

In-Context Example Selection Based On Error Structure For Grammatical Error Correction

Gustavo Sutter Pascal Poupart



UNIVERSITY OF
WATERLOO

DAVID R. CHERITON SCHOOL
OF COMPUTER SCIENCE

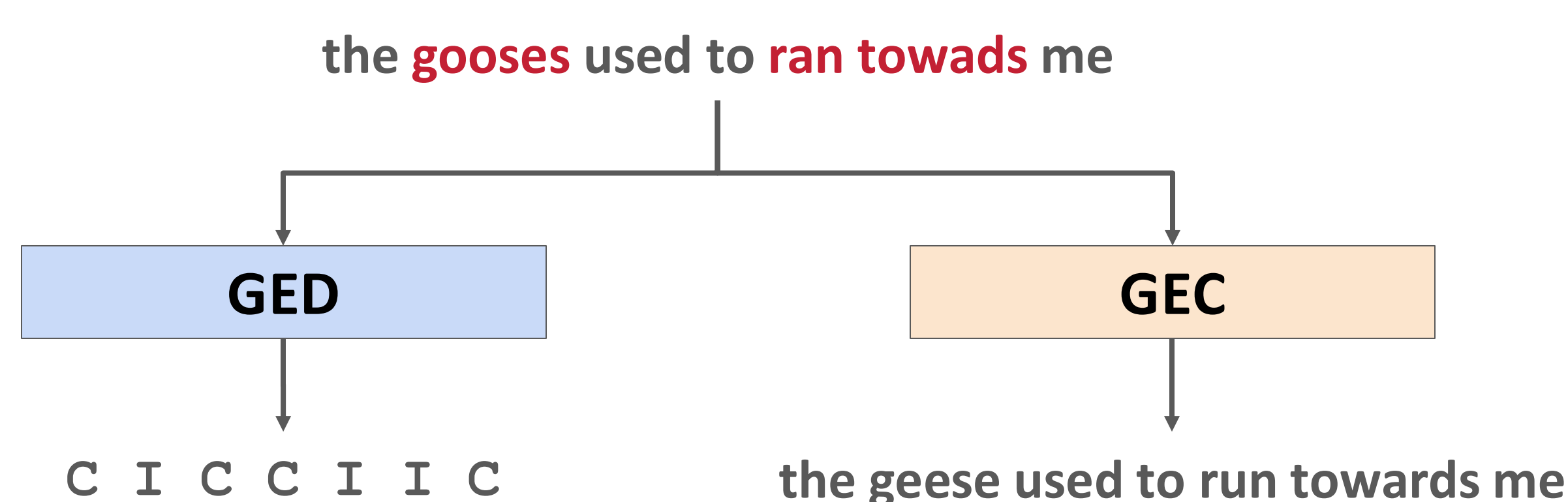
Introduction

- Grammatical Error Correction (GEC) is a key NLP task present in modern text editors, benefiting both language learners and native speakers.
- Large language models (LLMs) performance is still weaker than supervised models.
 - LLMs tend to have difficulty following GEC benchmark's correction guidelines
- In-context learning could help, but existing semantics-based example retrieval is not suited for GEC
- In this work we propose a method for in-context example selection based on predicted error structure crafted with GEC in mind

Main hypothesis: In GEC a sentence with a similar error structure should be a more informative example than a sentence with a similar meaning

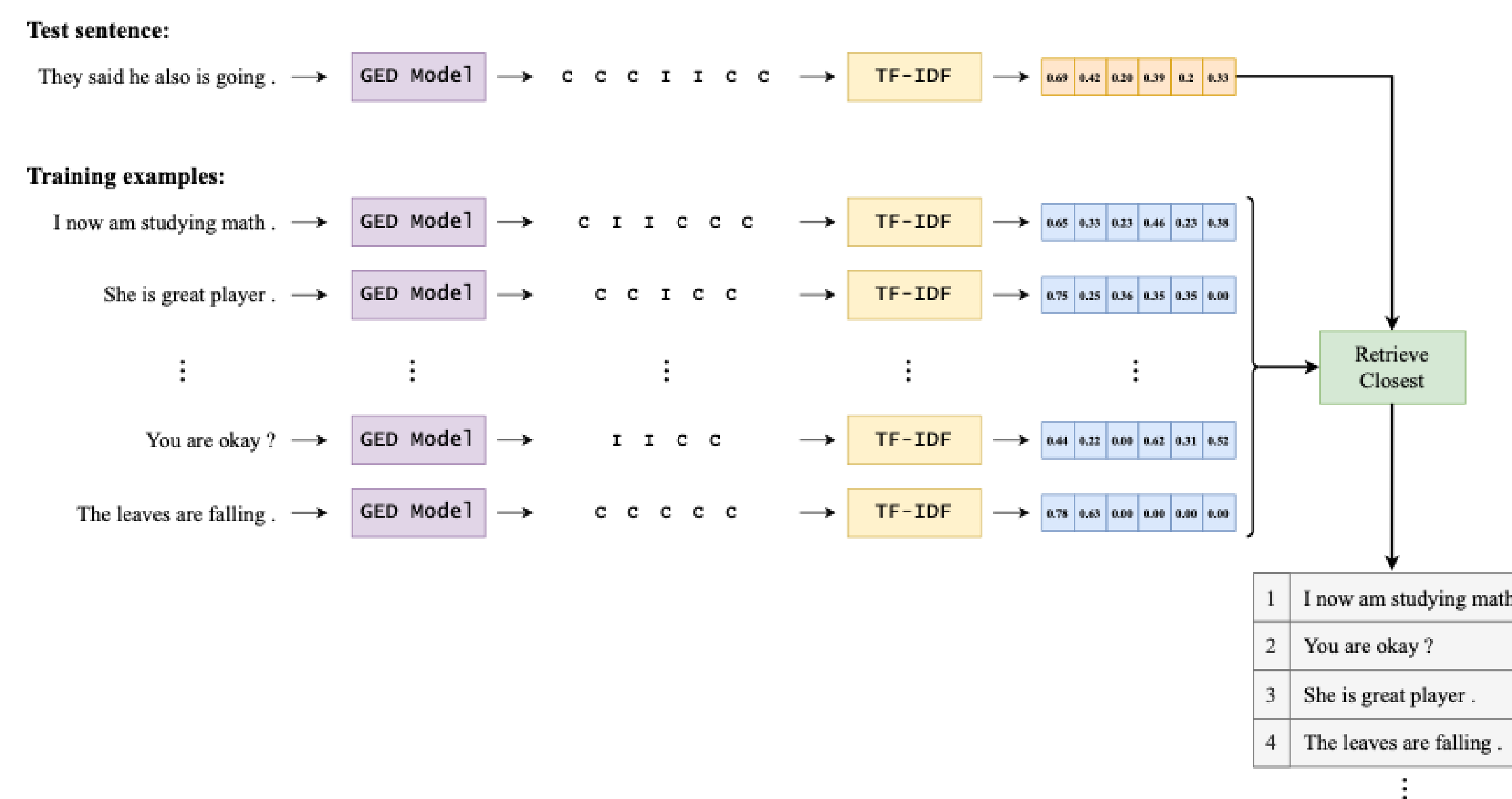
Grammatical Error Detection

- Grammatical Error Detection (GED) is the classification counterpart of GEC
- GED models require less data to be trained and achieve reasonable performance



Method

- We develop a method based on error detection labels features
- A GED model is used to predict the binary detections labels of the input sentence
- We use *TF-IDF* on the detection labels instead of the text to compute feature vectors used for retrieval

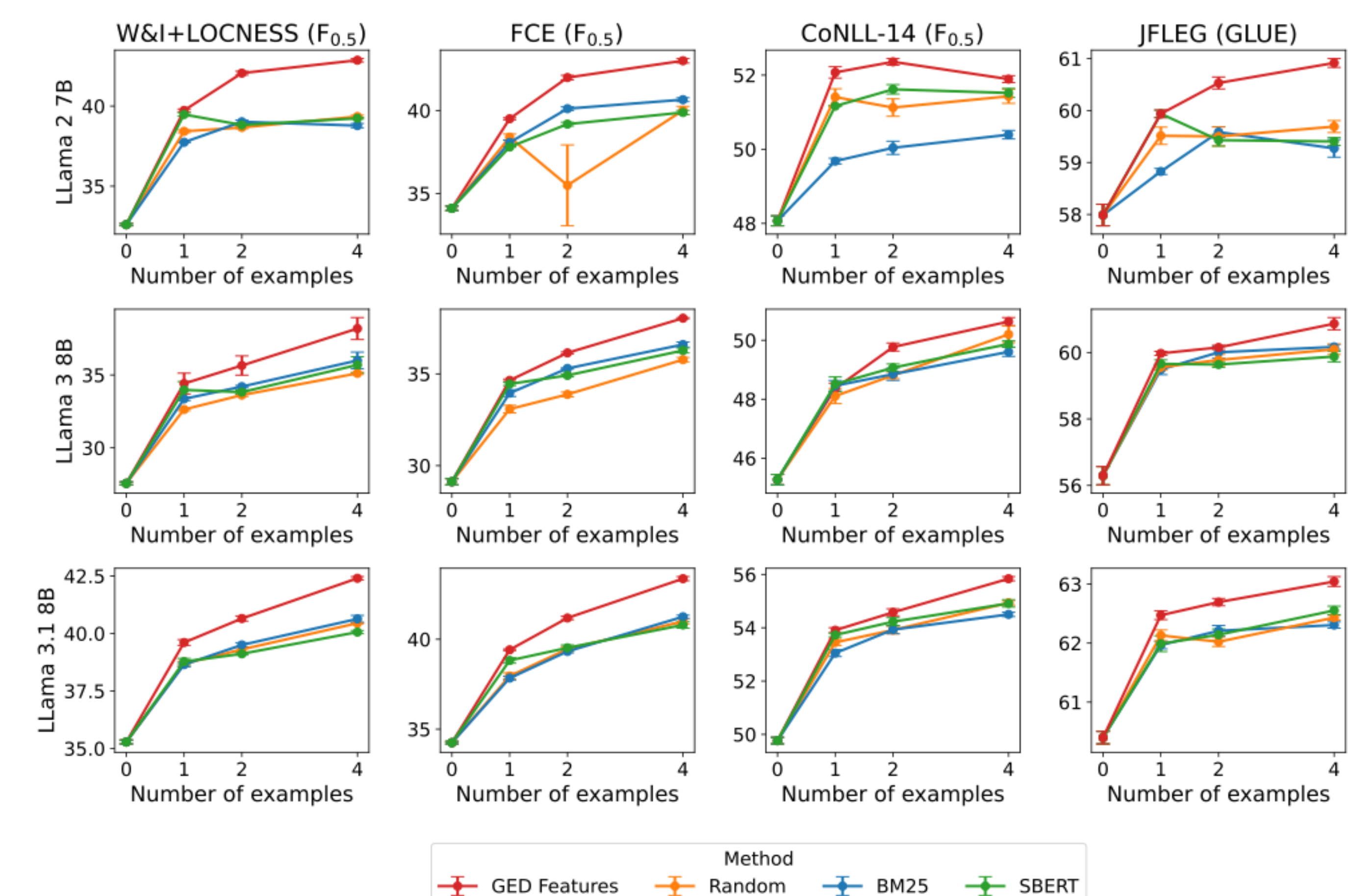


Experimental Setup

- Experiments on four main GEC benchmarks:
 - W&I+LOCNESS, FCE, CoNLL-2014, JFLEG
- We experiment using 1,2 and 4 in-context examples
- We compare against popular approaches for example selection:
 - Random, BM25, SBERT
- We evaluate the performance of the methods on different LLMs:
 - Llama 2 7B, Llama 3 8B, Llama 3.1 8B

Results

The results demonstrate that our method consistently outperforms other methods, many times requiring less in-context examples



Conclusion

We introduced a method for in-context example selection for GEC based on GED predictions. We demonstrate that our method outperforms methods based on semantic similarity across different datasets and LLMs.

Acknowledgements

This research was supported by the David R. Cheriton Graduate Scholarship, which significantly contributed to the completion of this research.