Injecting Knowledge into Sentence Embedding Models for Information Retrieval using Adapters

Kick-Off Presentation

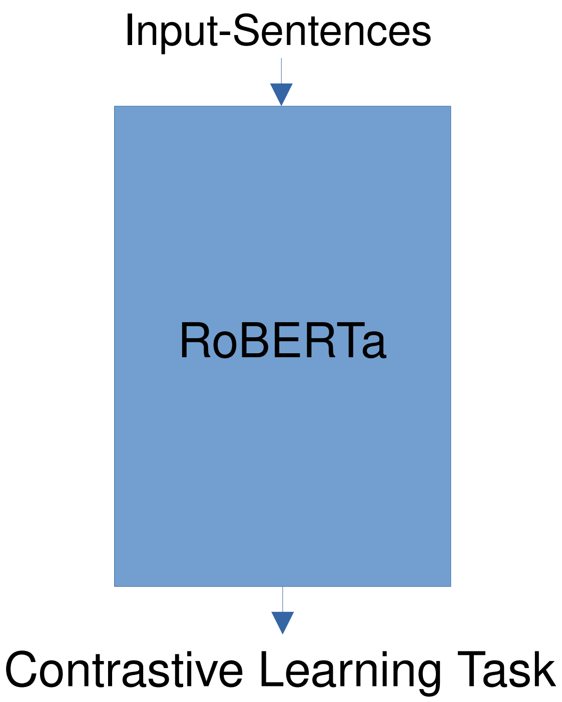
23.01.2023

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
Technical University of Munich (TUM)
wwwmatthes.in.tum.de

Outline

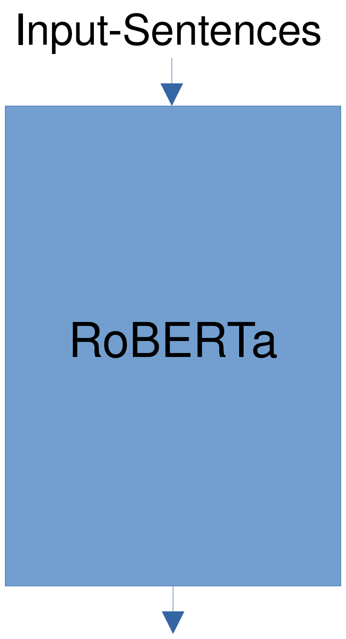
Motivation – Sentence Embedding Models

- Similarity comparison of sentences/texts
- Information Retrieval Tasks



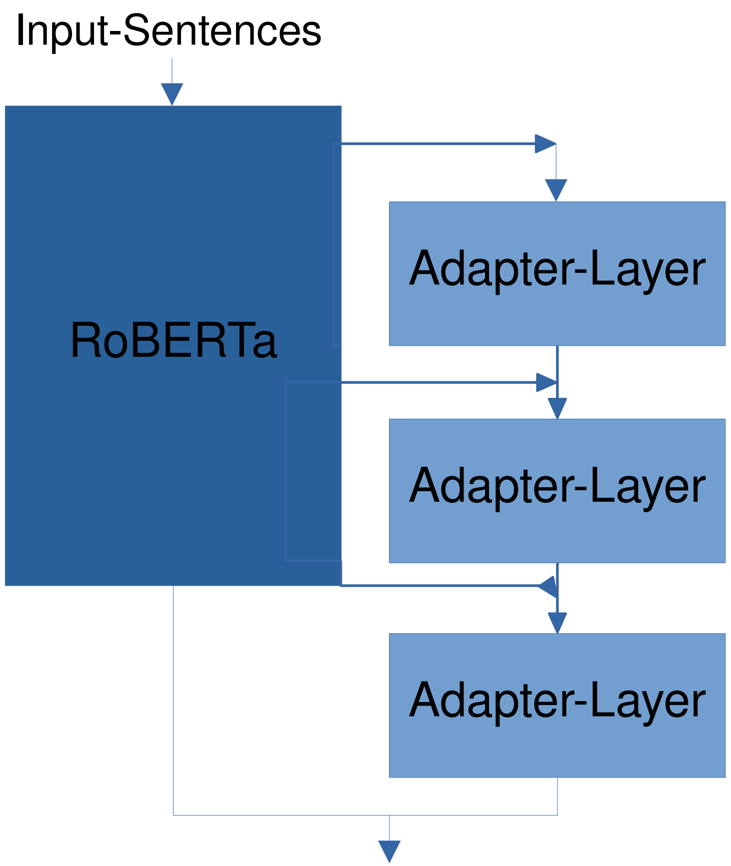
Motivation – Adapters

- Pretrained once, „Plug'n'Play“ for different use cases
- Lower training time due to less parameters



Motivation – Adapters

- Pretrained once, „Plug’n’Play“ for different use cases
- Lower training time due to less parameters



Problem Statement

- Sentence Embedding Models fail to encode factual knowledge [1]
- Enhance Sentence Embedding Models by injecting Knowledge through Adapters
    - Inject structured knowledge from Knowledge Graphs
- SotA: Contrastive Learning

Methodology – Injecting Knowledge

RQ1: How to inject structured Knowledge into Sentence Embedding models with adapters?

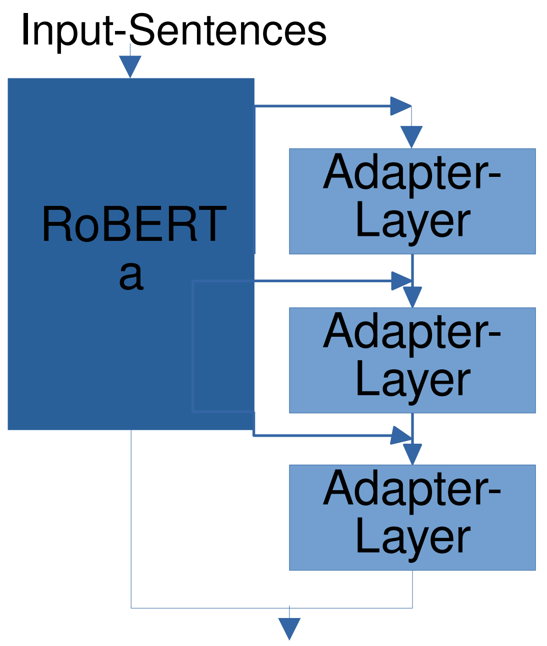
- Learning methods for Adapters
  - Contrastive Learning of Graph-based data
    - Anchor: (--, relation, target_a) „Amazon is an American tech company.“
    - Pos: (--, relation, target_a) „Facebook is an American tech company.“
    - Neg: (--, relation, target_b) „Mercedes-Benz is a German car manufacturer.“
  - Masked Language Modeling on Knowledge Graphs
    - (“Mercedes-Benz”, “is-a”, “car manufacturer”)
    - Mask out the object
    - Mercedes-Benz is a <MASK>

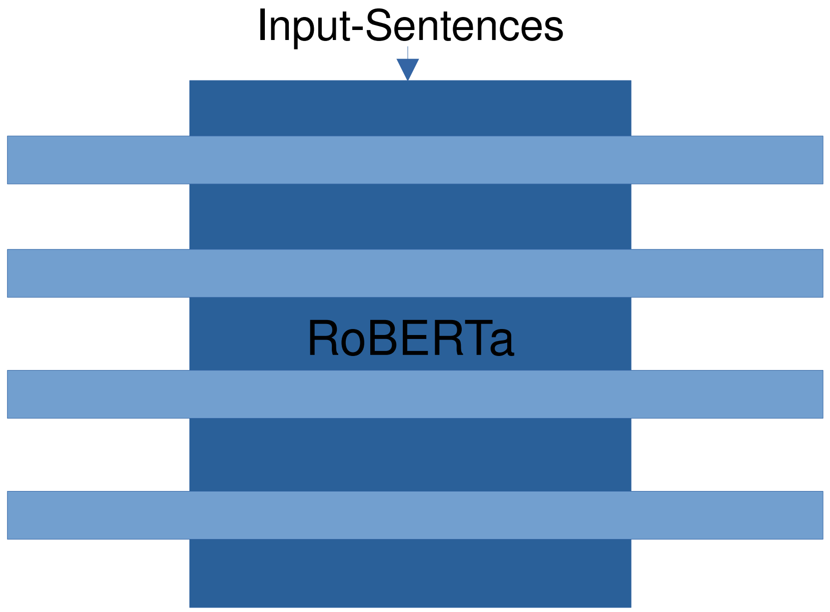
Methodology – Sentence Embedding Model Training

- Using SotA: SimCSE [2]
- Supervised SimCSE:
    - Contrastive Learning of Anchor, Positive, Negative Sentence
- Unsupervised SimCSE:
    - Contrastive Learning of Anchor, Dropout-Anchor, Different Sentence

Methodology – Adapter Architectures

- K-Adapter [1]
- Houlsby [3]
- ...





Preliminary Results

RQ2: Do Knowledge Adapters improve information retrieval tasks of Sentence Embedding models?

- K-Adapter (supervised SimCSE)
  - Adapter pretrained on different dataset, via Contrastive Learning
  - Resulting Knowledge-infused model trained on same dataset

Model	STS12	STS13	STS14	STS15	STS16	STSBenchmark	SICKRelatedness	Avg.
Roberta	76.33	87.36	82.10	86.03	83.82	86.35	80.69	83.24
K-Adapter	77.87	87.24	82.56	87.17	84.62	86.26	79.93	83.68

Preliminary Results

- K-Adapter (unsupervised SimCSE)
  - Adapter pretrained on different dataset, via Contrastive Learning (supervised)
  - Resulting Knowledge-infused model trained on same dataset (unsupervised)

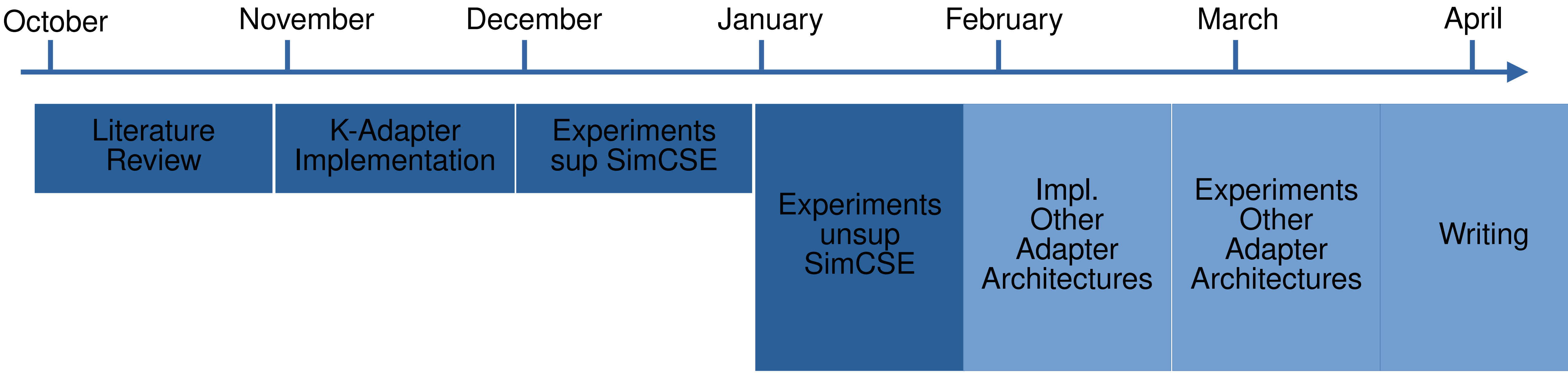
Model	STS12	STS13	STS14	STS15	STS16	STSBenchmark	SICKRelatedness	Avg.
Roberta	71.26	83.75	75.26	85.06	81.17	81.69	70.84	78.43
K-Adapter	73.37	84.89	76.21	87.24	83.17	81.79	72.63	79.90

Domain-specific Usage

RQ3: How to combine domain-specific Knowledge Adapters for the scholarly domain?

- Domain adaptation for training Sentence Embedding Models difficult
    - Usually no training data for specific domain available
- Train domain-specific adapters (MLM) and plug into generic STS-trained models

Timeline


October
November
December
January
February
March
April
Literature Review
K-Adapter Implementation
Experiments sup SimCSE
Experiments unsup SimCSE
Impl. Other Adapter Architectures
Experiments Other Adapter Architectures
Writing

Thank you for your attention

- Any questions?

References

- [1]: Kassner, Schuetze, 2020, Negated and misprimed probes for pretrained language models: Birds can talk, but cannot fly
- [2]: Goa et al. 2021, SimCSE: Simple Contrastive Learning of Sentence Embeddings
- [3]: Houlsby et al. 2019, Parameter-Efficient Transfer Learning for NLP