Advanced Machine Learning Project

Naive Zero-shot NER and Discussion of LUKE

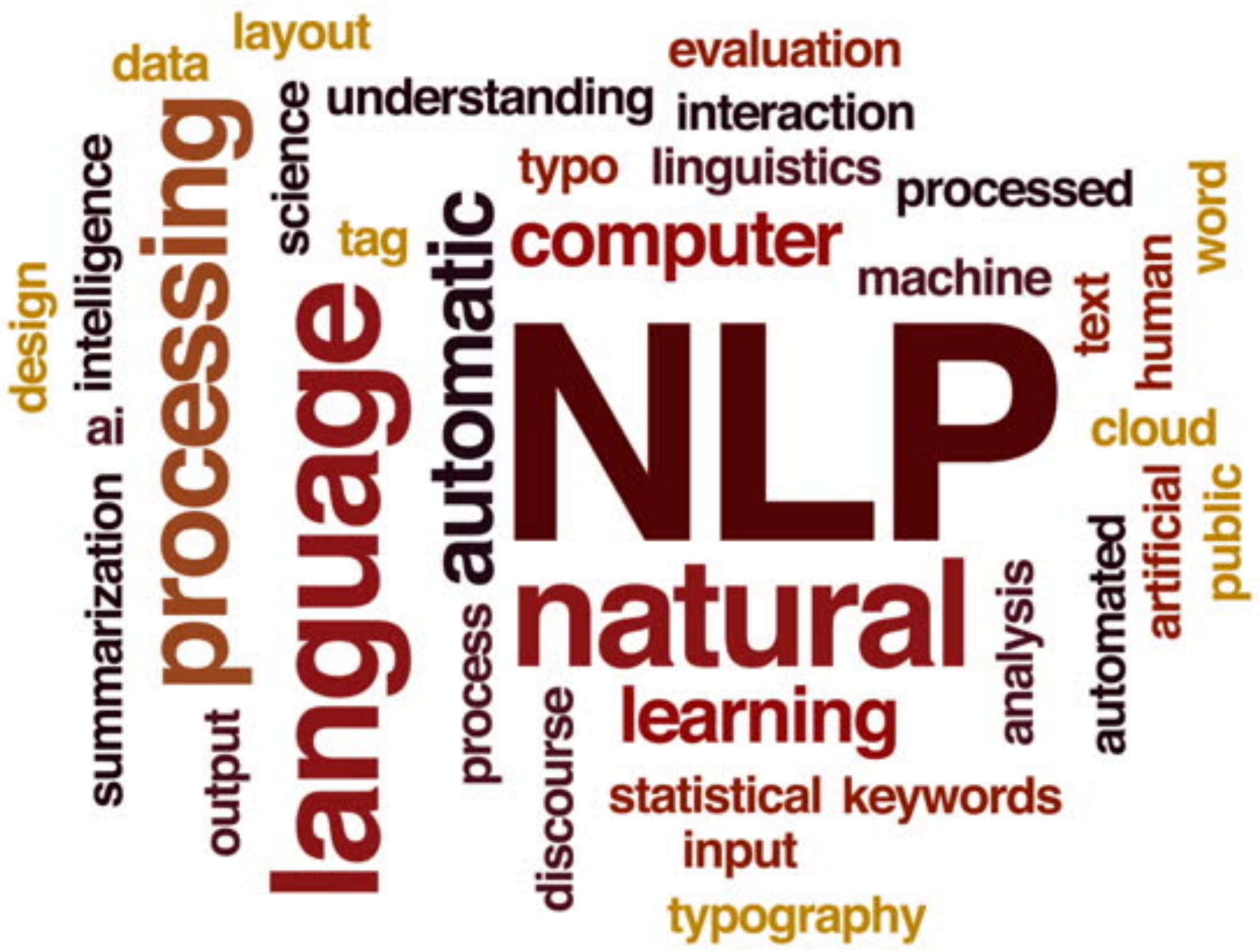
Ayan Gangopadhyay 24th February 2021

Introduction

What is NER?

Why NER is required for NLP.

- Information Retrieval.
- Question Answering systems.
- Modern Conversational Agents for slot filling.
- Feature Extraction.
- ASR systems.



Difficulties in NER tasks.

Why is NER hard?

- Free form text or unstructured text.
- Context - based entities.
- Nested Entities.
- Abbreviations and Acronyms.
- Homonyms.

CoNLL-2003.

Dataset and Task.

- One of the most popular dataset for NER benchmarking.
- Language Independence.
- Only English portion freely available.
- Drawbacks of IOB structure.

Word	POS	Syntactic Chunk	NE Tag
U.N.	NNP	I-NP	I-ORG
official	NN	I-NP	O
Ekeus	NNP	I-NP	I-PER
heads	VBZ	I-VP	O
for	IN	I-PP	O
Baghdad	NNP	I-NP	I-LOC
.	.	.	O

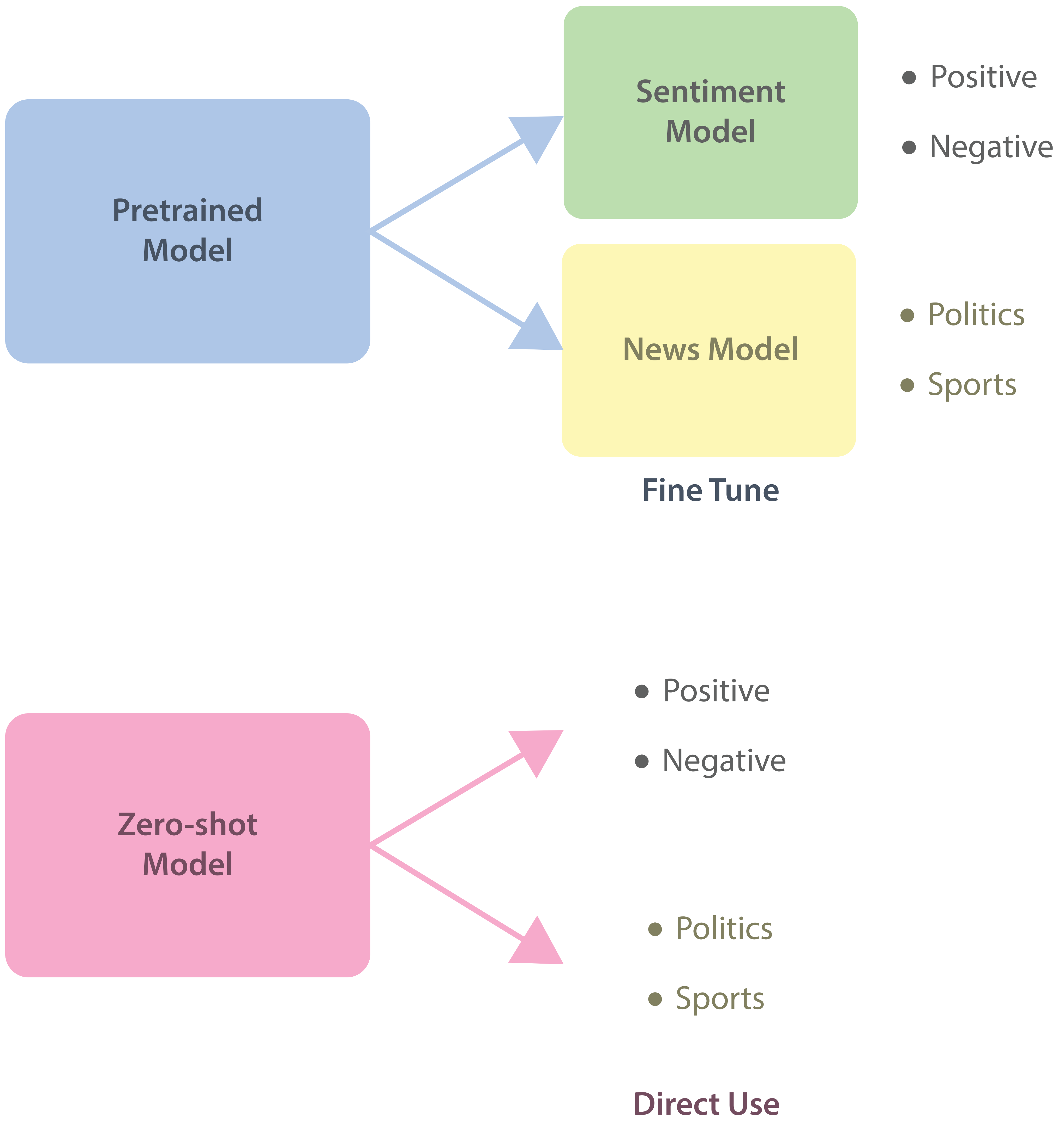
Prefix	Meaning
I	Start of a new entity.
B	Continuation of entity from previous line.

Naive Zero-shot approach

Zero-shot learning.

Setting and Use Cases.

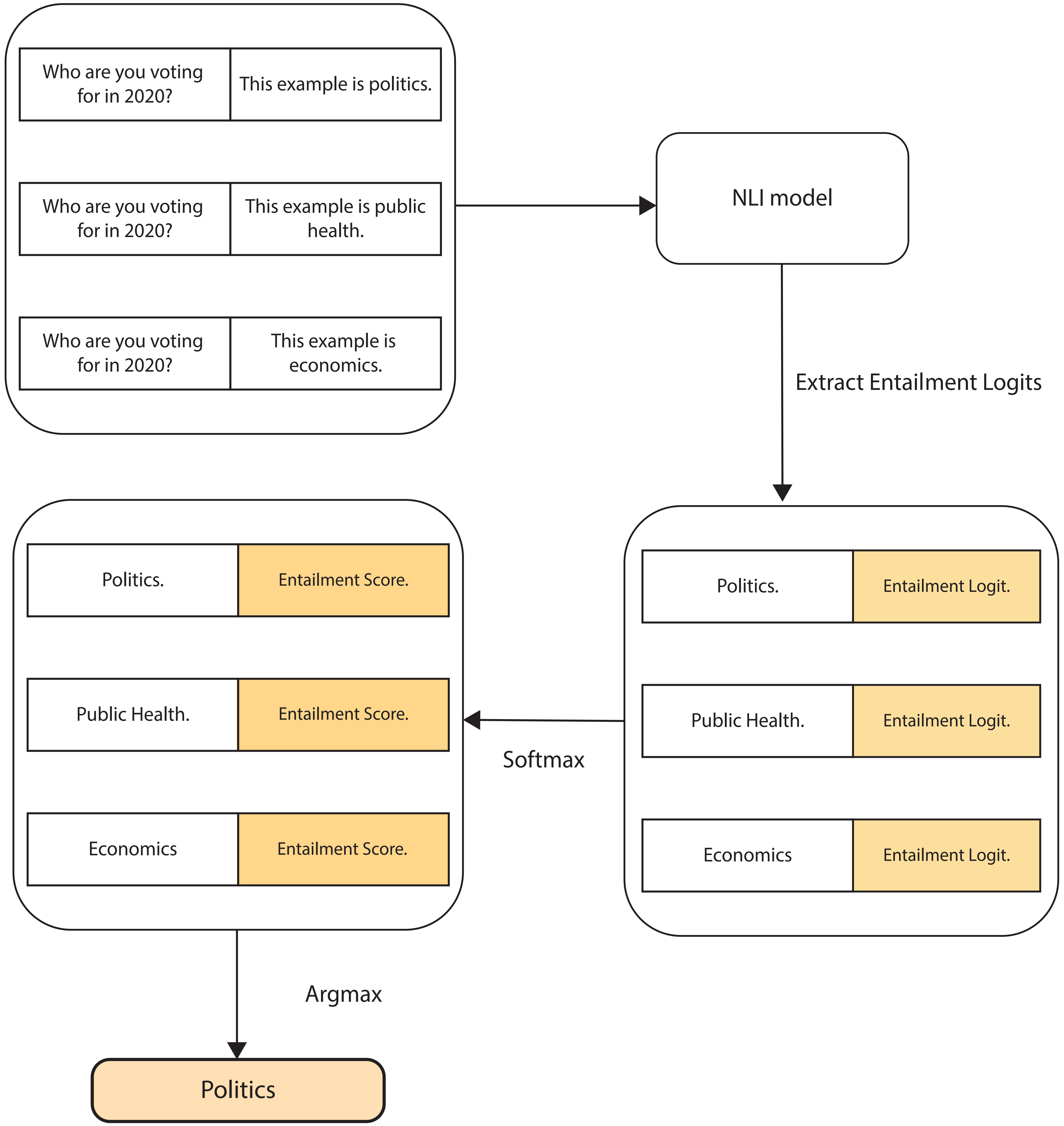
- **Pros :**
  - No fine tuning/training for downstream task.
  - Extremely useful in practical scenarios if done well.
- **Cons :**
  - Higher inference times in general.
  - Task is harder, and hence performance suffers.
  - Difficult to build effective zero-shot learners as model complexity is higher.



Zero-shot Classification.

Methodology.

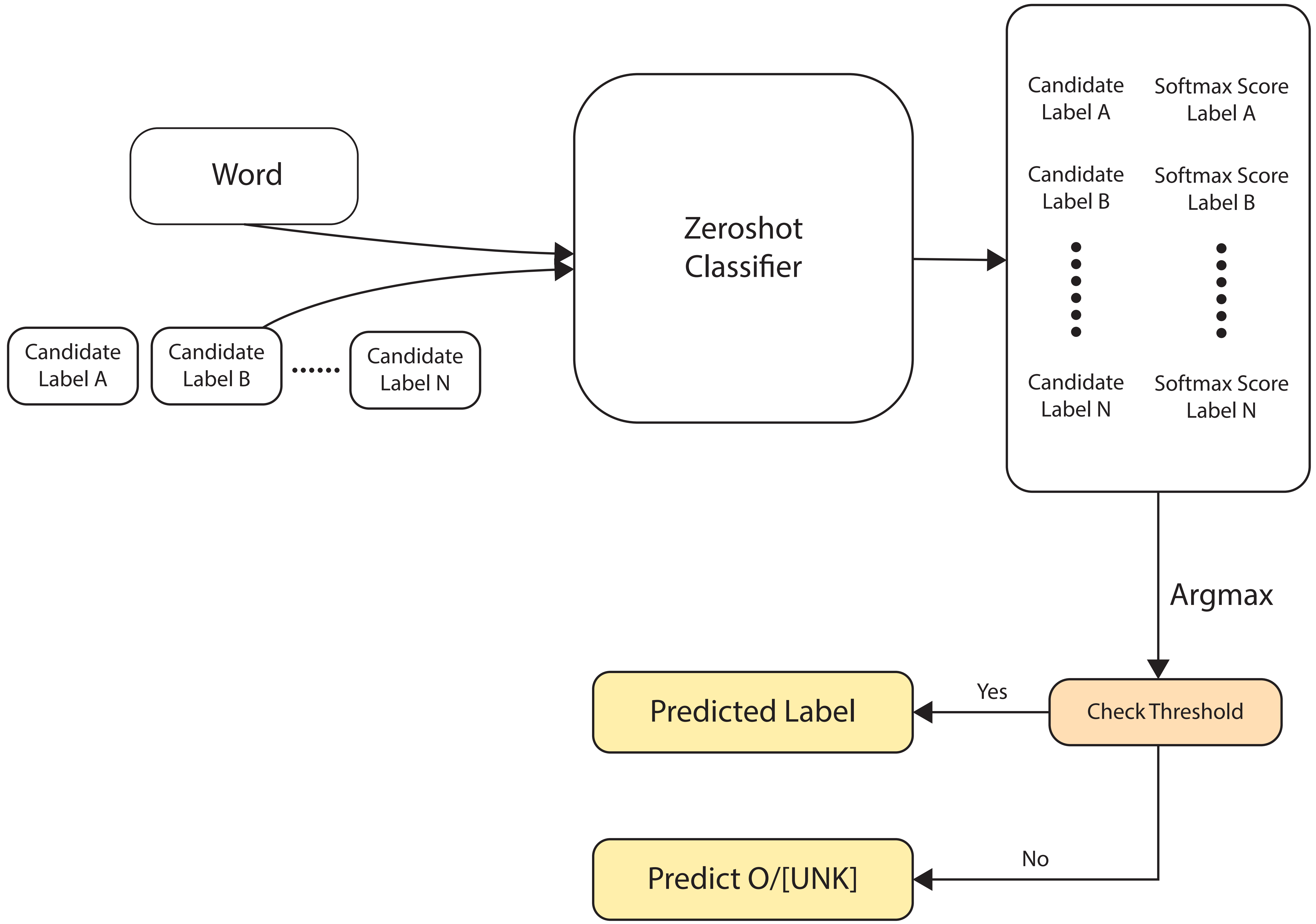
- Classification as Natural Language Inference.
- Proposed by Yin et. al. in 2019.
- Implemented using the BART model in Huggingface.



Naive Zero-shot NER.

Methodology.

- How can the zero-shot classification pipeline be used naively for NER?
- **Pros :**
  - Easy to implement.
  - Beats baseline for original task even without any kind of fine tuning.
- **Cons :**
  - Does not make use of contextual information.
  - Current methods beats this by a wide margin.



Naive Zero-shot NER.

Results and Discussions.

- Threshold of 0.7 seems good enough and passes the baseline.
  Increasing threshold causes call performance measures to drop.
- Using meaningful label names is important.
- Much better performance from modern Deep Learning NER approaches.
- Still this method shows promise and further work is necessary.
- Use the CoNLL-2003 processor to exhaustively enumerate all entity spans and try classification.

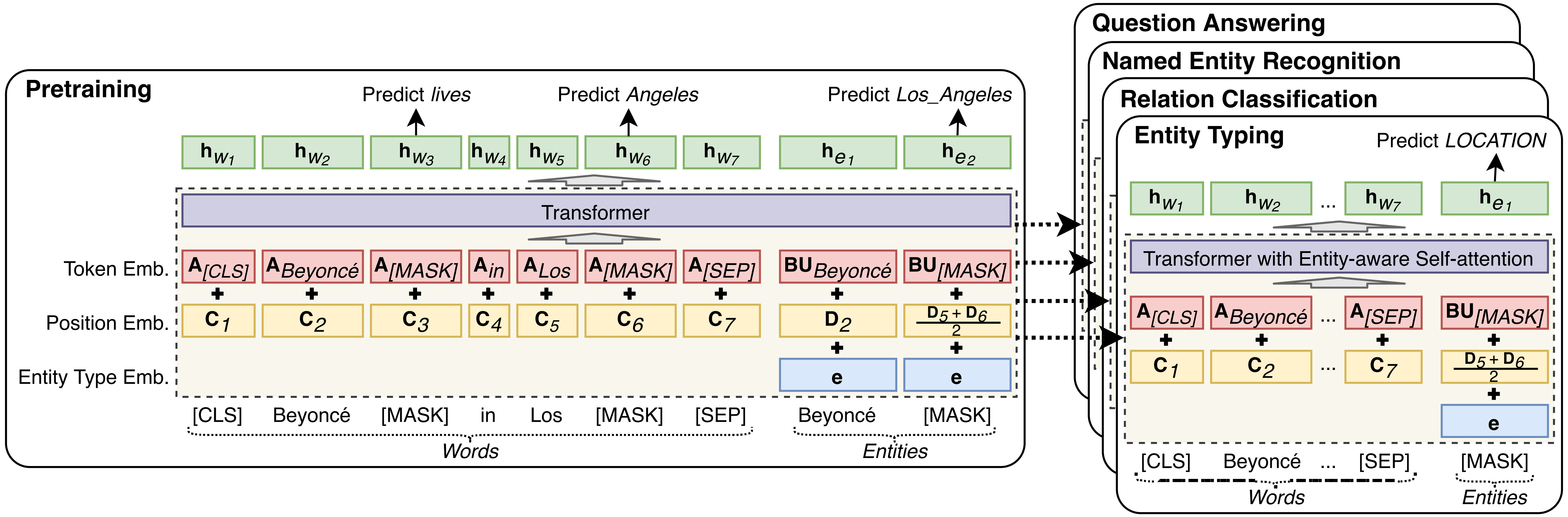
Label Type	Threshold	Precision	Recall	F1
Meaningful	0.5	57.56%	65%	59.72
	0.6	72.99%	67.52%	69.49
	0.7	75.83%	68.12%	70.27
Meaningless	0.5	1.72%	37.44%	3.21
	0.6	0.42%	38.01%	0.82
	0.7	0.07%	43.35%	0.15

LUKE

LUKE

What is new and different?

- Drawback of KB methods.
- Drawback of of old CWR methods.
- Differentiating entities and words.



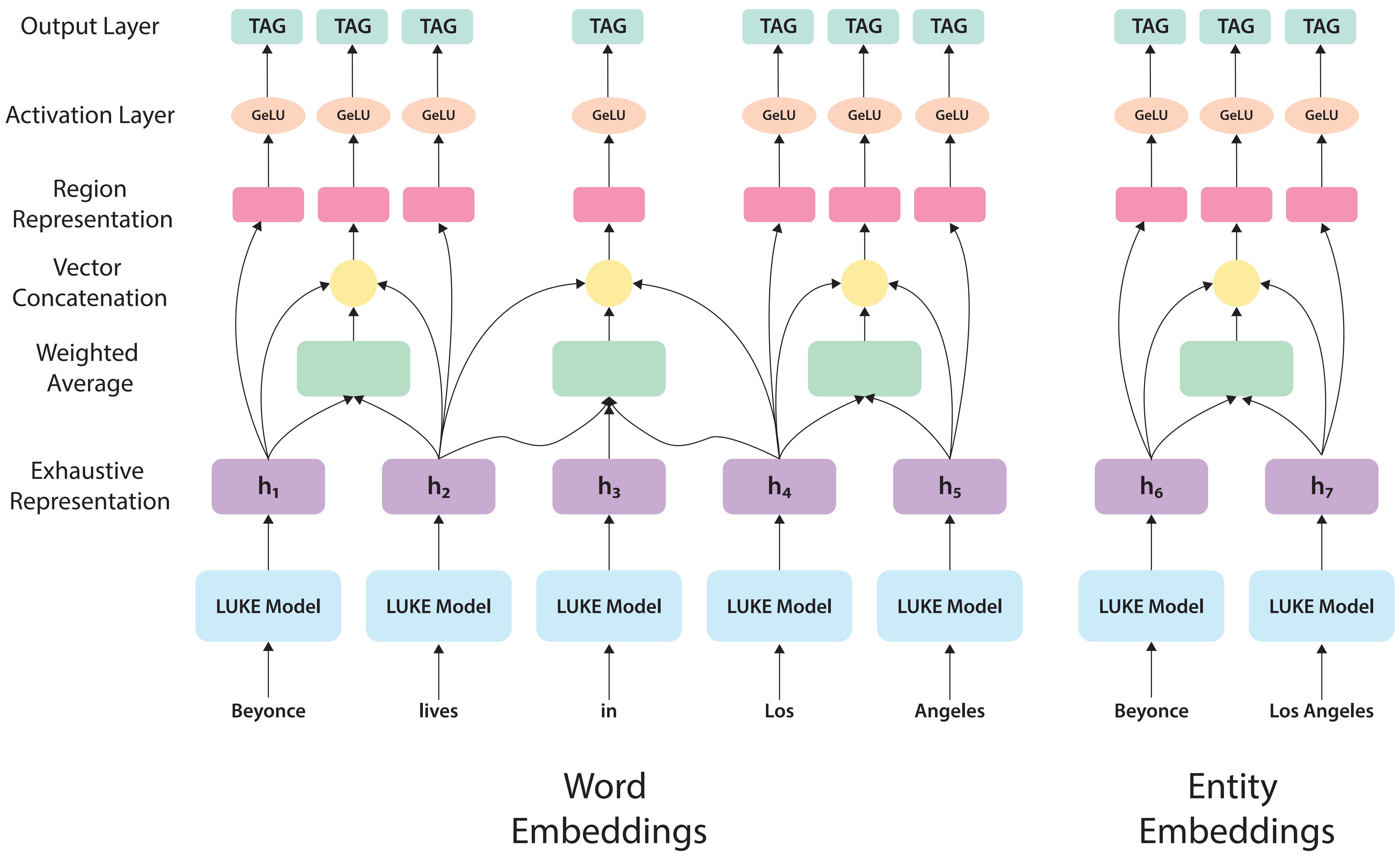
*Picture taken from the LUKE paper by Yamada *et. al.*

Exhaustive Enumeration for NER.

How we do NER using the entity aware LUKE model.

- The methodology for doing NER using the LUKE model was taken from the very effective entity representation scheme due to <u>Sohrab and Miwa.</u>

- Is able to detect both nested and flat entities.



*Picture adapted for LUKE from Sohrab and Miwa's paper.

LUKE.

Results.

- Achieves SOTA on 5 well known datasets.
    - Open Entity
    - TACRED
    - CoNLL-2003
    - ReCoRD
    - SQuAD
- Has an F1 score of 94.3 on the CoNLL-2003 dataset.
- Results reproduced.
- Fast and Effective.
- PyTorch implementation available on GitHub.

Conclusion

Conclusion and Future work

What was observed.

- Naive zero-shot NER promising but needs more work.
- Can try out the exhaustive enumeration scheme with this approach.
- Not much point in doing plain NER approaches because current methods are effective and tractable.
- Currently going through LUKE source code to understand how to incorporate exhaustive enumeration with zero-shot NER.

Acknowledgement and References

Acknowledgement

I am thankful to Dripta Mj. for all his help in this project and for the opportunity to work through the academic process on my own.

I also want to thank Purnendu Mukherjee who helped me come up with the idea for this project.

References

- John D. Lafferty , Andrew Kachites McCallum, Fernando C. N. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. ICML '01: Proceedings of the Eighteenth International Conference on Machine LearningJune 2001 (Pages 282–289).
- Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, Yuji Matsumoto. LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention. EMNLP 2020.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692.
- Mohammad Golam Sohrab, Makoto Miwa. Deep Exhaustive Model for Nested Named Entity Recognition. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (Pages 2843–2849).
- Erik F. Tjong Kim Sang, Fien De Meulder. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 (Pages 142–147).
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, Luke Zettlemoyer. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. arXiv:1910.13461.
- YonghuiWu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi,Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, Jeffrey Dean. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. arXiv:1609.08144