@ukhndlwl

Nearest Neighbor Machine Translation

Urvashi Khandelwal, Angela Fan, Dan Jurafsky, Luke Zettlemoyer, Mike Lewis

Stanford University, Facebook AI Research

Stanford | NLP

facebook Artificial Intelligence

ICLR 2021

Nearest Neighbor Retrieval:
“Generalization through Memorization”

Nearest Neighbor Retrieval: “Generalization through Memorization” for Machine Translation

Key Results

Memorizing the training data improves machine translation generalization, and allows a multilingual model to specialize.

A single translation model can adapt to multiple domains by memorizing domain-specific data, without any in-domain training.

Memorization can make model predictions more interpretable.

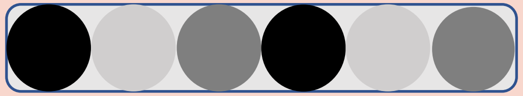
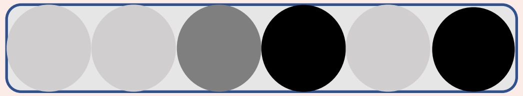
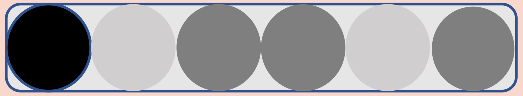
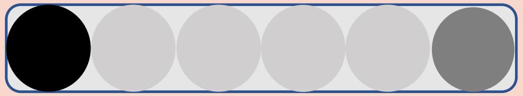
Nearest Neighbor Language Models (kNN-LM)

[Khandelwal et al., 2020]

Interpolate a pre-trained (autoregressive) language model with a k-nearest neighbors model, with NO additional training.

Urvashi Khandelwal, Omer Levy, Dan Jurafsky, Luke Zettlemoyer and Mike Lewis. Generalization through Memorization: Nearest Neighbor Language Models. In ICLR, 2020.

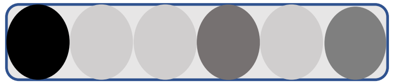
kNN-LM Datastore

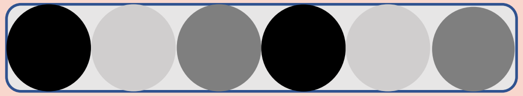
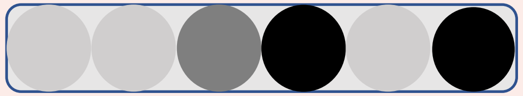
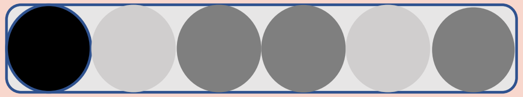
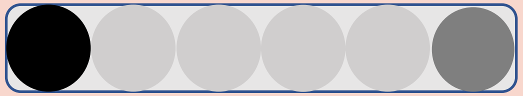
Training Contexts	*Keys=LM Context Representations*	*Values=Targets*
Tony Stark fought on		*Titan*
Tony Stark is married to		*Pepper*
Tony Stark lives in		*Malibu*
…	…	…
Tony Stark is a resident of		*Malibu*

Nearest Neighbor Retrieval

Test Context: *Tony Stark resides in* *???*

Query = Test Context Representation



Training Contexts	*Keys=LM Context Representations*	*Values=Targets*
Tony Stark fought on		*Titan*
Tony Stark is married to		*Pepper*
Tony Stark lives in		*Malibu*
…	…	…
Tony Stark is a resident of		*Malibu*

kNN-LM

Test Context: *Tony Stark resides in* *__???__*

Language Model	
Malibu	*0.2*
Titan	*0.2*
…	…

k-Nearest Neighbors	
Malibu	*0.8*
Titan	*0.2*

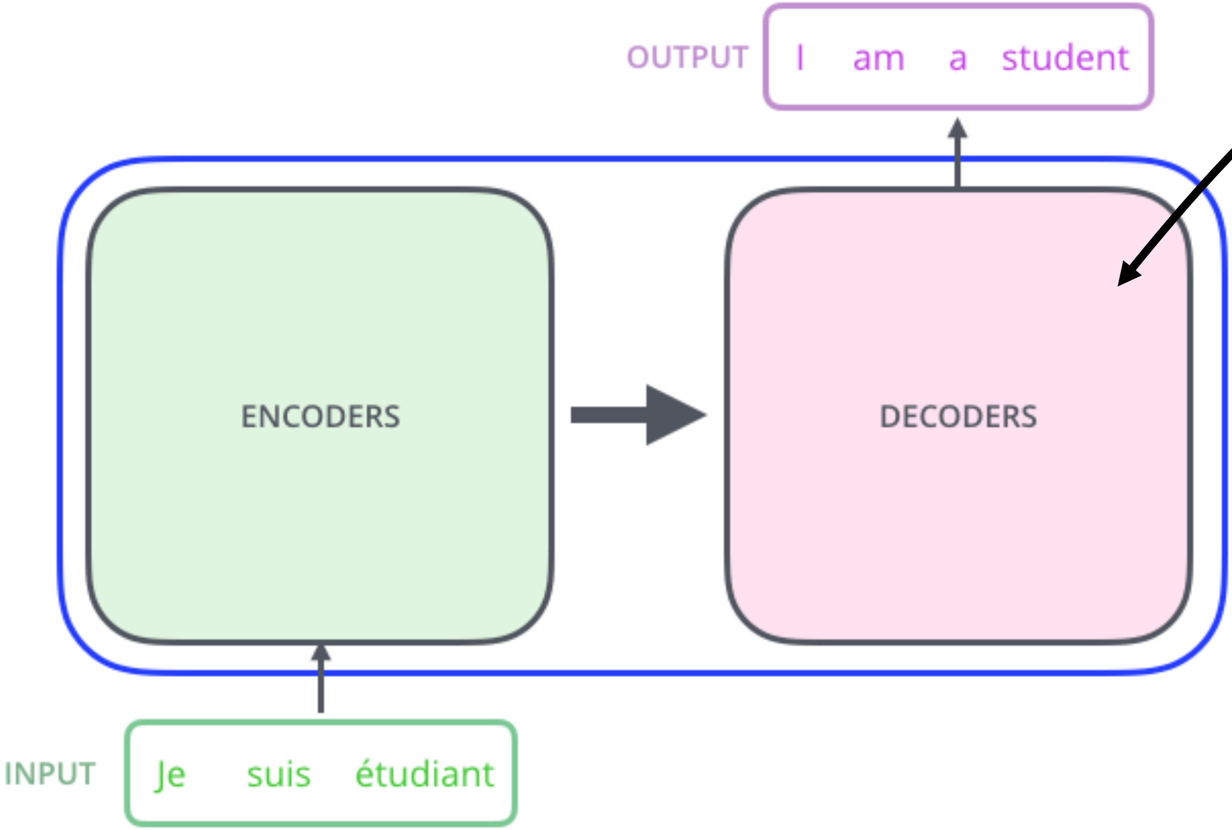
→

kNN-LM $(1-\lambda)\ p_{\mathrm{LM}} + \lambda\ p_{\mathrm{kNN}}$	
Malibu	*0.6*
Titan	*0.2*
…	…

Nearest Neighbor Machine Translation (kNN-MT)

Interpolate a pre-trained machine translation model with a k-nearest neighbors model, with NO additional training.

The MT decoder is a language model.



A language model!

Image from http://jalammar.github.io/illustrated-transformer/

kNN-MT

Stored representations rely on ground truth prior context as well as the source sequence.

kNN-MT

Stored representations rely on ground truth prior context as well as the source sequence.

Pride and Prejudice a été écrit par Jane Austen <>
Pride and Prejudice foi escrito por Jane Austen

kNN-MT

Stored representations rely on ground truth prior context as well as the source sequence.

Pride and Prejudice a été écrit par Jane Austen <>

Pride and Prejudice foi escrito por Jane Austen

kNN-MT

Stored representations rely on ground truth prior context as well as the source sequence.

Pride and Prejudice a été écrit par Jane Austen <>

Pride and Prejudice foi escrito por Jane ***Austen***

Experiments

Key Results

Memorizing the training data improves machine translation generalization, and allows a multilingual model to specialize.

A single translation model can adapt to multiple domains by memorizing domain-specific data, without any in-domain training.

Memorization can make model predictions more interpretable.

Memorizing MT training data

State-of-the-art German-English translation model. [Ng et al., 2019]

770 million key-value pairs memorized.

Model	*BLEU (↑)*
Base MT	*37.59*

Memorizing MT training data

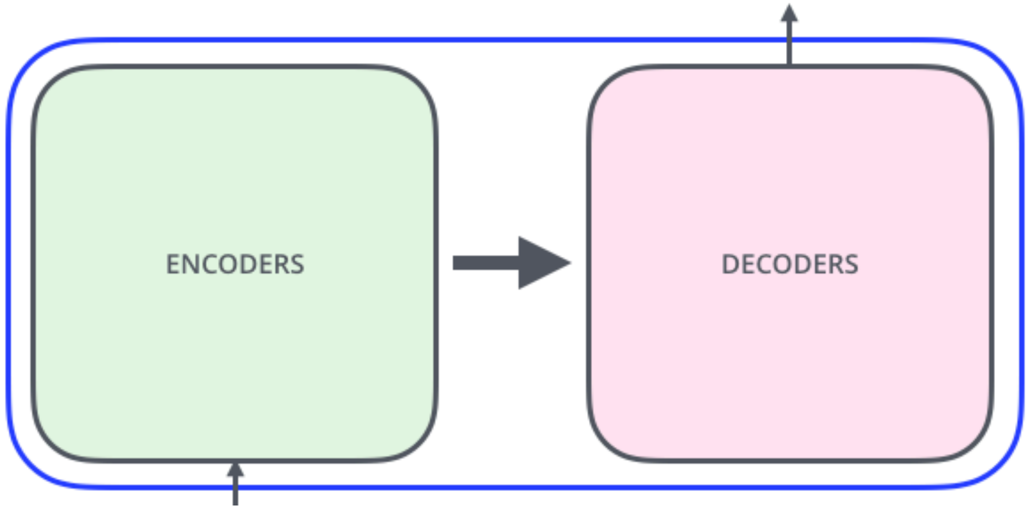
State-of-the-art German-English translation model.
[Ng et al., 2019]

770 million key-value pairs memorized.

Model	*BLEU (↑)*
Base MT	*37.59*
kNN-MT	*39.08*

Multilingual MT

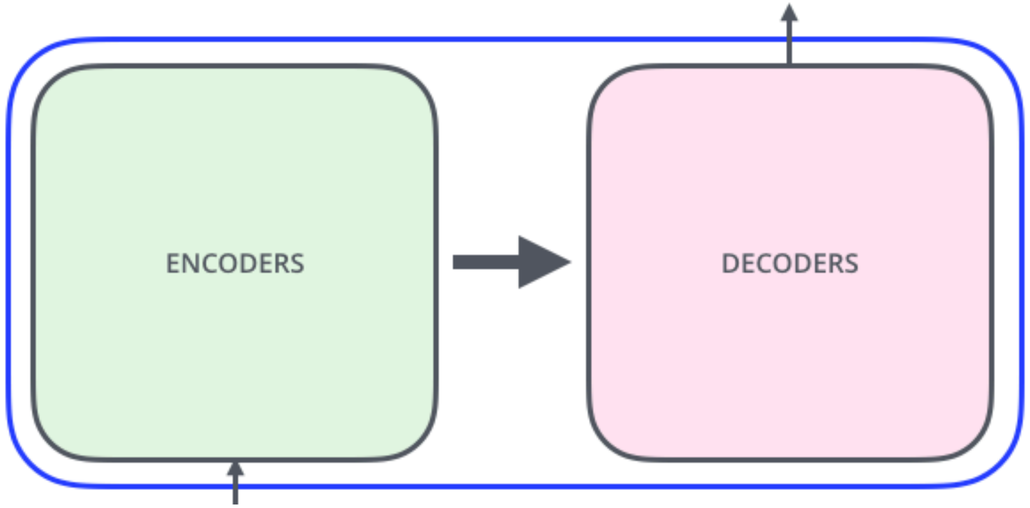
English / German / French /
Japanese / Chinese / Czech /
Korean / Hindi / … … …



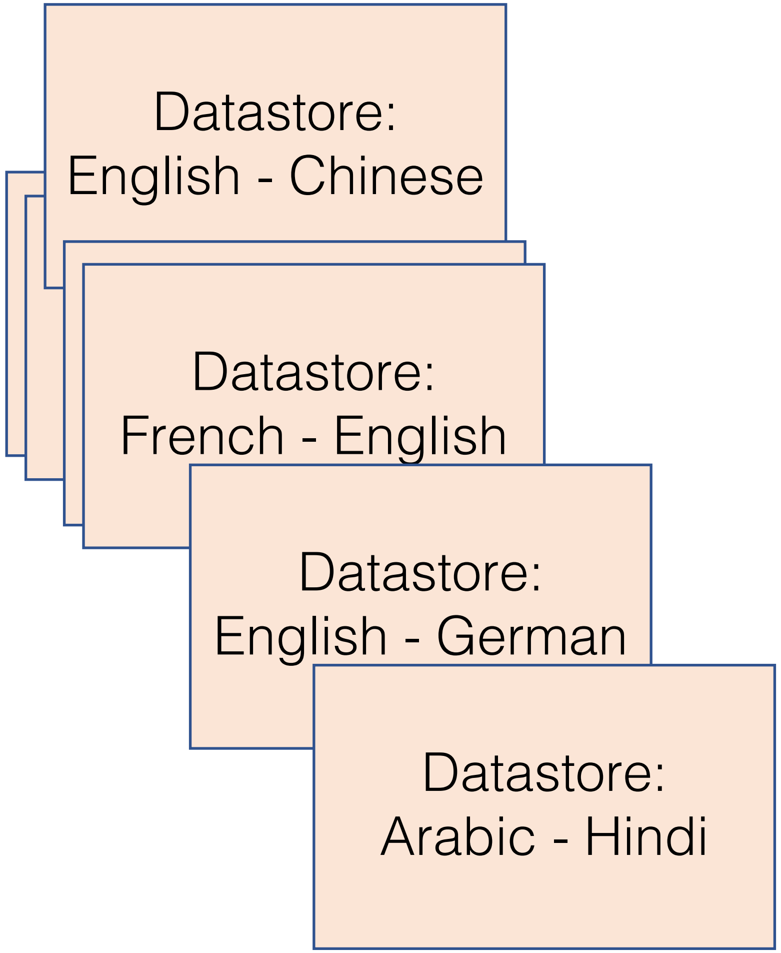
English / German / French /
Japanese / Chinese / Czech /
Korean / Hindi / … … …

Multilingual MT with kNN-MT

English / German / French /
Japanese / Chinese / Czech /
Korean / Hindi / … … …



English / German / French /
Japanese / Chinese / Czech /
Korean / Hindi / … … …



Datastore:
English - Chinese

Datastore:
French - English

Datastore:
English - German

Datastore:
Arabic - Hindi

Specialized multilingual MT models

Model	*English-German*	*Chinese-English*	*English-Chinese*
Base MT	*36.47*	*24.23*	*30.22*

Specialized multilingual MT models

Model	*English-German*	*Chinese-English*	*English-Chinese*
Base MT	*36.47*	*24.23*	*30.22*
kNN-MT	*39.49* ***(+3.02)***	*27.51* ***(+3.28)***	*33.63* ***(+3.41)***
Datastore Size	*6.50B*	*1.19B*	*1.13B*

Key Results

Memorizing the training data improves machine translation generalization, and allows a multilingual model to specialize.

A single translation model can adapt to multiple domains by memorizing domain-specific data, without any in-domain training.

Memorization can make model predictions more interpretable.

Domain Adaptation in MT: News to Medical

MT Training Data	*Datastore*	*BLEU on Medical (↑)*
Medical (Aharoni & Goldberg, 2020)	-	*56.65*
News	-	*39.91*

Domain Adaptation in MT: News to Medical

MT Training Data	*Datastore*	*BLEU on Medical (↑)*
Medical (Aharoni & Goldberg, 2020)	-	*56.65*
News	-	*39.91*
News	*Medical (5.7M)*	*54.35*

A single MT model can be useful in multiple domains by simply adding a domain-specific datastore!

Domain Adaptation in MT: News to Legal

MT Training Data	*Datastore*	*BLEU on Legal (↑)*
Legal (Aharoni & Goldberg, 2020)	-	*59.00*
News	-	*45.71*
News	*Legal (18.3M)*	*61.78*

A single MT model can be useful in multiple domains by simply adding a domain-specific datastore!

Key Results

Memorizing the training data improves machine translation generalization, and allows a multilingual model to specialize.

A single translation model can adapt to multiple domains by memorizing domain-specific data, without any in-domain training.

Memorization can make model predictions more interpretable.

German input: *Dabei schien es, als habe Erdogan das Militär gezähmt.*

English output so far: *In doing so, it seems as if Erdogan has tamed the*

German input: *Dabei schien es, als habe Erdogan das Militär gezähmt.*

English output so far: *In doing so, it seems as if Erdogan has tamed the*

German (Source)	*English (Prior Context)*	*Value*
Dem charismatischen Ministerpräsidenten Recep Tayyip Erdoğan, der drei aufeinanderfolgende Wahlen für sich entscheiden konnte, ist es gelungen seine ***Autorität gegenüber dem Militär*** *geltend zu machen.*	*The charismatic prime minister, Recep Tayyip Erdoğan, having won three consecutive elections, has been able to exert his* ***authority over the***	*military (p = 0.132)*
Ein bemerkenswerter Fall war die Ermordung des gemäßigten Premierministers Inukai Tsuyoshi im Jahre 1932, die das Ende jeder wirklichen zivilen ***Kontrolle des Militärs*** *markiert.*	*One notable case was the assassination of moderate Prime Minister Inukai Tsuyoshi in 1932, which marked the end of any real civilian* ***control of the***	*military (p = 0.130)*
Sie sind Teil eines Normalisierungsprozesses und der Herstellung der absoluten zivilen ***Kontrolle über das Militär*** *und bestätigen das Prinzip, dass niemand über dem Gesetz steht.*	*They are part of a process of normalization, of the establishment of absolute civilian* ***control of the***	*military (p = 0.129)*

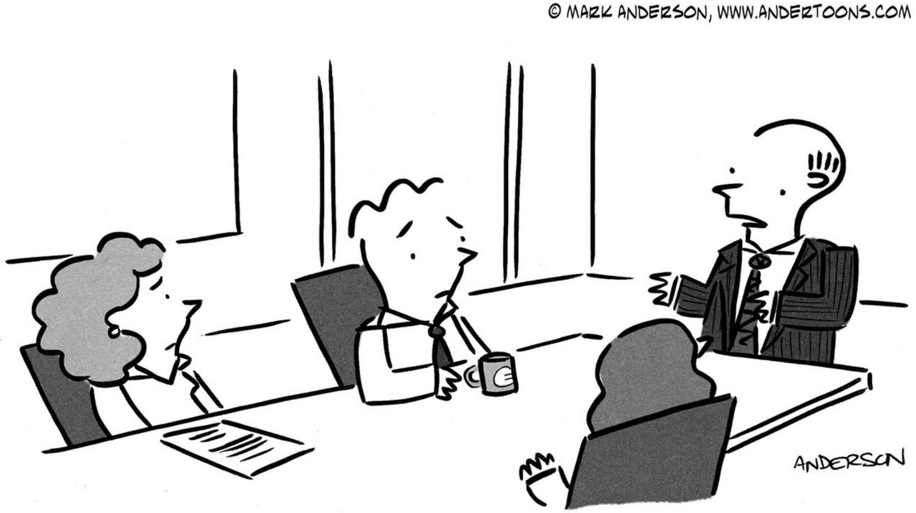
Key Results

Memorizing the training data improves machine translation generalization, and allows a multilingual model to specialize.

A single translation model can adapt to multiple domains by memorizing domain-specific data, without any in-domain training.

Memorization can make model predictions more interpretable.

Thanks!



"To make a long story short, what it all boils down to in the final analysis is that what you should take away from this is…"

Paper: https://arxiv.org/pdf/2010.00710.pdf

Memorizing the training data improves machine translation generalization, and allows a multilingual model to specialize.

A single translation model can adapt to multiple domains by memorizing domain-specific data, without any in-domain training.

Memorization can make model predictions more interpretable.