



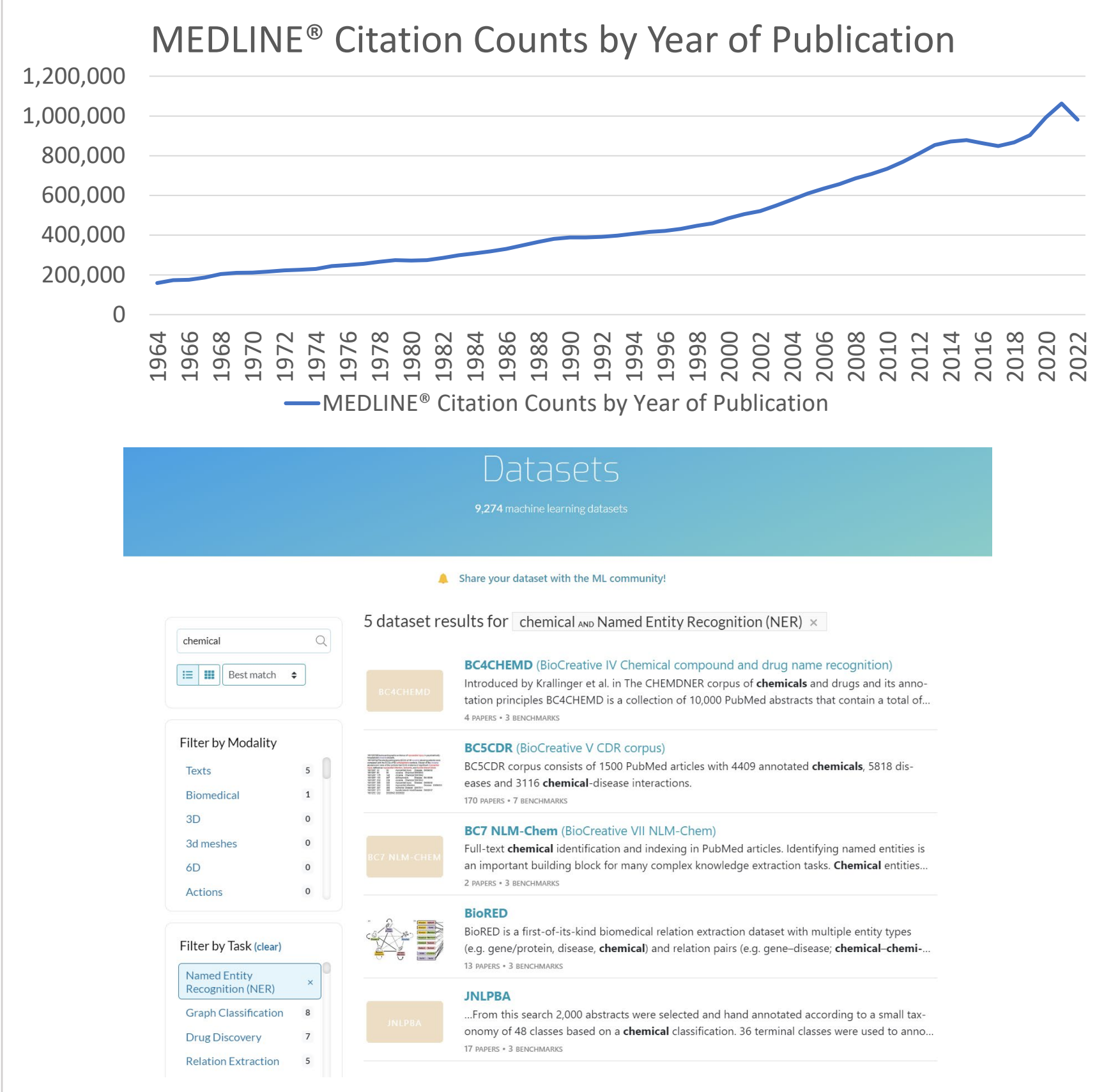
Chem-FINESE: Validating Fine-Grained Few-shot Entity Extraction through Text Reconstruction

Qingyun Wang, Zixuan Zhang, Hongxiang Li, Xuan Liu, Jiawei Han, Huimin Zhao, Heng Ji
University of Illinois at Urbana-Champaign



Background

- Millions of scientific papers are published every year.
- Human reading ability keeps almost the same over the years.
- Few benchmarks are publicly available



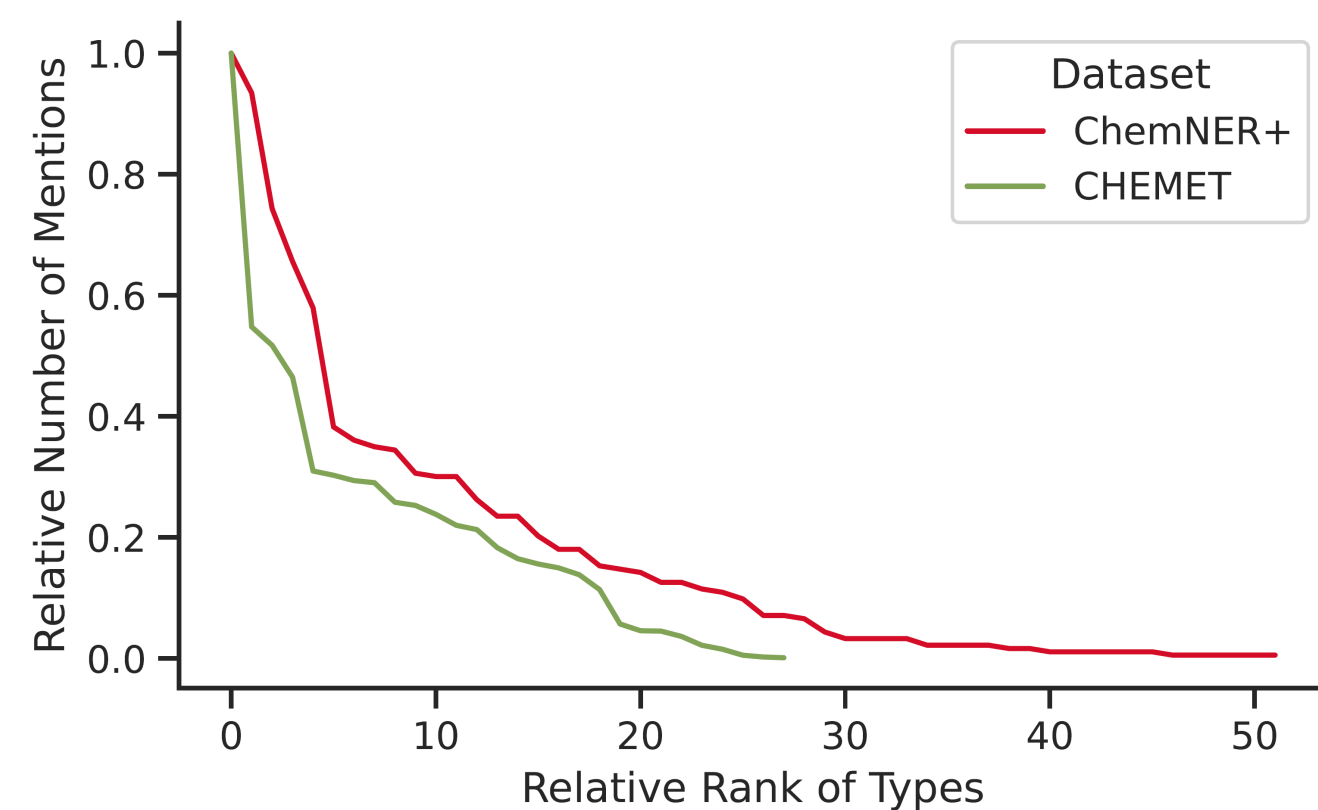
Challenges

Missing mentions

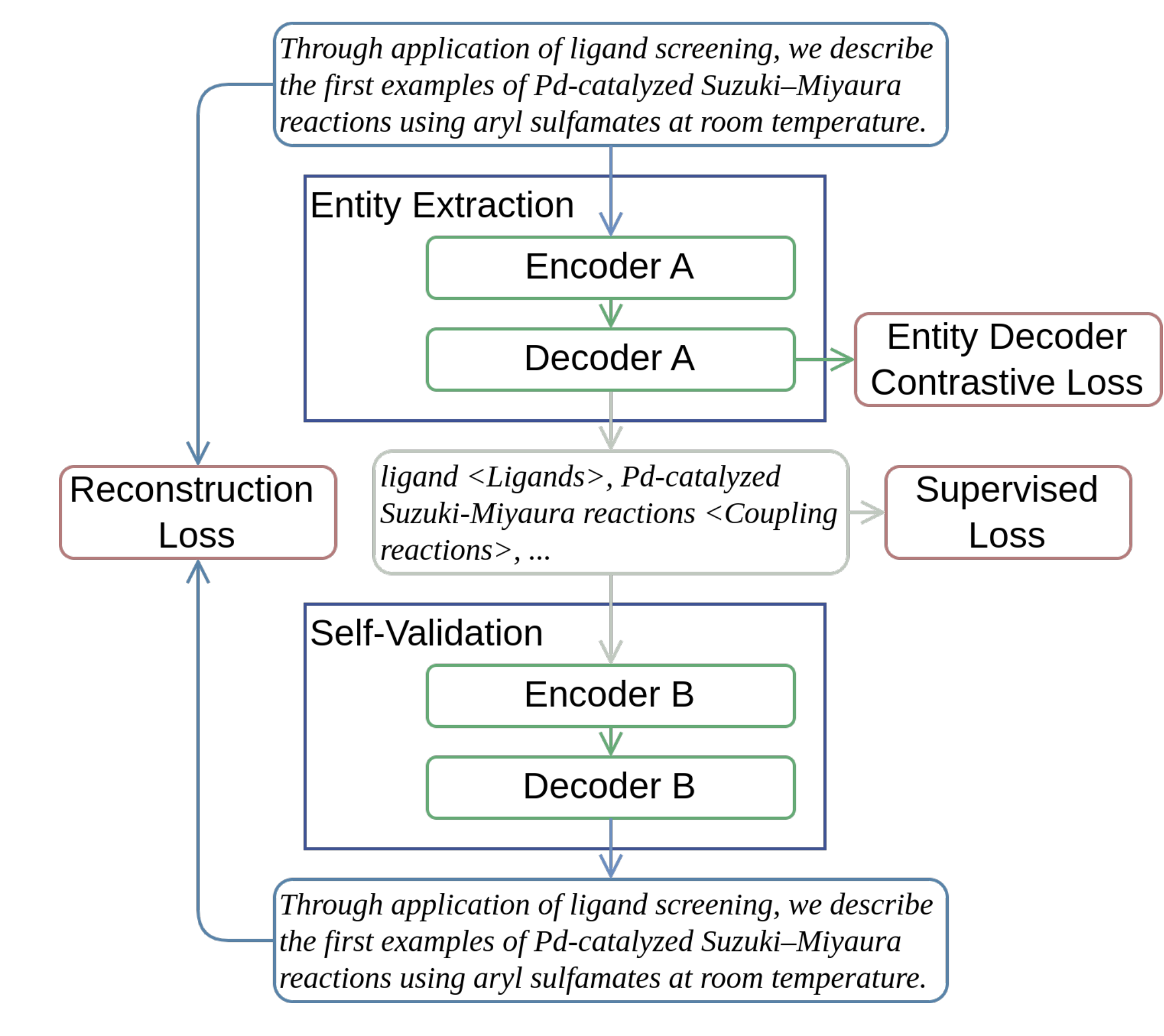
- Scientific documents contain more entities per sentence compared the sentence in general domain (3.1 in ChemNER+ vs 1.5 in CONLL2003)

Incorrect long-tail predictions

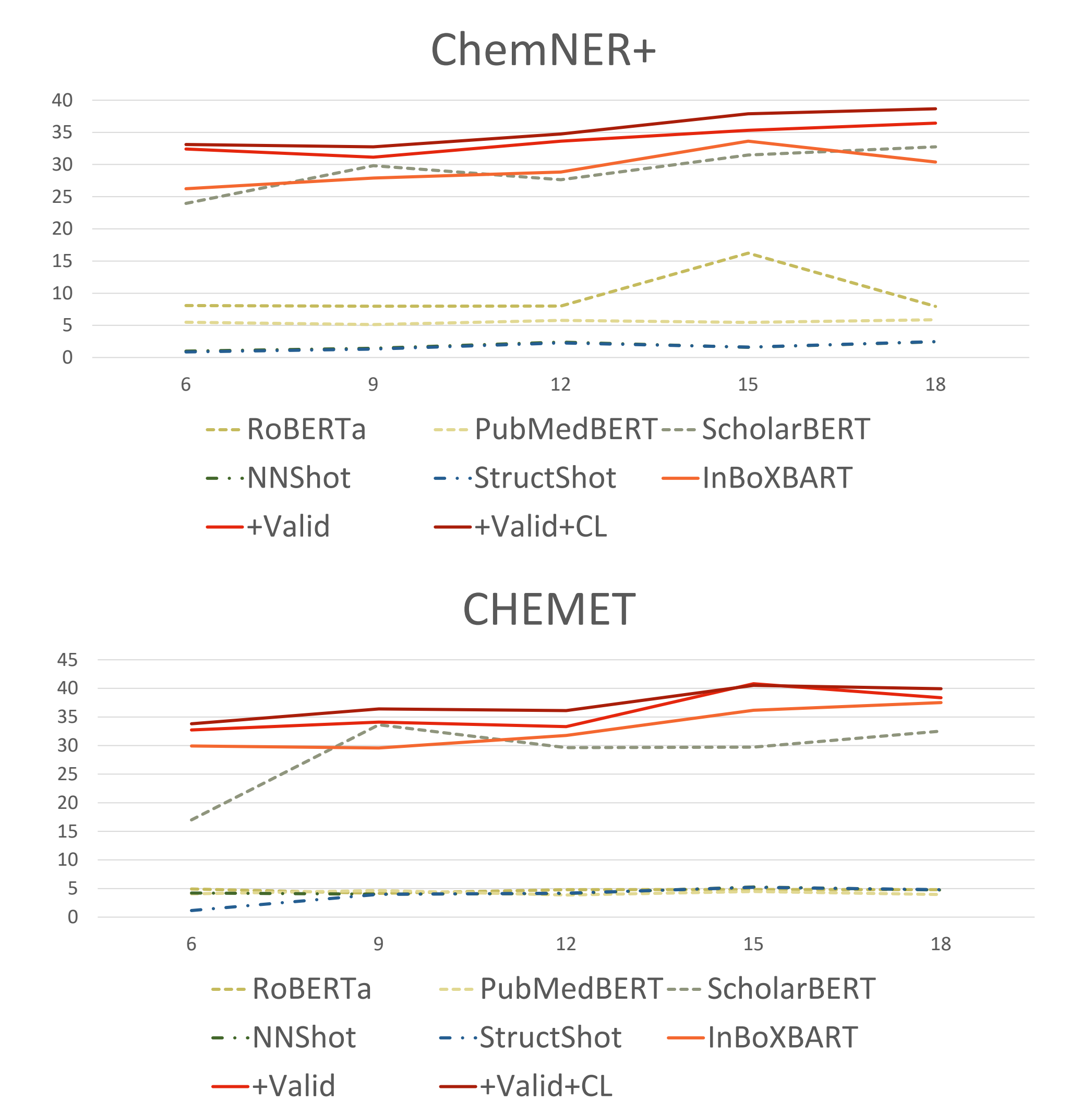
- Long-tail problems are more prevalent in the information extraction tasks in low-resource domains than in general domains.



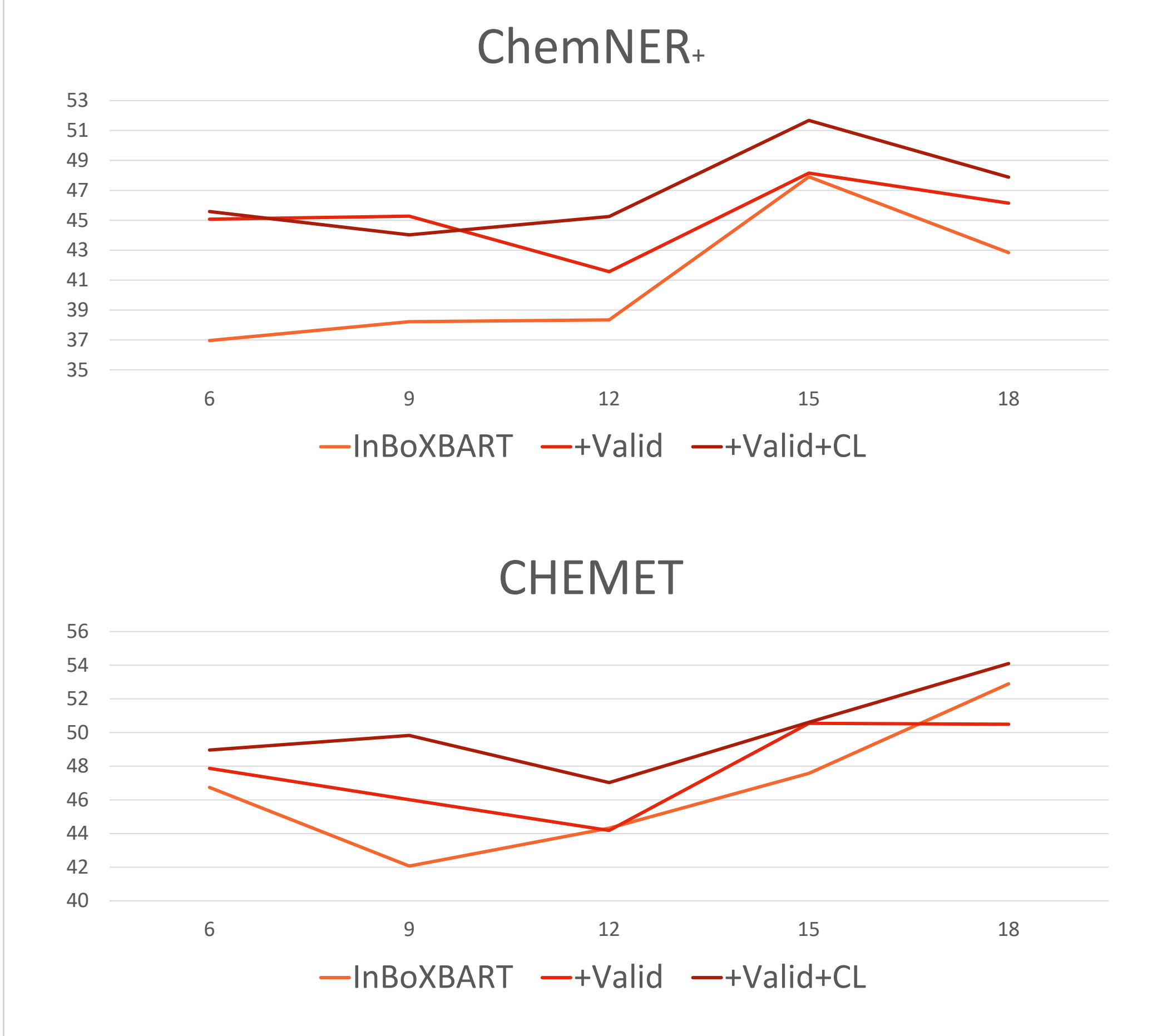
Approach Overview



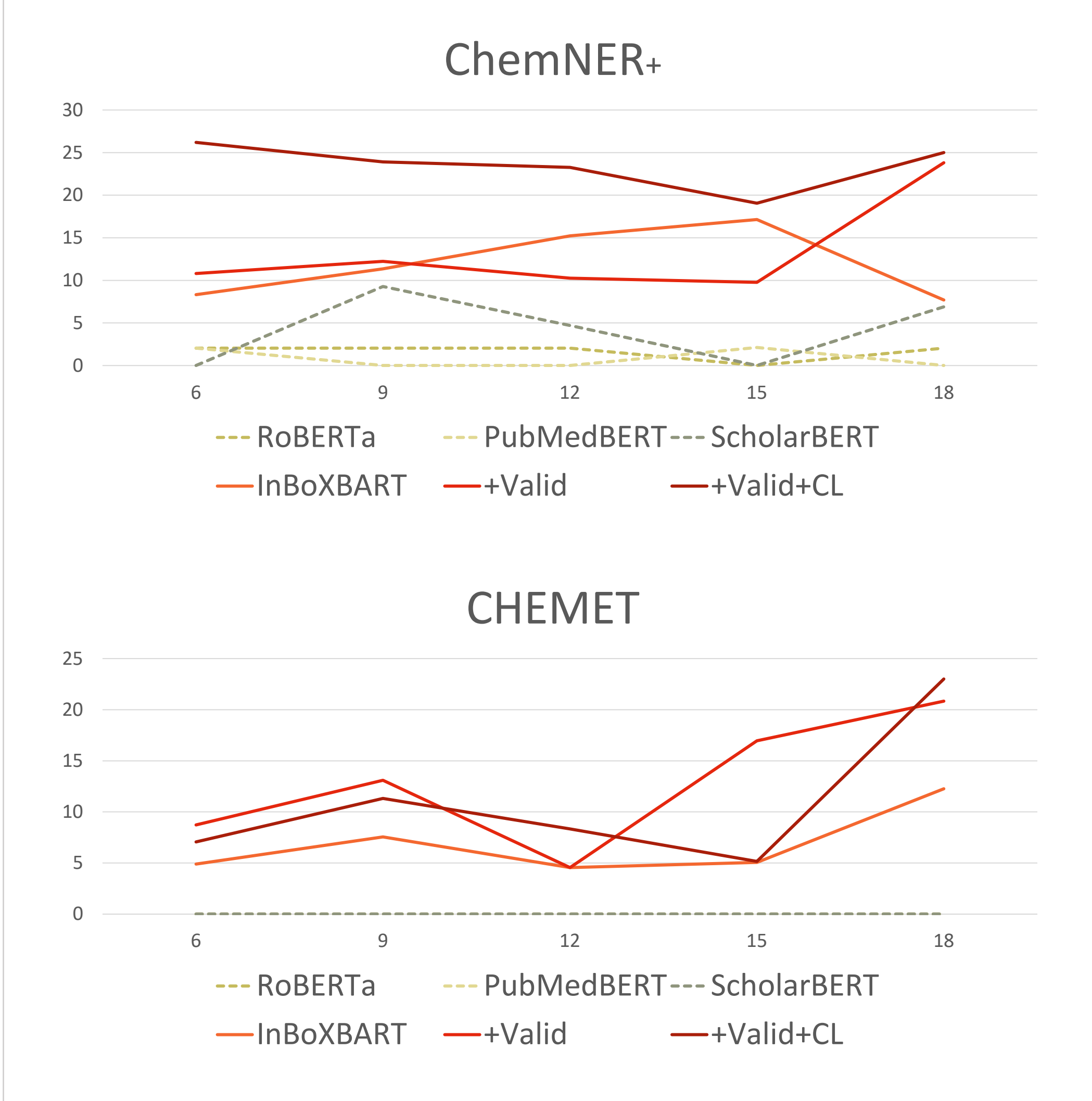
k-shot Micro-F1 Results



k-shot Mention Micro-F1 Results



k-shot Micro-F1 for Long-tail Entity



Motivation

- If the model extracts knowledge precisely, readers should be able to **reconstruct the original sentence** using the extraction results precisely
 - Self-validation module to reconstruct the original sentences based on entity extraction results
- Seq2Seq entity extraction models tend to **excessively copy** from original sentence
 - A new entity decoder contrastive loss to control the mention spans

Input

Through application of ligand screening, we describe the first examples of Pd-catalyzed Suzuki–Miyaura reactions using aryl sulfamates at room temperature.

Ground Truth

ligand <Ligands>, Pd-catalyzed Suzuki–Miyaura reactions <Coupling reactions>, aryl sulfamates <Aromatic compounds>, room temperature <Thermodynamic properties>

Sentence Reconstructed from Ground Truth

Ligands play a crucial role in Pd-catalyzed Suzuki–Miyaura reactions, which are coupling reactions that enable the synthesis of diverse organic compounds such as aryl sulfamates at room temperature, exploiting their favorable thermodynamic properties.

InBoXBART Entity Extraction Results

ligand screening <Ligands>, Pd-catalyzed Suzuki–Miyaura reactions <Coupling reactions>, aryl sulfamates <Catalysts> [Missing: room temperature <Thermodynamic properties>]

Sentence Reconstructed from Name Tagging Results

Ligand screening is conducted to identify suitable ligands for Pd-catalyzed Suzuki–Miyaura reactions, which are coupling reactions known for their efficacy in the synthesis of aryl sulfamates, acting as catalysts in the process. [Missing: room temperature <Thermodynamic properties>]

Example

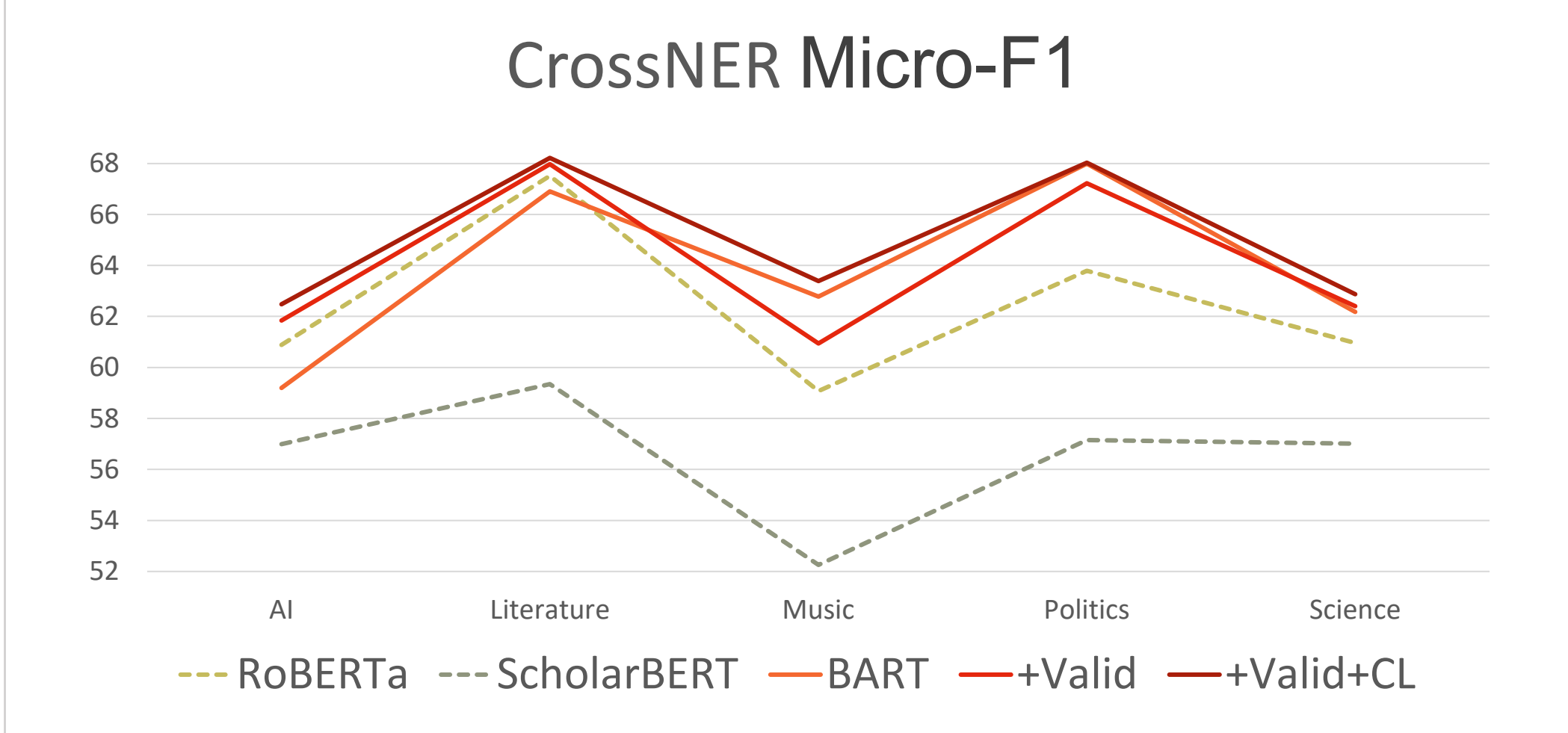
InBoXBART Several *cyclophanes*, *polycycles*, ... have been synthesized by employing a combination of *Suzuki cross-coupling and metathesis* Coupling reactions.

+ Valid Several *cyclophanes* Heterocyclic Compounds, *polycycles*, ... have been synthesized by employing a combination of *Suzuki cross-coupling and metathesis* Organic reactions.

+ Valid + CL Several *cyclophanes* Heterocyclic Compounds, *polycycles* Biomolecules, ... have been synthesized by employing a combination of *Suzuki cross-coupling* Coupling reactions and *metathesis* Chemical properties.

Ground Truth Several *cyclophanes* Aromatic Compounds, *polycycles* Organic polymers, ... have been synthesized by employing a combination of *Suzuki cross-coupling* Coupling reactions and *metathesis* Substitution reactions.

Case Study: CrossNER



Dataset

- ChemNER+
 - Based on available sentences from ChemNER (Wang et al., 2021)
 - Annotated by two Chemistry Ph.D. students
 - Covering 59 fine-grained chemistry types with 742 sentences
- CHEMET (Sun et al., 2021)
 - Consisting of 30 fine-grained organic chemical types
- Few-shot Setup
 - Randomly sample a subset based on the frequency of each type class
 - Choose the values 6, 9, 12, 15, 18 as the potential maximum entity mentions for k

Code and Dataset:

github.com/EagleW/Chem-FINESE