

Machine Translation in Argos Translate (2021)



Sequence to sequence model

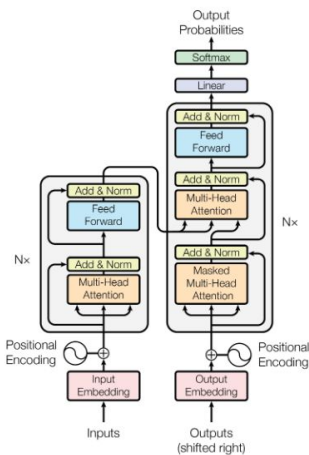


Figure 1: The Transformer - model architecture.

<Hello> < my> < name> < is> < John> <.>



<Hola> < me> < llamo> < John> <.>

NMT OpenNMT

An open source neural machine translation system.

arXiv.org > cs > arXiv:1706.03762

Search... All fields Search

Help | Advanced Search

Computer Science > Computation and Language

[Submitted on 12 Jun 2017 (v1), last revised 6 Dec 2017 (this version, v5)]

Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new

Download:

- PDF
- Other formats (license)

Current browse context: cs.CL

Inference

OpenNMT / CTranslate2

Watch 16 Unstar 142 Fork 39

<> Code Issues 10 Pull requests Actions Projects 1 Security Insights

master 6 branches 85 tags

Go to file

Add file

Code

About

Fast inference engine for OpenNMT models

opennmt.net/

guillaumekln Mark CPUInfo struct as constant ✓ 974f8db 3 days ago 1,201 commits

.github/workflows Rename CMake option WITH_TESTS to BUILD_TESTS (#430) 4 days ago



Data

OPUS ... the open parallel corpus

OPUS is a growing collection of translated texts from the web. In the OPUS project we try to convert and align free online data, to add linguistic annotation, and to provide the community with a publicly available parallel corpus. OPUS is based on open source products and the corpus is also delivered as an open content package. We used several tools to compile the current collection. All pre-processing is done automatically. No manual corrections have been carried out.

The OPUS collection is growing! Check this page from time to time to see new data arriving ... Contributions are very welcome! Please contact <jorg.tiedemann@helsinki.fi >

Search & download resources: ☐ show all versions

kaikki.org

Wiktionary



English <i>The free dictionary</i> 6 603 000+ entries	Русский <i>Свободный словарь</i> 1 115 000+ статей
Ελληνικά <i>Το ελεύθερο λεξικό</i> 798 000+ λήμματα	Français <i>Le dictionnaire libre</i> 4 123 000+ articles
Deutsch <i>Das freie Wörterbuch</i> 961 000+ Einträge	Polski <i>Wolny słownik</i> 736 000+ stron
Español <i>El diccionario libre</i> 905 000+ entradas	Nederlands <i>Het vrije woordenboek</i> 778 000+ artikelen
Italiano <i>Il dizionario libero</i> 544 000+ lemmi	日本語 <i>フリー多機能辞典</i> 264 000+ 項目

Data processing

argosopentech / onmt-models

Watch 1 Unstar 20 Fork 2

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

master 1 branch 0 tags

Go to file

Add file

Code



PJ-Finlay Added Polish overtraining threshold

6657d5e 3 days ago 106 commits



raw_data

Added README for raw_data

2 months ago



.gitignore

More work on training script, script now fully trains models!

8 months ago



LICENSE

Initial commit

10 months ago

About



No description, website, or topics provided.

Readme

MIT License

<define> veracious → being in the habit of telling the truth

Tokenization

google / sentencepiece

Watch 117 Unstar 4.9k Fork 671

<> Code Issues 29 Pull requests 3 Actions Projects Security Insights

master 3 branches 20 tags

Go to file

Add file

Code



taku910 Merge pull request #630 from sarubi/zwj_fix

bc53923 on Feb 25 667 commits

data	Removed codes where Zero Width Joiner replaced with whitespace.	last month
doc	Create options.md	6 months ago
python	fixed python module to check the id range.	3 months ago
src	Merge pull request #630 from sarubi/zwj_fix	last month
tensorflow	remove tensorflow module	6 months ago

About

Unsupervised text tokenizer for Neural Network-based text generation.

natural-language-processing

neural-machine-translation

word-segmentation

Readme

Apache-2.0 License

<H> <ello> <,> <_the> <_cat> <_is> <_un> <f> <riendly> <.>

Sentence boundary detection

 Stanza

Search Stanza

[Overview](#)
[Usage](#)
[Neural Pipeline](#)
[Models](#)
[Biomedical Models](#)
[Stanford CoreNLP Client](#)
[Resources](#)

Stanza – A Python NLP Package for Many Human Languages

`pypi v1.2` `conda v1.2.0` `python 3.6 | 3.7 | 3.8`

Stanza is a collection of accurate and efficient tools for many human languages in one place. Starting from raw text to syntactic analysis and entity recognition, Stanza brings state-of-the-art NLP models to languages of your choosing.

TABLE OF CONTENTS

- [About](#)
- [Getting Started](#)
- [License](#)

Dr. Kang and Ph.D. Doe are arriving at 2. Is P.J. going to be there?

Package Management

Readme	From name	To name	Package version	Argos version	From code	To code	Install
view	Arabic	English	1.0	1.0	ar	en	Installed
view	German	English	1.0	1.0	de	en	↓
view	English	Arabic	1.0	1.0	en	ar	Installed
view	English	German	1.0	1.0	en	de	↓
view	English	Spanish	1.0	1.0	en	es	↓
view	English	French	1.0	1.0	en	fr	Installed
view	English	Italian	1.0	1.0	en	it	↓
view	English	Portuguese	1.0	1.0	en	pt	↓
view	English	Russian	1.0	1.0	en	ru	↓
view	English	Chinese	1.1	1.1	en	zh	↓
view	Spanish	English	1.0	1.0	es	en	↓
view	French	English	1.0	1.0	fr	en	↓
view	Italian	English	1.0	1.0	it	en	↓
view	Portuguese	English	1.0	1.0	pt	en	↓
view	Russian	English	1.0	1.0	ru	en	↓
view	Chinese	English	1.1	1.1	zh	en	↓
view	English	Japanese	1.1	1.1	en	ja	↓
view	Japanese	English	1.1	1.1	ja	en	↓
view	English	Hindi	1.1	1.1	en	hi	↓
view	Hindi	English	1.1	1.1	hi	en	↓
view	English	Irish	1.1	1.1	en	ga	Installed
view	Irish	English	1.1	1.1	ga	en	Installed
view	English	Korean	1.1	1.1	en	ko	↓
view	Korean	English	1.1	1.1	ko	en	↓
view	English	Turkish	1.1	1.1	en	tr	Installed
view	Turkish	English	1.1	1.1	tr	en	↓
view	English	Polish	1.1	1.1	en	pl	Installed
view	Polish	English	1.1	1.1	pl	en	Installed

Peer to peer

InterPlanetary File System



Original author(s) Juan Benet and Protocol Labs^[1]

Developer(s) Protocol Labs

Initial release February 2015;
6 years ago^[1]



Original author(s) Bram Cohen

Developer(s) BitTorrent, Inc.

Initial release July 2, 2001; 19
years ago^[1]

Web

🚩 LibreTranslate

API Docs GitHub 🔑

Translation API

English

↔

French

Input Text

Alan Mathison Turing OBE FRS (/ˈtjʊərniː/; 23 June 1912 – 7 June 1954) was an English mathematician, computer scientist, logician, cryptanalyst, philosopher, and theoretical biologist.[6][7] Turing was highly influential in the development of theoretical computer science, providing a formalisation of the concepts of algorithm and computation with the Turing machine, which can be considered a model of a general-purpose computer.[8][9][10] Turing is widely considered to be the father of theoretical computer science and artificial intelligence.[11] Despite these accomplishments, he was never fully recognised in his home country, if only because much of his work was covered by the Official Secrets Act.

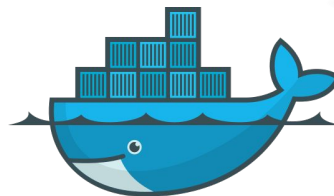
- Wikipedia

720 / 5000

Alan Mathison Turing OBE FRS (expédiéjuôr téléphonique; 23 juin 1912 – 7 juin 1954) était un mathématicien anglais, informaticien, logicien, cryptanalyste, philosophe et biologiste théorique. [6][7] La Turing a été très influente dans le développement de l'informatique théorique, fournissant une formalisation des concepts d'algorithme et de calcul avec la machine Turing, qui peut être considéré comme un modèle d'ordinateur général. [8][9][10] La Turing est largement considérée comme le père de l'informatique théorique et de l'intelligence artificielle. [11] Malgré ces réalisations, il n'a jamais été pleinement reconnu dans son pays d'origine, si seulement parce qu'une grande partie de son travail était couvert par la Loi sur les secrets officiels.

- Wikipedia

COPY TEXT





Argos Translate

Territories

112

