Disambiguation of Idiomatic Expressions with Images

Katsumi Ibaraki
Motivation

- **Emergence of CLIP**
  - Contrastive Language–Image Pre-training
  - Neural network that learns visual concepts from natural language supervision
  - Originally for visual classification
Uses of CLIP
- DALL-E and Stable Diffusion
  - Generate digital images from natural language descriptions

Issues
- Disambiguation of entities
  - e.g. financial bank vs river banks

DALL-E 2: An astronaut riding a horse in photorealistic style.
“Break a leg”
Explore approaches to generate images that express the figurative meaning
- Compare natural language-based approaches with text-image-based approaches

Generate images for 10 English idioms
Methodology

- **Baseline and Fine-tuned model**
  - Stable Diffusion model
  - Stable Diffusion model fine-tuned on other idioms

- **Rephrasing the expressions**
  - GPT-3.5
  - Convert idiom into simpler expressions
Experiments

▸ Generations
  ▷ Each model given the 10 idioms to generate 20 images

▸ Evaluation
  ▷ 2 independent evaluators asked to judge:
    ▷ Likely literal or likely figurative
## Results

<table>
<thead>
<tr>
<th>Idiomatic Expression</th>
<th>Untuned Literal%</th>
<th>Untuned Figurative%</th>
<th>Finetuned Literal%</th>
<th>Finetuned Figurative%</th>
<th>GPT-Reprrased Literal%</th>
<th>GPT-Reprrased Figurative%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bite the bullet</td>
<td>1.00</td>
<td>0.00</td>
<td>0.90</td>
<td>0.10</td>
<td>0.15</td>
<td>0.85</td>
</tr>
<tr>
<td>Break a leg</td>
<td>0.95</td>
<td>0.05</td>
<td>0.80</td>
<td>0.20</td>
<td>0.10</td>
<td>0.90</td>
</tr>
<tr>
<td>Break the ice</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Get out of hand</td>
<td>0.95</td>
<td>0.05</td>
<td>0.90</td>
<td>0.10</td>
<td>0.05</td>
<td>0.95</td>
</tr>
<tr>
<td>Hit the sack</td>
<td>1.00</td>
<td>0.00</td>
<td>0.85</td>
<td>0.15</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>On the ball</td>
<td>0.80</td>
<td>0.20</td>
<td>0.70</td>
<td>0.30</td>
<td>0.10</td>
<td>0.90</td>
</tr>
<tr>
<td>Pull yourself together</td>
<td>0.95</td>
<td>0.05</td>
<td>0.85</td>
<td>0.15</td>
<td>0.20</td>
<td>0.80</td>
</tr>
<tr>
<td>To get bent out of shape</td>
<td>0.90</td>
<td>0.10</td>
<td>0.85</td>
<td>0.15</td>
<td>0.10</td>
<td>0.90</td>
</tr>
<tr>
<td>Under the weather</td>
<td>1.00</td>
<td>0.00</td>
<td>0.75</td>
<td>0.25</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Wrap your head around something</td>
<td>1.00</td>
<td>0.00</td>
<td>0.90</td>
<td>0.10</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>0.955</strong></td>
<td><strong>0.045</strong></td>
<td><strong>0.850</strong></td>
<td><strong>0.150</strong></td>
<td><strong>0.070</strong></td>
<td><strong>0.930</strong></td>
</tr>
</tbody>
</table>
“Break a leg”

(a) Untuned

(b) Fine-tuned

(c) Rephrased
Discussion

- Effects of fine-tuning
  - “Under the weather” (sick)
Discussion

Future work

- Combine rephrasing with fine-tuning
- Compare DALL-E and Stable Diffusion
- Extend past idioms
  - Provide two similar expressions and differentiate by images