# Improvements of a local bibliographic citation recommendation model based on hyperparameter and query optimizations

Zoran Medić, mag. ing.
(zoran.medic@fer.hr)
mentor: prof. dr. sc. Jan Šnajder
University of Zagreb Faculty of Electrical Engineering and Computing

## 1. Introduction

- Rate of published articles is growing to the extent that researchers can't keep up with published work
- Models based on **machine learning (ML)** and **natural language processing (NLP)** can be used to help researchers find relevant articles
- **Local bibliographic citation recommendation (LCR)** is a task of finding relevant articles for a given input query composed of a citation context (*local information*)
- If LCR is solved successfully, researchers could use LCR models to get recommendations and find relevant work

## 2. Problem Description

- **Task:** given a piece of text (*citation context*) as input, model has to output relevant articles for that input

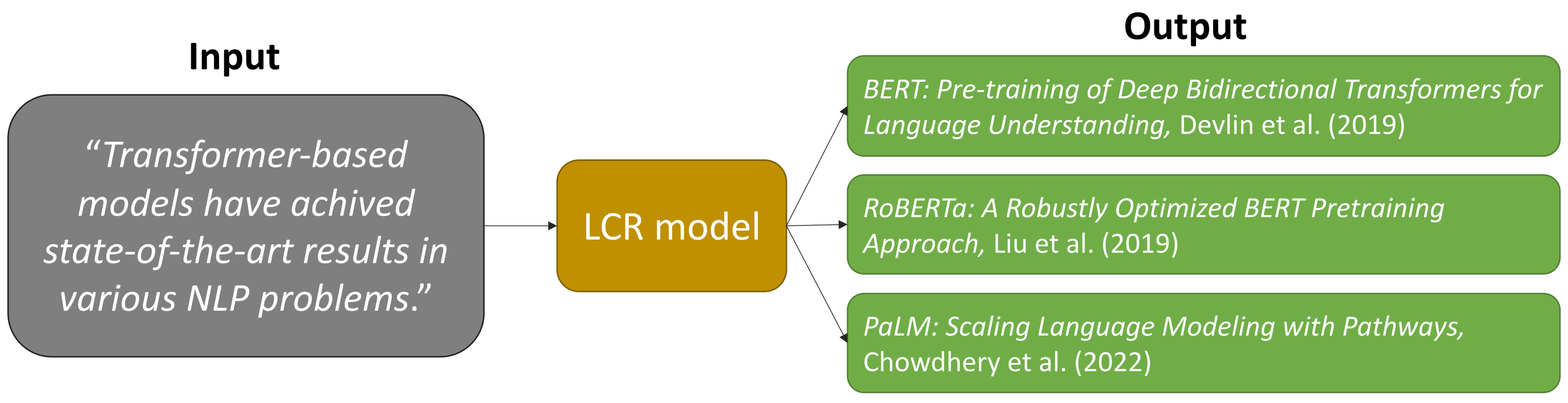


**Figure 1.** Standard approach to LCR task

- **Datasets:** easily obtainable in two steps: (1) PDF parsing of full article texts and (2) matching cited articles with those in the database
- Recently, LCR models have shifted to **deep learning (DL)**-based models
- In this thesis, the focus is on **three aspects of DL-based LCR models**:
  1. effect of citation context size on model's performance,
  2. design choices such as negative sampling,
  3. improving task usability

## 3. Citation Context Size

- Experiments on two datasets with differing context sizes: ACL-ARC (larger context) and RefSeer (smaller context)
- **Model:** a Bi-LSTM layer followed by an attention layer and trained with triplet loss
- **Triplet loss** minimizes distance between the context (query) and a cited paper (positive example), while maximizing distance between the context and a non-cited paper (negative example)
- Comparison of context only (DualCon) and globally enhanced (DualEnh) model (article's *global information*)

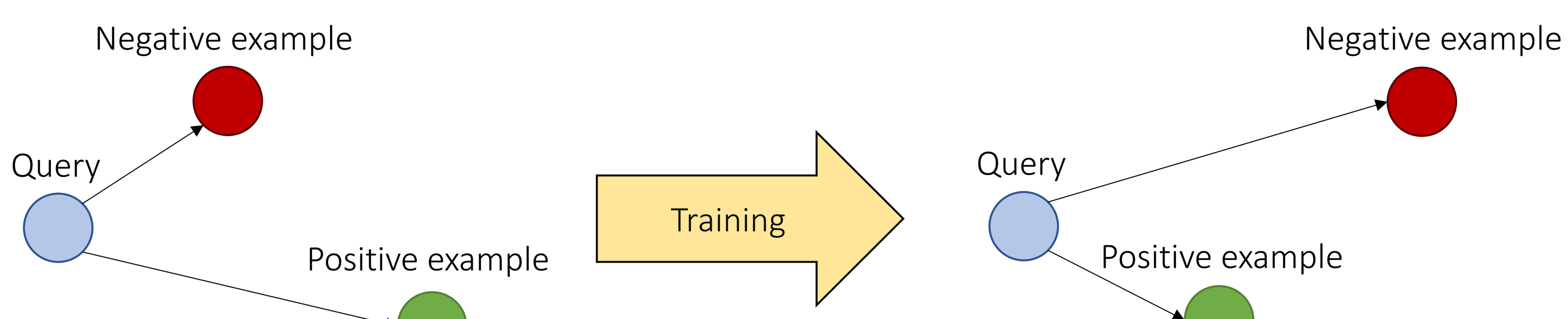


**Figure 2.** Training procedure when using triplet loss

- **Conclusion:** including global information with context improves performance for smaller context sizes

| | ACL-ARC | | RefSeer | |
|---|---|---|---|---|
| | **R@10** | **MRR** | **R@10** | **MRR** |
| DualCon | 0.69 | 0.37 | 0.41 | 0.21 |
| DualEnh | 0.70 | 0.36 | **0.53** | **0.28** |

## 4. Sampling of Negatives

- Ranking models' performance is affected by the **choice of negative examples in triplets**
- Negative examples should not be too hard or too easy
- Strategies for choosing negative examples:
  1. citation graph neighbors,
  2. candidates obtained using other model (BM25),
  3. most often cited articles

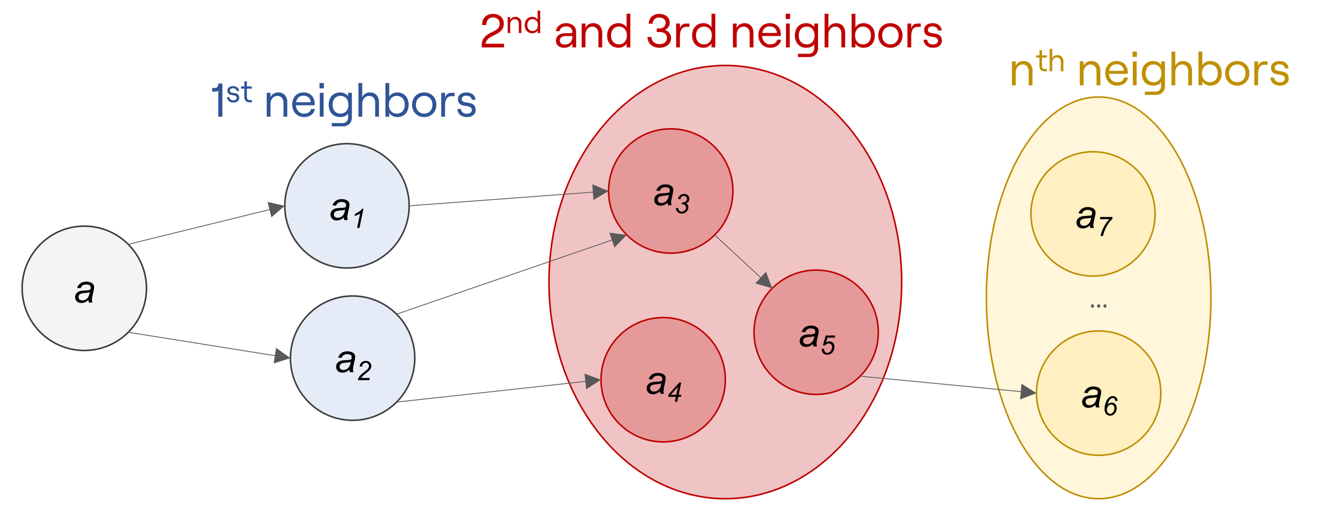


**Figure 3.** Citation graph. Node *a* represents a query article. Red and yellow nodes represent hard and easy negatives, respectively.

- **Conclusion:** the best strategy is dataset-dependent and a specific strategy can lead to a strong boost in performance

| Sampling strategy | ACL-ARC | | RefSeer | |
|---|---|---|---|---|
| | **R@10** | **MRR** | **R@10** | **MRR** |
| *Random* | 0.69 | 0.37 | 0.29 | 0.13 |
| Graph | 0.66 | 0.34 | **0.42** | **0.21** |
| BM25 | 0.68 | 0.37 | 0.29 | 0.15 |
| Most cited | **0.74** | **0.40** | 0.30 | 0.14 |

## 5. Task Usability

- Typically, users' need in a query is expressed with something in between global and local information
- This can be modeled using *topic sentences* as part of query
- **Topic sentences** introduce the topic of a paragraph, while the following text (including citations) provides more details

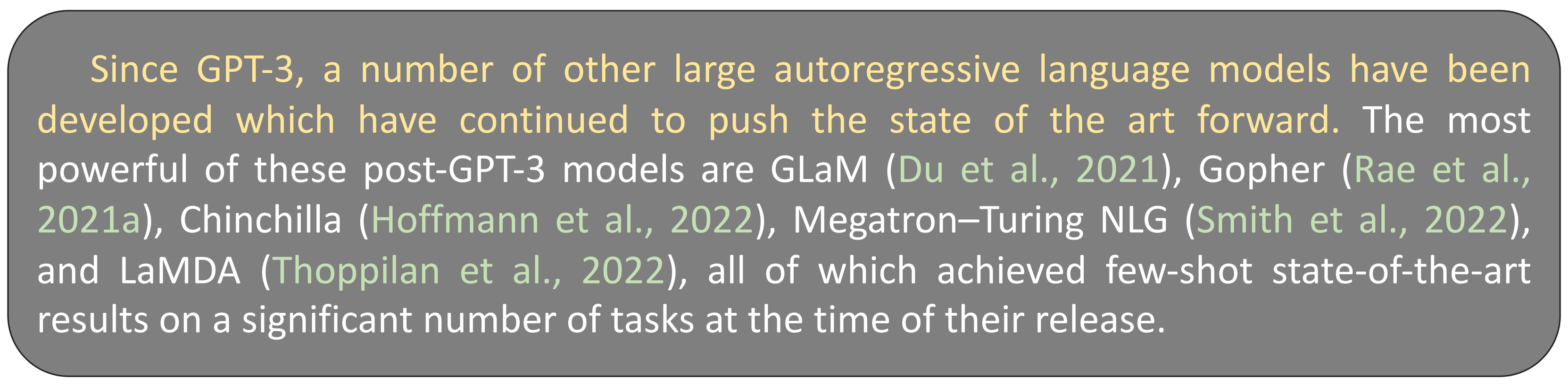
Since GPT-3, a number of other large autoregressive language models have been developed which have continued to push the state of the art forward. The most powerful of these post-GPT-3 models are GLaM (Du et al., 2021), Gopher (Rae et al., 2021a), Chinchilla (Hoffmann et al., 2022), Megatron–Turing NLG (Smith et al., 2022), and LaMDA (Thoppilan et al., 2022), all of which achieved few-shot state-of-the-art results on a significant number of tasks at the time of their release.

**Figure 4.** Example of a topic sentence (in yellow) in a paragraph.

- **Proposed approach:**
  1. construct a dataset of topic sentences and articles cited in the corresponding paragraphs,
  2. train a DL-based model for recommending relevant articles for a given topic sentence-enhanced query,
  3. evaluate the model and compare with recommendations from GCR and LCR models

## 6. Conclusion

- LCR models can be improved with a careful optimization of hyperparameters, that typically depend on the dataset
- Usability of CR task can be enhanced by using topic sentences as part of query

## 7. Project Acknowledgement

PhD candidate was supported by a grant from the Croatian Science Foundation (HRZZ-DOK-2018-09).

*PhD Day Faculty of Electrical Engineering and Computing, 2 June 2022.*