iDiR Department of Computer Science and Engineering
The University of Texas at Arlington

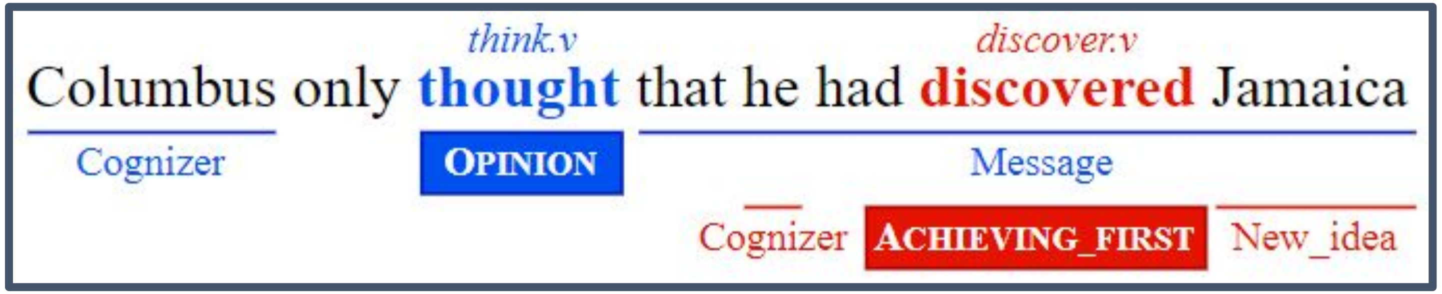
UNIVERSITY OF TEXAS ARLINGTON

Robust Frame-Semantic Models with Lexical Unit Trees and Negative Samples

Jacob Devasier,
Yogesh Gurjar,
Chengkai Li

Frame-Semantic Parsing

- 3 primary tasks
  - **Target** Identification - The identification of words which evoke **FRAMES** (**targets**)
    - Targets are instances of *lexical units*, a unique pairing of a word and meaning, e.g., *think.v*
  - **FRAME** Identification - The identification of the **FRAME** evoked by a target
    - Structures which describe a particular situation, object, or event
    - There can be many frames in a sentence, e.g., **OPINION** and **ACHIEVING_FIRST**
  - Argument Identification - Identification of a **FRAME**'s frame elements (FEs)
    - Each frame has its own set of FEs, e.g., Cognizer and Message



Target Identification

Predicting whether each word evokes a frame using a statistical/learned model does not work (1200+ classes, limited data), instead…

- **Candidate Generation** - Lexical Unit Tree
  - Frames are only evoked by certain words
  - We can limit search space for targets to words that ***can*** evoke frames (*candidate targets*)

- **Candidate Target Filtering** - RoBERTa binary-classification model
  - Filter out candidate targets that do not evoke a frame (false positives)
  - Previous systems used manually filters based on linguistic features

Target Identification - Candidate Generation

iDiR | Department of Computer Science and Engineering
The University of Texas at Arlington

- Frames can only be evoked by certain words
- Previous works enumerate all morphological variants of all lexical units in FrameNet
  - i.e., searching a given sentence for: “find”, “found”, “finds”, “finding”, “discover”, …

OPINION: *opinion.n, view.n, take.n, feel.v, figure.v, think.v, …*

ACHIEVING_FIRST: *discovery.n, inventor.n, pioneer.n, find.v, discover.v, originate.v, pioneer.v, invent.v, pioneering.a, …*

IDIR
Department of Computer Science and Engineering
The University of Texas at Arlington

Target Identification - Candidate Generation

- Our method works *bottom-up* by identifying lexical units a given word can belong to
  - Example: “I held a stray cat”
    - Bottom-up: “held” -> hold -> *hold.v*

- Built a ***Lexical Unit Tree*** for efficient search and enablement of part-of-speech wildcard search for disjoint lexical units
  - Wildcards used for multi-word verb LUs
  - Supported wildcards:
    - Pronouns, proper nouns, nouns, conjunctions, determiners

Single-word LU:

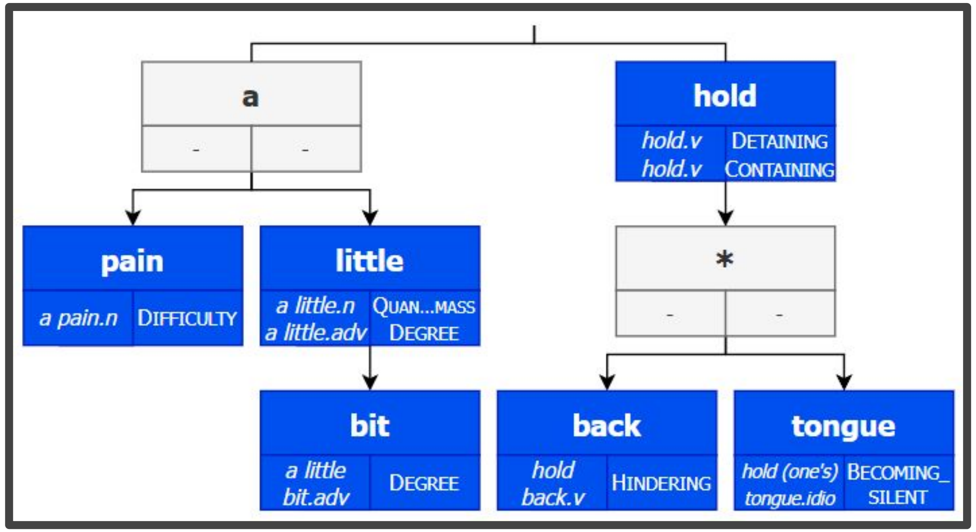
hold.v

Multi-word LU:

a little bit.adv

Disjoint LU:

hold (one’s) tongue.v



Frame Identification - Challenges

- About half of the frames do not have annotated sentences

- Lexicon filtering - Targets can only evoke certain frames
  - For example, “think” can only evoke 4 frames
  - Reduces possible frames by over **99%**, from roughly 1,220 to 1.29 (population) or 2.12 (sample)

- “Trivial” targets only evoke a single frame, e.g., “abuse”
  - Multiple-choice models cannot learn from targets with 1 possible frame (1 class)
    - We refer to these as under-utilized samples
      - e.g., “They suffered constant **abuse**”

think

Awareness

Cogitation

Opinion

Regard

discover

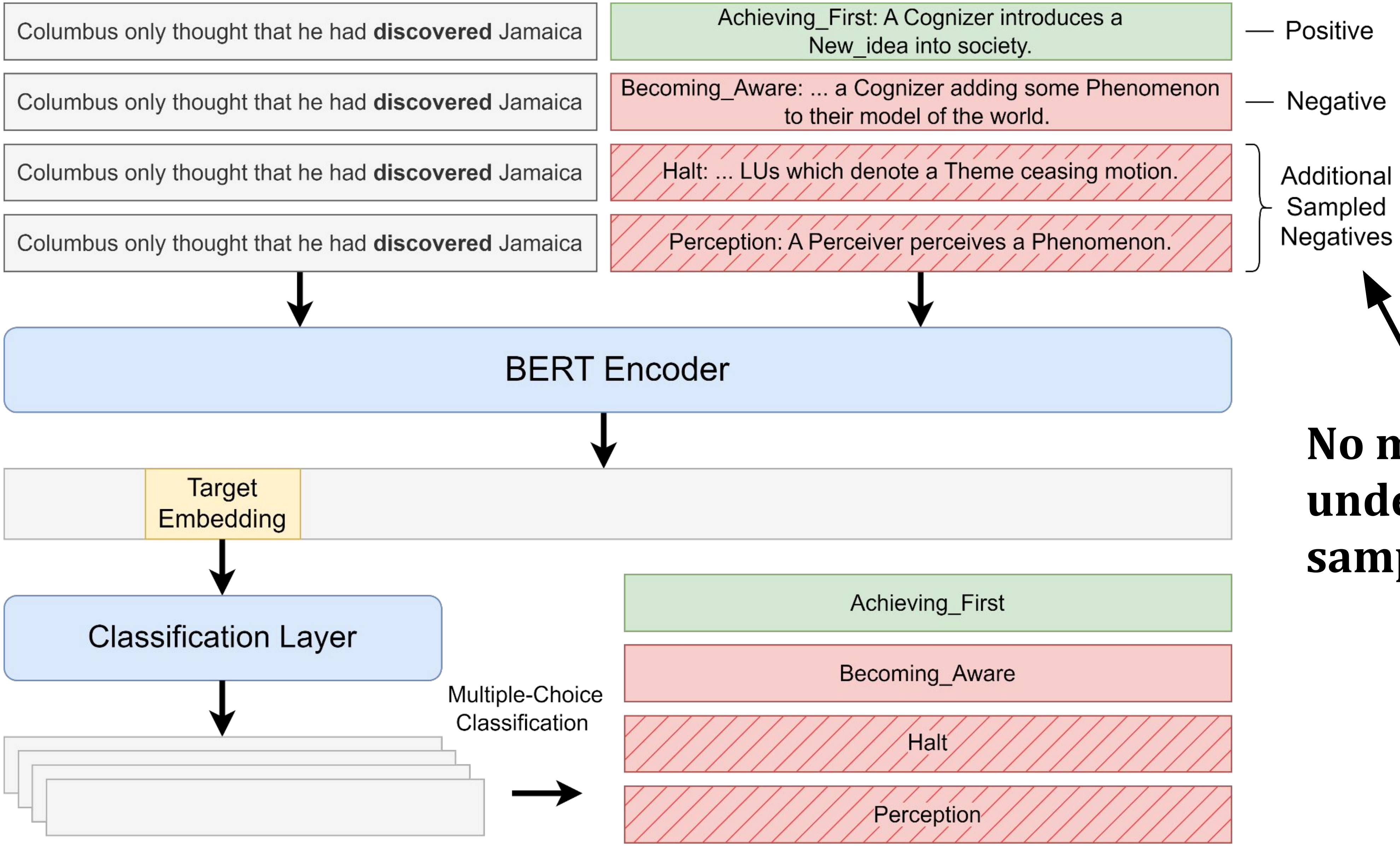
Achieving_first

Becoming_aware

abuse

Abusing

Frame Identification - Model Training



No more under-utilized samples

Experiments - Target Identification

- Q1. What is the performance of our candidate target generation algorithm?
  - Sets an upper bound on system's coverage / recall
  - Our system achieved 0.994 recall and 0.145 precision before filtering

Model	FN1.5	FN1.7
Das et al. (2014)	0.454	-
Swayamdipta et al. (2017)	0.732	0.733
Bastianelli et al. (2020)	0.768	-
Lin et al. (2021)	0.769	0.763
Our model	**0.773**	**0.775**
Our model (manually filtered)	0.388	0.392

Table 1: F1 score of our target identification model compared against previous approaches.

- Q2. Are contextualized embeddings beneficial for filtering candidate targets?
  - No previous study has evaluated the utility of contextualized embeddings on candidate filtering
  - Using contextualized embeddings improved our target identification model by +38.3% on FN1.7, improving the state-of-the-art by +1.2%

iDiR
Department of Computer Science and Engineering
The University of Texas at Arlington

Experiments - Frame Identification

- Q3. Do additional negative frames improve frame identification model performance on under-utilized samples?
  - FIDO is a system with very similar model architecture but lacks additional negative samples needed to learn from trivial targets
  - We see significant improvements on two of our benchmark datasets derived from FrameNet 1.7

Model	Test-1CF	Test-UU
FIDO (Jiang and Riloff, 2021)	0.754	0.538
Our model	**0.893**	**0.603**

Table 3: Accuracy of our frame identification model on Test-1CF, a subset of the test dataset containing targets that only have one candidate frame, and Test-UU, an augmentation of Test-1CF with additional difficult negative samples for the evaluation of multiple-choice classification models.

- Q4. Does additional negative sampling improve frame identification performance on rare frames?
  - We found that our model is better on frames with very few samples at all compared thresholds

K	# Frames	Our Model	FIDO	Δ
1	94	**0.781**	0.753	+0.028
3	235	**0.810**	0.778	+0.032
5	316	**0.853**	0.809	+0.044
10	426	**0.850**	0.826	+0.024

Table 4: Accuracy of our frame identification model compared with FIDO on frames which only appear K or fewer times in the FrameNet 1.7 training set.

Experiments - Frame Identification cont.

iDiR Department of Computer Science and Engineering
The University of Texas at Arlington

- Q5. Can our frame identification model substitute the target filtering model?
  - Candidate target filtering and Frame Identification have very similar task description and model architecture
- How does our frame identification model compare with previous works?
  - Compared with FIDO, the other semantic-based method, our system performs better
  - Graph-based methods still perform better, but lack potential benefits of semantic methods, namely performance on frames with few samples

Model	Acc	F1
Our model (candidate filter)	**0.788**	**0.775**
FIDO (Jiang and Riloff, 2021)	0.653	0.644
Our model	0.664	0.678

Model	FN1.5		FN1.7	
	All	Amb	All	Amb
Das et al. (2014)	0.836	0.692	-	-
Hermann et al. (2014)	0.887	0.737	-	-
Hartmann et al. (2017)	0.876	0.738	-	-
Yang and Mitchell (2017)	0.882	0.757	-	-
Swayamdipta et al. (2017)	0.864	-	0.866	-
Peng et al. (2018)	0.900	0.780	0.891	0.775
Bastianelli et al. (2020)	0.901	-	-	-
Lin et al. (2021)	0.906	-	0.906	-
Su et al. (2021)*	0.919	0.823	**0.924**	**0.844**
Tamburini (2022)*	**0.922**	**0.831**	0.922	0.843
Zheng et al. (2022a)	0.917	-	-	-
Jiang and Riloff (2021)	0.913	0.810	0.921	0.836
Jiang and Riloff (2021) (frame)	0.901	-	0.911	-
Our model (binary)	0.877	0.785	0.887	0.816
Our model	0.917	0.818	0.923	0.841

* Performance can not be verified due to private source code.

Future Work

- Unify target identification and frame identification
- Full frame-semantic parsing in a single step
- Applications in structured NLP
  - Initial focus on explainability in automated fact-checking and information retrieval systems