Learning Prototypical Functions for Physical Artifacts

Tianyu Jiang and Ellen Riloff
University of Utah
ACL-IJCNLP, August 2021
Humans create things for a reason.

- spears for hunting
- knives for cutting
- pots for cooking
Introduction

- We refer to the most typical way a physical object is often being used as its **prototypical function**.

- The prototypical function of human-made physical artifacts is a kind of **commonsense knowledge** that often plays a role in natural language understanding.
Examples

a) He killed the mayor with a *gun*.

b) He killed the mayor with a *knife*.

c) He killed the mayor with a *bomb*.
Examples

a) He killed the mayor with a **gun**.

b) He killed the mayor with a **knife**.

c) He killed the mayor with a **bomb**.
It is essential for NLP systems to “read between the lines”.

The goal of this work is to learn the prototypical functions of human-made physical artifacts.

We define a new NLP task to associate physical artifacts with frames from FrameNet as the representation for the function.
Dataset Creation

01 Artifact Selection

02 Function Representation

03 Human Annotation
Artifact Selection

- We focus on artifacts that are
  - Physical objects
  - Created by people

- WordNet synset “artifact.n.01”

- Concreteness dictionary \[^1\]
  - We apply a threshold >= 4.5

- Intersection produces a set of 1,017 concrete physical artifacts.
Function Representation

- FrameNet 1.7 contains 1,221 frame definitions.

Cooking_creation
Def. This frame describes food and meal preparation...
LU. bake.v, cook.v, grill.v, ...

Protecting
Def. Some protection prevents a danger from harming an asset.
LU. guard.v, protect.v, ...

- Not all frames are suitable for representing functions of physical artifacts, e.g.,
  - Abstract nominals: Calendric_unit
  - High-level abstractions: Intentionally_act
  - Misc: Judgement

- We manually selected 42 frames in FrameNet
### Human Annotation

<table>
<thead>
<tr>
<th>Frames</th>
<th>Artifact examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wearing</td>
<td>hat, shirt</td>
</tr>
<tr>
<td>Containing</td>
<td>basket, luggage</td>
</tr>
<tr>
<td>Self_motion</td>
<td>bicycle, yacht</td>
</tr>
<tr>
<td>Protecting</td>
<td>armor, helmet</td>
</tr>
<tr>
<td>Supporting</td>
<td>chair, scaffolding</td>
</tr>
<tr>
<td>Cause_harm</td>
<td>cannon, spear</td>
</tr>
<tr>
<td>Perception_exp</td>
<td>earphone, eyeglass</td>
</tr>
<tr>
<td>Make_noise</td>
<td>bell, violin</td>
</tr>
<tr>
<td>Cause_motion</td>
<td>engine, propeller</td>
</tr>
<tr>
<td>Cutting</td>
<td>knife, scissors</td>
</tr>
</tbody>
</table>

**Introduction**

**Dataset**

**Methods**

**Evaluation**

**Conclusion**
Task Definition

- Multiclass classification

- Denote
  - artifacts as $a_i (i = 1..m)$,
  - frames as $f_j (j = 1..n)$,

- The task is to select $f_j$ that represents the prototypical function for a given $a_i$. 
Baselines

- For each artifact in Conceptnet [2], we extract the first word under **UsedFor** relation.

- For COMET, we use its beam-10 setting to generate 10 phrases of the UsedFor relation for each artifact.

- Use the extracted words to rank candidate frames through *lexical units*.
Baselines

- We use 3 dependency patterns to extract <N, V> pairs from Wikipedia.

- 3.8 million <N, V> pairs for our 938 artifacts
Baselines

- Use masked language model [4] to get prediction score for every artifact and lexical unit pair.

- 6 sentence templates:
  
  ___ can be used to ___.
  I used ___ to ___.
  ___ can be used for ___.
  I used ___ for ___.
  The purpose of ___ is to ___.
  If I had ___ , I could ___.

1. Existing knowledge base - ConceptNet and COMET
2. Co-occurrence information
3. Masked language model (MLM)
Instead of using the predicted word for the mask token, we retrieve the last hidden state vector of the [MASK] token.

Then pass it through a linear layer for classification.
Learning from Definitions - PF_{def}

- Each green block represents an artifact $a_i$ paired with one of the candidate frames.
- Model produces a probability distribution over all candidate frames.
Combining $P_{\text{mask}}$ and $P_{\text{def}}$

- Left is output of $P_{\text{def}}$, a matrix of dimension (# of frames, hidden vector size)
- Right is output of $P_{\text{mask}}$, a matrix of dimension (1, hidden vector size)
# Experimental Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Acc</th>
<th>Pre</th>
<th>Rec</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConceptNet</td>
<td>17.5</td>
<td>33.6</td>
<td>13.5</td>
<td>16.4</td>
</tr>
<tr>
<td>Co-occurrence</td>
<td>31.9</td>
<td>24.1</td>
<td>23.9</td>
<td>19.9</td>
</tr>
<tr>
<td>COMET</td>
<td>30.7</td>
<td>29.7</td>
<td>35.6</td>
<td>28.2</td>
</tr>
<tr>
<td>MLM</td>
<td>42.8</td>
<td>29.5</td>
<td>33.8</td>
<td>28.2</td>
</tr>
</tbody>
</table>
## Experimental Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Acc</th>
<th>Pre</th>
<th>Rec</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConceptNet</td>
<td>17.5</td>
<td>33.6</td>
<td>13.5</td>
<td>16.4</td>
</tr>
<tr>
<td>Co-occurrence</td>
<td>31.9</td>
<td>24.1</td>
<td>23.9</td>
<td>19.9</td>
</tr>
<tr>
<td>COMET</td>
<td>30.7</td>
<td>29.7</td>
<td>35.6</td>
<td>28.2</td>
</tr>
<tr>
<td><strong>MLM</strong></td>
<td><strong>42.8</strong></td>
<td><strong>29.5</strong></td>
<td><strong>33.8</strong></td>
<td><strong>28.2</strong></td>
</tr>
<tr>
<td>$\text{PF}_{\text{mask}}$</td>
<td>58.5</td>
<td>35.7</td>
<td>36.5</td>
<td>35.4</td>
</tr>
<tr>
<td>$\text{PF}_{\text{def}}$</td>
<td>74.7</td>
<td>63.5</td>
<td>57.6</td>
<td>59.3</td>
</tr>
<tr>
<td>$\text{PF}_{\text{def}+\text{mask}}$</td>
<td><strong>76.8</strong></td>
<td><strong>65.2</strong></td>
<td><strong>61.1</strong></td>
<td><strong>62.4</strong></td>
</tr>
</tbody>
</table>
# Experimental Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Acc</th>
<th>Pre</th>
<th>Rec</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConceptNet</td>
<td>17.5</td>
<td>33.6</td>
<td>13.5</td>
<td>16.4</td>
</tr>
<tr>
<td>Co-occurrence</td>
<td>31.9</td>
<td>24.1</td>
<td>23.9</td>
<td>19.9</td>
</tr>
<tr>
<td>COMET</td>
<td>30.7</td>
<td>29.7</td>
<td>35.6</td>
<td>28.2</td>
</tr>
<tr>
<td>MLM</td>
<td>42.8</td>
<td>29.5</td>
<td>33.8</td>
<td>28.2</td>
</tr>
<tr>
<td>(PF_{mask})</td>
<td>58.5</td>
<td>35.7</td>
<td>36.5</td>
<td>35.4</td>
</tr>
<tr>
<td>(PF_{def})</td>
<td>74.7</td>
<td>63.5</td>
<td>57.6</td>
<td>59.3</td>
</tr>
<tr>
<td>(PF_{def+mask})</td>
<td><strong>76.8</strong></td>
<td><strong>65.2</strong></td>
<td><strong>61.1</strong></td>
<td><strong>62.4</strong></td>
</tr>
</tbody>
</table>
Conclusion

- We introduced the new task of learning prototypical functions for human-made physical artifacts.

- A manually annotated data set.

- We built a transformer-based model using both artifact and frame definitions as well as masked pattern predictions, that can learn the function of a given artifact.

- We will release our data at [https://github.com/tyjiangU/physical_artifacts_function](https://github.com/tyjiangU/physical_artifacts_function)
THANKS!

Do you have any questions?
Reference


