Skip to main content

[![Cornell University](/static/browse/0.3.4/images/icons/cu/cornell-reduced-white-SMALL.svg)](https://www.cornell.edu/)

We gratefully acknowledge support from the Simons Foundation, [member institutions](https://info.arxiv.org/about/ourmembers.html), and all contributors. [Donate](https://info.arxiv.org/about/donate.html)

[](/IgnoreMe)

[![arxiv logo](/static/browse/0.3.4/images/arxiv-logo-one-color-white.svg)](/) > [cs](/list/cs/recent) > arXiv:2106.09700

Computer Science > Computation and Language

arXiv:2106.09700 (cs)

[Submitted on 17 Jun 2021 ([v1](https://arxiv.org/abs/2106.09700v1)), last revised 21 Sep 2021 (this version, v2)]

Title:Scientific Language Models for Biomedical Knowledge Base Completion: An Empirical Study

Authors:[Rahul Nadkarni](https://arxiv.org/search/cs?searchtype=author&query=Nadkarni,+R), [David Wadden](https://arxiv.org/search/cs?searchtype=author&query=Wadden,+D), [Iz Beltagy](https://arxiv.org/search/cs?searchtype=author&query=Beltagy,+I),

[Noah A.

Smith](https://arxiv.org/search/cs?searchtype=author&query=Smith,+N+A),

[Hannaneh

Hajishirzi](https://arxiv.org/search/cs?searchtype=author&query=Hajishirzi,+H),

[Tom Hope](https://arxiv.org/search/cs?searchtype=author&query=Hope,+T)

View a PDF of the paper titled Scientific Language Models for Biomedical Knowledge Base Completion: An Empirical Study, by Rahul Nadkarni and 5 other authors

[View PDF](/pdf/2106.09700)

> Abstract:Biomedical knowledge graphs (KGs) hold rich information on entities
> such as diseases, drugs, and genes. Predicting missing links in these graphs
> can boost many important applications, such as drug design and repurposing.
> Recent work has shown that general-domain language models (LMs) can serve as
> "soft" KGs, and that they can be fine-tuned for the task of KG completion.
> In this work, we study scientific LMs for KG completion, exploring whether
> we can tap into their latent knowledge to enhance biomedical link
> prediction. We evaluate several domain-specific LMs, fine-tuning them on
> datasets centered on drugs and diseases that we represent as KGs and enrich
> with textual entity descriptions. We integrate the LM-based models with KG
> embedding models, using a router method that learns to assign each input
> example to either type of model and provides a substantial boost in
> performance. Finally, we demonstrate the advantage of LM models in the
> inductive setting with novel scientific entities. Our datasets and code are
> made publicly available.

Comments: | AKBC 2021 camera-ready

---|---

Subjects: | Computation and Language (cs.CL); Machine Learning (cs.LG)

Cite as: | [arXiv:2106.09700](https://arxiv.org/abs/2106.09700) [cs.CL]

| (or [arXiv:2106.09700v2](https://arxiv.org/abs/2106.09700v2) [cs.CL] for this version)

| <https://doi.org/10.48550/arXiv.2106.09700> Focus to learn more arXiv-issued DOI via DataCite

Submission history

From: Rahul Nadkarni [[view email](/show-email/c3f9cabd/2106.09700)]

[[v1]](/abs/2106.09700v1) Thu, 17 Jun 2021 17:55:33 UTC (1,873 KB)

[v2] Tue, 21 Sep 2021 04:38:40 UTC (1,875 KB)

Full-text links:

Access Paper:

View a PDF of the paper titled Scientific Language Models for Biomedical Knowledge Base Completion: An Empirical Study, by Rahul Nadkarni and 5 other authors

* [View PDF](/pdf/2106.09700)
* [TeX Source](/src/2106.09700)
* [Other Formats](/format/2106.09700)

[view license](http://arxiv.org/licenses/nonexclusive-distrib/1.0/ "Rights to

this article")

Current browse context:

cs.CL

[< prev](/prevnext?id=2106.09700&function=prev&context=cs.CL "previous in cs.CL \(accesskey p\)") | [next >](/prevnext?id=2106.09700&function=next&context=cs.CL "next in cs.CL \(accesskey n\)")

[new](/list/cs.CL/new) | [recent](/list/cs.CL/recent) | [2021-06](/list/cs.CL/2021-06)

Change to browse by:

[cs](/abs/2106.09700?context=cs)

[cs.LG](/abs/2106.09700?context=cs.LG)

References & Citations

* [NASA ADS](https://ui.adsabs.harvard.edu/abs/arXiv:2106.09700)
* [Google Scholar](https://scholar.google.com/scholar_lookup?arxiv_id=2106.09700)
* [Semantic Scholar](https://api.semanticscholar.org/arXiv:2106.09700)

[DBLP](https://dblp.uni-trier.de) \- CS Bibliography

[listing](https://dblp.uni-trier.de/db/journals/corr/corr2106.html#abs-2106-09700 "listing on DBLP") |

[bibtex](https://dblp.uni-trier.de/rec/bibtex/journals/corr/abs-2106-09700 "DBLP bibtex record")

[Iz Beltagy](https://dblp.uni-trier.de/search/author?author=Iz%20Beltagy "DBLP
author search")
[Noah A. Smith](https://dblp.uni-
trier.de/search/author?author=Noah%20A.%20Smith "DBLP author search")
[Hannaneh Hajishirzi](https://dblp.uni-
trier.de/search/author?author=Hannaneh%20Hajishirzi "DBLP author search")
[Tom Hope](https://dblp.uni-trier.de/search/author?author=Tom%20Hope "DBLP
author search")

[a](/static/browse/0.3.4/css/cite.css) export BibTeX citation Loading...

BibTeX formatted citation

×

loading...

Data provided by:

Bookmark

[![BibSonomy logo](/static/browse/0.3.4/images/icons/social/bibsonomy.png)
](http://www.bibsonomy.org/BibtexHandler?requTask=upload&url=https://arxiv.org/abs/2106.09700&
description=Scientific
Language Models for Biomedical Knowledge Base Completion: An Empirical Study

"Bookmark on BibSonomy") [![Reddit

logo](/static/browse/0.3.4/images/icons/social/reddit.png)

](https://reddit.com/submit?url=https://arxiv.org/abs/2106.09700&title=Scientific

Language Models for Biomedical Knowledge Base Completion: An Empirical Study

"Bookmark on Reddit")

Bibliographic Tools

Bibliographic and Citation Tools

Bibliographic Explorer Toggle

Bibliographic Explorer _([What is the

Explorer?](https://info.arxiv.org/labs/showcase.html#arxiv-bibliographic-

explorer))_

Connected Papers Toggle

Connected Papers _([What is Connected

Papers?](https://www.connectedpapers.com/about))_

Litmaps Toggle

Litmaps _([What is Litmaps?](https://www.litmaps.co/))_

scite.ai Toggle

scite Smart Citations _([What are Smart Citations?](https://www.scite.ai/))_

Code, Data, Media

Code, Data and Media Associated with this Article

alphaXiv Toggle

alphaXiv _([What is alphaXiv?](https://alphaxiv.org/))_

Links to Code Toggle

CatalyzeX Code Finder for Papers _([What is CatalyzeX?](https://www.catalyzex.com))_

DagsHub Toggle

DagsHub _([What is DagsHub?](https://dagshub.com/))_

GotitPub Toggle

Gotit.pub _([What is GotitPub?](http://gotit.pub/faq))_

Huggingface Toggle

Hugging Face _([What is Huggingface?](https://huggingface.co/huggingface))_

Links to Code Toggle

Papers with Code _([What is Papers with Code?](https://paperswithcode.com/))_

ScienceCast Toggle

ScienceCast _([What is ScienceCast?](https://sciencecast.org/welcome))_

Demos

Demos

Replicate Toggle

Replicate _([What is Replicate?](https://replicate.com/docs/arxiv/about))_

Spaces Toggle

Hugging Face Spaces _([What is

Spaces?](https://huggingface.co/docs/hub/spaces))_

Spaces Toggle

TXYZ.AI _([What is TXYZ.AI?](https://txyz.ai))_

Related Papers

Recommenders and Search Tools

Link to Influence Flower

Influence Flower _([What are Influence

Flowers?](https://influencemap.cmlab.dev/))_

Core recommender toggle

CORE Recommender _([What is CORE?](https://core.ac.uk/services/recommender))_

* Author
* Venue
* Institution
* Topic

About arXivLabs

arXivLabs: experimental projects with community collaborators

arXivLabs is a framework that allows collaborators to develop and share new arXiv features directly on our website.

Both individuals and organizations that work with arXivLabs have embraced and accepted our values of openness, community, excellence, and user data privacy. arXiv is committed to these values and only works with partners that adhere to them.

Have an idea for a project that will add value for arXiv's community? [**Learn more about arXivLabs**](https://info.arxiv.org/labs/index.html).

[Which authors of this paper are endorsers?](/auth/show-endorsers/2106.09700) | [Disable MathJax](javascript:setMathjaxCookie\(\)) ([What is MathJax?](https://info.arxiv.org/help/mathjax.html))

* [About](https://info.arxiv.org/about)
* [Help](https://info.arxiv.org/help)

* contact arXivClick here to contact arXiv [Contact](https://info.arxiv.org/help/contact.html)
* subscribe to arXiv mailingsClick here to subscribe [Subscribe](https://info.arxiv.org/help/subscribe)

* [Copyright](https://info.arxiv.org/help/license/index.html)
* [Privacy Policy](https://info.arxiv.org/help/policies/privacy_policy.html)

* [Web Accessibility Assistance](https://info.arxiv.org/help/web_accessibility.html)
* [arXiv Operational Status](https://status.arxiv.org)

Get status notifications via [email](https://subscribe.sorryapp.com/24846f03/email/new) or [slack](https://subscribe.sorryapp.com/24846f03/slack/new)