

CiteBART: Learning to Generate Citations for Local Citation Recommendation

Ege Yiğit ÇELİK^[0000-0001-5138-6116] and Selma TEKİR^[0000-0002-0488-9682]

İzmir Institute of Technology, İzmir, Turkey

A Qualitative Analysis on Prompting Large Language Models

We conducted experiments on Large Language Models to evaluate their performances in local citation recommendation. We prompted the open-source "Llama-2-70b-chat" model for our trials. In each prompt, we first list a set of citation tokens (200, due to the limits of chat windows) from our dataset, followed by a few examples of masked contexts with the corresponding ground truth mask values. Subsequently, we ask the model to fill in the mask for a new context by selecting a citation from the initially provided list.

We present four examples in Figures 1 and 2 to illustrate the workings of the base and global pre-training schemes, respectively. Due to space constraints, we partially display the list of citations, example contexts, and citing abstracts in the prompts. Each example consists of three parts: the prompt, the LLM's answer, and the ground truth value of the masked citation token provided at the end of the prompt.

Figure 1 includes a correct prediction in Part (a) and an incorrect one in (b). Indeed, the correct prediction is the only successful example in several trials using the base approach. The model responds to the prompt by "Shwartz et al., 2016" explaining its choice. On the other hand, the model fills in the mask by "Bahdanau et al., 2016" in Part (b), where "Bluche, 2016" is expected. Its reasoning sheds light on its wrong choice as it strongly associates the term "attention-based mechanisms" in the local context with Bahdanau et al.'s seminal paper on attention-based sequence modeling.

In Figure 2, Part (a) presents a successful example based on the global dataset where the prompt includes the citing paper's title and abstract with the local citation context. The LLM generates the correct citation without an explanation, unlike other predictions. The second example in Part (b) belongs to an incorrect prediction, yet the LLM makes a plausible choice here, judging from its grounding. We can conclude from the observed behavior that LLMs need custom pre-training for the citation tokens to perform well in the task of local citation recommendation.

Our further trials with LLMs demonstrate that they tend not to restrict their predictions to the provided list of citations but to recommend the best choice based on their prior knowledge. They also exhibit a known deficiency. They sometimes ask for confirmation when they provide an answer, and even if you

confirm, they lean towards changing the answer. In conclusion, they suffer from hallucinations.

PROMPT:

Here is a list of citations:

"Stalzer, 2017"

"Langford et al., 2008"

"Zhao et al., 2015"

...

There are example context windows taken from scientific papers below. Each of them has a <mask> inside. Actual values of masks have been provided as well. They are filled using one the citations listed above.

An example context window: "... compared with hand-crafted networks. The recent work MetaQNN <mask> and ral Architecture Search with reinforcement learning reported surprising ..."

Actual value of the mask: "Baker et al., 2016"

...

Now, find the following context window's actual mask using one of the citations listed above:

An example context window: "... relation between the words . Recently, presented xNET, an extension of HypeNET <mask>. This method integrates both path-based and ..."

What is the actual value of the mask?

ANSWER:

The actual value of the mask in the given context window is "Shwartz et al., 2016".

Here's how you can identify the correct citation: The context window mentions "xNET" and "HypeNET", which are both neural network models used for semantic relation classification. The authors mention that the approach presented in the paper integrates both path-based and distributional information for semantic relation classification. The paper was published recently, relative to the publication date of the paper containing the context window. All of these clues point to the citation being "Shwartz et al., 2016", which is indeed the correct answer.

Ground Truth Citation: Shwartz et al., 2016

(a) Base approach - Successful example

PROMPT:

Here is a list of citations:

"Stalzer, 2017"

"Langford et al., 2008"

"Zhao et al., 2015"

...

There are example context windows taken from scientific papers below. Each of them has a <mask> inside. Actual values of masks have been provided as well. They are filled using one the citations listed above.

An example context window: "... compared with hand-crafted networks. The recent work MetaQNN <mask> and ral Architecture Search with reinforcement learning reported surprising ..."

Actual value of the mask: "Baker et al., 2016"

...

Now, find the following context window's actual mask using one of the citations listed above:

An example context window: "... follow-up work solves this problem with attention based mechanisms <mask> , this will be investigated in future work. The recognizer is trained on both ..."

What is the actual value of the mask?

ANSWER:

The actual value of the mask in the given context window is likely to be "Bahdanau et al., 2016". This citation is mentioned in the list of references provided earlier and fits the context of the sentence where it mentions "attention-based mechanisms". Therefore, the complete reference for the context window would be:

"... follow-up work solves this problem with attention-based mechanisms Bahdanau et al., 2016, this will be investigated in future work. The recognizer is trained on both ..."

Ground Truth Citation: Bluche, 2016

(b) Base approach - Unsuccessful example

Fig. 1. Prompt examples on Large Language Models for Base approach.

PROMPT:

Here is a list of citations:

"Mirowski et al., 2016"

"Lowe et al., 2017"

"Walsh, 2014"

...

There are example context windows taken from scientific papers below. Each of them has a <mask> inside. Alongside the context windows, the titles and abstracts of citing papers, which contain the context windows, are provided. Actual values of masks have been provided as well. They are filled using one the citations listed above.

An example context window, citing title and abstract: "... that it searches for critical points rather than local optima <mask> . Therefore, it is more common to use natural gradient or updates... </s> A Kronecker-factored approximate Fisher matrix for convolution layers </s> Second-order optimization methods such as natural gradient descent have the potential to speed up ..."
Actual value of the mask: "Pascanu et al., 2014"

...

Now, find the following context window's actual mask using one of the citations listed above:

An example context window, citing title and abstract: "... for generating specific responses, inspired by the semi-supervised deep generative model <mask> . The principle idea is to generate the next response ... </s> A Conditional Variational Framework for Dialog Generation </s> Deep latent variable models have been shown to facilitate the response generation for open-domain dialog ..."
What is the actual value of the mask?

ANSWER:

The actual value of the mask is "Kingma et al., 2014".

Ground Truth Citation: Kingma et al., 2014

(a) Global approach - Successful example

PROMPT:

Here is a list of citations:

"Mirowski et al., 2016"

"Lowe et al., 2017"

"Walsh, 2014"

...

There are example context windows taken from scientific papers below. Each of them has a <mask> inside. Alongside the context windows, the titles and abstracts of citing papers, which contain the context windows, are provided. Actual values of masks have been provided as well. They are filled using one the citations listed above.

An example context window, citing title and abstract: "... that it searches for critical points rather than local optima <mask> . Therefore, it is more common to use natural gradient or updates... </s> A Kronecker-factored approximate Fisher matrix for convolution layers </s> Second-order optimization methods such as natural gradient descent have the potential to speed up ..."
Actual value of the mask: "Pascanu et al., 2014"

...

Now, find the following context window's actual mask using one of the citations listed above:

An example context window, citing title and abstract: "... stacked Denoising Auto-encoders to extract meaningful representations. <mask> main adaptation from multiple sources. For domain adaptation ... </s> Knowledge Adaptation: Teaching to Adapt </s> Domain adaptation is crucial in many real-world applications where the distribution of the training data differs from the distribution of ..."
What is the actual value of the mask?

ANSWER:

Based on the context window, I'm going to take a guess that the actual value of the mask is "Nguyen et al., 2016". My reasoning is as follows:

The context window mentions "domain adaptation" and "learning domain-invariant representations", which suggests that the paper is related to domain adaptation.

The sentence with the mask mentions "main adaptation from multiple sources", which suggests that the paper is discussing a method for adapting to multiple sources or domains.

The citation "Nguyen et al., 2016" is listed in the original list of citations, and the title of the paper is not provided, but the abstract mentions "domain adaptation" and "multiple sources", which matches the context of the context window.

Ground Truth Citation: Chen et al., 2012

(b) Global approach - Unsuccessful example

Fig. 2. Prompt examples on Large Language Models for Global approach.