Representing Relational Knowledge with Language Models

Asahi Ushio

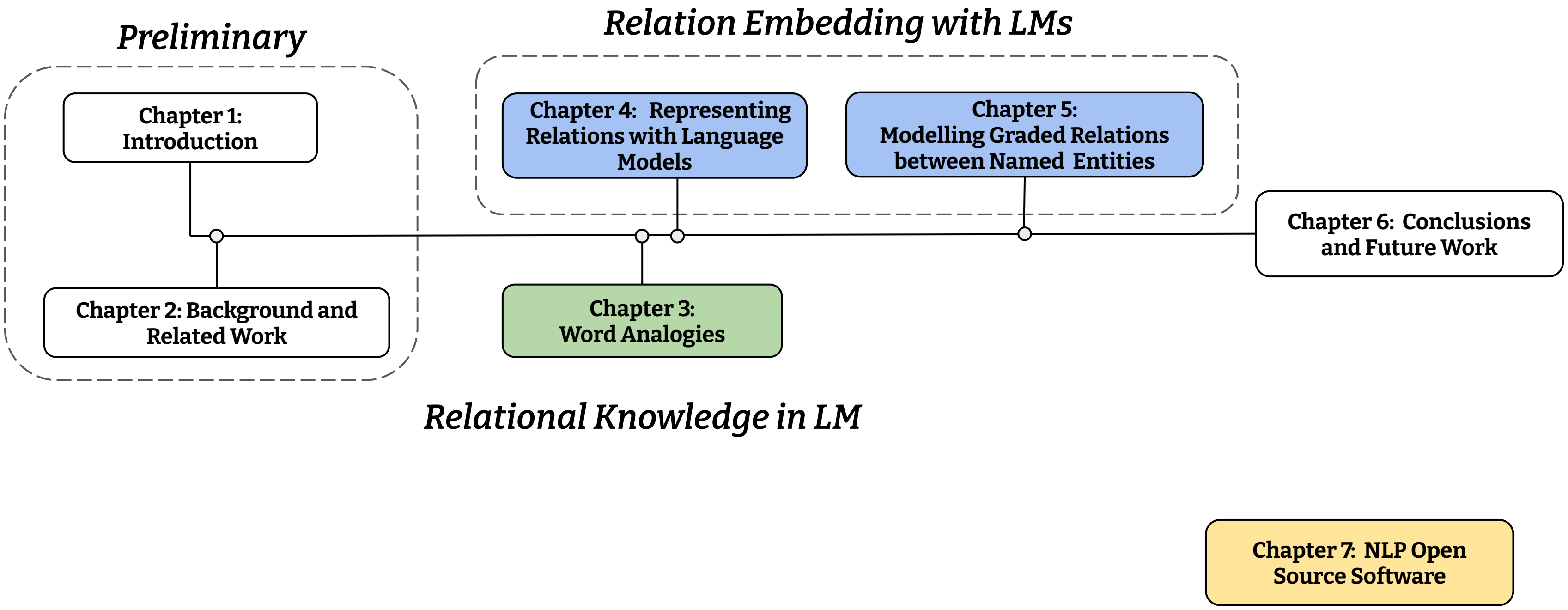
PhD viva

12th December 2023

Cardiff NLP

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Thesis Main Outline

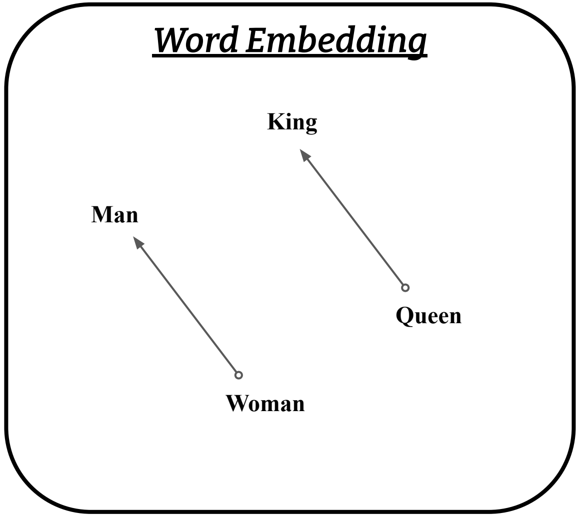


Relation Representation

Word Embedding [Mikolov (2013)]

Pair2Vec [Joshi (2019)]

Relative [Camacho-Collados (2019)]



X	Y	Contexts
hot	*cold*	with **X** and **Y** baths too **X** or too **Y** neither **X** nor **Y**
Portland	*Oregon*	in **X**, **Y** the **X** metropolitan area in **Y** **X** International Airport in **Y**
crop	*wheat*	food **X** are maize, **Y**, etc dry **X**, such as **Y**, more **X** circles appeared in **Y** fields
Android	*Google*	**X** OS comes with **Y** play the **X** team at **Y** **X** is developed by **Y**

Language Models Probing

Language model (LM) pretraining is a breakthrough in NLP, while being **black box**.

LM Probing

- ➢ Syntactic Knowledge: Hewitt 2019, Tenney 2019, Clark 2020
- ➢ Factual Knowledge: Petroni 2019, Kassner 2020, Jiang 2020
- ➢ Relational Knowledge: Bouraoui 2020

Research Questions

Language model (LM) pretraining is a breakthrough in NLP, while being **black box**.

LM Probing

- ➢ Syntactic Knowledge: Hewitt 2019, Tenney 2019, Clark 2020
- ➢ Factual Knowledge: Petroni 2019, Kassner 2020, Jiang 2020
- ➢ Relational Knowledge: Bouraoui 2020

Q1: Can we evaluate relational knowledge of LM without training?

Research Questions

Language model (LM) pretraining is a breakthrough in NLP, while being **black box**.

LM Probing

- ➢ Syntactic Knowledge: Hewitt 2019, Tenney 2019, Clark 2020
- ➢ Factual Knowledge: Petroni 2019, Kassner 2020, Jiang 2020
- ➢ Relational Knowledge: Bouraoui 2020

Q1: Can we evaluate relational knowledge of LM without training?
Q2: Can we obtain a better relation representation with LM?

Relational Knowledge in LM

Word analogy as a probing task of relational knowledge.

- ➢ Solvable without training.
- ➢ Different Levels
  - ○ Primary school to college
- ➢ Various Relation Types
  - ○ Named entity, common noun

Query:		word:language
Candidates:	(1)	paint:portrait
	(2)	poetry:rhythm
	(3)	**note:music**
	(4)	tale:story
	(5)	week:year

Sample from SAT analogy dataset.

Asahi Ushio, et al. "BERT is to NLP what AlexNet is to CV: Can Pre-Trained Language Models Identify Analogies?" (ACL 2021)

Relational Knowledge in LM

Word analogy as a probing task of relational knowledge.

- ➢ Solvable without training.
- ➢ Different Levels
  - ○ Primary school to college
- ➢ Various Relation Types
  - ○ Named entity, common noun

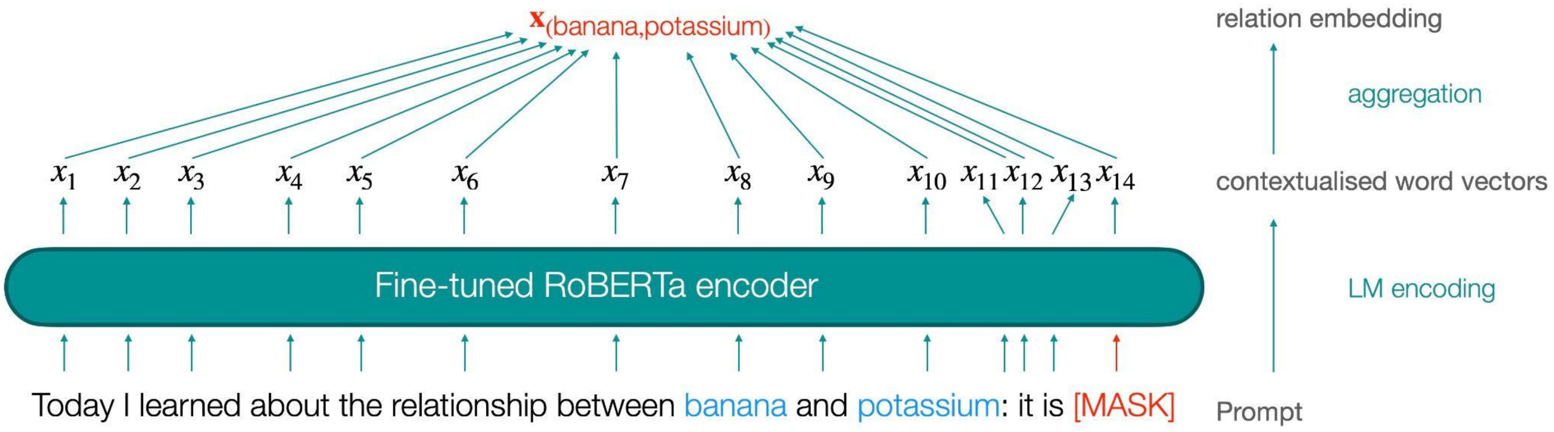
Query:		word:language
Candidates:	(1)	paint:portrait
	(2)	poetry:rhythm
	(3)	**note:music**
	(4)	tale:story
	(5)	week:year

Sample from SAT analogy dataset.

- **Poor capability of LMs at solving word analogies.**
- **Improvement with tailored scoring function (with validation).**

Asahi Ushio, et al. "BERT is to NLP what AlexNet is to CV: Can Pre-Trained Language Models Identify Analogies?" (ACL 2021)

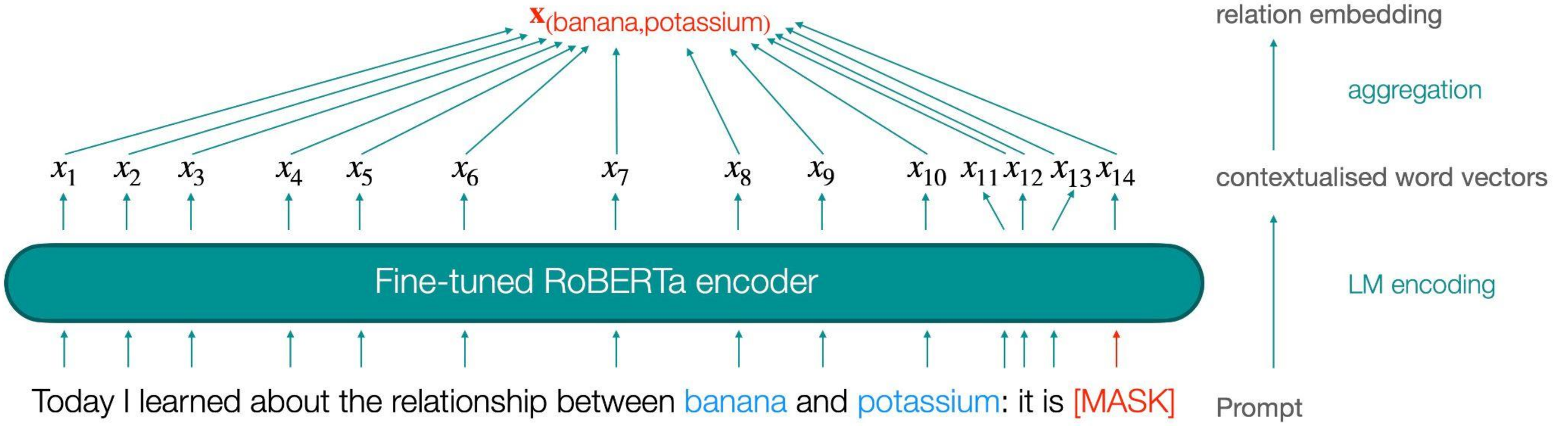
Relation Embedding with LMs
RelBERT



- ➢ Relation embedding obtained via prompting word pair
- ➢ Fine-tuning LM with contrastive learning

Asahi Ushio, et al. "Distilling Relation Embeddings from Pretrained Language Models" (EMNLP 2021)

Relation Embedding with LMs
RelBERT



- ➢ Relation embedding obtained via prompting word pair
- ➢ Fine-tuning LM with contrastive learning

- **SoTA in analogies (outperform GPT-3), relation classification.**
- **Better generalization (i.e. relations not in the training data).**

Asahi Ushio, et al. "Distilling Relation Embeddings from Pretrained Language Models" (EMNLP 2021)

Is RelBERT perfect?

RelBERT (and some LLMs) can solve word analogies.

➢ Does this mean they understand relations completely?

Graded Relation Ranking

Relation Types	Examples (Ordered by Prototypicality)
competitor of	*[Dell, HP]* > *[Neoclassicism, Romanticism]* > *[Steve Jobs, Atlanta]*
friend of	*[Australia, New Zealand]* > *[The Beatles, Queen]* > *[KGB, CIA]*
influenced by	*[Plato, Socrates]* > *[Hip Hop, Jazz]* > *[Sauron, Shiba Inu]*
known for	*[Apple, iPhone]* > *[Apple, Apple Watch]* > *[Pixar, Novosibirsk]*
similar to	*[Coca-Cola, Peps]* > *[Christmas, Easter]* > *[Italy, Superman]*

New challenging tasks.

- **5** relation types.
- Pairs of **named entities**.
- **Rank** the pairs based on **prototypicality**.

Asahi Ushio, et al. “A RELENTLESS Benchmark for Modelling Graded Relations between Named Entities” (EACL 2024 under review)

Graded Relation Ranking

Relation Types	Examples (Ordered by Prototypicality)
competitor of	***[Dell, HP]*** **>** ***[Neoclassicism, Romanticism]*** **>** ***[Steve Jobs, Atlanta]***
friend of	***[Australia, New Zealand]*** **>** ***[The Beatles, Queen]*** **>** ***[KGB, CIA]***
influenced by	***[Plato, Socrates]*** **>** ***[Hip Hop, Jazz]*** **>** ***[Sauron, Shiba Inu]***
known for	***[Apple, iPhone]*** **>** ***[Apple, Apple Watch]*** **>** ***[Pixar, Novosibirsk]***
similar to	***[Coca-Cola, Peps]*** **>** ***[Christmas, Easter]*** **>** ***[Italy, Superman]***

New challenging tasks.

➢ **5** relation types

➢ **Rank** the pairs based on **prototypicality**.

LLMs (eg. GPT-4) and RelBERT are underperforming human baselines.

Asahi Ushio, et al. "A RELENTLESS Benchmark for Modelling Graded Relations between Named Entities" (EACL 2024 under review)

Conclusion

Can we evaluate relational knowledge of LM without training?

➢ Yes! We can leverage word analogies to assess the relational knowledge without training it.

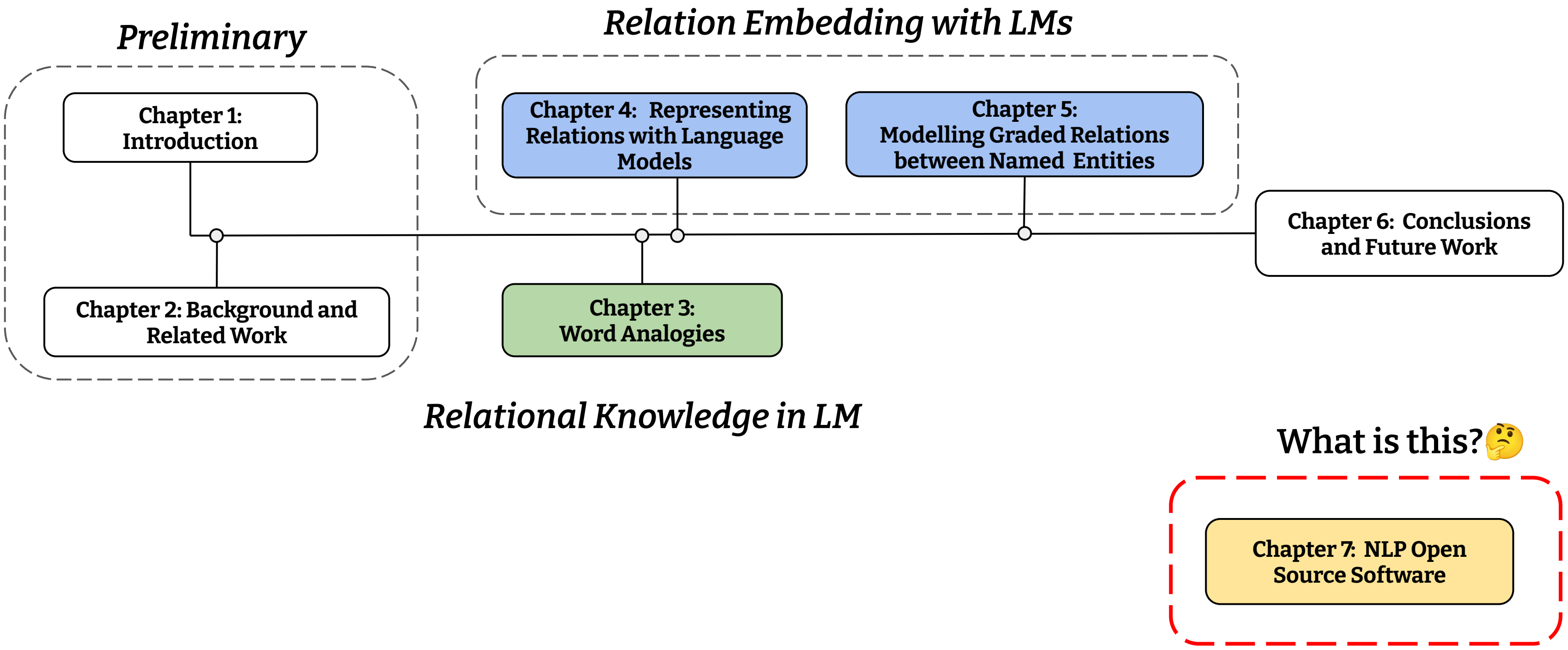
Can we obtain a better relation representation with LM?

➢ Yes! RelBERT, LM based relation representation model can achieve SoTA on analogy benchmarks.

Are there any challenges/limitations for RelBERT?

➢ Yes… We proposed a new task, ranking named-entity pairs based on prototypicality, where RelBERT underperforms human baselines largely.

NLP Open Source Software



NLP Open Source Software

Python libraries to facilitate NLP applications.

- *KEX*: Keyword Extraction (⭐***51***)
- *T−NER*: Named-Entity Recognition (⭐***332***)
- *LMQG*: Question & Answer Generation (⭐***200***)
- *TweetNLP*: NLP on Twitter (⭐***229***)

Presented at ACL conferences (demo or main).

🌳Thank you!🌳

QA-1

Prompt to solve the analogy.

- ➢ We use custom prompt, because prompt optimization is not easy for analogy question. Automatic prompt optimization such as p-tuning and autoprompt require iterative update based on the large training set, and need seed templates. We don’t know what seed templates should be used for analogy as there are four words to be included in the template at least, and no training set is available either.

Permutation of Analogical Proportion

- ➢ The experiment shows that mixing the domain (eg. a:b::c:d -> a:d::c:b) leads worst results.

QA-2

RelBERT applications (which cannot be done with KG)

- ➢ (i) Fine-grained relation clustering: RelBERT can generate embeddings on pairs, and one can cluster the embedding to investigate fine-grained relation types under a single relation type.
- ➢ (ii) Few-shot relation classification: Given a new word pair, one can compute the similarities over the given demonstration and choose the closest one as the relation prediction.

Difference from KG (in addition to the above)

- ➢ RelBERT can deal with any pairs in theory, while KG cannot be generalized beyond the fixed nodes/relation types.

QA-3

Why named entity for Relentless?

➢ RelBERT shows a good performance on entity analogy (T-REX, NELL-One), but the performance is relatively low than the others. To further confirm the capability of RelBERT of representation for the entity pairs, Relentless focuses on named entities.

➢ Named entity is challenging for LMs.

➢ More abstract and nuanced than lexical/syntactic relations.

QA-4

Why graded relation (prototypicality) for Relentless?

- ➢ It’s one of the benefit of continuous representation to compute the numeric score that reflects the nuanced difference instead of binary signal of symbolic model, so we want to directly evaluate the sensitive to the subtle difference of relations.

QA-5

What is the connection of chapter 7 to the others?

➢ They are independent. Relation representation is the primal research goal, but concurrently I am interested in the NLP OSS, and build those packages.