# Taxonomy-Guided Zero-Shot Recommendations with LLMs
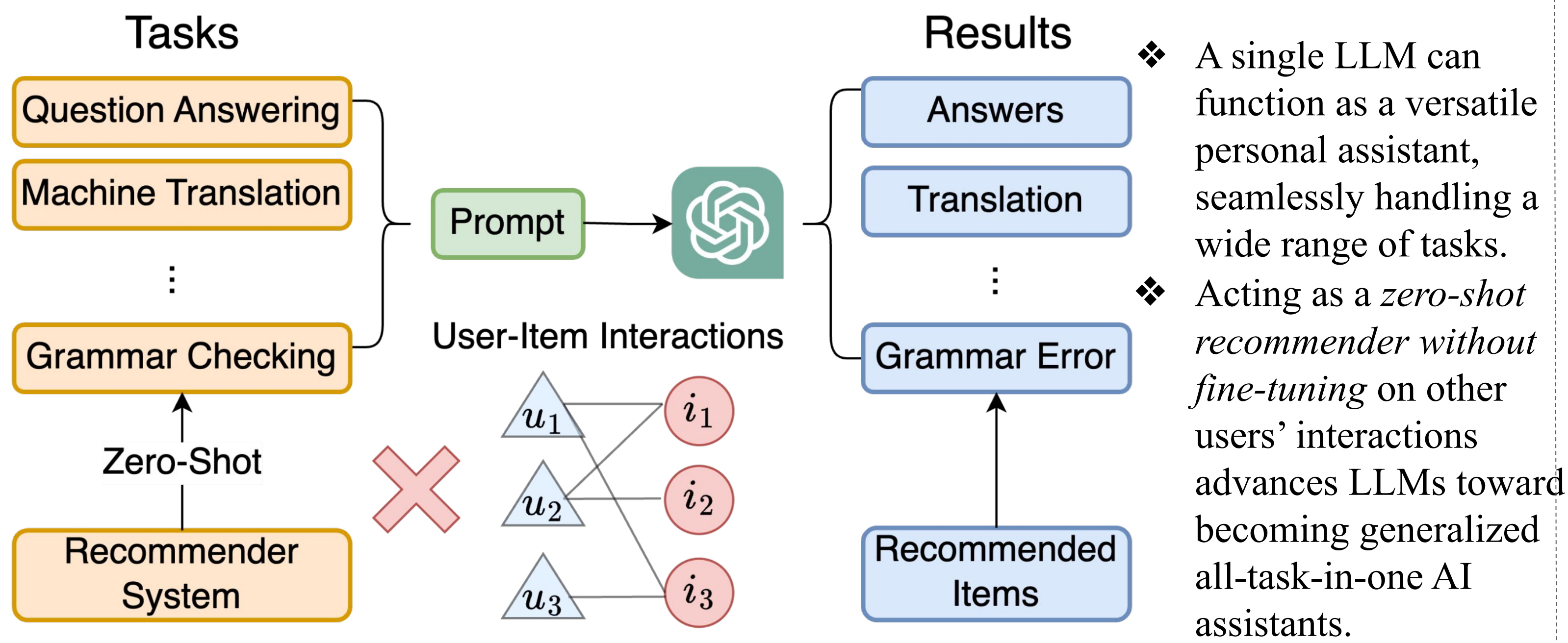
Yueqing Liang[1]  Liangwei Yang[2]  Chen Wang[2]  Xiongxiao Xu[1]  Philip S. Yu[2]  Kai Shu[3]

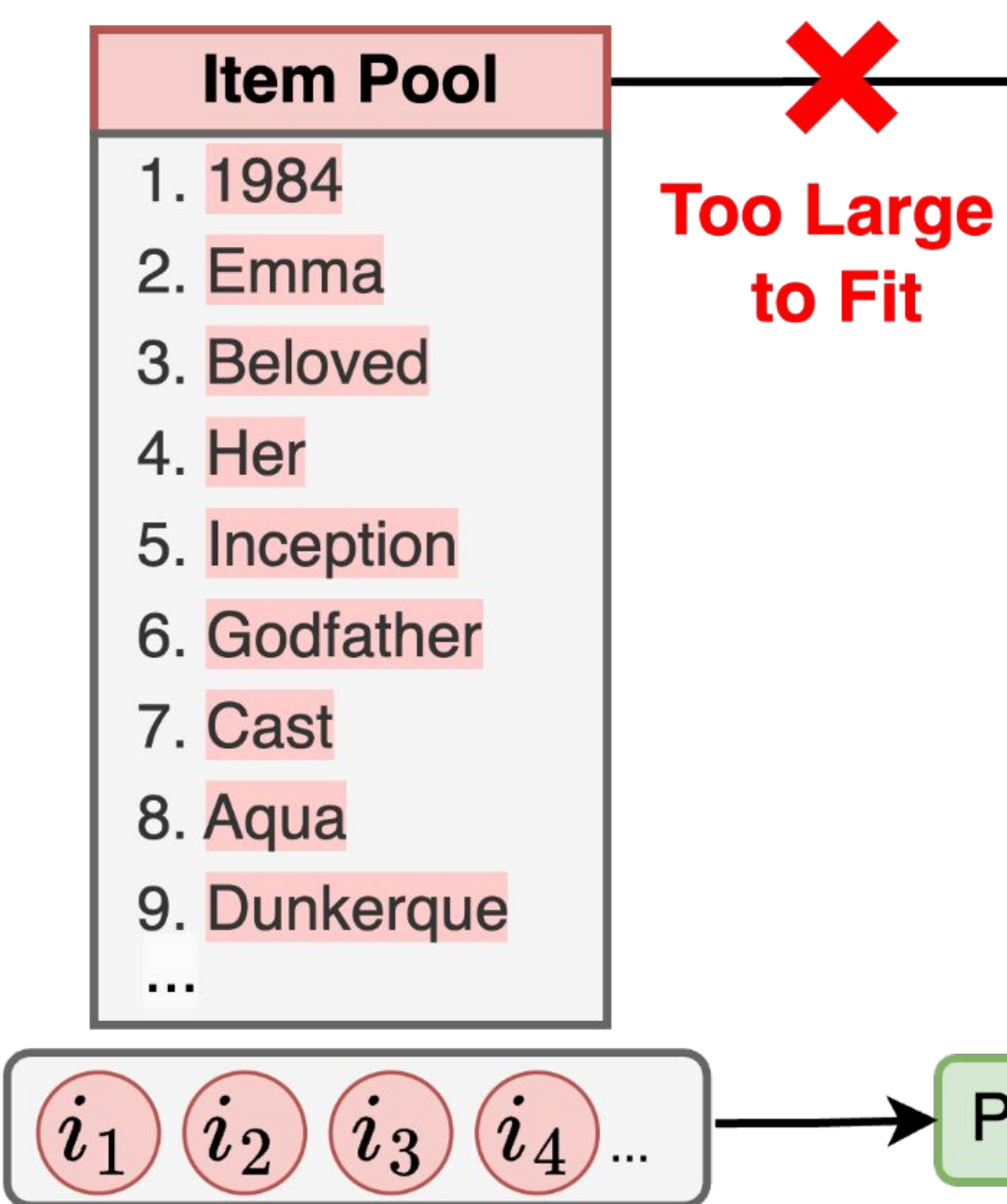ILLINOIS TECH [1]   UIC [2]   EMORY UNIVERSITY [3]
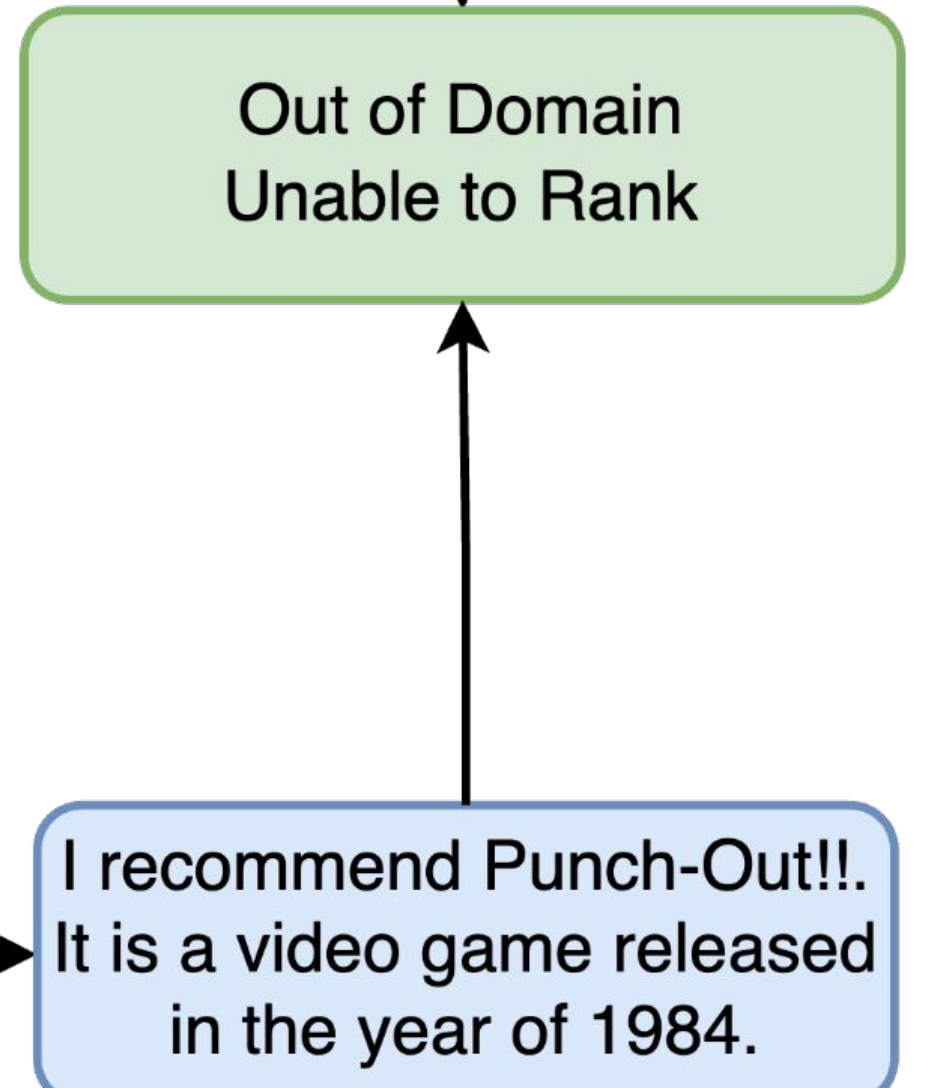
## Motivation

Tasks
- Question Answering
- Machine Translation
- ⋮
- Grammar Checking

Zero-Shot

- Recommender System

Prompt →

User-Item Interactions

Results
- Answers
- Translation
- ⋮
- Grammar Error
- Recommended Items

- ❖ A single LLM can function as a versatile personal assistant, seamlessly handling a wide range of tasks.
- ❖ Acting as a *zero-shot recommender without fine-tuning* on other users' interactions advances LLMs toward becoming generalized all-task-in-one AI assistants.

## Challenges

(1) Limited Prompt Length

Item Pool
1. 1984
2. Emma
3. Beloved
4. Her
5. Inception
6. Godfather
7. Cast
8. Aqua
9. Dunkerque
...

Too Large to Fit ✗

(2) Vague Item Title

| 1984: | Year? |
| Dunkerque: | Place? |
| Emma: | People? |
| Her: | Pronoun? |

(3) Unconstrained Generation

Out of Domain Unable to Rank

$i_1$ $i_2$ $i_3$ $i_4$ ... → Prompt → ? → I recommend Punch-Out!!. It is a video game released in the year of 1984.

## Proposed Method: TaxRec

### (a) One-time Taxonomy Categorization

Taxonomy Generation Prompt

Item Pool
1. 1984
2. Emma
3. Beloved
...

→ Categorization Prompt → Retrieve Knowledge →

Retrieve Knowledge →

Taxonomy
- Genre: Fiction, Thriller, Romance, ...
- Theme: Love, Science, Art, ...
- Language: English, French, ...

Categorized Item Pool

| | Genre | Theme | ... |
| 1. 1984: | Fiction, | Power | ... |
| 2. Emma: | Romance, | Love, | ... |
| 3. Beloved: | Fiction, | Slavery, | ... |
| ... | | | |

### (b) LLM-based Recommendation

$i_1$ $i_2$ $i_3$ $i_4$ ... →

$i_1$ Genre: ..., Theme: ...,
$i_2$ Genre: ..., Theme: ...,
$i_3$ Genre: ..., Theme: ...,
$i_4$ Genre: ..., Theme: ...,
...

→ Recommendation Prompt → Genre: ..., Theme: ..., ... → Parse → Match →

🥇 To Kill a Mockingbird
🥈 The Catcher in the Rye
🥉 The Great Gatsby

(1) Historical Interactions   (2) Categorized Historical Interactions   (3) Categorized Recommendation   (4) Recommended Items

## Contributions

- ❖ We propose **TaxRec**, a taxonomy-guided method to retrieve LLM's knowledge and enhance LLM's ability as personal recommenders.
- ❖ By systematically categorizing and organizing items through a taxonomy framework, we address the key challenges faced by LLM-based recommendation systems.
- ❖ Experiments show significant improvements of TaxRec over current zero-shot recommenders, proving the effectiveness of our proposed taxonomy-guided method.

## Experiments

### Major experiments

| Datasets | Methods | R@1 | R@5 | R@10 | N@1 | N@5 | N@10 |
|---|---|---|---|---|---|---|---|
| Movie | Popularity | 0.005 | 0.035 | 0.160 | 0.005 | 0.020 | 0.061 |
| | AvgEmb | 0.000 | 0.040 | 0.100 | 0.000 | 0.020 | 0.039 |
| | ZESRec | 0.032 | 0.095 | 0.222 | 0.032 | 0.059 | 0.099 |
| | UniSRec | 0.032 | 0.063 | 0.143 | 0.032 | 0.048 | 0.074 |
| | RecFormer | 0.016 | 0.141 | 0.219 | 0.016 | 0.077 | 0.103 |
| | DirectRec-LLaMA2 | 0.033 | 0.058 | 0.085 | 0.033 | 0.042 | 0.051 |
| | TAXREC-LLaMA2 | 0.045 | 0.126 | 0.190 | 0.045 | 0.095 | 0.148 |
| | DirectRec-GPT4 | 0.045 | 0.100 | 0.180 | 0.045 | 0.074 | 0.099 |
| | TAXREC-GPT4 | **0.060** | **0.175** | **0.300** | **0.060** | **0.117** | **0.157** |
| Book | Popularity | 0.030 | 0.070 | 0.155 | 0.030 | 0.046 | 0.073 |
| | AvgEmb | 0.005 | 0.075 | 0.115 | 0.005 | 0.038 | 0.051 |
| | ZESRec | 0.005 | 0.070 | 0.115 | 0.005 | 0.037 | 0.051 |
| | UniSRec | 0.000 | 0.050 | 0.085 | 0.000 | 0.025 | 0.035 |
| | RecFormer | 0.010 | 0.060 | 0.125 | 0.010 | 0.033 | 0.054 |
| | DirectRec-LLaMA2 | 0.001 | 0.010 | 0.015 | 0.001 | 0.004 | 0.006 |
| | TAXREC-LLaMA2 | 0.040 | 0.099 | 0.150 | 0.040 | 0.072 | 0.109 |
| | DirectRec-GPT4 | 0.000 | 0.015 | 0.025 | 0.000 | 0.006 | 0.010 |
| | TAXREC-GPT4 | **0.070** | **0.150** | **0.240** | **0.070** | **0.109** | **0.138** |

Our proposed TAXREC significantly outperforms all baseline methods, showing the effectiveness of prompting LLMs with our taxonomy framework in a zero-shot scenario.
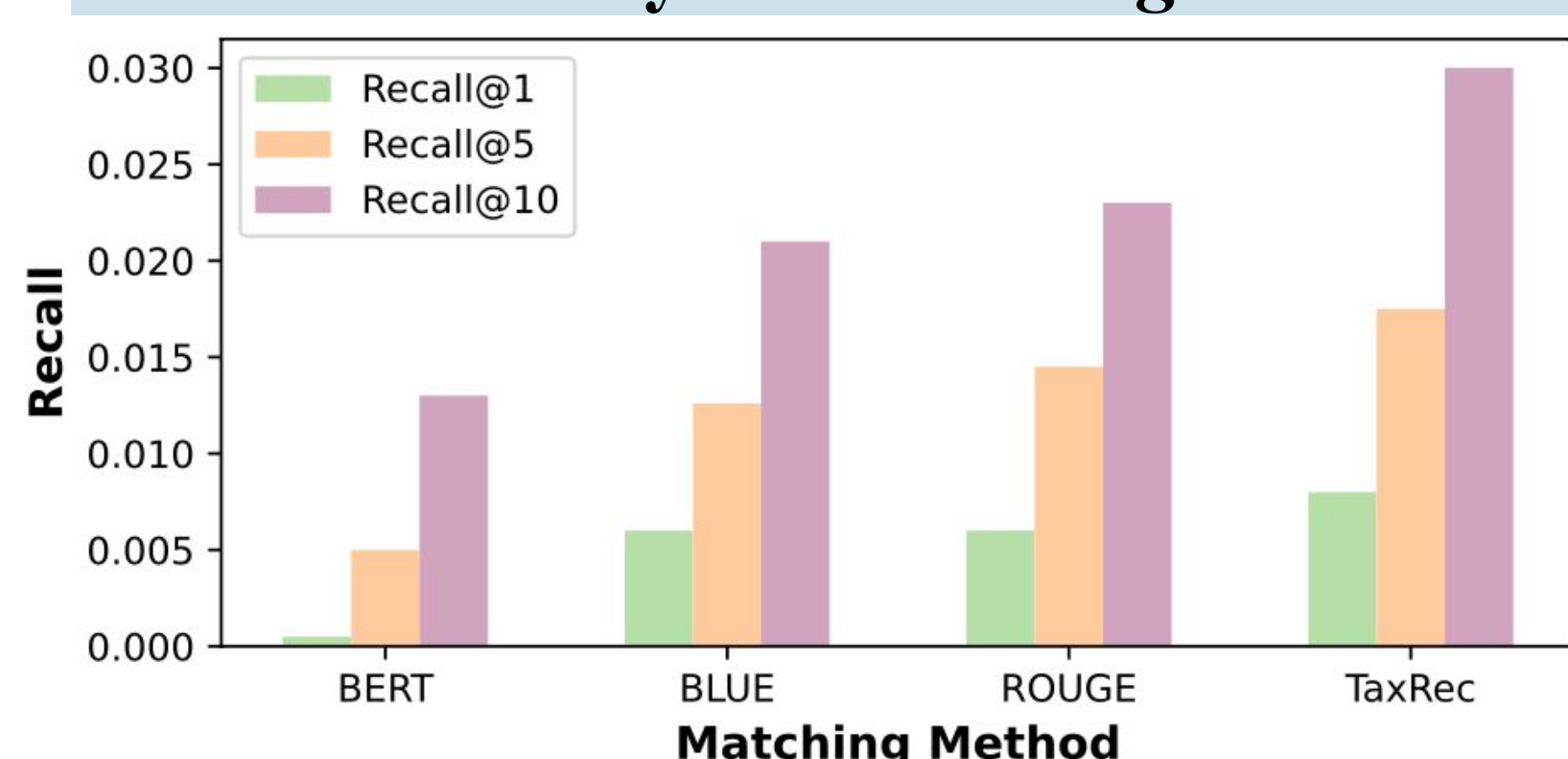
### Ablation

| Variant | Movie | | Book | |
|---|---|---|---|---|
| | R@10 | N@10 | R@10 | N@10 |
| w/o Tax | 0.112 | 0.078 | 0.025 | 0.010 |
| w/o Match | 0.254 | 0.127 | 0.165 | 0.100 |
| TAXREC | **0.300** | **0.157** | **0.265** | **0.132** |

### Parameter Analysis - # *Features in Taxonomy*
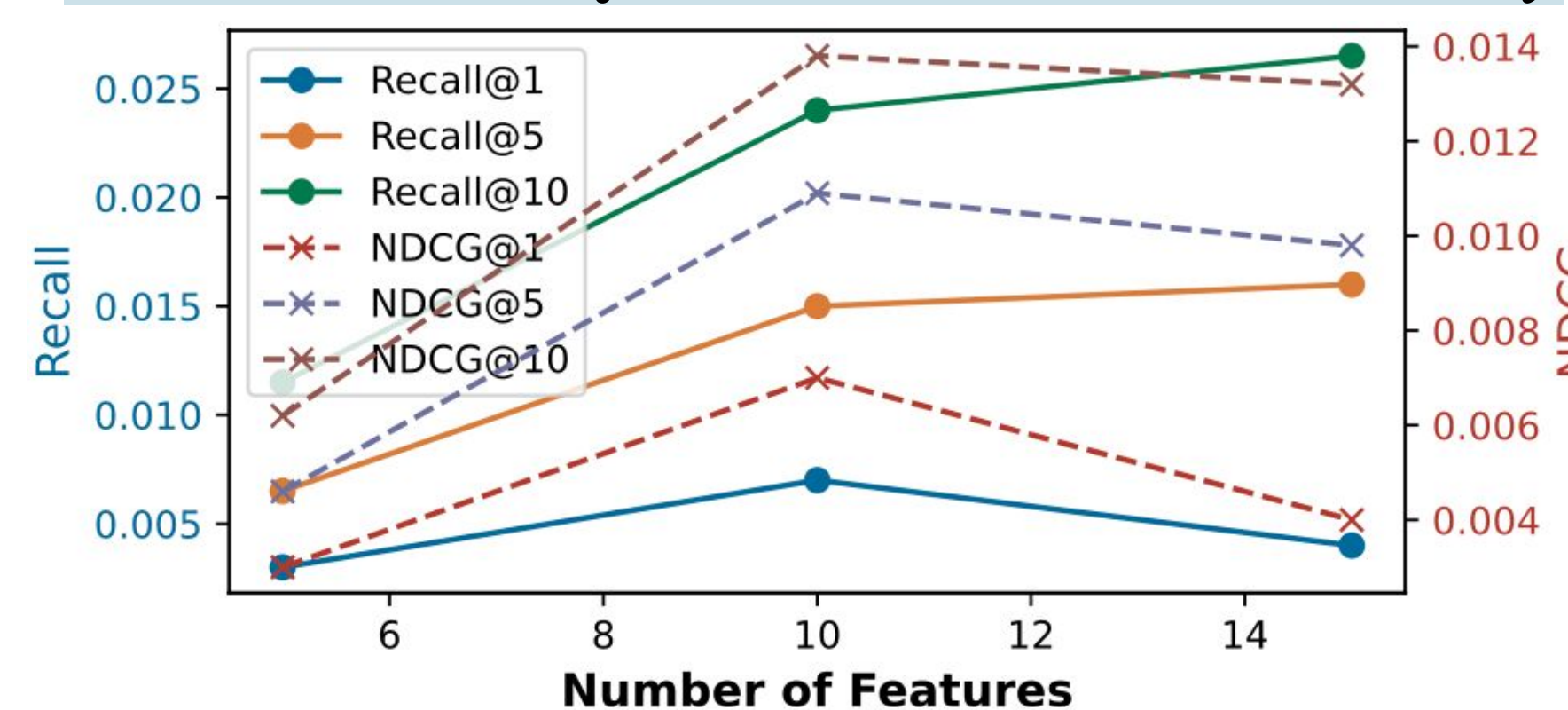
- ❖ More features lead to better performance.
- ❖ However, excessive features may introduce noise which lead to performance reduction.

### Parameter Analysis - *Matching Score Method*

- ❖ Learning-based method (BERT embedding) performs poorly on both datasets.
- ❖ Rule-based methods are better suited to TAXREC.

## More Information

Paper

Code

The 31st International Conference on Computational Linguistics

COLING 2025 · Abu Dhabi