad Niculae
MPI-SWS
Cornell University

Cristian Danescu-Niculescu-Mizil
MPI-SWS
Cornell University
Comparisons

Assert likeness of two things;
Point out shared properties.
Aren’t always meant to be taken literally!

“Sterling is much cheaper than gold.”

“Her voice makes the song shine brighter than gold.”

Literal comparison •

Figurative comparison (simile) •
This talk:
What makes comparisons figurative?

"Sterling is much cheaper than gold."

"Her voice makes the song shine brighter than gold."

Literal comparison •

Figurative comparison (simile) •
Figurative language

Metaphor, simile, etc.

Most research is on metaphor.

- Conceptual interpretation (Shutova et al., 2013)
- Identification
  tree kernels (Hovy et al., 2013)
  conceptual features (Tsvetkov et al., 2014)

Less (computational) research on simile.

- Irony, humor (Veale et al., 2012)
Role of comparisons

• Describe new things in terms of given things.
  “A unicorn is like a horse, but it has a horn.”

• Rank with respect to a property.
  “The cord is more durable than the adapter.”

Applications in Knowledge Extraction
(Tandon et al., 2014, Lofi et al., 2014, inter alia)
Assumption: comparisons are literal.

But can we tell if a comparison is literal?
our main insight:

For figurative language, context is important!

social, topical...
applying our insight:

Does **topic** information differentiate literal and figurative comparisons?

Can we take advantage of the **structure** of the comparison?

How does **social context** drive figurative language use?
Data we analyze: Amazon.com product reviews

Rich in comparisons need to describe products

Rich in **social context** helpfulness, star ratings

**Topical context** comes for free product categories
Figurative Comparison Dataset

1400 sentences with marked comparisons.

Figurative scores from three Amazon Mechanical Turk annotators.

Freely available: http://vene.ro/figurative-comparisons/
Characterizing figurative comparisons
Topical context

Is figurativeness more likely in certain product categories?
Topical context

Is figurative word choice topic-specific?

The song feels like a boring melody (music)

Her book is like a sweet melody (books)

specificity(word, topic) = \frac{\#(word appears in topic)}{\#(word appears in all topics)}
Topical context
Is figurative word choice topic-specific?

specificity(word, topic) =

\[
\frac{\text{#(word appears in topic)}}{\text{#(word appears in all topics)}}
\]

\(96\%\)

\(3\%\)

\(song: \text{melody}\) (music)

\(book: \text{melody}\) (books)
Comparison structure

Similarity between the things being compared.

The song shines brighter than gold.
Sterling is cheaper than gold.

*cosine similarity in a vector space model
Metaphor features
Inspired by Tsvetkov et al. (2014)

- Concreteness
  - kiss vs. happiness

- Imageability
  - cinnamon vs. casque

- Supersenses: coarse categories
  - annotate: \{communication, creation\}

- Vector space embeddings
  - Inspired by Tsvetkov et al. (2014)
Are these insights predictive?
Classification (logistic regression)

Linguistic insight features:

Topic-aware
- Review category
- Word specificity

Topic-agnostic
- Structural similarity
- Metaphor features (concreteness, imageability, supersenses, embeddings)
- Definiteness
## Classification (logistic regression)

<table>
<thead>
<tr>
<th>method</th>
<th>acc.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>bag-of-words</td>
<td>80%</td>
<td>75%</td>
</tr>
</tbody>
</table>

Why so strong?

- “the hero is truly larger than life”
- “the headphones sounded like crap”
## Classification (logistic regression)

<table>
<thead>
<tr>
<th>method</th>
<th>acc.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>bag-of-words</td>
<td>80%</td>
<td>75%</td>
</tr>
<tr>
<td>linguistic insight</td>
<td>86%</td>
<td>79%</td>
</tr>
</tbody>
</table>

Top features:

1. **Structural** similarity
   
   “the older **man** was wiser than the **boy**”

2. **Topic** specificity
   
   “the cord’s more durable than the **adapter**”

3. Imageability
   
   “the explanations are as clear as **mud**”
Classification (logistic regression)

<table>
<thead>
<tr>
<th>method</th>
<th>acc.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>bag-of-words</td>
<td>80%</td>
<td>75%</td>
</tr>
<tr>
<td>linguistic insight</td>
<td>86%</td>
<td>79%</td>
</tr>
<tr>
<td>both</td>
<td>88%</td>
<td>82%</td>
</tr>
</tbody>
</table>

The two approaches are complementary.

(p < 0.05)
## Classification (logistic regression)

<table>
<thead>
<tr>
<th>method</th>
<th>acc.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>bag-of-words</td>
<td>80%</td>
<td>75%</td>
</tr>
<tr>
<td>linguistic insight</td>
<td>86%</td>
<td>79%</td>
</tr>
<tr>
<td>both</td>
<td>88%</td>
<td>82%</td>
</tr>
<tr>
<td>human</td>
<td>96%</td>
<td>94%</td>
</tr>
</tbody>
</table>

- Optimistic estimation:
- Majority voting among three Mechanical Turk workers
Why not just use a metaphor system?

(p < 0.05)
Practical importance

Verified Purchase (What's this?)
This review is from: Sony MDR-NC6 Noise Canceling Headphones (Discontinued by Manufacturer)
Initially the headphones sounded like crap (no bass and no high end,
Figurative language and social context
Social context of reviews

Verified Purchase (What's this?)
This review is from: Sony MDR-NC6 Noise Canceling Headphones (Discontinued by Manufacturer)
Initially the headphones sounded like crap (no bass and no high end, no middle range, it's just a flat sound)

8 of 10 people found the following review helpful

🌟🌟🌟🌟 Not bad for the price, December 18, 2004
Intrinsic opinion

![Graph showing the ratio of figurative comparisons against star ratings. The graph indicates a significant decrease in ratio from 1 to 3 stars, followed by an increase from 3 to 5 stars. The p-value is less than 0.05.](image)
Extrinsic opinion
(how the review is perceived)

(p < 0.05)
Conclusions

Compiled a **dataset** of figurative & literal comparisons.

Figurative language is tied to a **social** context.

**Topic** information is relevant for figurativeness.

Similes benefit from specific computational treatment.

“He uses statistics as a drunken man uses lamp-posts... for support rather than illumination.”

— Andrew Lang
“Getting an education was a bit like a communicable sexual disease. It made you unsuitable for a lot of jobs and then you had the urge to pass it on.”

— Terry Pratchett, Hogfather
Extra slides
<table>
<thead>
<tr>
<th>method</th>
<th># feat</th>
<th>acc.</th>
<th>P</th>
<th>R</th>
<th>F</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>bag-of-words</td>
<td>1840</td>
<td>.80</td>
<td>.64</td>
<td>.90</td>
<td>.75</td>
<td>.89</td>
</tr>
<tr>
<td>metaphor-inspired</td>
<td>345</td>
<td>.75</td>
<td>.60</td>
<td>.72</td>
<td>.65</td>
<td>.84</td>
</tr>
<tr>
<td>domain-specific</td>
<td>8</td>
<td>.69</td>
<td>.51</td>
<td>.81</td>
<td>.63</td>
<td>.76</td>
</tr>
<tr>
<td>all insight features</td>
<td>365</td>
<td>.86</td>
<td>.76</td>
<td>.83</td>
<td>.79</td>
<td>.92</td>
</tr>
<tr>
<td>all of the above</td>
<td>2202</td>
<td>.88</td>
<td>.80</td>
<td>.84</td>
<td>.82</td>
<td>.94</td>
</tr>
<tr>
<td>human</td>
<td></td>
<td>.96</td>
<td>.92</td>
<td>.96</td>
<td>.94</td>
<td></td>
</tr>
</tbody>
</table>

Training set size = test set size
(408 comparisons, 66% literal)
Matching comparisons

The [song Topic] [shines Event] [brighter Property] [than Comparator] [gold Vehicle]

3 types of patterns:

- T is like V
- T is as P as V
- T is P-er than V
Data collection

Amazon.com reviews

extract comparisons

- - -

Validate (3x MTurk)

Annotate 1-4 (3x MTurk)

consensus, binarize

- -

-(3, 4, 4)
-(4, 2, 1)
-(1, 2, 1)
Concreteness and imageability

<table>
<thead>
<tr>
<th></th>
<th>more imageable</th>
<th>less imageable</th>
</tr>
</thead>
<tbody>
<tr>
<td>more concrete</td>
<td><em>cinnamon,</em></td>
<td><em>casque,</em></td>
</tr>
<tr>
<td></td>
<td><em>kiss</em></td>
<td><em>pugilist</em></td>
</tr>
<tr>
<td>less concrete</td>
<td><em>devil,</em></td>
<td><em>aspect,</em></td>
</tr>
<tr>
<td></td>
<td><em>happiness</em></td>
<td><em>however</em></td>
</tr>
</tbody>
</table>

Concreteness and imageability (Tsvetkov et al., 2014 for metaphor)
Concreteness, imageability: MRC Psycholinguistic Database

Supersenses:
  WordNet for nouns, verbs
  Tsvetkov et al. (2014) for adjectives

Logistic regression on word embeddings.

(Tsvetkov et al., 2014 for metaphor)
Metaphor vs. Simile


Metaphors that can’t be rephrased as similes:

• “Her argument was somewhat murky.”
• “The house had great bones.”

Similes that can’t be rephrased as metaphors:

• “The windshield wipers made a great clatter like two idiots clapping in church.”
• “Two bears sat facing each other like two matrons having tea.”