Factiverse and IAI at CheckThat! 2025

Adaptive ICL for Claim Normalization

Pratuat Amatya
PhD Student
Factiverse AS, University of Stavanger (IAI Group)
Stavanger, Norway

Vinay Setty
Associate Professor
University of Stavanger (IAI Group)
Stavanger, Norway

Problem

- This task aims to develop methods to simplify noisy, unstructured social media posts into a concise form.
- Task consisted of two settings: **monolingual and zero-shot**.
- We focused mainly on **monolingual** settings under time constraints.
- We explored **Fine-tuning** and **In-Context Learning** based method.

Dataset

Datasets divided into two sub-tasks

1. Monolingual
    - 13 languages
    - **train**, **dev** and **test** data
    - diverse dataset size
2. Zero-shot
    - 8 languages
    - only **test** data

Language	Train	Dev	Test
Monolingual Languages			
Arabic (ara)	470	118	100
German (deu)	386	101	100
English (eng)	11374	1171	1285
French (fra)	1174	147	148
Hindi (hi)	1081	50	100
Marathi (mr)	137	50	100
Indonesian (msa)	540	137	100
Pubjabi (pa)	445	50	100
Polish (pol)	163	41	100
Portuguese (por)	1735	223	225
Spanish (spa)	3458	439	439
Tamil (ta)	102	50	100
Thai (tha)	244	61	100
Zero-shot Languages			
Bengali (bn)	–	–	81
Czech (ces)	–	–	123
Greek (ell)	–	–	156
Korean (kor)	–	–	274
Dutch (nld)	–	–	177
Romanian (ron)	–	–	141
Telgu (te)	–	–	116

Proposed Solution

1. **Zero-shot Prompting**
2. In-context Learning
   - Fixed ICL
   - Adaptive ICL
3. Fine-tuning

This approach involves using a carefully designed prompt with explicit instructions to perform claim normalization.

```
You are a helpful AI assistant. Given a noisy and unstructured social media
post, rewrite it as a simple and concise statement.


Produce concise statement for the following post (delimited by ###).
The original language of the post is {language}.
###
{post}
###

Always produce a valid json string as a final output using the format below.
{{
    "normalized_claim" < generated normalized claim translated in {language}
    language>
}}
```

Proposed Solution

This approach leverages LLM's ability to learn from examples embedded within the input prompt.

1. Zero-shot Prompting

2. **In-context Learning**

    - Fixed ICL
    - Adaptive ICL

3. Fine-tuning

```
You are a helpful AI assistant. Given a noisy and unstructured social media
post, rewrite it as a simple and concise statement.

Below are some examples of the task intended with input post and expected
outcome.

------------- Examples -------------

{examples}

--------- End of Examples ----------

Produce concise statement for the following post (delimited by ###).
The original language of the post is {language}.
###
{post}
###

Always produce a valid json string as a final output using the format below.
{{
    "normalized_claim" < generated normalized claim translated in {language}
    language>
}}
```

Proposed Solution

1. Zero-shot Prompting
2. In-context Learning
   - **Fixed ICL**
   - Adaptive ICL
3. Fine-tuning

Fixed-ICL approach inserts a fixed number (**K**) of examples selected from the training set based on their similarity to the input.

Formally, the posterior probability of generating the true claim can be expressed as

$$P(\, y \mid x, k\,) = f\big(x, E_k(x); \phi_{\mathrm{LLM}}\big)$$

x is input text,
$E_k(x)$ is an example set of k numbers of documents that are most similar to x,
ϕ_{LLM} are the decoder parameters of the pretrained LLM,

Proposed Solution

1. Zero-shot Prompting
2. In-context Learning
   - Fixed ICL
   - **Adaptive ICL**
3. Fine-tuning

The main idea behind the **Adaptive-ICL** approach is to dynamically determine the number of examples based on similarity metrics between the input and candidate examples, in contrast to fixed size of the example set in FICL.

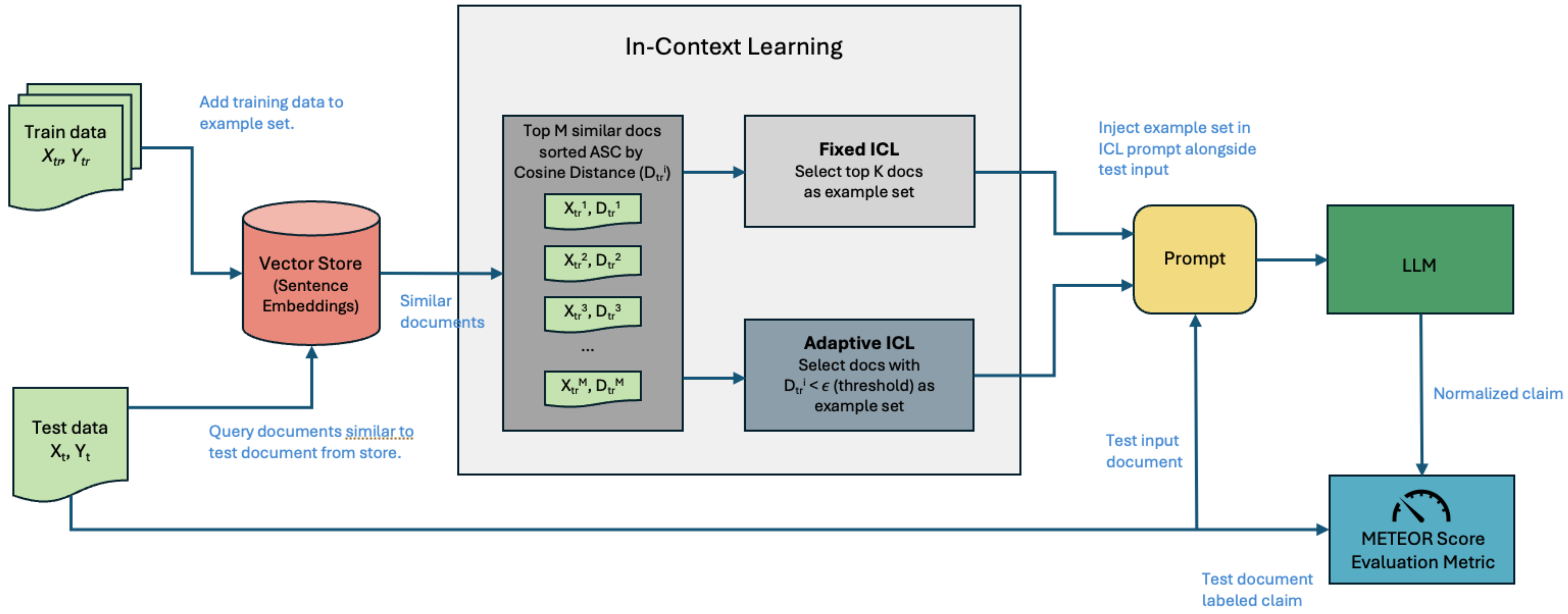
Formally, the posterior probability of generating the true claim can be expressed as

$$P(y \mid x, \epsilon) = f\big(x, E_\epsilon(x); \phi_{\mathrm{LLM}}\big)$$

x is input text,
$E_\epsilon(x)$ is an example set of varying numbers of documents that has cosine distance below threshold ϵ,
ϕ_{LLM} are the decoder parameters of the pretrained LLM,

Proposed Solution

In-Context Learning based method architecture



Proposed Solution

1. Zero-shot Prompting
2. In-context Learning
    - Fixed ICL
    - Adaptive ICL
3. **Fine-tuning**

- **google/flan-t5-large** model fine tuned on available training data
- key challenge was training data scarcity of some of the languages
- to address the data scarcity, we used translations of high-resource language data into low-resource languages

Results: Monolingual setting

- Method focuses on five languages from **monolingual** setting: English, German French, Spanish and Portuguese

- Result evaluation is done using average METEOR score for all normalized claim against hidden golden claims.

- Among four variants of models, **Fine-tuned model** showed superior performance across majority of languages except for Portuguese, whereas **Zero-shot** model performed worst

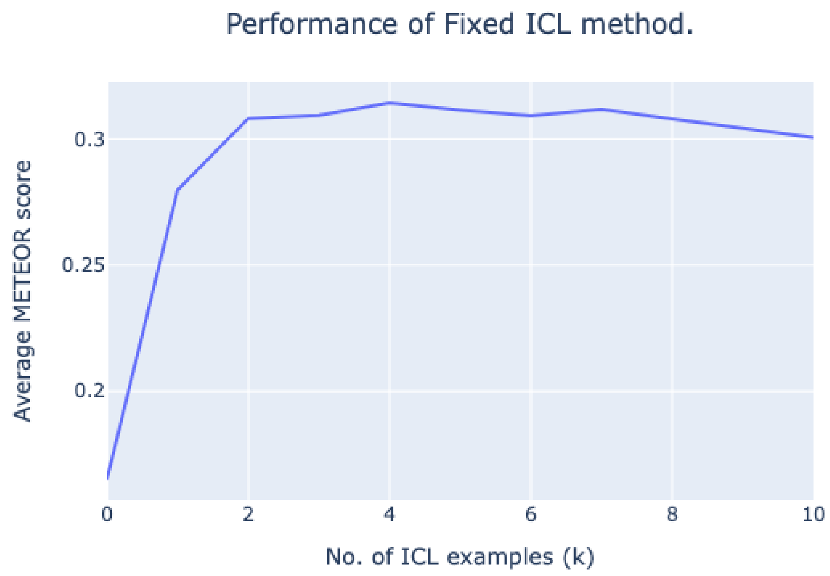
- Among **ICL** based method, **Fixed ICL** performed better over **Adaptive ICL** except of Portuguese

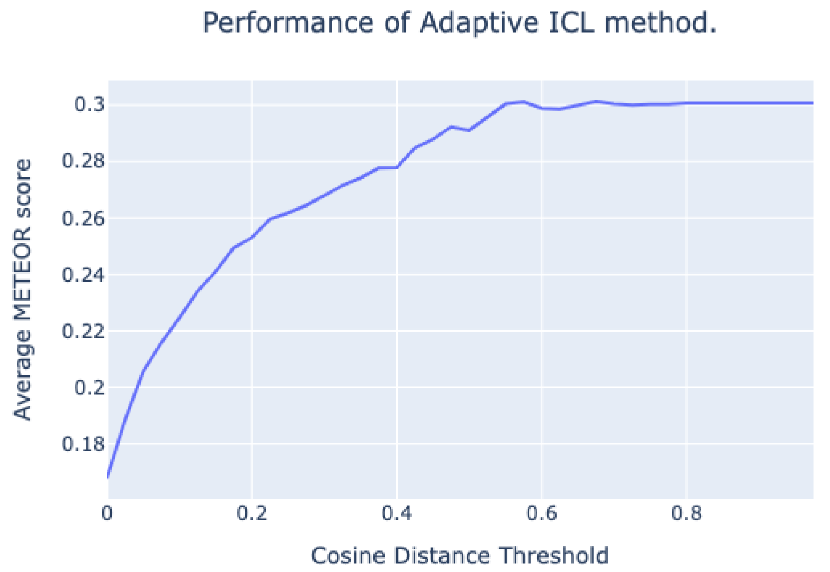
- Observing the overall trend, English and Spanish have higher scores across all models, suggesting better generalization due to more training data

Approaches	Average METEOR score by languages				
	English	German	French	Spanish	Portuguese
Zero-shot	0.21	0.15	0.14	0.21	0.20
Fine-tuned model	**0.40**	**0.26**	**0.37**	**0.38**	0.33
FICL (Mistral-7B)	0.39	0.22	0.32	0.36	**0.36**
AICL (Mistral-7B)	0.37	0.2	0.3	0.33	0.35

Results : In-Context Learning

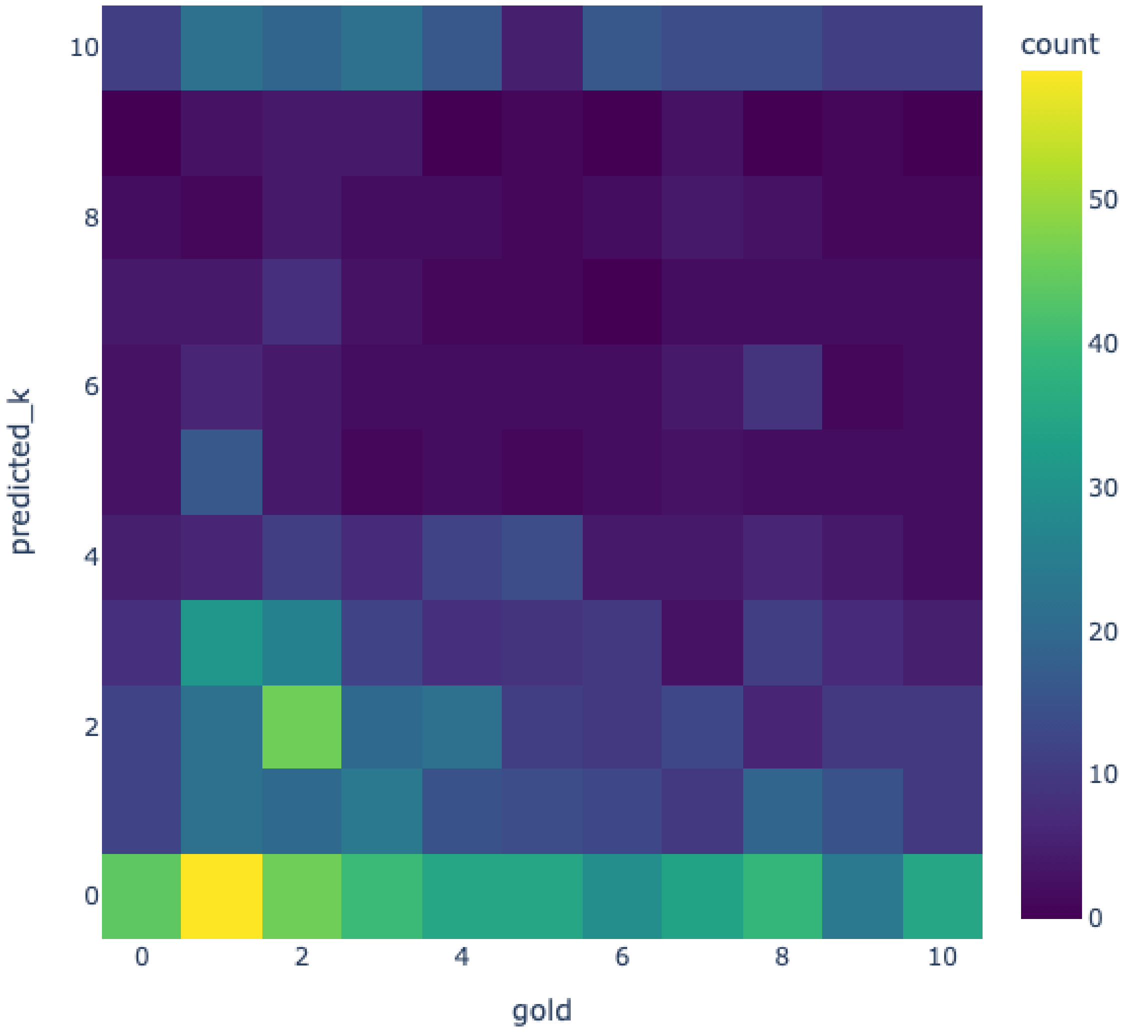
- More detailed observations of **ICL** based method using **dev** dataset were made *(for English)*.
- In comparison to **Zero-shot method**, both **ICL** based methods performs much better, above 0.4 in each method compared to 0.24 for **Zero-shot**.
- Incorporating a small number or relevant examples leads to substantial gains in generation quality.
- However, further increasing the number of examples beyond an optimal point results in diminishing performance, likely due to noise introduced by less relevant examples.





Results : In-Context Learning

- More detailed observations of **ICL** based method using **dev** dataset were made *(for English)*.
- In comparison to **Zero-shot method**, both **ICL** based methods performs much better, above 0.4 in each method compared to 0.24 for **Zero-shot**.
- Incorporating a small number or relevant examples leads to substantial gains in generation quality.
- However, further increasing the number of examples beyond an optimal point results in diminishing performance, likely due to noise introduced by less relevant examples.



Conclusion

- Fine-tuned model performed best, ICL methods were comparable.
- Fine-tuning worked best for language with larger training dataset.
- Although Fixed-ICL outperformed Adaptive-ICL, we still believe Adaptive-ICL could be a interesting approach to explore further

Future work

- Conduct systematic hyperparameter tuning of key training parameters which were not exhaustively optimized.
- Fine-tune bigger/better models like **flan-t5-xl, flan-t5-xxl** other LLMs
- Entity expansions

Thank You