Sentence simplification with character-level transformer

Kuan Yu[1]

October 29, 2018

[1]kuanyu@uni-potsdam.de

sentence simplification

- as machine translation
- at the character-level
- using a transformer

"It is located in Potsdam ." $\longmapsto$ "It is in Potsdam ."

https://github.com/ysmiraak/explicharr

sentence simplification as machine translation

Wubben, Van Den Bosch, and Krahmer (2012)

"Sentence simplification by monolingual machine translation"

Narayan and Gardent (2014)

"Hybrid simplification using deep semantics and machine translation"

Xu, Napoles, et al. (2016)

"Optimizing statistical machine translation for text simplification"

modelling level

characters (Kalchbrenner et al. 2016)

"Neural machine translation in linear time"

subwords/wordpieces (Sennrich, Haddow, and Birch 2015)

"Neural Machine Translation of Rare Words with Subword Units"

word-character hybrid (Luong and Manning 2016)

"Achieving open vocabulary neural machine translation with hybrid word-character models"

simple english wikipedia

Napoles and Dredze (2010)

"Learning simple Wikipedia: A cogitation in ascertaining abecedarian language"

Coster and Kauchak (2011)

"Simple English Wikipedia: a new text simplification task"

Hwang et al. (2015)

"Aligning sentences from standard wikipedia to simple wikipedia"

Xu, Callison-Burch, and Napoles (2015)

"Problems in current text simplification research: New data can help"

aligned wikipedia sentences (Hwang et al. 2015)

226,208 instances after cleaning up

- pairs with identical source and target
- overly long sentences (length $>$ 256)

256 characters modelled

- over 2,000 characters in either parts
- the top 256 make up 99.97% of the data

1% random split for validation

evaluation metrics

loss and accuracy

- cross-entropy loss
- with label smoothing (Szegedy et al. 2016)

BLEU score

- not suitable for text simplification (Štajner, Béchara, and Saggion 2015)
- ~78% character pairs by positions are identical

manual inspection

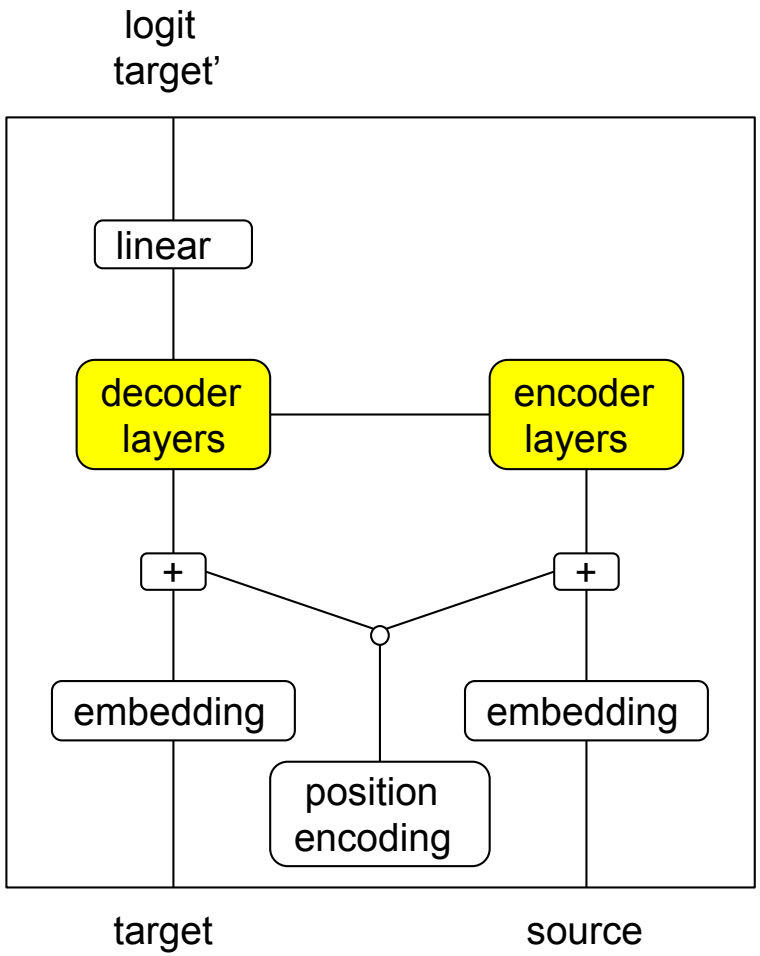
transformer

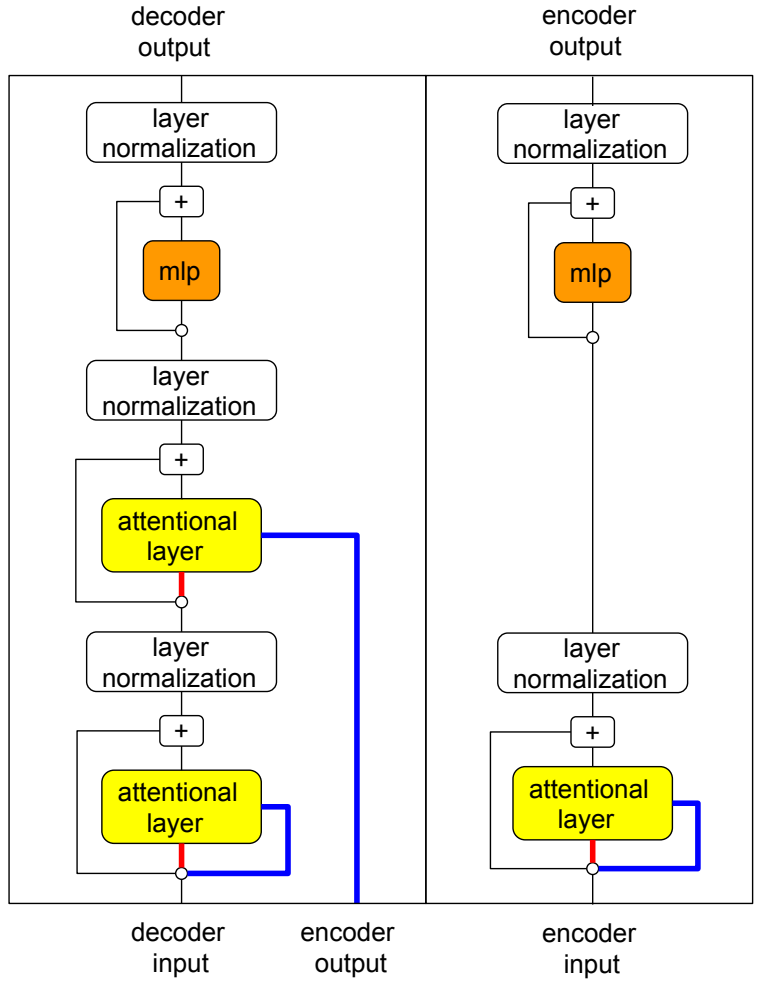
"Attention is all you need" (Vaswani et al. 2017)

other applications

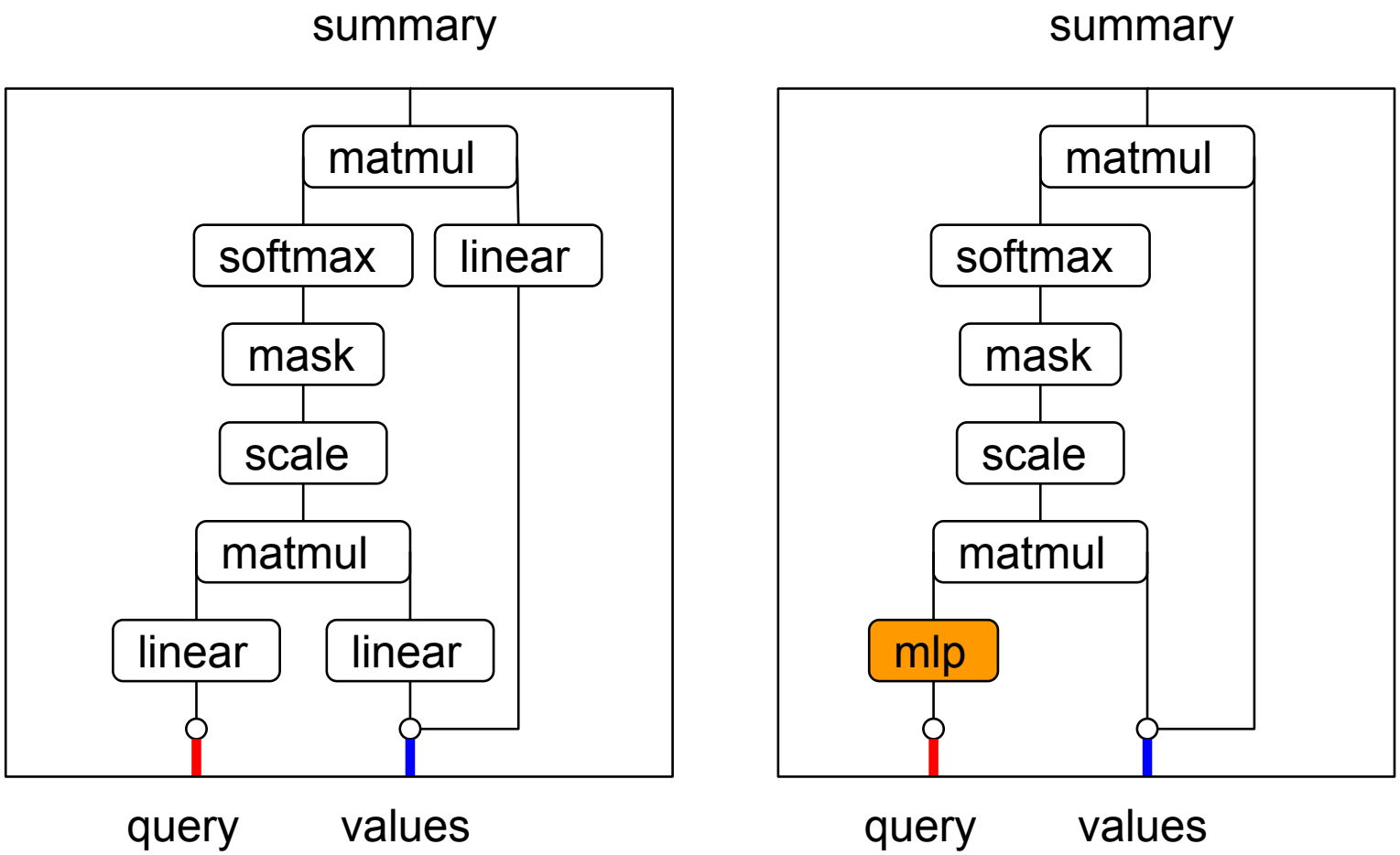
- constituency parsing (Kitaev and Klein 2018)
- language generation (Liu et al. 2018)
- image generation (Parmar et al. 2018)
- speech recognition (Zhou, Dong, et al. 2018; Zhou, Xu, and Xu 2018)
- speech synthesis (Li et al. 2018)

architecture





attentional layer



left: original scaled dot-product attention. right: our variant.

examples from the validation set

```
source from the english wikipedia
prediction from our system
aligned target from the simple english wikipedia
```

- over 80% of the predictions are identical to the source

example 1

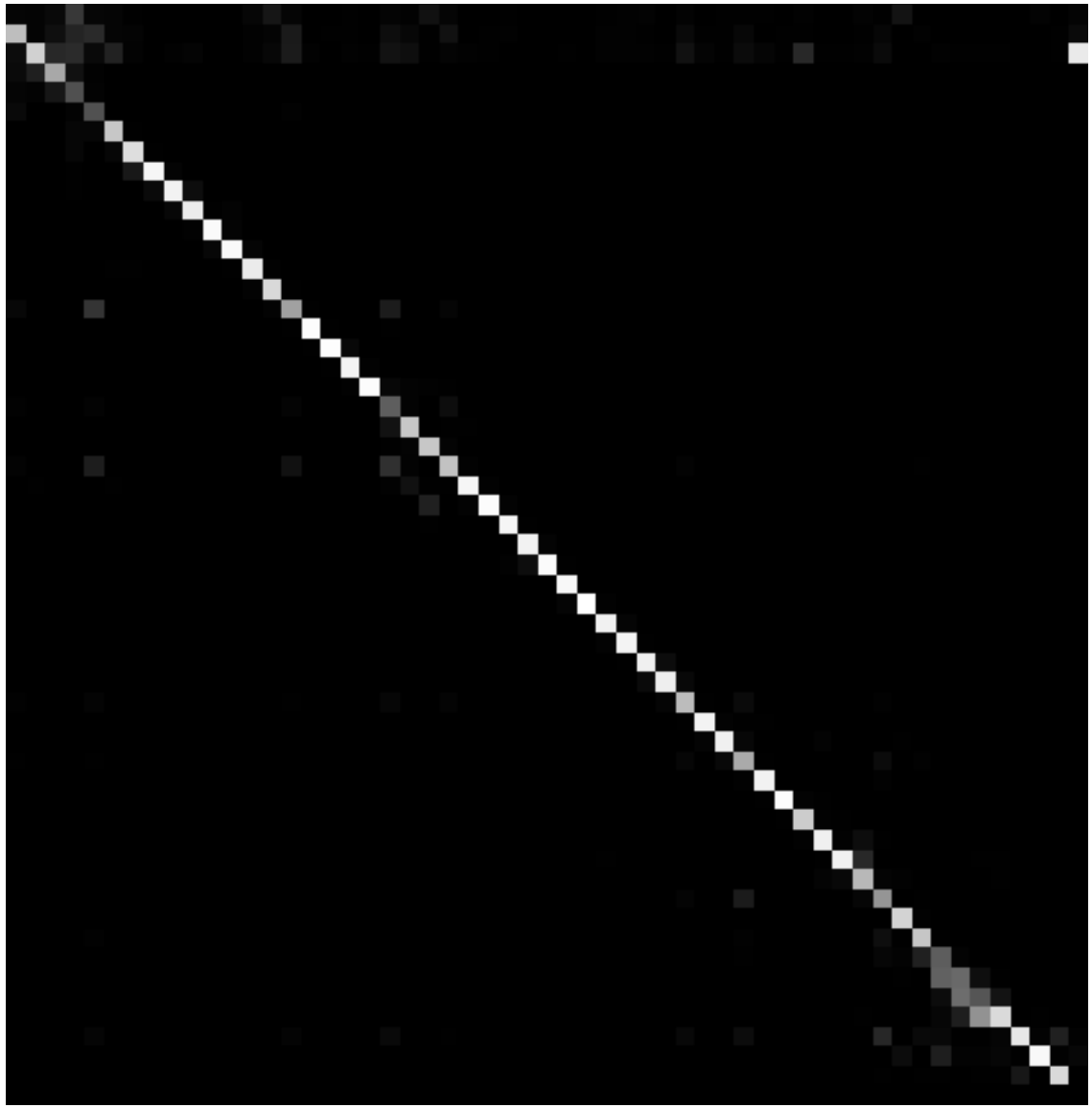
```
The enshrined kami is Isonotakeru
 no mikoto ( 五十猛命 ? )

The enshrined kami is Isonotakeru
 no mikoto (      ? )

This place is special to the kami
 named Isonotakeru no mikoto ( 五十猛命 ? )
```

- text normalization: ignores unknown characters

attention alignment: copy



example 2

Edo (江户 ?

Edo (?)

Edo (江户 ?)

- text normalization: fills missing brackets

example 3

```
Buddha-Bhagavan ) .

Buddha-Bhagavan .

Buddhists believe there have been many Buddhas
 who were alone , called pacceka-buddhas .
```

- text normalization: removes brackets

example 4

```
Punch 's wife was originally called `` Joan . ''
Punch 's wife was originally called `` Joan '' .
Punch 's wife was originally `` Joan '' .
```

- text normalization: prefers not to end sentences with quotes

example 5

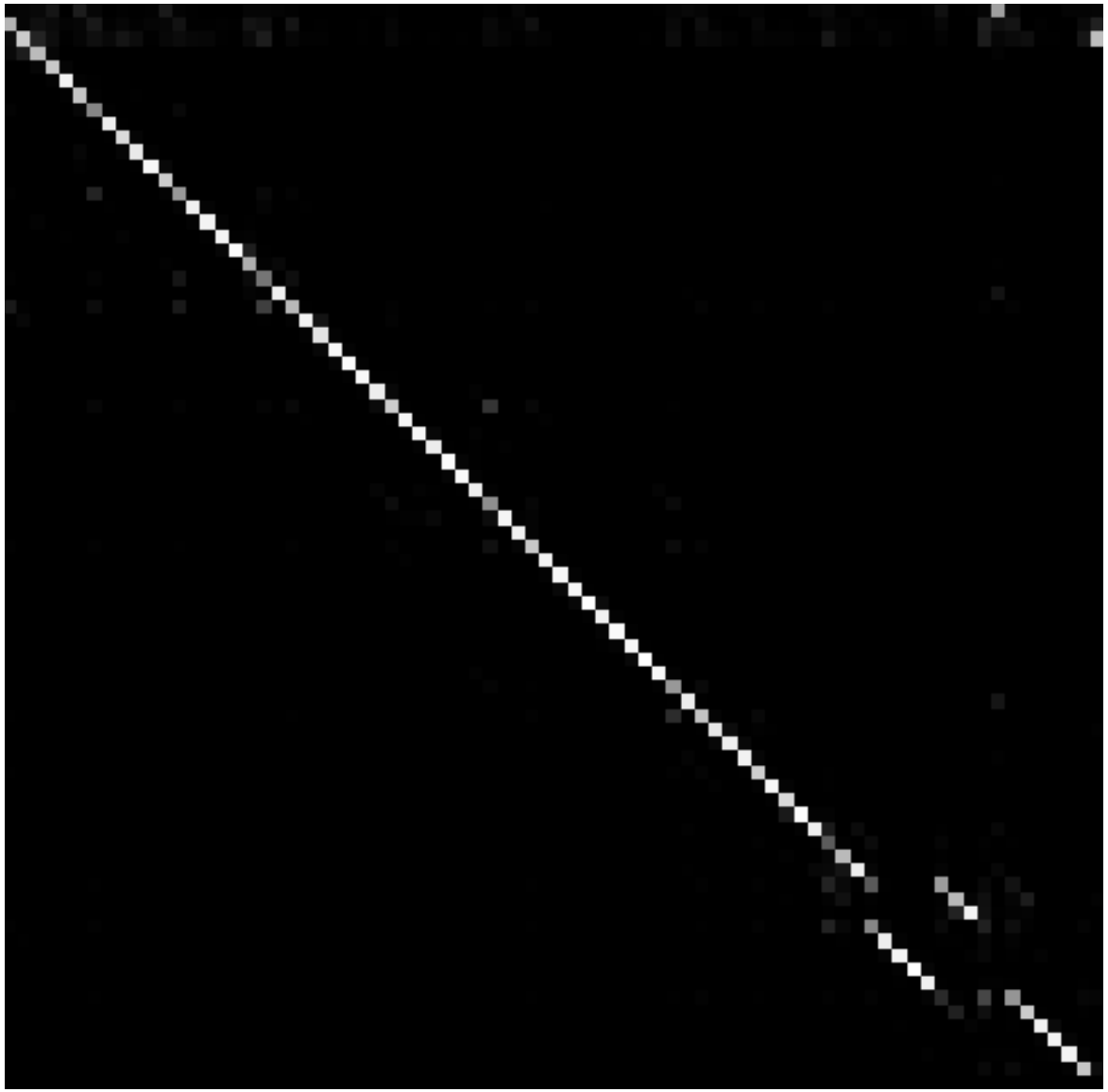
```
Their first child , Prince George of Cambridge
 , was born on 22 July 2013 .

Their first child , Prince George of Cambridge
 , was born on July 22 , 2013 .

On July 24 , it was announced that the baby
 would be called Prince George of Cambridge
 , with the full given name George Alexander Louis .
```

- text normalization: changes date format

attention alignment: swap



example 6

```
is a 1982 arcade-style platform
 video game by Nintendo .

In 1982 , the style platform
 video game by Nintendo .

Donkey Kong Jr. is an arcade
 video game made by Nintendo in 1982 .
```

- incomplete sentences: reduces to noun phrases

example 7

```
reflects the influence of the sexagenary cycle
 as a count of years .
```

```
The influence of the sexagenary cycle
 as a count of years .
```

```
Ancient records show that the sexagenary cycle
 was used to count years in the 3rd century B.C. .
```

- incomplete sentences: reduces to noun phrases

example 8

```
and Mario 's Time Machine .

The machine is a machine .

In the Dr. Mario games first seen in 1990
 , Mario is a doctor .
```

- incomplete sentences: reduces to trivial sentences

example 9

```
and Airplane II : The Sequel .
The Sequel is a sequel .
( 1980 ) and Airplane II : The Sequel ( 1982 ) .
```

- incomplete sentences: reduces to trivial sentences

example 10

```
and it is because they do n't understand
 what Parkour is ; ` Who is the best ? '

They are the best ?

Star Jumping Parkour does involve risks
 and many people get injured every day .
```

- incomplete sentences: attempts to complete

example 11

```
It is officially known as Avtomat Kalashnikova
 ( Russian : Автомат Калашникова ) .

It is known as Avtomat Kalashnikova
 ( Russian : : 1950 : 1998 ) .

The letters AK stand for Avtomat Kalashnikova
 , which is Russian for Kalashnikov 's
 Automatic Rifle .
```

- incomplete sentences: attempts to complete

example 12

```
With one huge blow from his olive-wood club
 , Hercules killed the watchdog .

Hercules killed the watchdog .

Herakles killed her .
```

- simplification: removes initial adverbials

example 13

```
For example , the speed of sound in gases
 depends on temperature .

The speed of sound in gases depends on temperature .

Loudness depends on sound intensity
 , sound frequency , and the person 's hearing .
```

- simplification: removes initial adverbials

example 14

```
In it , Goldilocks is re-imagined
 as a girl in her 20s .

Goldilocks is re-imagined as a girl in her 20s .

She finally became Goldilocks sometime
 in the early 20th century .
```

- removes the initial adverbial

example 15

```
In 2008 , she starred in the comedy film Baby Mama
 , alongside former SNL co-star Amy Poehler .
In 2008 , she starred in the comedy movie Baby Mama
 , alongside former SNL co-star Amy Poehler .
In 2008 , she starred in the comedy movie Baby Mama
 , with Amy Poehler .
```

- does not always remove the initial adverbial

example 16

```
The spines , which may be up to 50 mm ( 2 in ) long
 , are modified hairs , mostly made of keratin .

The spines , which may be up to 50 mm long
 , are modified hairs , mostly made of keratin .

The spines , which may be up to 50 mm long
 , are modified hairs , mostly made of keratin .
```

- simplification: removes technical details

example 17

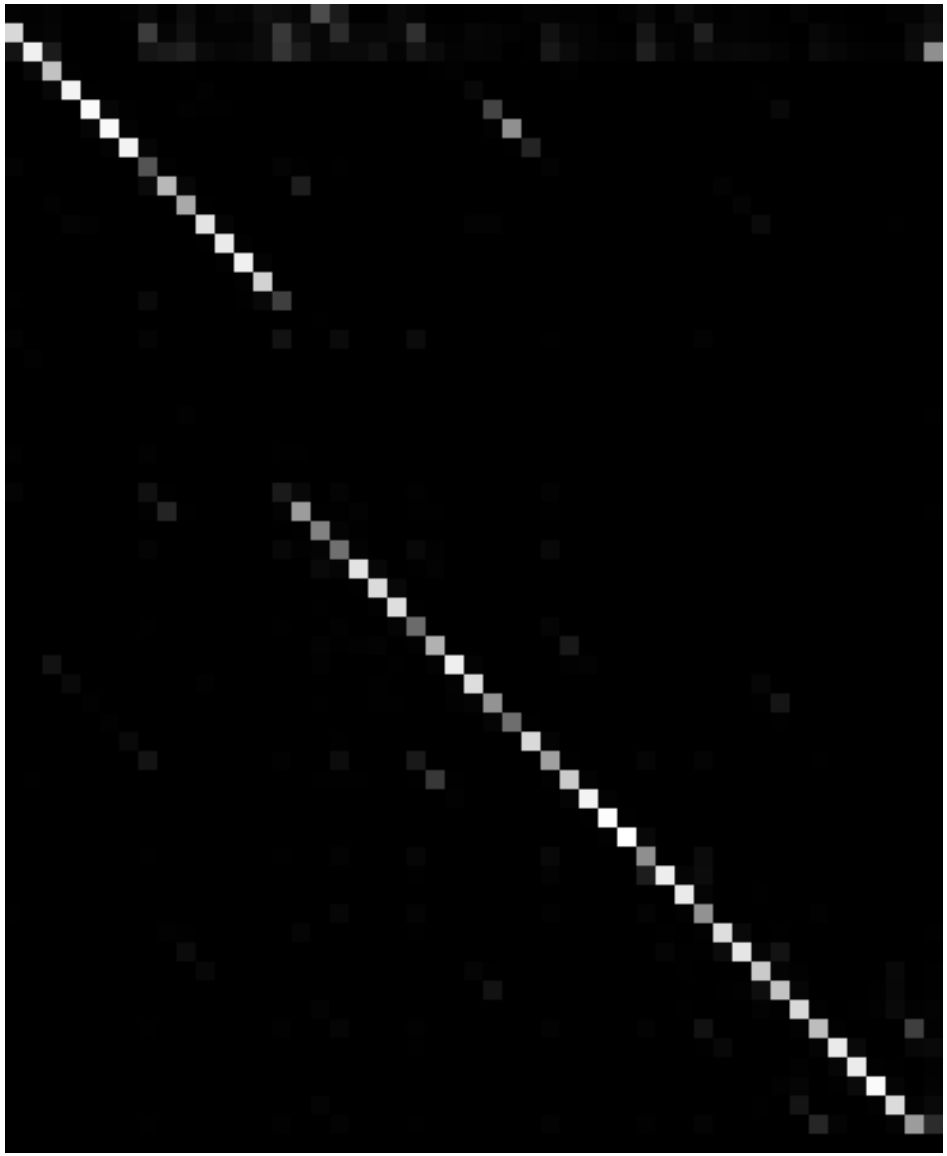
```
Sodium iodate ( NaIO3 ) is
 the sodium salt of iodic acid .

Sodium iodate is the sodium salt of iodic acid .

Sodium iodate is a chemical compound .
```

- simplification: removes technical details

attention alignment: skip



example 18

```
The palate / 'pælɨt / is the roof of the mouth
 in humans and other mammals .

The palate is the roof of the mouth
 in humans and other mammals .

The palate is the roof of the mouth
 in humans and other mammals .
```

- simplification: removes technical details

example 19

```
It is located in the Market Square
 , Aylesbury , Buckinghamshire .

It is in the Market Square
 , Aylesbury , Buckinghamshire .

The King 's Head is a National Trust property
 in Aylesbury , Buckinghamshire , England , UK .
```

- simplification: removes words with little information

example 20

```
Home Alone received mixed reviews from film critics .

Home Alone received mixed reviews from critics .

Home Alone 3 ( 1997 ) and Home Alone 5
 : The Holiday Heist ( 2012 ) have completely
 different characters and actors .
```

- simplification: removes words with little information

example 21

```
However , low frequency also causes noticeable
 flicker in arc lamps and incandescent light bulbs .

The low frequency also causes noticeable
 flicker in arc lamps and incandescent light bulbs .

This happens so fast that a light bulb
 does not stop glowing .
```

- may also add words

example 22

```
In the film 's story the location
 was named as Footscray station .

In the movie 's story the location
 was named as Footscray station .

In the movie it was called Footscray station .
```

- simplification: substitutes words

example 23

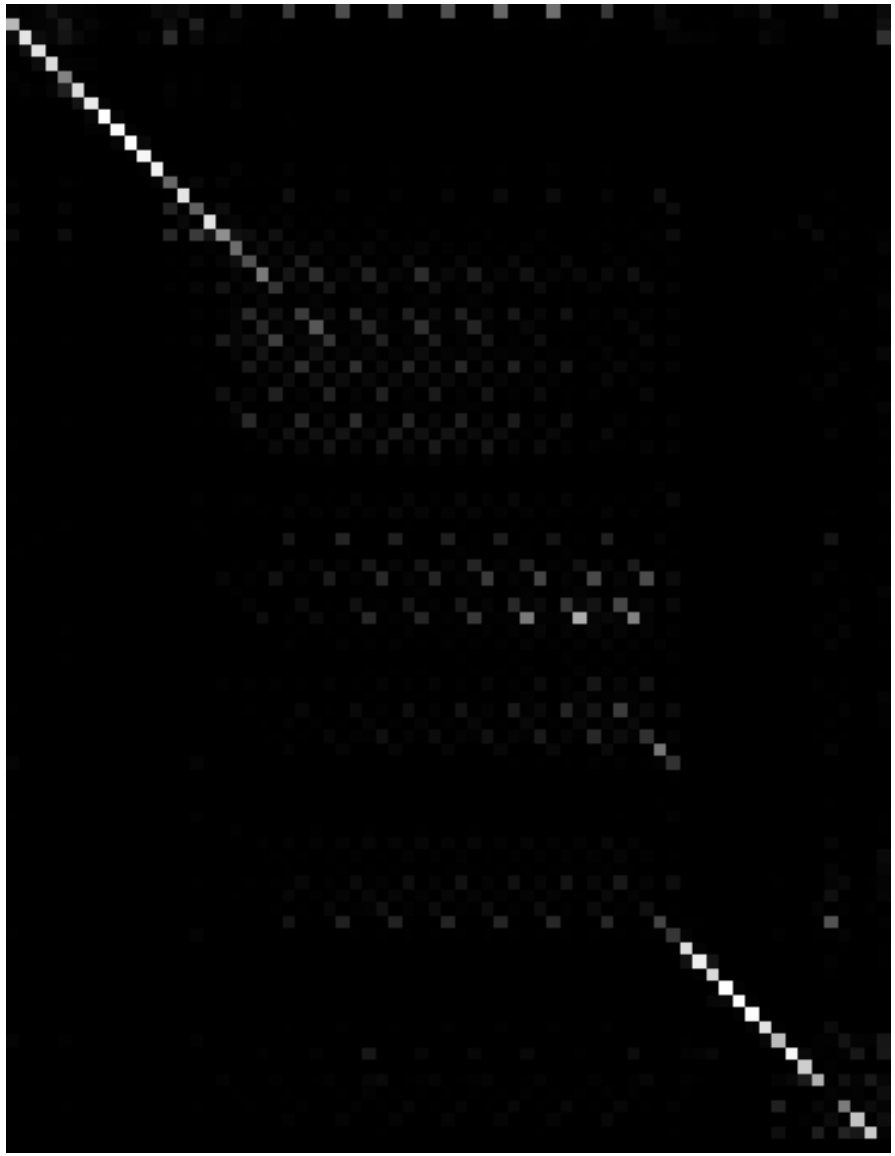
```
For example , 2 + 3 = 3 + 2 = 5
 and 2 · 3 = 3 · 2 = 6 , but 23 = 8
 , whereas 32 = 9 .

For example , 2 + 3 = 3 = 3 = 3 = 3 = 3 = 3 = 3
 , whereas 32 = 9 .

Then people write powers using the ^ sign
 : 2 ^ 3 means .
```

- gets confused by math

attention alignment: confused



example 24

```
1 and - 1 divide ( are divisors of ) every integer .

In the 19th century , the state of the state
 of the state of the state of the state
 of the state of the state .

For example , if the last digit is even
 ( 0 , 2 , 4 , 6 or 8 ) , then 2 is a divisor .
```

- gets really confused by math

attention alignment: really confused

our model learned

- the common forms of english sentences
- the general differences between the two wikipedias
- limited abilities to simplify constituents, often by deletion
- no ability for syntactic transformations
- no understanding for semantic structures

remedies

still as supervised machine translation

- better datasets are needed
- with suitable examples of simplification
- and interesting linguistic properties

still with wikipedia datasets

- aligned sentences are not enough
- unsupervised machine translation (Lample, Denoyer, and Ranzato 2017)

references I


Napoles, Courtney and Mark Dredze (2010). "Learning simple Wikipedia: A cogitation in ascertaining abecedarian language". In: *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing: Writing Processes and Authoring Aids*. Association for Computational Linguistics, pp. 42–50.


Coster, William and David Kauchak (2011). "Simple English Wikipedia: a new text simplification task". In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, pp. 665–669.


Wubben, Sander, Antal Van Den Bosch, and Emiel Krahmer (2012). "Sentence simplification by monolingual machine translation". In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics, pp. 1015–1024.

references II


Narayan, Shashi and Claire Gardent (2014). "Hybrid simplification using deep semantics and machine translation". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vol. 1, pp. 435–445.


Hwang, William et al. (2015). "Aligning sentences from standard wikipedia to simple wikipedia". In: *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 211–217.


Sennrich, Rico, Barry Haddow, and Alexandra Birch (2015). "Neural machine translation of rare words with subword units". In: *arXiv preprint arXiv:1508.07909*.

references III


Štajner, Sanja, Hannah Béchara, and Horacio Saggion (2015). "A deeper exploration of the standard PB-SMT approach to text simplification and its evaluation". In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Vol. 2, pp. 823–828.


Xu, Wei, Chris Callison-Burch, and Courtney Napoles (2015). "Problems in current text simplification research: New data can help". In: *Transactions of the Association of Computational Linguistics* 3.1, pp. 283–297.


Kalchbrenner, Nal et al. (2016). "Neural machine translation in linear time". In: *arXiv preprint arXiv:1610.10099*.


Luong, Minh-Thang and Christopher D Manning (2016). "Achieving open vocabulary neural machine translation with hybrid word-character models". In: *arXiv preprint arXiv:1604.00788*.

references IV

- Szegedy, Christian et al. (2016). "Rethinking the inception architecture for computer vision". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.
- Xu, Wei, Courtney Napoles, et al. (2016). "Optimizing statistical machine translation for text simplification". In: *Transactions of the Association for Computational Linguistics* 4, pp. 401–415.
- Lample, Guillaume, Ludovic Denoyer, and Marc'Aurelio Ranzato (2017). "Unsupervised machine translation using monolingual corpora only". In: *arXiv preprint arXiv:1711.00043*.
- Vaswani, Ashish et al. (2017). "Attention is all you need". In: *Advances in Neural Information Processing Systems*, pp. 5998–6008.
- Kitaev, Nikita and Dan Klein (2018). "Constituency Parsing with a Self-Attentive Encoder". In: *arXiv preprint arXiv:1805.01052*.

references V

- Li, Naihan et al. (2018). "Close to Human Quality TTS with Transformer". In: *arXiv preprint arXiv:1809.08895*.
- Liu, Peter J et al. (2018). "Generating wikipedia by summarizing long sequences". In: *arXiv preprint arXiv:1801.10198*.
- Parmar, Niki et al. (2018). "Image Transformer". In: *arXiv preprint arXiv:1802.05751*.
- Zhou, Shiyu, Linhao Dong, et al. (2018). "Syllable-Based Sequence-to-Sequence Speech Recognition with the Transformer in Mandarin Chinese". In: *arXiv preprint arXiv:1804.10752*.
- Zhou, Shiyu, Shuang Xu, and Bo Xu (2018). "Multilingual End-to-End Speech Recognition with A Single Transformer on Low-Resource Languages". In: *arXiv preprint arXiv:1806.05059*.