# Evaluating pre-trained transformer models for similar article recommendation in PubMed

Li Zhang

May 14, 2021

**Abstract**

The work presented here promises to make the exploration of scholarly material faster and more accurate.

# 1 Introduction

Simialr article recomendation is an important feature in many academic seaching database/digital libraries. It enable users to go though relative researher quackly, play an impor6tant role in improving user searching experimence, while more importantly, is can help to accurate dismi the valuable biomedical findings. Improved literature search engines can save researchers time and effort by making it easier to locate the most important and relevant literature. [**?**]

How long has PubMed supplied "Simialr article" function? Like other system, ResearchGate, Scopus, NCBI's PubMed system, has integrated this feature since *, However, the their method to find the similar article is still very *. To date, with the fast developemnt of nature language proccessing, the cutting-edge techniques have prove an opputunity to improve simialr article recommendation performance.

In this paper, we show the simialr article performance are be largely promoted by using large-heavy pertrained language model.

PubMed related article links identify closely related articles and enhance our ability to navigate the biomedical literature [**?**].

PubMed has integrated the "Simialr article" function for a long time,

show the case here.

significence: Fig * shows the similar article recommendation functionality, this function is very helpful for biomedical scholar, as a recent works by NLM/NCBI team suggest user needs can be largely improved while users explore related articles. can power PubMed user experience,...

Why similar article recommendation is very important for further improveing search exerperience, why it is a necessary functionality? Related works: how PubMed improving use searching experience. To improving user searching experience, NCBI has provided many in place measurements from serverl aspects. such as autocomplition, ..., . refrequent search terms recommendation. However, this measures can be not necessary enough, user may expore other What did europen Pubmed did? and what did other platform did?

In many academic service platform, such as web of science,..., they

How did others find similar article?

Our contribution are three parts. we provide an method to automaticilly build similar article dataset for development models we evaluated the most well-known pre-trained models on four dataset, and emperical evaluation shows fine-tuned * model shows state-of-the-art result. Using this method, we obtain the paper distribution vector for whole PubMed papers.

# 2 Related works

show detailed is there any evaluation dataset?

# 3 Method

## 3.1 dataset building

In this section, we show how we did to bulid the dataset.

Note that we did not consider the publication timeframe, as we can a later-published article can exist in the related article list.

## 3.2 fine-tuning

# 4 Results

# 5 Discussion

## 5.1 how this works can be intregtared into PubMed system?

fast, very short words embeddings, inductive-infering.

## 5.2 user study

how it can be

## 5.3 limitation

## 5.4 future protential imporovement

This study only recomnedation paper from semantic perspective, however, in many commercial recommender system, the recommeded items may not ony semantic relatedness, other relationship the relationships that a mature system should consider is not equal to semantic relatedness. other crucial aspect such as ... also need consideration. However, we can not obtain such real word dataset, i.e, integrating PubMed searching log to develop more powerful recommneder system. Thus, we can image the paper recommnedation system in PubMed can be more powerful by leveraging state-of-the-art techinique in recomendation system and information retrival. for example, recommendation with intrepretation,

exploring more user intellience that can be available in NCBI

## 6 Conclusions

In this works, we shows an effective and effecient way to locate similar PubMed for power user search experpence in PubMed system. This study provide initial, portary investigation on simiar article for PubMed system. ...

Our future works intent more relationship in paper recomendation.