Computer Science > Computation and Language

arXiv:2010.00904 (cs)

[Submitted on 2 Oct 2020 ([v1](https://arxiv.org/abs/2010.00904v1)), last revised 24 Mar 2021 (this version, v3)]

Title:Autoregressive Entity Retrieval

Authors:[Nicola De Cao](https://arxiv.org/search/cs?searchtype=author&query=De+Cao,+N), [Gautier Izacard](https://arxiv.org/search/cs?searchtype=author&query=Izacard,+G), [Sebastian Riedel](https://arxiv.org/search/cs?searchtype=author&query=Riedel,+S), [Fabio

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

View a PDF of the paper titled Autoregressive Entity Retrieval, by Nicola De Cao and 3 other authors

[View PDF](/pdf/2010.00904)

> Abstract:Entities are at the center of how we represent and aggregate knowledge. For instance, Encyclopedias such as Wikipedia are structured by entities (e.g., one per Wikipedia article). The ability to retrieve such entities given a query is fundamental for knowledge-intensive tasks such as entity linking and open-domain question answering. Current approaches can be understood as classifiers among atomic labels, one for each entity. Their weight vectors are dense entity representations produced by encoding entity meta information such as their descriptions. This approach has several shortcomings: (i) context and entity affinity is mainly captured through a vector dot product, potentially missing fine-grained interactions; (ii) a large memory footprint is needed to store dense representations when considering large entity sets; (iii) an appropriately hard set of negative data has to be subsampled at training time. In this work, we propose GENRE, the first system that retrieves entities by generating their unique names, left to right, token-by-token in an autoregressive fashion. This mitigates the aforementioned technical issues since: (i) the autoregressive formulation directly captures relations between context and entity name, effectively cross encoding both; (ii) the memory footprint is greatly reduced because the parameters of our encoder-decoder architecture scale with vocabulary size, not entity count; (iii) the softmax loss is computed

> without subsampling negative data. We experiment with more than 20 datasets
> on entity disambiguation, end-to-end entity linking and document retrieval
> tasks, achieving new state-of-the-art or very competitive results while
> using a tiny fraction of the memory footprint of competing systems. Finally,
> we demonstrate that new entities can be added by simply specifying their
> names. Code and pre-trained models at [this https
> URL](https://github.com/facebookresearch/GENRE).

Comments: | Accepted (spotlight) at International Conference on Learning Representations (ICLR) 2021. Code at [this https URL](https://github.com/facebookresearch/GENRE). 20 pages, 9 figures, 8 tables

---|---

Subjects: | Computation and Language (cs.CL); Information Retrieval (cs.IR); Machine Learning (cs.LG); Machine Learning (stat.ML)

Cite as: | [arXiv:2010.00904](https://arxiv.org/abs/2010.00904) [cs.CL]

| (or [arXiv:2010.00904v3](https://arxiv.org/abs/2010.00904v3) [cs.CL] for this version)

| <https://doi.org/10.48550/arXiv.2010.00904> Focus to learn more arXiv-issued DOI via DataCite

Submission history

From: Nicola De Cao [[view email](/show-email/1e49b680/2010.00904)]

[[v1]](/abs/2010.00904v1) Fri, 2 Oct 2020 10:13:31 UTC (1,006 KB)

[[v2]](/abs/2010.00904v2) Sat, 27 Feb 2021 15:20:52 UTC (1,073 KB)

[v3] Wed, 24 Mar 2021 07:21:07 UTC (1,057 KB)

Full-text links:

Access Paper:

View a PDF of the paper titled Autoregressive Entity Retrieval, by Nicola De Cao and 3 other authors

* [View PDF](/pdf/2010.00904)
* [TeX Source](/src/2010.00904)
* [Other Formats](/format/2010.00904)

Current browse context:

cs.CL

[< prev](/prevnext?id=2010.00904&function=prev&context=cs.CL "previous in cs.CL \(accesskey p\)") | [next >](/prevnext?id=2010.00904&function=next&context=cs.CL "next in cs.CL \(accesskey n\)")

[new](/list/cs.CL/new) | [recent](/list/cs.CL/recent) | [2020-10](/list/cs.CL/2020-10)

Change to browse by:

[cs](/abs/2010.00904?context=cs)
[cs.IR](/abs/2010.00904?context=cs.IR)

[cs.LG](/abs/2010.00904?context=cs.LG)

[stat](/abs/2010.00904?context=stat)

[stat.ML](/abs/2010.00904?context=stat.ML)

References & Citations

* [NASA ADS](https://ui.adsabs.harvard.edu/abs/arXiv:2010.00904)
* [Google Scholar](https://scholar.google.com/scholar_lookup?arxiv_id=2010.00904)
* [Semantic Scholar](https://api.semanticscholar.org/arXiv:2010.00904)

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

[listing](https://dblp.uni-trier.de/db/journals/corr/corr2010.html#abs-2010-00904 "listing on DBLP") | [bibtex](https://dblp.uni-trier.de/rec/bibtex/journals/corr/abs-2010-00904 "DBLP bibtex record")

[Nicola De Cao](https://dblp.uni-trier.de/search/author?author=Nicola%20De%20Cao "DBLP author search")
[Gautier Izacard](https://dblp.uni-trier.de/search/author?author=Gautier%20Izacard "DBLP author search")
[Sebastian Riedel](https://dblp.uni-trier.de/search/author?author=Sebastian%20Riedel "DBLP author search")
[Fabio Petroni](https://dblp.uni-trier.de/search/author?author=Fabio%20Petroni "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/2010.00904&
description=Autoregressive
Entity Retrieval "Bookmark on BibSonomy") [![Reddit
logo](/static/browse/0.3.4/images/icons/social/reddit.png)
](https://reddit.com/submit?url=https://arxiv.org/abs/2010.00904&title=Autoregressive
Entity Retrieval "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/2010.00904) | [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)