Selecting the Right Experts: Generalizing Information Extraction for Unseen Scenarios via Task-Aware Expert Weighting

University of Glasgow
School of Computing Science

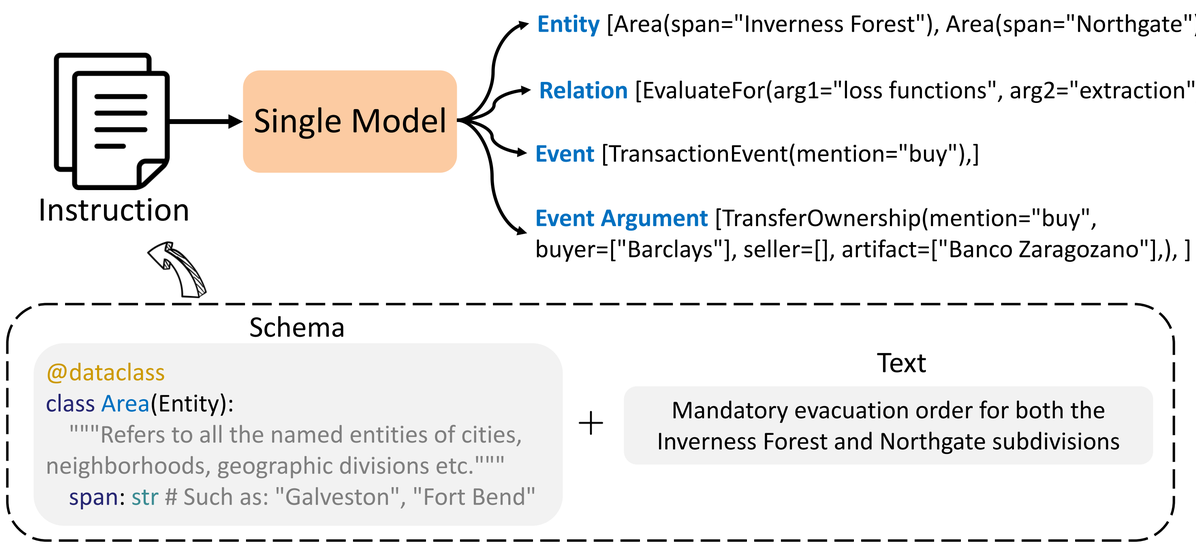
Lubingzhi Guo, Javier Sanz-Cruzado, Richard McCreadie
l.guo.1@research.gla.ac.uk,
{javier.sanz-cruzadopuig, richard.mccreadie}@glasgow.ac.uk

Paper Code

Overview

Information Extraction (IE) systems aim to convert free text into structured knowledge that can be easily consumed by downstream tasks. However, as text sources and information needs have diversified, resulting in a need for generalization solutions for unseen data. In this work, we introduce Task-Aware MoELoRA method that embeds an additional task signal into the IE process via a mixture-of-experts router to enhance the unseen performance.

1. Task and Motivation



Universal Information Extraction (UIE)

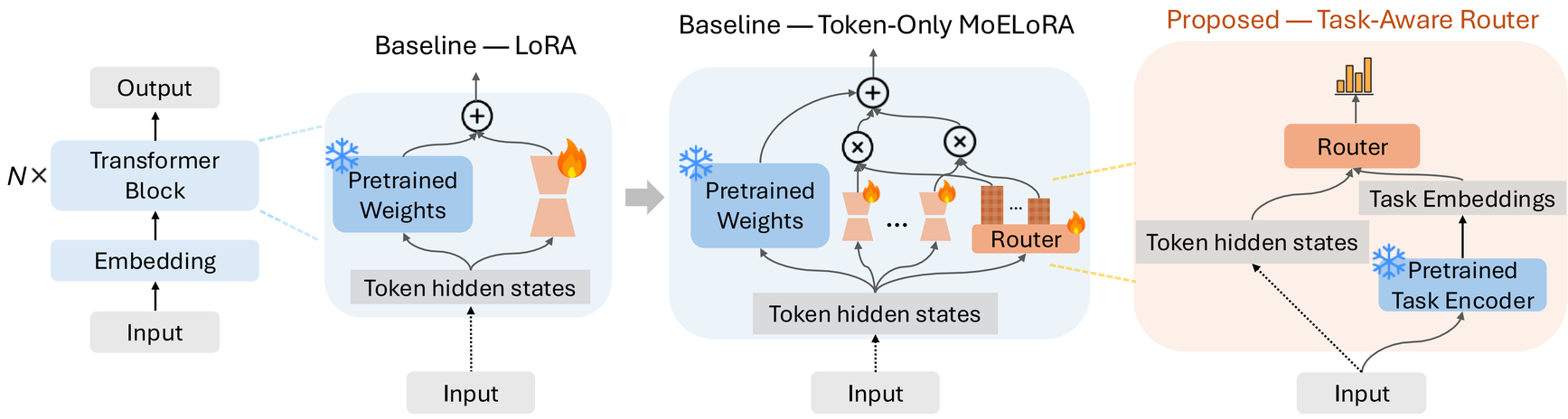
- Encodes diverse label types and structures in a unified format
- Enables a single model to perform multiple extraction tasks
  - NER -- Named Entity Recognition
  - EE&EAE -- Event Extraction and Event Argument Extraction
  - RE -- Relation Extraction
- Recently, LLMs have been adapted via LoRA fine-tuning for UIE tasks

Limitations of LoRA for UIE

- Lacks task awareness – one parameter set fits all
- Leans task patterns tied to the training data
- Limited generalization to unseen schemas / domains

How can we improve unseen performance?

2. Task-Aware MoELoRA

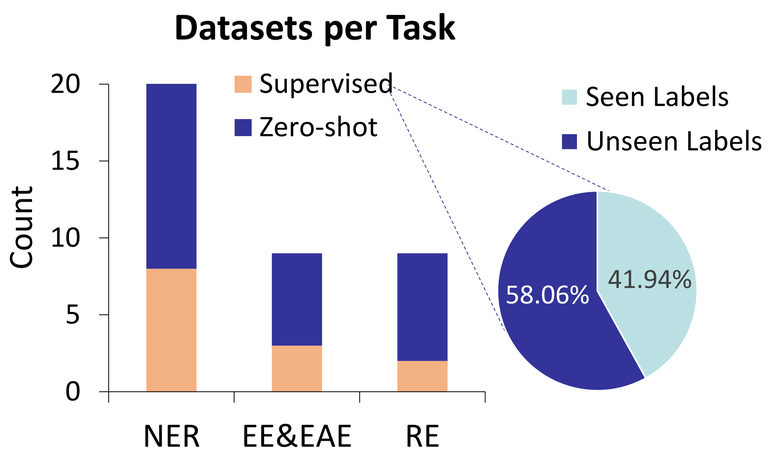


Core Idea – Task-Aware MoELoRA for **UIE**

- Baseline: MoELoRA Architecture
  - Several smaller LoRA adapters act as experts
  - A router dynamically determines expert weights
  - Limitation: routing relies only on token-level features
- **Our Approach:** Task-Aware Router
  - Token hidden states (local context)
  - Task embeddings (semantic task signal)
  - ✓ Router uses both signals to better differentiate tasks

Why use task-aware routing? It allows the router adjusts expert weights for unseen tasks by learning from similarities to previously seen tasks

3. Experimental setup



Benchmarks

- Datasets covering NER, EE&EAE, RE
- **Two splits**: Supervised (used for training) and Zero-shot (for generalization)
- **Zero-shot datasets** include:
  - Seen labels – semantically similar to labels in supervised data
  - Unseen labels – entirely new labels
- **Evaluation metric**: Average F1 score

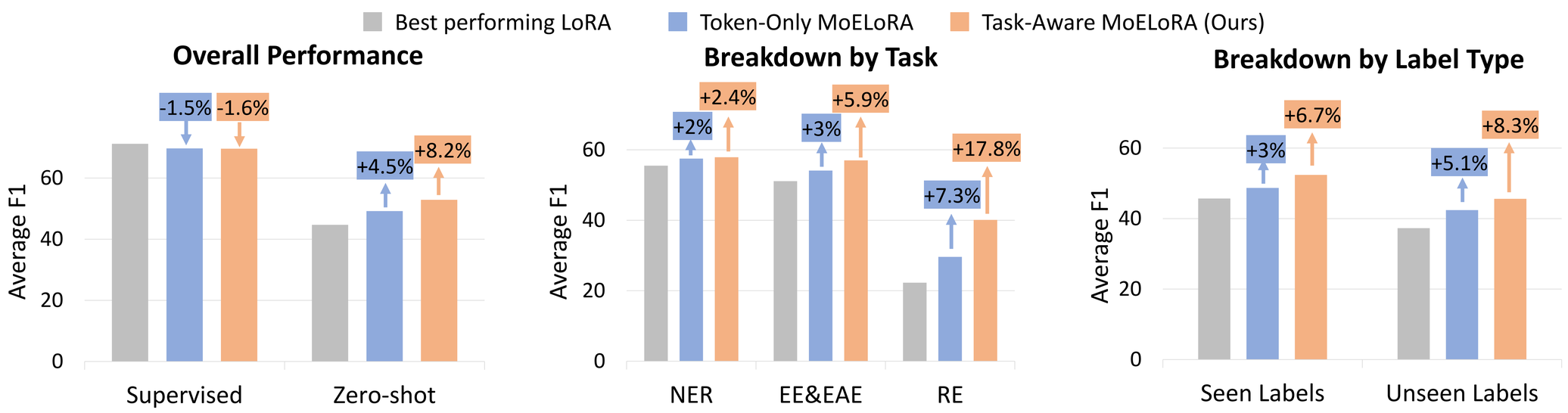
Model Configurations

- Baseline: LoRA with rank 32 and 36, ensuring comparable rank and parameter counts
- MoELoRA: 8 × rank-4 LoRA experts + router
  - Token-Only: router uses token information
  - Task-Aware: router uses the full inference prompt + token information

Backbone Pretrained Models

- LLM: Code Llama 7B
- Task Encoder: Code T5 *(see ablations in paper)*

4. Results & Conclusions



Key Insights

- The model **bridges the gap between seen and unseen tasks** by learning task-level semantics, allowing it to **generalize well** beyond the training distribution.
- It facilitates accurate information extraction in **real-world low-resource and zero-shot** scenarios, where annotated data is limited and new labels frequently appear.