NewsMet: A 'Do It All' dataset of contemporary Metaphors in News headlines

Rohan Joseph¹, Timothy Liu², Aik Beng Ng², Simon See^{1,2}, Sunny Rai^{1,3}

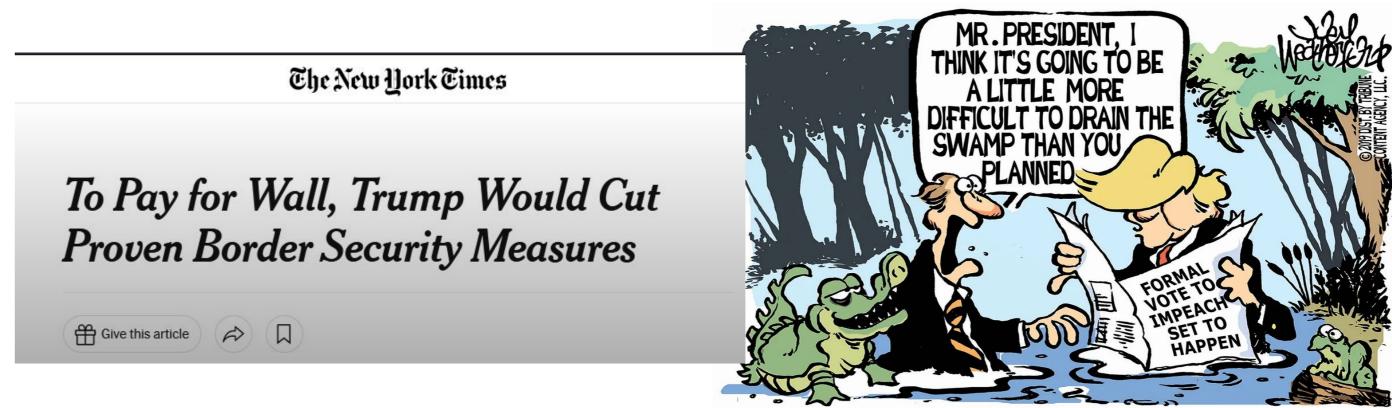
rohan18545@mechyd.ac.in

What is NewsMet?

- High quality dataset of 2669 headlines to detect, interpret and generate metaphors.
- Headlines from sources such as NYT, Rawstory and Breitbart.
- Contemporary headlines published between 2017 to 2018.
- Labelled with SVO triplet + Metaphoricity of Verb

Why is Metaphor Processing important?

- Metaphors are everywhere!
- Highly creative constructs.
- Continuously evolving usage.
 - Love **heals** soul → Amtrak
 Dining Car **heals** nation
- Culture-dependent. Want to test Multilingual LLMs? Use Metaphors!



This is a crucial time in which the killer disease is almost on verge of spiraling to an uncontrolled state like a chain reaction.

Group

Invalid

Hm & Vm

 $H_{m} \& V_{l}$

 $H_m \& (V_m || V_l)$

H₁ & V₁

7:30 AM · Apr 20, 2021 · Twitter Web App

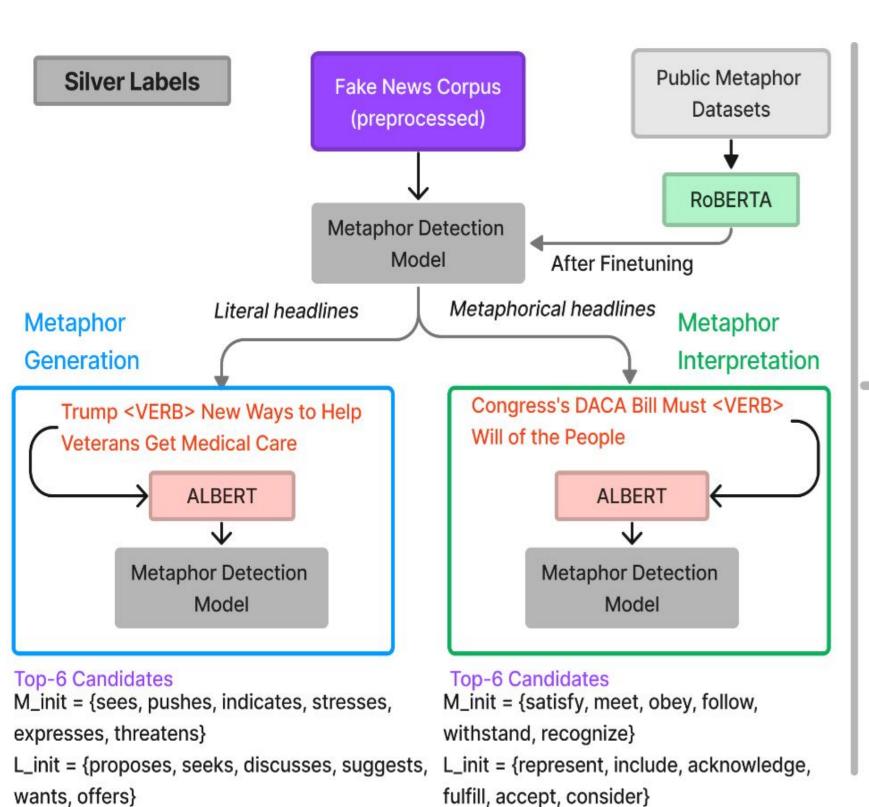


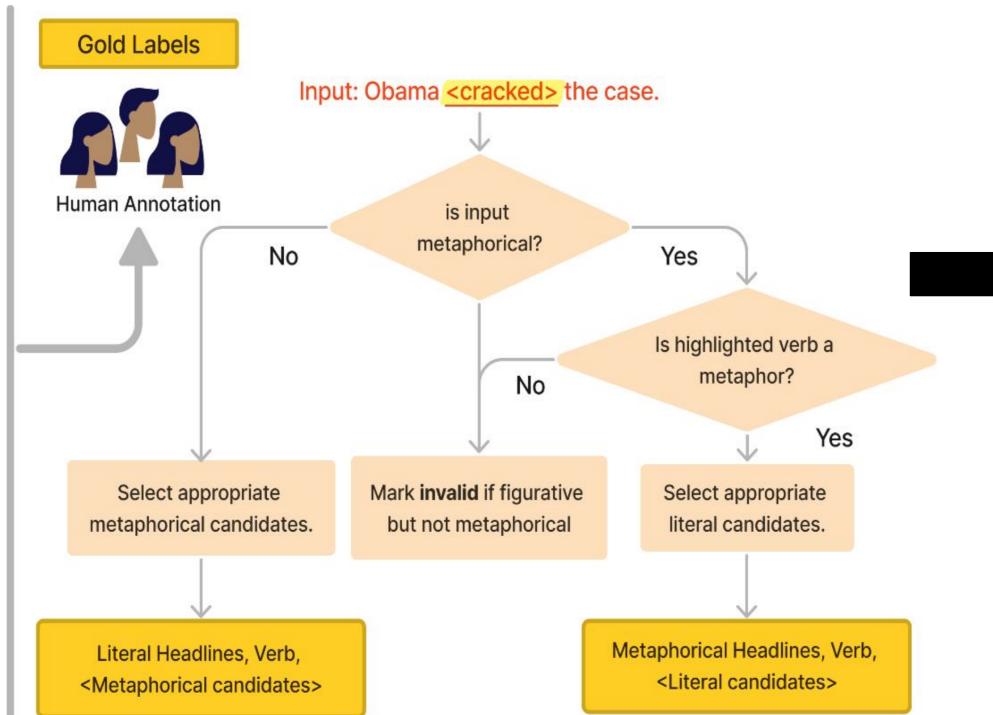
=1603

=1066

= 2669

Proposed Approach





ALBERT Metaphor Detection Model Top-6 Candidates init = {sees, pushes, indicates, stresses, presses, threatens} nit = {proposes, seeks, discusses, suggests, include, acknowledge, init = {proposes, seeks, discusses, suggests, include, acknowledge, init, soffers} ALBERT Select appropriate metaphorical candidates. Mark invalid if figurative but not metaphorical literal candidates. Metaphorical Headlines, Verb, <Metaphorical candidates> Literal Headlines, Verb, <Metaphorical candidates> Metaphorical candidates>

등 500

5 300

Number 100

200

-0.3 - 0.2 - 0.1 0

400

59.7%

literal

Type of headline

min(cos(s, c), cos(o, c))

Examples Examples Fig. 4: Union allies as

Assigned label

Candidate Set

M_init

M_final

Incorrect

Correct

72.6%

27.4%

metaphorical

0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8

Ex 1: Handling <subject> of Police Killing Spurs <metaphor> Grand Jury Inquiry <object> Into Prosecutor

D gold

314

389

205

594

611

=1205

D_{gold+}

1009

1009

455

=1464

Literal Candidates: [prompting, causing]

Ex 2: The Pneumonic Plague <subject> Killed literal verb> 50 Million People <object> in the Middle Ages and Now It's Back (VIDEO)

Metaphorical Candidates: [strangled, murdered, wiped]

Findings and Results

- Finetuned LLM model has a high (>50%) false positive rate when labelling metaphors!
- LLMs offer candidates with low metaphoricity however, semantic appropriateness is a challenge!
- LLMs are good at generating candidates from diverse SOURCE DOMAINS. 84% of metaphorical headlines with diversity score >=4.

An example for calculating *diversity*Headline (focus verb)

Headline (focus verb)	Clustersi	diversity
Far Cry 5': Coop Mission has a massive	{ confronts, faces} 1, { creates, causes} 2,	4
problem with mission progress	{finds} ₃ , {tackles} ₄	

Even a small Gold Labelled Dataset
 from new corpora improved quality
 of predictions!

Model	A	P	R	F1
M met_det	0.54	0.70	0.35	0.46
M Dimbal+ Doold+	0.61	0.68	0.57	0.62

Takeaways

- Use caution when deploying finetuned LLMs for labelling new corpora.
- Proposed metrics such as diversity and metaphoricity serve as heuristics to measure quality of generated metaphors.
- Use our Annotation pipeline to semi-automate annotating new corpora.
- When to use NewsMet? Evaluate
 Metaphor processing systems/ LLMs,
 Hyprepartisan News, Figurative text
 translation (Read Paper for more details)

Dataset available at https://github.com/AxleBlaze3/NewsMet Metaphor Dataset



¹Ecole Centrale School of Engineering, Mahindra University

²NVIDIA AI Technology Center

³Department of Computer & Information Science, University of Pennsylvania.

