THE UNIVERSITY OF CHICAGO

Metaphor Detection using LSTM and BERT

Jacob Zweifler[1] Owen Karpf[1] Wesley Chen[1]

[1]University of Chicago

Metaphors in Language

Lakoff and Johnson (1980) define *metaphor* as
"a mapping between a source and target domain".

Example

Consider the use of the words **loop** and **emerge** in the following sentences:

- Suddenly, the proprietor **emerged** from his office. (literal usage)
- Some nice results **emerged** from the study. (metaphorical usage)
- The bicycle **looped** around the tree. (??? usage)
- The stunt pilot **looped** his plane. (??? usage)

What do you think the labels should be for the last two sentences? Even for humans, the classification task is not always clear.

Why is Metaphor Detection Important in NLP?

- Essential for computers to understand human language
- Necessary for NLP tasks like translation and generative modeling to achieve good results

Why is Metaphor Detection Difficult?

- Not always clear if a word is used literally or metaphorically
- Many different kinds of mappings between concepts
- Getting annotated data is especially difficult

Matephor Masters: Our Approach

- LSTM to classify words as literal or metaphorical and to identify sequences (adapted from Gao)
- Running DistilBERT to classify each word within a sentence as literal or metaphorical (new approach)

The Data Sets

- **VUA:** Publicly available dataset of academic texts, news texts, fiction, and conversations. Hand-annotated for metaphors.
- **MOH-X:** Subset of WordNet corpus, only contains sentences with verb–direct object and verb–subject relations.
- **TroFi:** Contains samples of text containing usages of 50 english words. Samples come from The 1987-89 WSJ Corpus Release 1.
- **ElMo vectors:** Take into account context of entire sentence and crates word embeddings for specific meanings of words
- **GloVe:** Learns word embeddings based on the probabilities that words co-occur

LSTM for Classification and Sequence Tasks

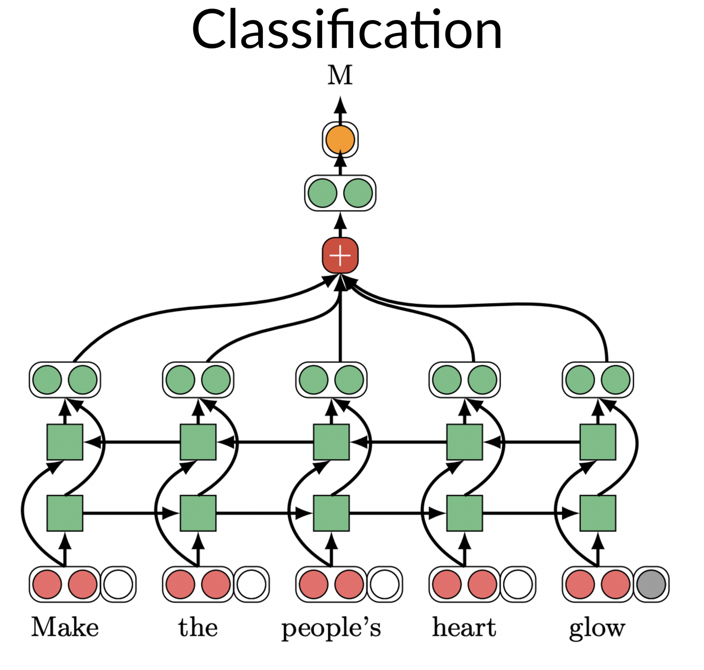

Figure 1. Example network of a model that classifies one verb per sentence. (taken from Gao et al)

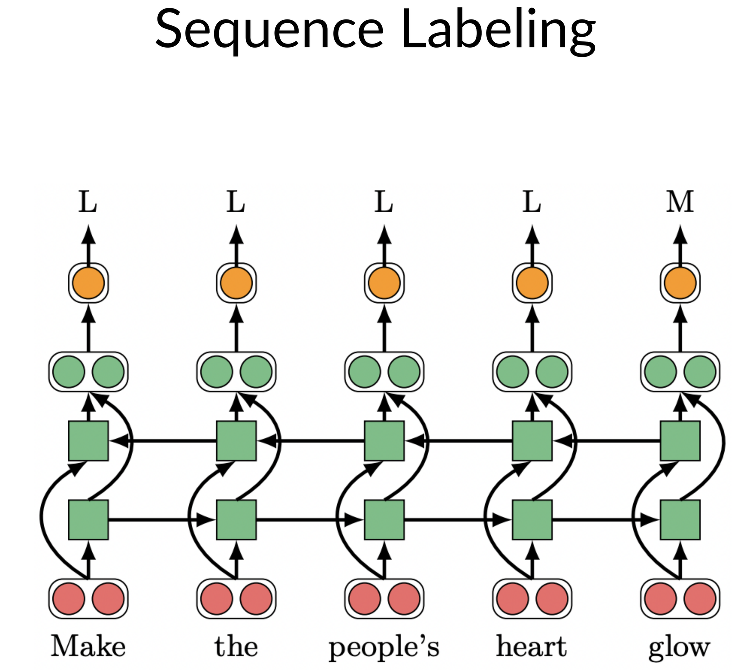

Figure 2. Example network for a model that sequentially labels words in a sentence. (taken from Gao et al)

Bidirectional Encoder Representations from Transformers (BERT)

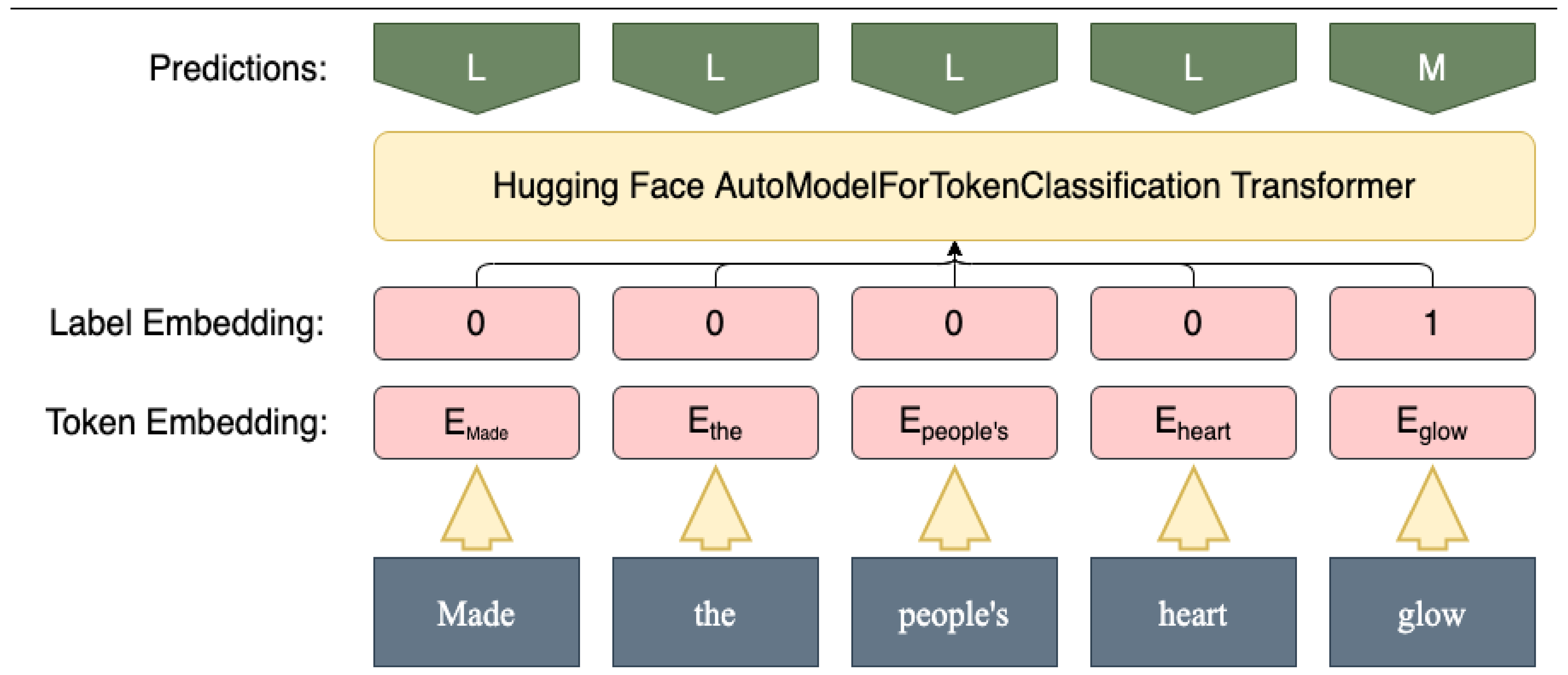

Figure 3. Architecture for the DistilBert Model using Hugging Face's automatic model for classifying tokens within a sequence

Results

Table 1. POS data for VUA Sequential with LSTM and BERT

POS	#	% metaphor	LSTM			BERT		
			P	R	F1	P	R	F1
VERB	20K	18.1	68.32	70.64	69.46	67.92	63.90	65.85
NOUN	20K	13.6	66.22	58.02	61.85	75.14	70.65	72.83
ADP	13K	28.0	87.86	88.38	88.11	63.12	60.30	61.68
ADJ	9K	11.5	63.65	58.60	61.02	61.30	50.00	55.08
PART	3K	10.1	60.53	61.75	61.13	51.95	49.59	50.74

Table 2. Performance of the Baseline, Classification, Sequence and BERT models

Model	MOH				Trofi				Vua			
	P	R	F1	Acc.	P	R	F1	Acc.	P	R	F1	Acc.
Baseline	39.09	26.70	31.29	43.56	72.39	55.74	62.92	71.42	67.90	40.72	50.91	76.45
Classification	78.46	77	77.33	77.89	68.45	76.53	72.19	74.26	72.27	61.73	66.58	81.42
Sequence	74.73	75.37	74.54	75.16	68.33	73.29	70.51	73.38	68.51	69.79	69.14	81.32
BERT (Seq)	58.39	55.41	56.86	94.90	60.73	58.30	59.49	98.89	67.77	64.38	66.03	94.12

Error analysis (VUA)

LSTM sample sentences:

(highlighted words are false positives and underlined words are false negatives)

- Other British towns , Croydon and Southampton **among** them , are also **considering** modern tramways .
- Dead , ragged **heads** of the climbing hydrangea can be removed .
- If , **on** the other **hand** , you **allow** rationality **to** children , you can't use their lack of it as criterion to distinguish them **from** adults .

BERT sample sentences:

- Other British towns , Croydon and Southampton among them , are also **considering** modern tramways .
- Dead , ragged **heads** of the climbing hydrangea can be removed .
- If , **on** the other **hand** , you allow rationality **to** children , you can't use their lack of it as criterion to distinguish them **from** adults .

Future Possibilities

Given our findings so far, here is how we can move forwards:

1. Try other BERT-based models like RoBERTa, which have been shown to perform better than BERT on nearly all, if not all, tasks
2. Look into why BERT seems to miss metaphors (lower recall) than LSTM.

References

- Julia Birke and Anoop Sarkar. A clustering approach for nearly unsupervised recognition of nonliteral language. In *11th Conference of the European Chapter of the Association for Computational Linguistics*, pages 329–336, Trento, Italy, April 2006. Association for Computational Linguistics.
- Julia Birke and Anoop Sarkar. Active learning for the identification of nonliteral language. In *Proceedings of the Workshop on Computational Approaches to Figurative Language*, pages 21–28, Rochester, New York, April 2007. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Association for Computational Linguistics, 2019.
- Erik-Lân Do Dinh and Iryna Gurevych. Token-level metaphor detection using neural networks. In *Proceedings of the Fourth Workshop on Metaphor in NLP*, pages 28–33, San Diego, California, June 2016. Association for Computational Linguistics.
- Ge Gao, Eunsol Choi, Yejin Choi, and Luke Zettlemoyer. Neural metaphor detection in context. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 607–613, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.
- Rui Mao, Chenghua Lin, and Frank Guerin. Word embedding and wordnet based metaphor identification and interpretation. In *ACL*, 2018.
- Saif Mohammad, Ekaterina Shutova, and Peter Turney. Metaphor as a medium for emotion: An empirical study. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 23–33, Berlin, Germany, August 2016. Association for Computational Linguistics.
- Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
- Anna Rogers, Olga Kovaleva, and Anna Rumshisky. A primer in BERTology: What we know about how BERT works. volume 8, pages 842–866, Cambridge, MA, 2020. MIT Press.
- Ekaterina Shutova, Douwe Kiela, and Jean Maillard. Black holes and white rabbits: Metaphor identification with visual features. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 160–170, San Diego, California, June 2016. Association for Computational Linguistics.
- Gerard Steen, Lettie Dorst, J. Herrmann, Anna Kaal, Tina Krennmayr, and Trijntje Pasma. *A method for linguistic metaphor identification: From MIP to MIPVU*. 06 2010.
- Krishnkant Swarnkar and Anil Kumar Singh. Di-LSTM contrast : A deep neural network for metaphor detection. In *Proceedings of the Workshop on Figurative Language Processing*, pages 115–120, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.