Detected Names in Red

GLiNER: Generalist Model for Named Entity Recognition using Bidirectional Transformer

Urchade Zaratiana[1,2], Nadi Tomeh[2], Pierre Holat[1,2], Thierry Charnois[2]
[1] FI Group, [2] LIPN, CNRS UMR 7030, France
zaratiana@lipn.fr
https://github.com/urchade/GLiNER

arXiv:2311.08526v1 [cs.CL] 14 Nov 2023

Abstract

Named Entity Recognition (NER) is essential in various Natural Language Processing (NLP) applications. Traditional NER models are effective but limited to a set of predefined entity types. In contrast, Large Language Models (LLMs) can extract arbitrary entities through natural language instructions, offering greater flexibility. However, their size and cost, particularly for those accessed via APIs like ChatGPT, make them impractical in resource-limited scenarios. In this paper, we introduce a compact NER model trained to identify any type of entity. Leveraging a bidirectional transformer encoder, our model, GLiNER, facilitates parallel entity extraction, an advantage over the slow sequential token generation of LLMs. Through comprehensive testing, GLiNER demonstrate strong performance, outperforming both ChatGPT and fine-tuned LLMs in zero-shot evaluations on various NER benchmarks.

1 Introduction

Named Entity Recognition plays a crucial role in various real-world applications, such as constructing knowledge graphs. Traditional NER models are limited to a predefined set of entity types. Expanding the number of entity types can be beneficial for many applications but may involve labeling additional datasets. The emergence of Large Language Models, like GPT-3 (Brown et al., 2020), has introduced a new era for open-type NER by enabling the identification of any types of entity types only by natural language instruction. This shift signifies a significant departure from the inflexibility observed in traditional models. However, powerful LLMs typically consist of billions of parameters and thus require substantial computing resources. Although it is possible to access some LLMs via APIs (OpenAI, 2023), using them at scale can incur high costs.

Recently, researchers have explored the fine-tuning of open source language models such as

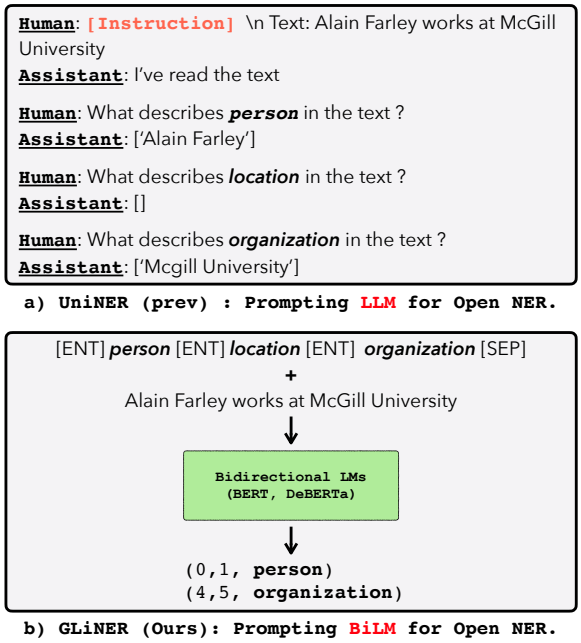

Figure 1: **BiLM for Open NER.** Previous models, such as UniNER (Zhou et al., 2023) (Fig. a), approach the task of open type NER by prompting LLMs with natural language instructions (using a multi-turn dialogue style). Our proposed GLiNER utilizes small bidirectional LMs and treats the NER task as matching entity types with textual spans in a latent space.

LLaMa (Touvron et al., 2023) for named entity recognition tasks. Wang et al. (2023), for example, introduced InstructUIE, a fine-tuned FlanT5-11B (Raffel et al., 2019; Chung et al., 2022) model on existing information extraction datasets, achieving excellent performance in zero-shot settings. Additionally, GoLLIE (Sainz et al., 2023) was introduced as an extension of InstructUIE work by by finetuning a CodeLLaMa (Rozière et al., 2023) using detailed annotation guidelines, resulting in significant performance improvements. Another recent proposal by Zhou et al. (2023), called UniversalNER, involves the fine-tuning of LLMs using diverse data from various domains annotated with ChatGPT instead of relying on standard NER