Controlling Text Generation

Alexander Rush

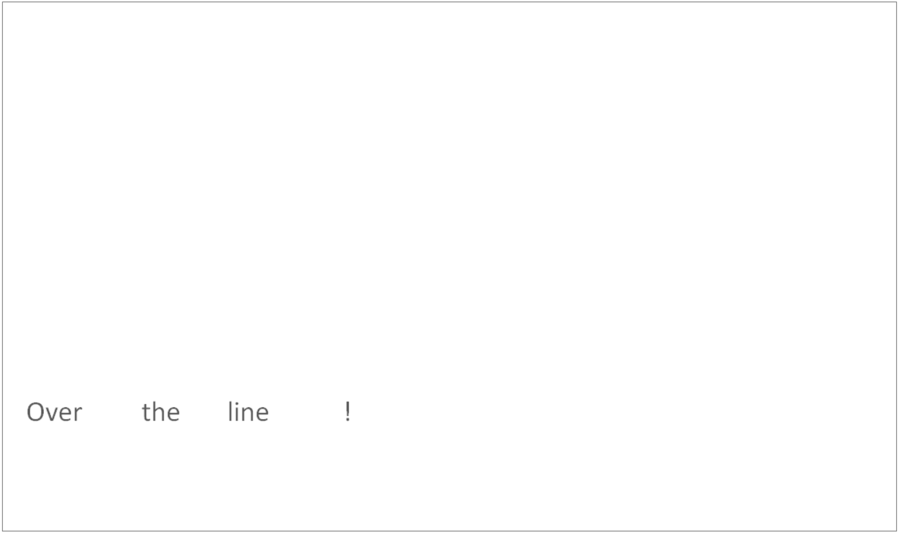
(based on work by Sam Wiseman,
Sebastian Gehrmann, and Yuntian Deng)

HarvardNLP

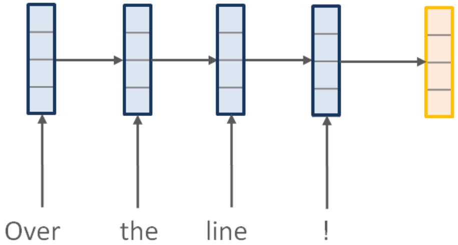
1. Review: Neural Machine Translation
2. Past Work: Some Progress on Text Generation
    - Summary
    - Image-to-Markup
    - Data-to-Text
3. Current Research: Uphill Battles in Generation
    - Generation Analysis
    - Controlling Generation

1) Preface: End-to-End Models for NLP

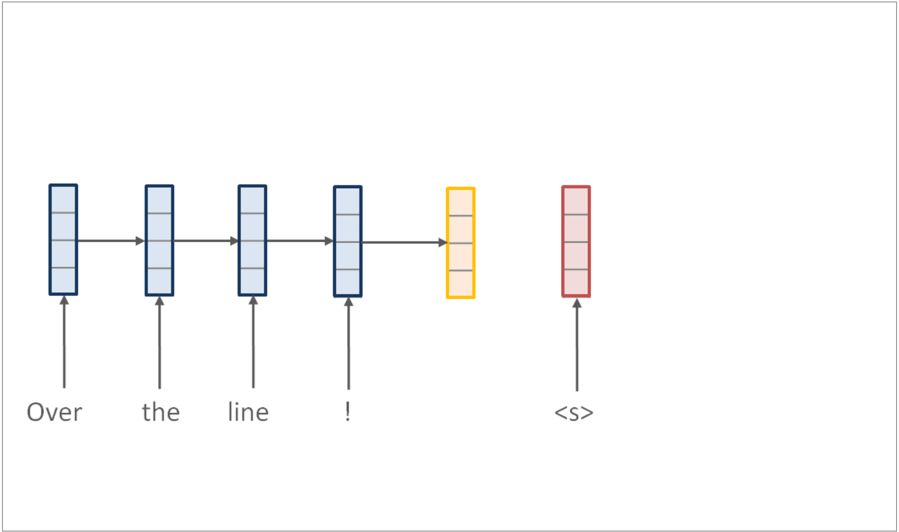
Example: Neural Machine Translation (Sutskever et al., 2014)



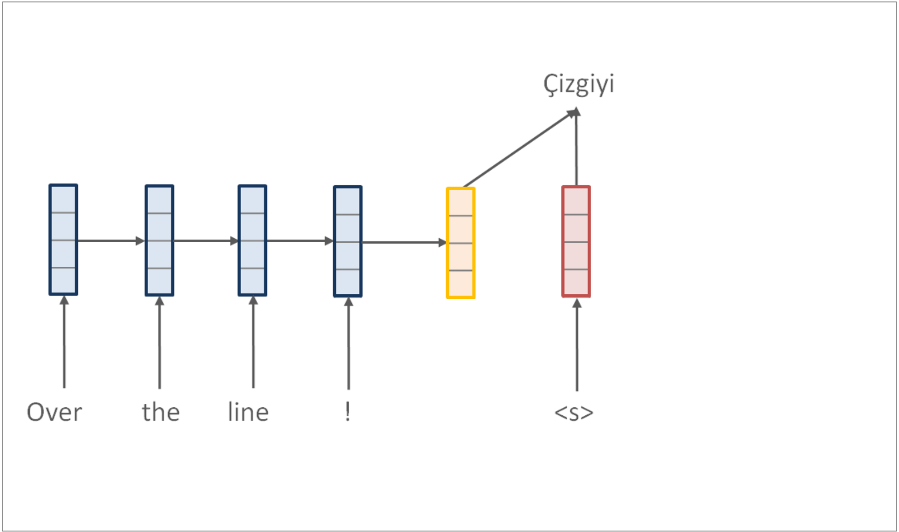
Example: Neural Machine Translation (Sutskever et al., 2014)



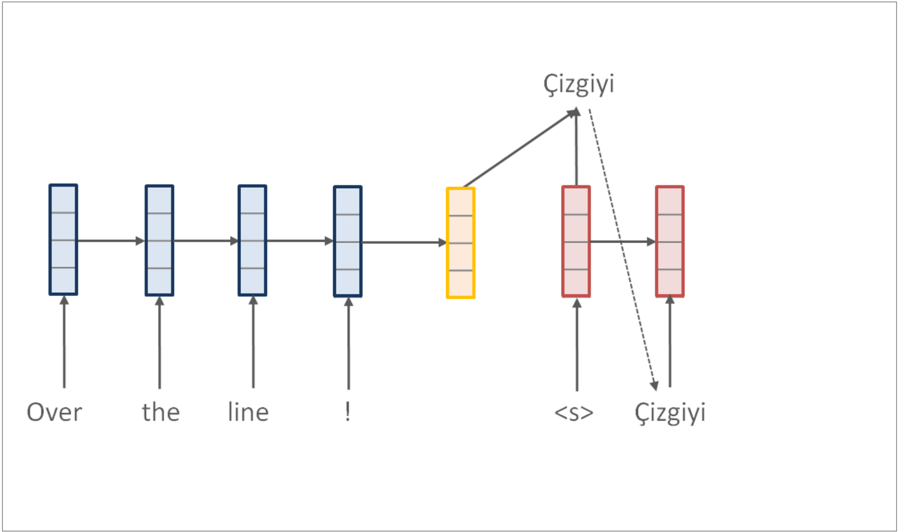
Example: Neural Machine Translation (Sutskever et al., 2014)



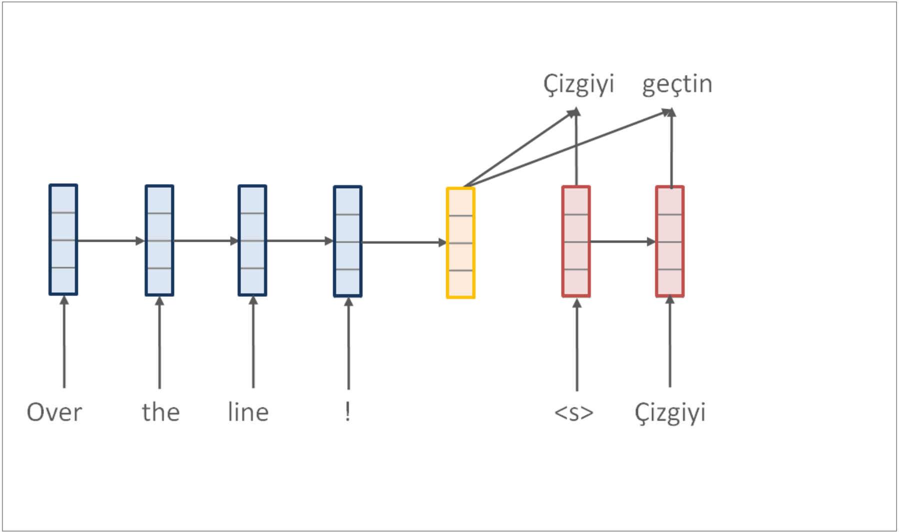
Example: Neural Machine Translation (Sutskever et al., 2014)



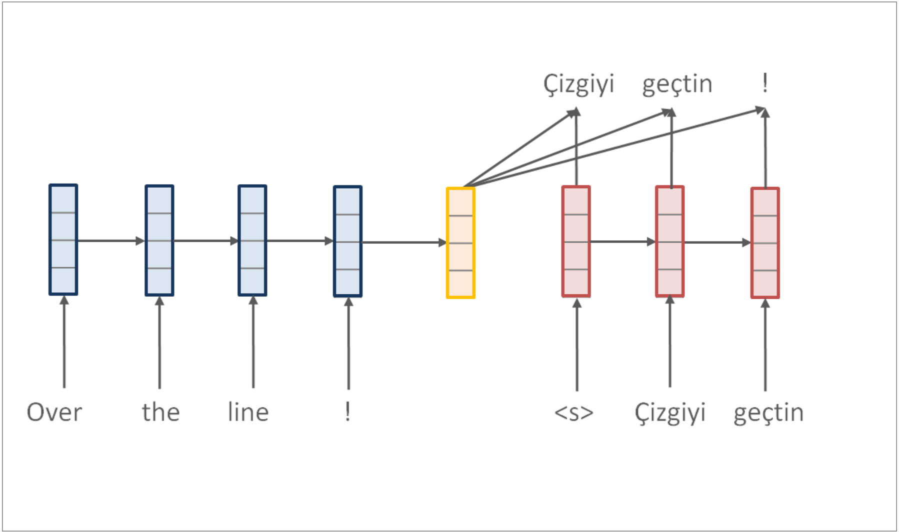
Example: Neural Machine Translation (Sutskever et al., 2014)



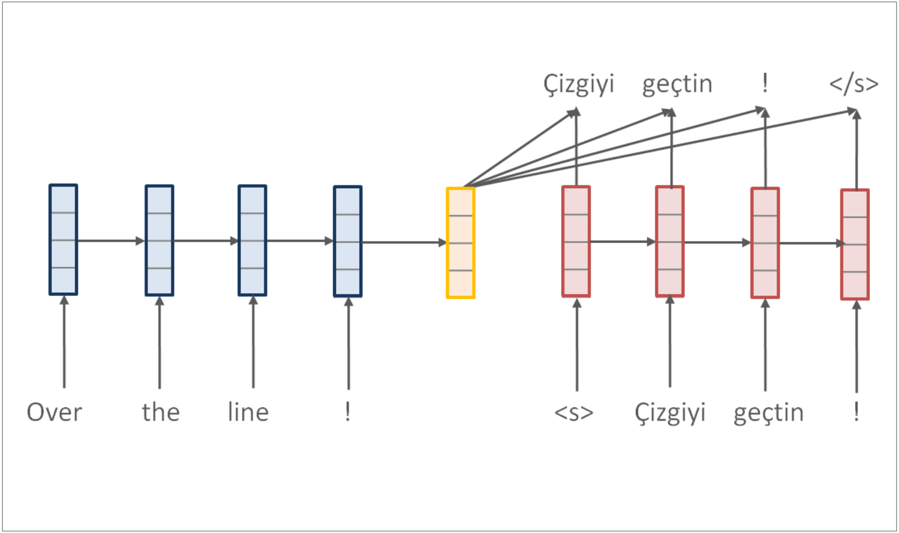
Example: Neural Machine Translation (Sutskever et al., 2014)



Example: Neural Machine Translation (Sutskever et al., 2014)



Example: Neural Machine Translation (Sutskever et al., 2014)



Model 1: Seq2Seq Model

Encoder ($enc(x)$):

$$\mathbf{h}_m^x \leftarrow \mathrm{RNN}(\mathbf{h}_{m-1}^x, x_m)$$

Context:

$$\mathbf{c}_n = \mathbf{h}_M^x$$

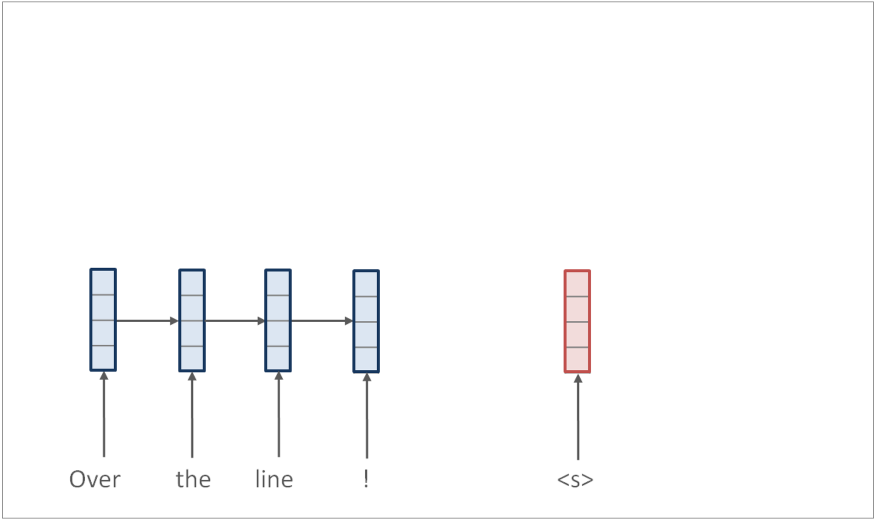
Decoder ($dec(\mathbf{c}_n)$):

$$\mathbf{h}_n \leftarrow \mathrm{RNN}(\mathbf{h}_{n-1}, w_n)$$

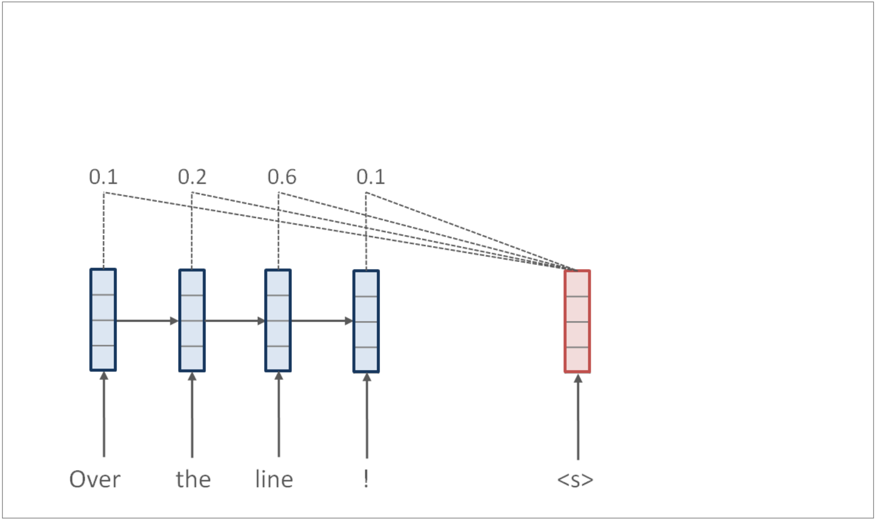
Prediction:

$$p(w_{n+1} \mid w_{1:n}, x_{1:M}) = \mathrm{softmax}(\mathbf{W}[\mathbf{h}_n, \mathbf{c}_n])$$

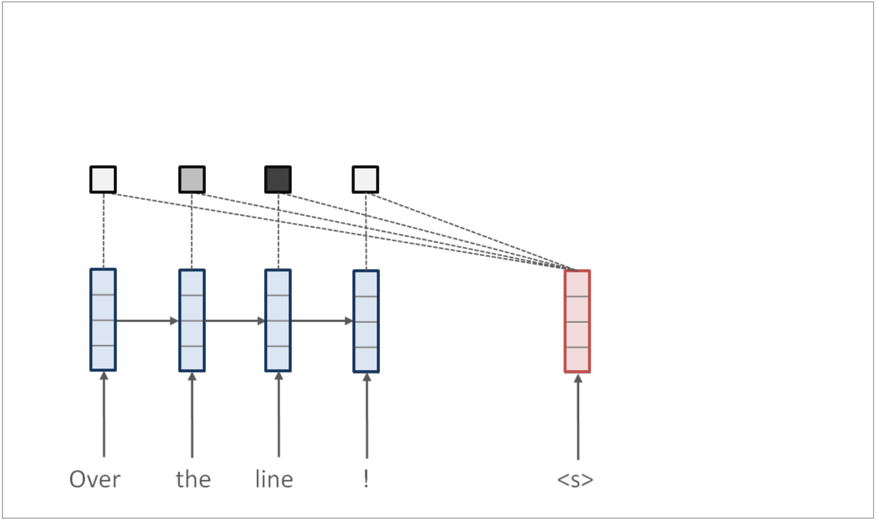
Attention-based Neural Machine Translation (Bahdanau et al., 2015)



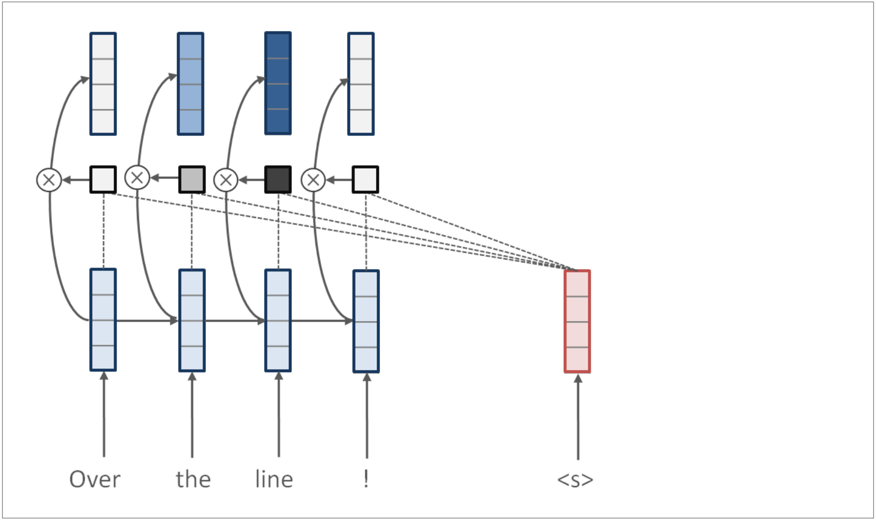
Attention-based Neural Machine Translation (Bahdanau et al., 2015)



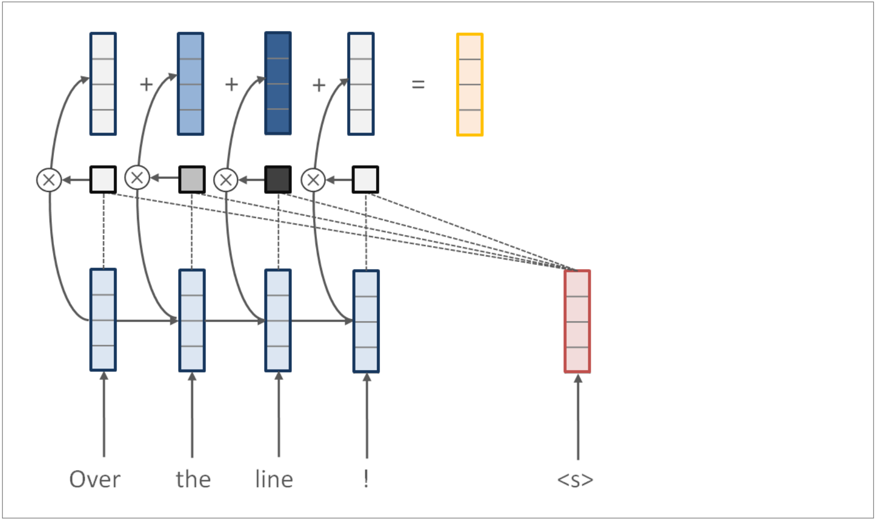
Attention-based Neural Machine Translation (Bahdanau et al., 2015)



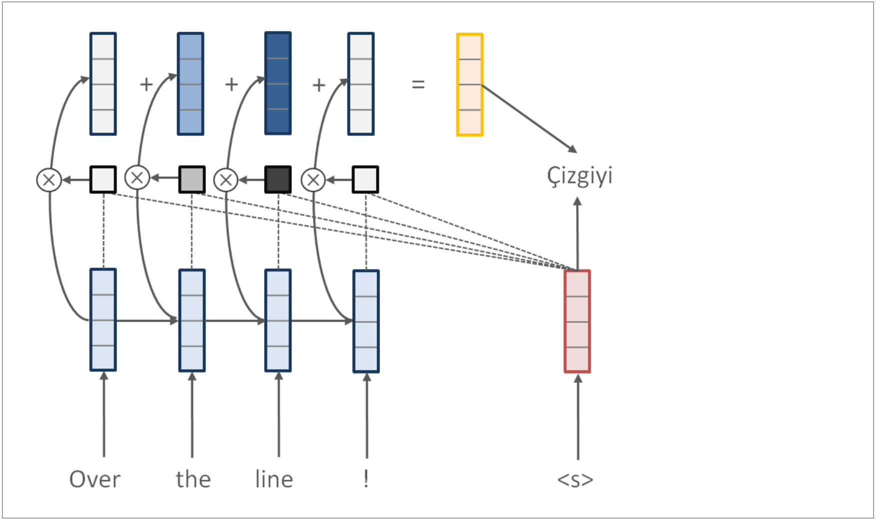
Attention-based Neural Machine Translation (Bahdanau et al., 2015)



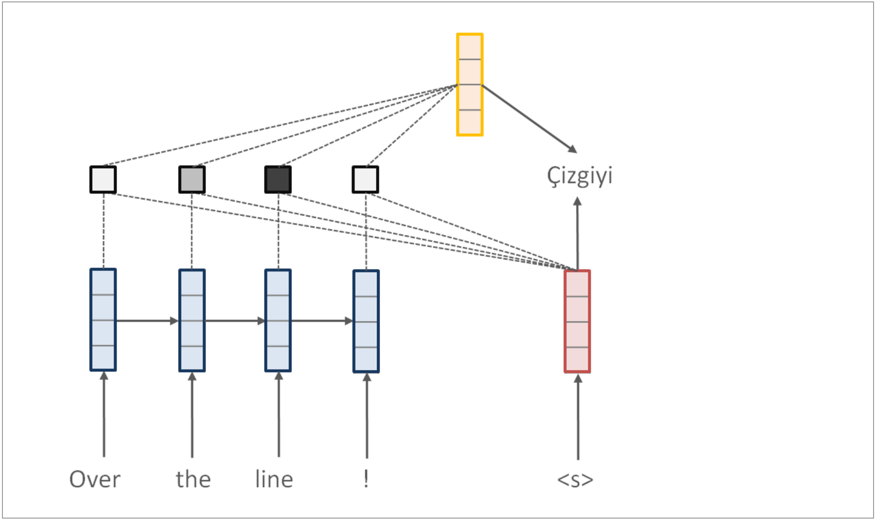
Attention-based Neural Machine Translation (Bahdanau et al., 2015)



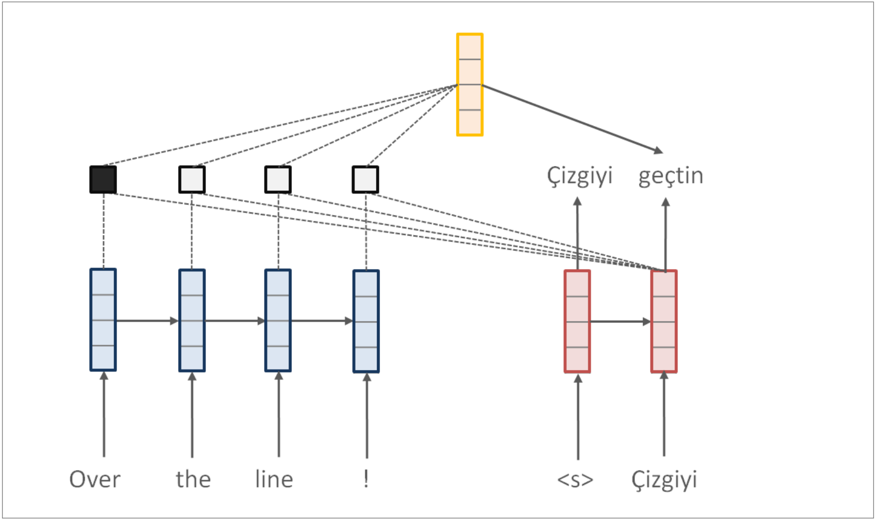
Attention-based Neural Machine Translation (Bahdanau et al., 2015)



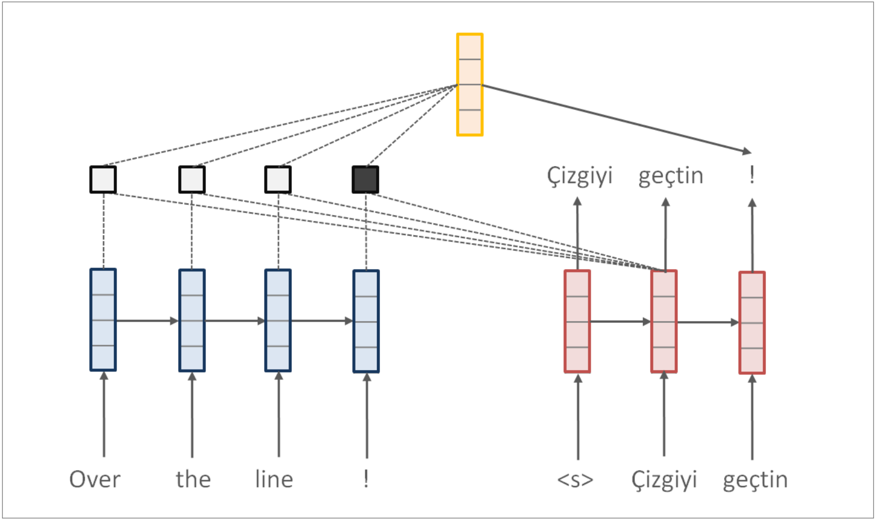
Attention-based Neural Machine Translation (Bahdanau et al., 2015)



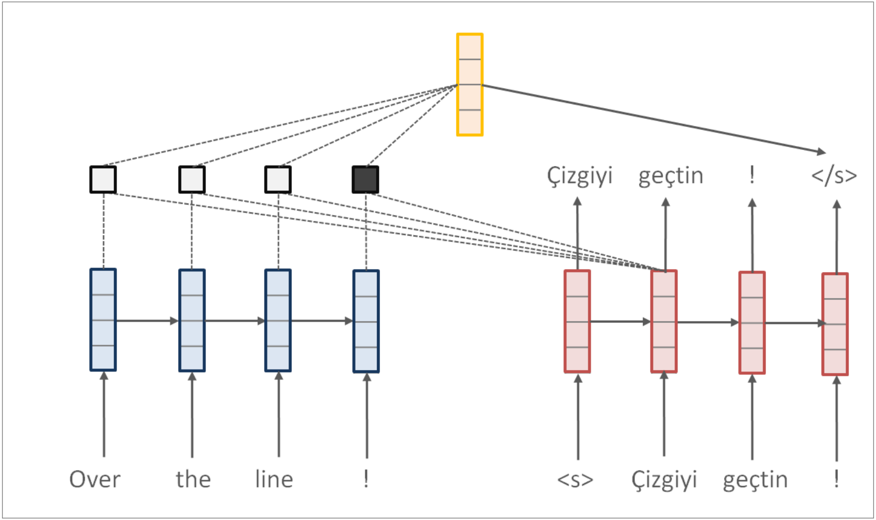
Attention-based Neural Machine Translation (Bahdanau et al., 2015)



Attention-based Neural Machine Translation (Bahdanau et al., 2015)



Attention-based Neural Machine Translation (Bahdanau et al., 2015)



Model 2: Seq2Seq+Attention Model

Encoder ($enc(x)$):

$$\mathbf{h}^x_m \leftarrow \mathrm{RNN}(\mathbf{h}^x_{m-1}, x_m)$$

Attention

$$p_{att}(m) \leftarrow \mathrm{softmax}([\mathbf{h}^x_1; \ldots; \mathbf{h}^x_M]^\top \mathbf{h}_n)$$

$$\mathbf{c}_n \leftarrow \mathbb{E}_{m \sim p_{att}}[\mathbf{h}^x_m] = \sum_{m=1}^{M} p_{att}(m)\mathbf{h}^x_m$$

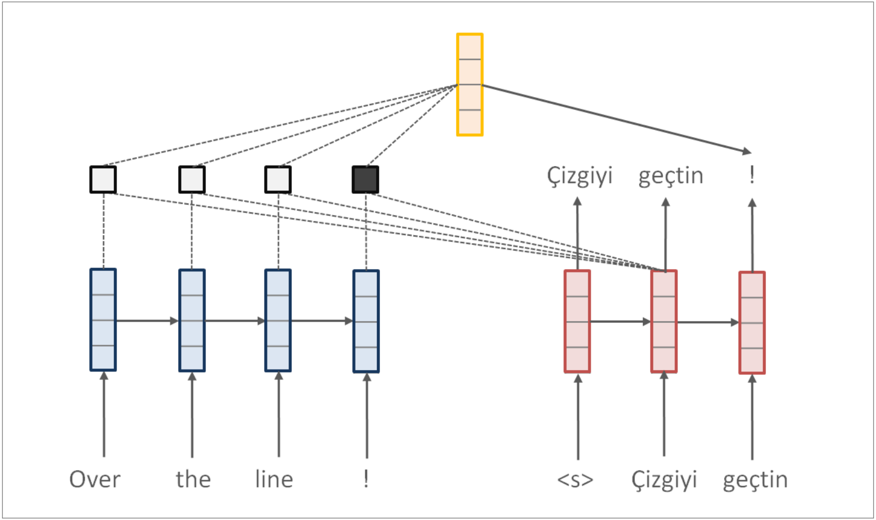
Decoder ($dec(\mathbf{c}_n)$):

$$\mathbf{h}_n \leftarrow \mathrm{RNN}(\mathbf{h}_{n-1}, w_n)$$

Prediction

$$p(w_{n+1} | w_{1:n}, x_{1:M}) = \mathrm{softmax}(\mathbf{W}[\mathbf{h}_n, \mathbf{c}_n])$$

Attention-based Neural Machine Translation (Bahdanau et al., 2015)



Model 3: Seq2Seq+Attention+Copy Model (Gulcehre et al, 2016) ...

Encoder $(enc(x))$:

$$\mathbf{h}_m^x \leftarrow \mathrm{RNN}(\mathbf{h}_{m-1}^x, x_m)$$

Attention

$$p_{att}(m) \leftarrow \mathrm{softmax}([\mathbf{h}_1^x; \ldots; \mathbf{h}_M^x]^\top \mathbf{h}_n)$$

$$\mathbf{c}_n \leftarrow \mathbb{E}_{m \sim p_{att}}[\mathbf{h}_m^x] = \sum_{m=1}^{M} p_{att}(m) \mathbf{h}_m^x$$

Decoder $(dec(\mathbf{c}_n))$:

$$\mathbf{h}_n \leftarrow \mathrm{RNN}(\mathbf{h}_{n-1}, w_n)$$

Prediction:

$$\begin{aligned} p_{gen} &= \sigma(\mathbf{U}[\mathbf{h}_n, \mathbf{c}_n]) \\ p(w_{n+1}|w_{1:n}, x_{1:M}) &= p_{gen} \times \mathrm{softmax}(\mathbf{W}[\mathbf{h}_n, \mathbf{c}_n]) \\ &\quad + (1 - p_{gen}) \times \mathbb{E}_{m \sim p_{att}}[\mathbf{1}(w_{n+1} = x_m)] \end{aligned}$$

Model 3: Seq2Seq+Attention+Copy Model (Gulcehre et al, 2016) ...

Encoder $(enc(x))$:

$$\mathbf{h}_m^x \leftarrow \mathrm{RNN}(\mathbf{h}_{m-1}^x, x_m)$$

Attention

$$p_{att}(m) \leftarrow \mathrm{softmax}([\mathbf{h}_1^x; \ldots; \mathbf{h}_M^x]^\top \mathbf{h}_n)$$

$$\mathbf{c}_n \leftarrow \mathbb{E}_{m \sim p_{att}}[\mathbf{h}_m^x] = \sum_{m=1}^{M} p_{att}(m)\mathbf{h}_m^x$$

Decoder $(dec(\mathbf{c}_n))$:

$$\mathbf{h}_n \leftarrow \mathrm{RNN}(\mathbf{h}_{n-1}, w_n)$$

Prediction:

$$\begin{aligned} p_{gen} &= \sigma(\mathbf{U}[\mathbf{h}_n, \mathbf{c}_n]) \\ p(w_{n+1}|w_{1:n}, x_{1:M}) &= p_{gen} \times \mathrm{softmax}(\mathbf{W}[\mathbf{h}_n, \mathbf{c}_n]) \\ &\quad + (1 - p_{gen}) \times \mathbb{E}_{m \sim p_{att}}[\mathbf{1}(w_{n+1} = x_m)] \end{aligned}$$

Model 3: Seq2Seq+Attention+Copy Model (Gulcehre et al, 2016) ...

Encoder ($enc(x)$):

$$\mathbf{h}^x_m \leftarrow \mathrm{RNN}(\mathbf{h}^x_{m-1}, x_m)$$

Attention

$$p_{att}(m) \leftarrow \mathrm{softmax}([\mathbf{h}^x_1; \ldots; \mathbf{h}^x_M]^\top \mathbf{h}_n)$$

$$\mathbf{c}_n \leftarrow \mathbb{E}_{m \sim p_{att}}[\mathbf{h}^x_m] = \sum_{m=1}^{M} p_{att}(m)\mathbf{h}^x_m$$

Decoder ($dec(\mathbf{c}_n)$):

$$\mathbf{h}_n \leftarrow \mathrm{RNN}(\mathbf{h}_{n-1}, w_n)$$

Prediction:

$$\begin{aligned} p_{gen} &= \sigma(\mathbf{U}[\mathbf{h}_n, \mathbf{c}_n]) \\ p(w_{n+1}|w_{1:n}, x_{1:M}) &= p_{gen} \times \mathrm{softmax}(\mathbf{W}[\mathbf{h}_n, \mathbf{c}_n]) \\ &\quad + (1 - p_{gen}) \times \mathbb{E}_{m \sim p_{att}}[\mathbf{1}(w_{n+1} = x_m)] \end{aligned}$$

Model 3: Seq2Seq+Attention+Copy Model (Gulcehre et al, 2016) ...

Encoder ($enc(x)$):

$$\mathbf{h}^x_m \leftarrow \mathrm{RNN}(\mathbf{h}^x_{m-1}, x_m)$$

Attention

$$p_{att}(m) \leftarrow \mathrm{softmax}([\mathbf{h}^x_1; \ldots; \mathbf{h}^x_M]^\top \mathbf{h}_n)$$

$$\mathbf{c}_n \leftarrow \mathbb{E}_{m \sim p_{att}}[\mathbf{h}^x_m] = \sum_{m=1}^{M} p_{att}(m)\mathbf{h}^x_m$$

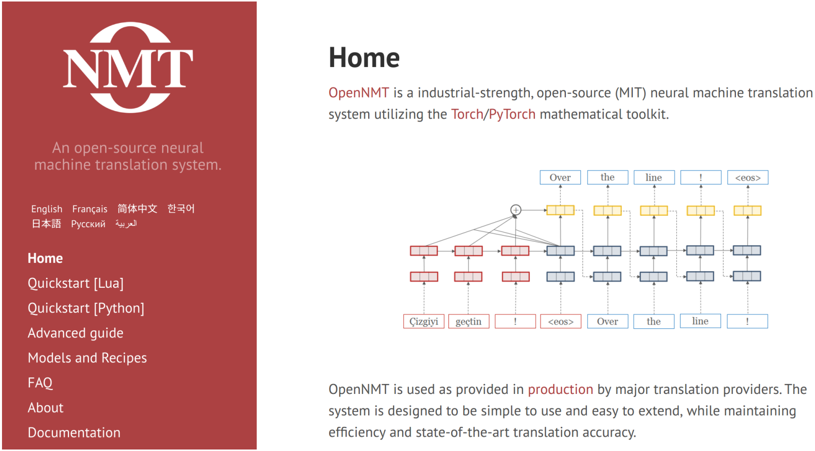
Decoder ($dec(\mathbf{c}_n)$):

$$\mathbf{h}_n \leftarrow \mathrm{RNN}(\mathbf{h}_{n-1}, w_n)$$

Prediction:

$$\begin{aligned}
p_{gen} &= \sigma(\mathbf{U}[\mathbf{h}_n, \mathbf{c}_n]) \\
p(w_{n+1}|w_{1:n}, x_{1:M}) &= p_{gen} \times \mathrm{softmax}(\mathbf{W}[\mathbf{h}_n, \mathbf{c}_n]) \\
&\quad + (1 - p_{gen}) \times \mathbb{E}_{m \sim p_{att}}[\mathbf{1}(w_{n+1} = x_m)]
\end{aligned}$$

Applications From HarvardNLP: OpenNMT



An open-source neural machine translation system.

English Français 简体中文 한국어
日本語 Русский العربية

Home

Quickstart [Lua]

Quickstart [Python]

Advanced guide

Models and Recipes

FAQ

About

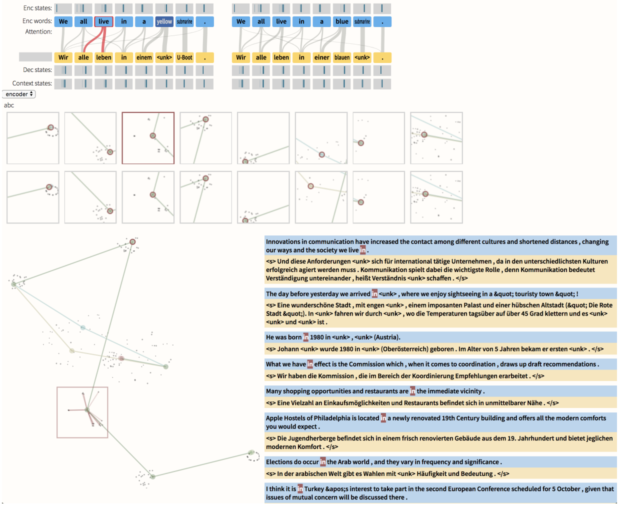
Documentation

Home

OpenNMT is a industrial-strength, open-source (MIT) neural machine translation system utilizing the Torch/PyTorch mathematical toolkit.

OpenNMT is used as provided in production by major translation providers. The system is designed to be simple to use and easy to extend, while maintaining efficiency and state-of-the-art translation accuracy.

Applications From HarvardNLP: Seq2Seq-Vis



Innovations in communication have increased the contact among different cultures and shortened distances , changing our ways and the society we live in .

<s> Und diese Anforderungen <unk> sich für international tätige Unternehmen , da in den unterschiedlichsten Kulturen erfolgreich agiert werden muss . Kommunikation spielt dabei die wichtigste Rolle , denn Kommunikation bedeutet Verständigung untereinander , heißt Verständnis <unk> schaffen . </s>

The day before yesterday we arrived in <unk> , where we enjoy sightseeing in a &quot; touristy town &quot; !

<s> Eine wunderschöne Stadt , mit engen <unk> , einem imposanten Palast und einer hübschen Altstadt (&quot; Die Rote Stadt &quot;). In <unk> fahren wir durch <unk> , wo die Temperaturen tagsüber auf über 45 Grad klettern und es <unk> <unk> und <unk> ist .

He was born in 1980 in <unk> , <unk> (Austria).

<s> Johann <unk> wurde 1980 in <unk> (Oberösterreich) geboren . Im Alter von 5 Jahren bekam er ersten <unk> . </s>

What we have in effect is the Commission which , when it comes to coordination , draws up draft recommendations .

<s> Wir haben die Kommission , die im Bereich der Koordinierung Empfehlungen erarbeitet . </s>

Many shopping opportunities and restaurants are in the immediate vicinity .

<s> Eine Vielzahl an Einkaufsmöglichkeiten und Restaurants befindet sich in unmittelbarer Nähe . </s>

Apple Hostels of Philadelphia is located in a newly renovated 19th Century building and offers all the modern comforts you would expect .

<s> Die Jugendherberge befindet sich in einem frisch renovierten Gebäude aus dem 19. Jahrhundert und bietet jeglichen modernen Komfort . </s>

Elections do occur in the Arab world , and they vary in frequency and significance .

<s> In der arabischen Welt gibt es Wahlen mit <unk> Häufigkeit und Bedeutung . </s>

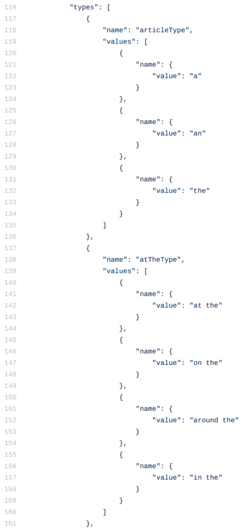
I think it is in Turkey &apos;s interest to take part in the second European Conference scheduled for 5 October , given that issues of mutual concern will be discussed there .

seq2seq-vis.io

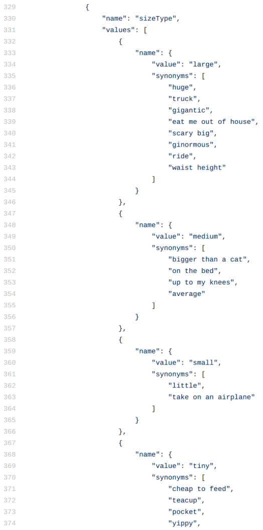
2) Applications: End-to-End Natural Language Generation

Natural language generation is the process of deliberately constructing a natural language text in order to meet specified communicative goals. - MacDonald (1987)

Common Practice Templated Generation / Intents



```
            "types": [
                {
                    "name": "articleType",
                    "values": [
                        {
                            "name": {
                                "value": "a"
                            }
                        },
                        {
                            "name": {
                                "value": "an"
                            }
                        },
                        {
                            "name": {
                                "value": "the"
                            }
                        }
                    ]
                },
                {
                    "name": "atTheType",
                    "values": [
                        {
                            "name": {
                                "value": "at the"
                            }
                        },
                        {
                            "name": {
                                "value": "on the"
                            }
                        },
                        {
                            "name": {
                                "value": "around the"
                            }
                        },
                        {
                            "name": {
                                "value": "in the"
                            }
                        }
                    ]
                },
```



```
                {
                    "name": "sizeType",
                    "values": [
                        {
                            "name": {
                                "value": "large",
                                "synonyms": [
                                    "huge",
                                    "truck",
                                    "gigantic",
                                    "eat me out of house",
                                    "scary big",
                                    "ginormous",
                                    "ride",
                                    "waist height"
                                ]
                            }
                        },
                        {
                            "name": {
                                "value": "medium",
                                "synonyms": [
                                    "bigger than a cat",
                                    "on the bed",
                                    "up to my knees",
                                    "average"
                                ]
                            }
                        },
                        {
                            "name": {
                                "value": "small",
                                "synonyms": [
                                    "little",
                                    "take on an airplane"
                                ]
                            }
                        },
                        {
                            "name": {
                                "value": "tiny",
                                "synonyms": [
                                    "cheap to feed",
                                    "teacup",
                                    "pocket",
                                    "yippy",
```

End-to-End Generation

- Neural MT has inspired interest in generation with E2E models.
- Differs significantly from much past work in NLG.
- Three areas we have worked on:
  - Summarization
  - Image-to-Markup
  - Data-to-Text
- Many others, e.g. image/video captioning, chatbots, dialogue response generation.

E2E Text Generation: Talk about Text (Summarization)

mexico city , mexico -lrb- cnn -rrb- – heavy rains and flooding have forced hundreds of thousands of people from homes in southern mexico 's state of tabasco over the past four days , with nearly as many trapped by the rising waters , state officials said thursday . officials say about 300,000 people are still trapped by the worst flooding in the region for 50 years . the grijalva river pushed over its banks through the state capital of villahermosa on thursday , forcing government workers to evacuate and leaving up to 80 percent of the city flooded , gov. andres granier 's office told cnn . about 700,000 people have seen their homes flooded , with about 300,000 of those still trapped there , granier 's office reported . one death had been blamed on the floods , which followed weeks of heavy rain in the largely swampy state . tabasco borders guatemala to the south and the gulf of mexico to the north . . . .

E2E Text Generation:
Talk about Text (Summarization)

mexico city , mexico -lrb- cnn -rrb- – heavy rains and flooding have forced hundreds of thousands of people from homes in southern mexico 's state of tabasco over the past four days , with nearly as many trapped by the rising waters , state officials said thursday . officials say about 300,000 people are still trapped by the worst flooding in the region for 50 years . the grijalva river pushed over its banks through the state capital of villahermosa on thursday , forcing government workers to evacuate and leaving up to 80 percent of the city flooded , gov. andres granier 's office told cnn . about 700,000 people have seen their homes flooded , with about 300,000 of those still trapped there , granier 's office reported . one death had been blamed on the floods , which followed weeks of heavy rain in the largely swampy state . tabasco borders guatemala to the south and the gulf of mexico to the north . . . .

tabasco and chiapas states hardest hit. authorities say 700,000 affected . . .

Abstractive E2E Sentence Summarization (Rush et al, 2015)

Input (First Sentence)

Russian Defense Minister Ivanov called Sunday for the creation of a joint front for combating global terrorism.

Output (Title)

Russia calls for joint front against terrorism.

Abstractive E2E Sentence Summarization (Rush et al, 2015)

Input (First Sentence)

Russian Defense Minister Ivanov called Sunday for the creation of a joint front for combating global terrorism.

Output (Title)

Russia calls for joint front against terrorism.

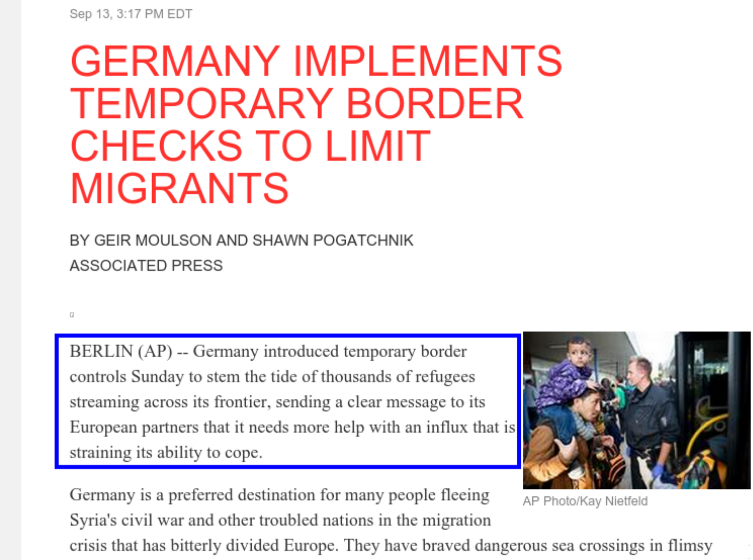
Abstractive E2E Sentence Summarization (Rush et al, 2015)

Input (First Sentence)

Russian Defense Minister Ivanov called Sunday for the creation of a joint front for combating global terrorism.

Output (Title)

Russia calls for joint front against terrorism.



Sep 13, 3:17 PM EDT

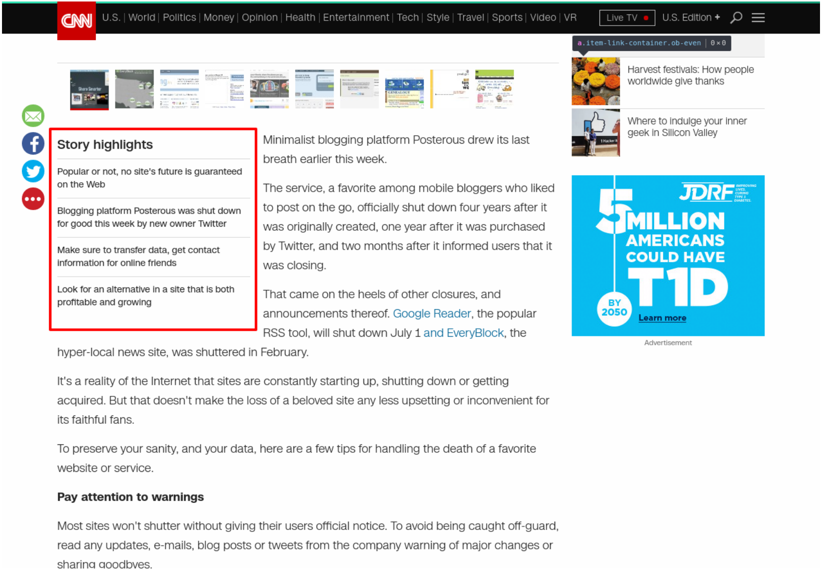
GERMANY IMPLEMENTS TEMPORARY BORDER CHECKS TO LIMIT MIGRANTS

BY GEIR MOULSON AND SHAWN POGATCHNIK
ASSOCIATED PRESS

BERLIN (AP) -- Germany introduced temporary border controls Sunday to stem the tide of thousands of refugees streaming across its frontier, sending a clear message to its European partners that it needs more help with an influx that is straining its ability to cope.

AP Photo/Kay Nietfeld

Germany is a preferred destination for many people fleeing Syria's civil war and other troubled nations in the migration crisis that has bitterly divided Europe. They have braved dangerous sea crossings in flimsy



Story highlights

- Popular or not, no site's future is guaranteed on the Web
- Blogging platform Posterous was shut down for good this week by new owner Twitter
- Make sure to transfer data, get contact information for online friends
- Look for an alternative in a site that is both profitable and growing

Minimalist blogging platform Posterous drew its last breath earlier this week.

The service, a favorite among mobile bloggers who liked to post on the go, officially shut down four years after it was originally created, one year after it was purchased by Twitter, and two months after it informed users that it was closing.

That came on the heels of other closures, and announcements thereof. Google Reader, the popular RSS tool, will shut down July 1 and EveryBlock, the hyper-local news site, was shuttered in February.

It's a reality of the Internet that sites are constantly starting up, shutting down or getting acquired. But that doesn't make the loss of a beloved site any less upsetting or inconvenient for its faithful fans.

To preserve your sanity, and your data, here are a few tips for handling the death of a favorite website or service.

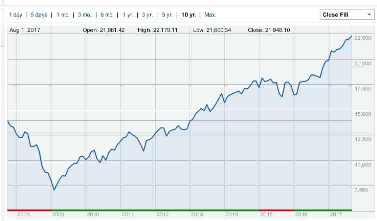
Pay attention to warnings

Most sites won't shutter without giving their users official notice. To avoid being caught off-guard, read any updates, e-mails, blog posts or tweets from the company warning of major changes or sharing goodbyes.

Harvest festivals: How people worldwide give thanks

Where to indulge your inner geek in Silicon Valley

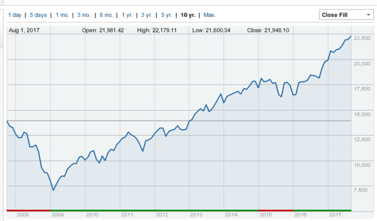
E2E Generation Challenge:
Talk about the Environment (Multimodal)



Aug 1, 2017 Open: 21,961.42 High: 22,179.11 Low: 21,600.34 Close: 21,948.10

E2E Generation Challenge:

Talk about the Environment (Multimodal)



Aug 1, 2017 Open: 21,961.42 High: 22,179.11 Low: 21,600.34 Close: 21,948.10

Dow and S&P 500 close out week at all-time highs ...

Image-to-Latex Dataset (Deng et al, 2017)

$$A_0^3(\alpha'\to 0)=2g_d\,\varepsilon_\lambda^{(1)}\varepsilon_\mu^{(2)}\varepsilon_\nu^{(3)}\left\{\eta^{\lambda\mu}\left(p_1^\nu-p_2^\nu\right)+\eta^{\lambda\nu}\left(p_3^\mu-p_1^\mu\right)+\eta^{\mu\nu}\left(p_2^\lambda-p_3^\lambda\right)\right\}.$$

```
(A_{0}^{3}(\alpha^{\prime }\rightarrow 0)=2g_{d}\,\,\varepsilon^{(1)}_{\lambda}\varepsilon^{(2)}
_{\mu }\varepsilon^{(3)}_{\nu }\left\{ \eta ^{\lambda \mu}\left( p_{1}^{\nu }-p_{2}^{\nu }\right) +
\eta ^{\lambda \nu }\left(p_{3}^{\mu }-p_{1}^{\mu }\right)+\eta ^{\mu \nu }\left( p_{2}^{\lambda}
-p_{3}^{\lambda }\right) \right\} . \label{17}
```

$$\left\{\begin{array}{rcl}\delta_\epsilon B&\sim&\epsilon F\,,\\ \delta_\epsilon F&\sim&\partial\epsilon+\epsilon B\,,\end{array}\right.$$

```
\left\{\begin{array}{rcl}\delta_{\epsilon} B & \sim & \epsilon F
\, , \\\delta_{\epsilon} F & \sim & \partial\epsilon + \epsilon B
\, , \\\end{array}\right.
```

$$\int\limits_{\mathcal{L}_{d-1}^{d}} f(H)d\nu_{d-1}(H)=c_3\int\limits_{\mathcal{L}_{2}^{A}}\int\limits_{\mathcal{L}_{d-1}^{L}} f(H)[H,A]^2d\nu_{d-1}^{L}(H)d\nu_2^A(L).$$

```
\int \limits_{{\cal L}^{d}_{d-1}}f(H)d\nu_{d-1}(H) =  c_{3} \int \limits_{{\cal L}^{A}_{2}}
\int \limits_{{\cal L}^{L}_{d-1}}f(H)[H,A]^{2}d\nu_{d-1}^{L}(H)d\nu_{2}^{A}(L).
```

$$J=\left(\begin{array}{cc}\alpha^t&\tilde{f}_2\\ f_1&\tilde{A}\end{array}\right)\left(\begin{array}{ll}0&0\\ 0&L\end{array}\right)\left(\begin{array}{cc}\alpha&\tilde{f}_1\\ f_2&A\end{array}\right)=\left(\begin{array}{ll}\tilde{f}_2Lf_2&\tilde{f}_2LA\\ \tilde{A}Lf_2&\tilde{A}LA\end{array}\right)$$

```
J=\left( \begin{array}{cc}\alpha ^{t} & \tilde{f}_{2} \\ f_{1} & \tilde{A}
\end{array}\right) \left( \begin{array}{ll}0 & 0 \\ 0 & L\end{array}\right)
\left( \begin{array}{cc}\alpha & \tilde{f}_{1} \\ f_{2} & A\end{array}\right) =
\left( \begin{array}{ll}\tilde{f}_{2}Lf_{2} & \tilde{f}_{2}LA \\ \tilde{A}Lf_{2} &
\tilde{A}LA\end{array}\right)
```

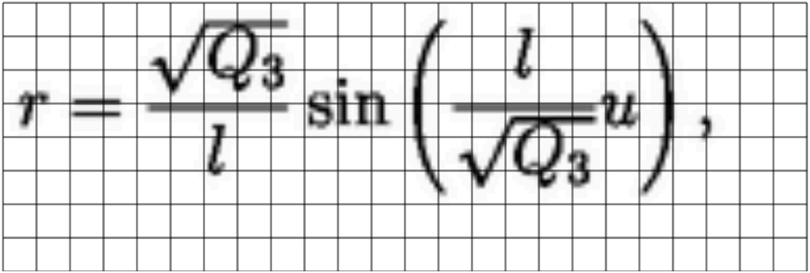
$$\lambda_{n,1}^{(2)}=\frac{\partial\overline{H}_0}{\partial q_{n,0}}\ ,lambda_{n,j_n}^{(2)}=\frac{\partial\overline{H}_0}{\partial q_{n,j_n-1}}-\mu_{n,j_n-1}\ ,\quad j_n=2,3,\cdots,m_n-1\ .$$

```
\lambda_{n,1}^{(2)}=\frac{\partial\overline{H}_0}{\partial q_{n,0}}\ ,\ \lambda_{n,j_n}^{(2)}=\frac{
\partial\overline{H}_0}{\partial q_{n,j_n-1}}-\mu_{n,j_n-1}\ ,\       \ j_n=2,3,\cdots,m_n-1\ .
```

$$(P_{ll'}-K_{ll'})\phi'(z_q)|\chi>=0$$

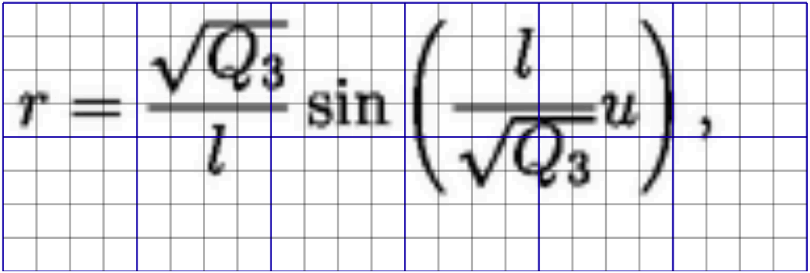
```
(P_{ll'} - K_{ll'}) \phi '(z_{q})|\chi > = 0
```

Coarse-to-Fine Attention



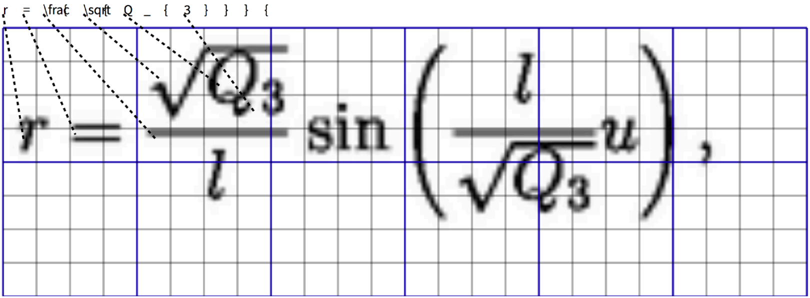
$$r = \frac{\sqrt{Q_3}}{l} \sin\left(\frac{l}{\sqrt{Q_3}} u\right),$$

Coarse-to-Fine Attention

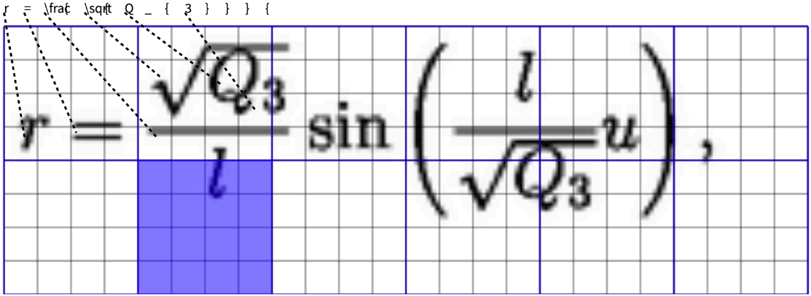


$$r = \frac{\sqrt{Q_3}}{l} \sin\left(\frac{l}{\sqrt{Q_3}} u\right),$$

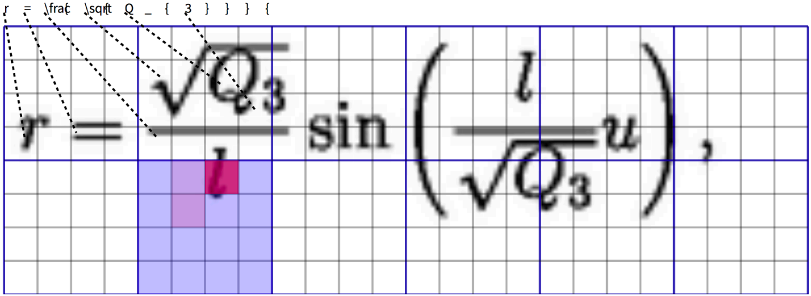
Coarse-to-Fine Attention



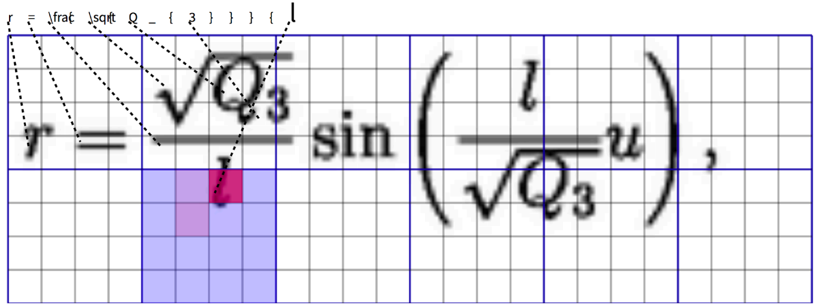
Coarse-to-Fine Attention



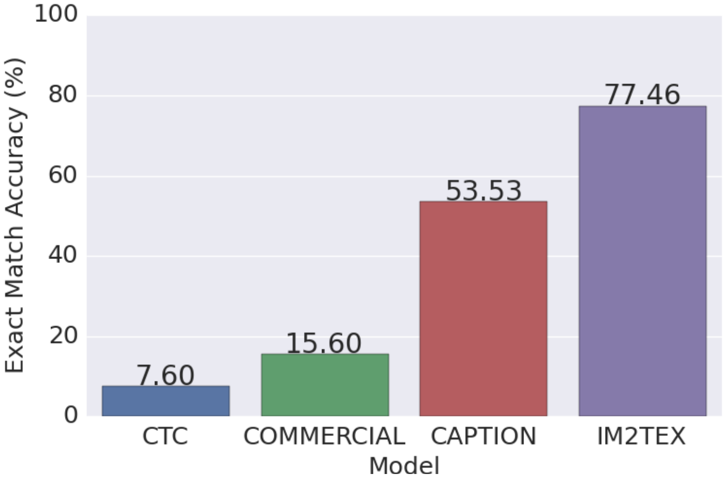
Coarse-to-Fine Attention



Coarse-to-Fine Attention

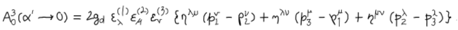


Baseline Results

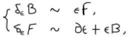


Handwritten Formulas

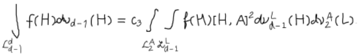
- Synthetic handwritten formulas by using handwritten characters [Kirsch, 2010] as font, used for pretraining
- Finetune and evaluate on CROHME 13 and 14 (8K training set)



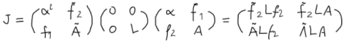
```
(A_{0}^{3}(\alpha^{\prime }\rightarrow 0)=2g_{d}\,\,\varepsilon^{(1)}_{\lambda}\varepsilon^{(2)}
_{\mu }\varepsilon^{(3)}_{\nu }\left\{ \eta ^{\lambda \mu}\left( p_{1}^{\nu }-p_{2}^{\nu }\right) +
\eta ^{\lambda \nu }\left(p_{3}^{\mu }-p_{1}^{\mu }\right)+\eta ^{\mu \nu }\left( p_{2}^{\lambda}
-p_{3}^{\lambda }\right) \right\} . \label{17}
```



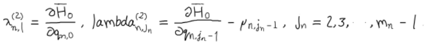
```
\left\{\begin{array}{rcl}\delta_{\epsilon} B & \sim & \epsilon F
\, , \\\delta_{\epsilon} F & \sim & \partial\epsilon + \epsilon B
\, , \\\end{array}\right.
```



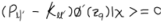
```
\int \limits_{{\cal L}^{d}_{d-1}}f(H)d\nu_{d-1}(H) =  c_{3} \int \limits_{{\cal L}^{A}_{2}}
\int \limits_{{\cal L}^{L}_{d-1}}f(H)[H,A]^{2}d\nu_{d-1}^{L}(H)d\nu_{2}^{A}(L).
```



```
J=\left( \begin{array}{cc}\alpha ^{t} & \tilde{f}_{2} \\ f_{1} & \tilde{A}
\end{array}\right) \left( \begin{array}{ll}0 & 0 \\ 0 & L\end{array}\right)
\left( \begin{array}{cc}\alpha & \tilde{f}_{1} \\ f_{2} & A\end{array}\right) =
\left( \begin{array}{ll}\tilde{f}_{2}Lf_{2} & \tilde{f}_{2}LA \\ \tilde{A}Lf_{2} &
\tilde{A}LA\end{array}\right)
```



```
\lambda_{n,1}^{(2)}=\frac{\partial\overline{H}_0}{\partial q_{n,0}}\ ,\ \\lambda_{n,j_n}^{(2)}=\frac{
\partial\overline{H}_0}{\partial q_{n,j_n-1}}-\mu_{n,j_n-1}\ ,\        \ j_n=2,3,\cdots,m_n-1\ .
```

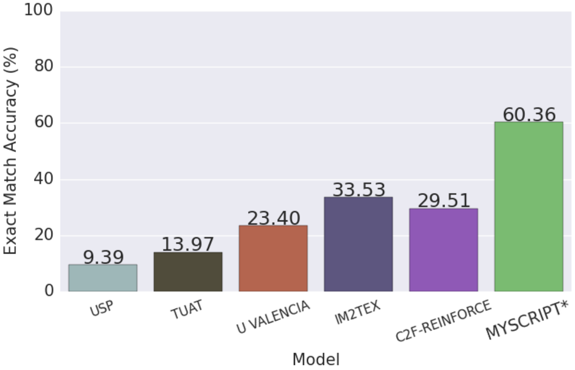


```
(P_{ll'} - K_{ll'}) \phi '(z_{q})|\chi > = 0
```

Handwritten Formulas

- Synthetic handwritten formulas by using handwritten characters [Kirsch, 2010] as font, used for pretraining
- Finetune and evaluate on CROHME 13 and 14 (8K training set)

CROHME 13 (*uses private in-domain handwritten training data)



E2E Text Generation:
Talk about Information (Generation)

	W	L	PTS	...
TEAM				
Heat	11	12	103	...
Hawks	7	15	95	...

E2E Text Generation:
Talk about Information (Generation)

TEAM	W	L	PTS	...
Heat	11	12	103	...
Hawks	7	15	95	...

The Atlanta Hawks defeated the Miami Heat, 103 - 95, at Philips Arena on Wednesday. Atlanta ...

MR	name[The Golden Palace], eatType[coffee shop], food[Fast food], priceRange[cheap], customer rating[5 out of 5], area[riverside]
Reference	A coffee shop located on the riverside called The Golden Palace, has a 5 out of 5 customer rating. Its price range are fairly cheap for its excellent Fast food.

Harvard System (Gehrmann et al, 2018)

Details:

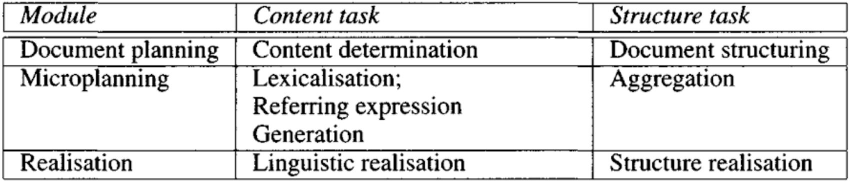
- Copy mechanism
- Coverage and length beam search
- Built in OpenNMT-py
- Transformer and LSTM model
- Diverse Ensembling (Lee et al, 2016)

Results: 2018 E2E Challenge (Novikova et al, 2017)

- Total of 60 submissions by 16 institutions with about 1/3 of these submissions coming from industry.
- Harvard systems finished **first** in ROUGE, CIDEr, and **second** in METEOR
- Best system: 2pts ROUGE improvement of baseline model.
- Human evaluation results more mixed, all systems similar.

Generation and Summarization beyond Translation

Building Natural Language Generation Systems (1999)



Module	*Content task*	*Structure task*
Document planning	Content determination	Document structuring
Microplanning	Lexicalisation; Referring expression Generation	Aggregation
Realisation	Linguistic realisation	Structure realisation

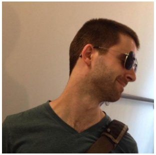
Figure 3.1 Modules and tasks.

- How can we maintain coherence through long-form text outputs?
- How can we discover complex relationships in source input?
- When is textual improvisation allowed versus literal mappings?

3) Uphill Battles in E2E Generation

1. **Challenges in Data-to-Document Generation**

   (with Sam Wiseman)



2. Controllable Generation

Case Study: Data-to-Document Generation

	WIN	LOSS	PTS	FG_PCT	RB	AS ...
TEAM						
Heat	11	12	103	49	47	27
Hawks	7	15	95	43	33	20

	AS	RB	PT	FG	FGA	CITY ...
PLAYER						
Tyler Johnson	5	2	27	8		
Dwight Howard	11	17	23	9		
Paul Millsap	2	9	21	8		
Goran Dragic	4	2	21	8		
Wayne Ellington	2	3	19	7		
Dennis Schroder	7	4	17	8		
Rodney McGruder	5	5	11	3		
...						

The Atlanta Hawks defeated the Miami Heat, 103 - 95, at Philips Arena on Wednesday. Atlanta was in desperate need of a win and they were able to take care of a shorthanded Miami team here. Defense was key for the Hawks, as they held the Heat to 42 percent shooting and forced them to commit 16 turnovers. Atlanta also dominated in the paint, winning the rebounding battle, 47 - 34, and outscoring them in the paint 58 - 26. The Hawks shot 49 percent from the field and assisted on 27 of their 43 made baskets. This was a near wire-to-wire win for the Hawks, as Miami held just one lead in the first five minutes. Miami (7 - 15) are as beat-up as anyone right now and it's taking a toll on the heavily used starters. Hassan Whiteside really struggled in this game, as he amassed eight points, 12 rebounds and one blocks on 4 - of - 12 shooting ...

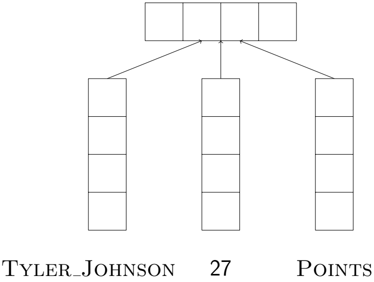
	RoboCup	WeatherGov	RotoWire	SBNation
Vocab	409	394	11,331	68,574
Tokens	11K	0.9M	1.6M	8.8M
Examples	1,919	22,146	4,853	10,903
Avg Len	5.7	28.7	337.1	805.4
Field Types	4	10	39	39
Avg Records	2.2	191	628	628

Player Types								
POSN	MIN	PTS	FGM	FGA	FG-PCT	FG3M	FG3A	FG3-PCT
FTM	FTA	FT-PCT	OREB	DREB	REB	AST	TOV	STL
BLK	PF	NAME1	NAME2					
Team Types								
PTS-QTR1	PTS-QTR2	PTS-QTR3	PTS-QTR4	PTS	FG-PCT	FG3-PCT	FT-PCT	REB
AST	TOV	WINS	LOSSES	CITY	NAME			

Content Encoding with Cell Embeddings

- $\{r_1, \ldots, r_S\}$
- $r.t = \textsc{points}$, and such that entity $r.e =$ (Tyler Johnson) and value $r.m = 27$ (Liang et al, 2009)
- $\boldsymbol{s}_j = E(\boldsymbol{r}_j)$ for $j \in \{1, \ldots S\}$

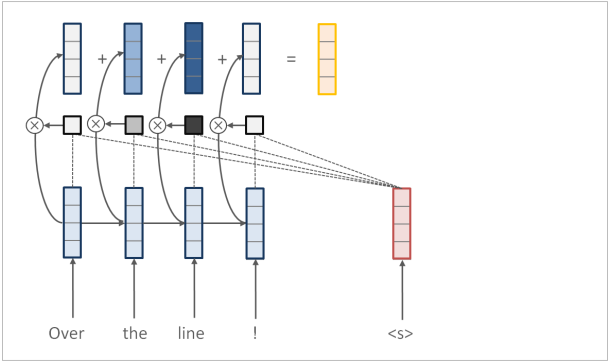
	AS	RB	PT	FG	FGA	CITY ...
PLAYER						
Tyler Johnson	5	2	27	8	16	Miami
Dwight Howard	11	17	23	9	11	Atlanta
Paul Millsap	2	9	21	8	12	Atlanta
Goran Dragic	4	2	21	8	17	Miami
Wayne Ellington	2	3	19	7	15	Miami
Dennis Schroder	7	4	17	8	15	Atlanta
Rodney McGruder	5	5	11	3	8	Miami
⋮						



Additional Extensions: Machine Translation

Architectural changes focusing on source selection.

- Copy Attention, Pointer Networks, and Reconstruction Models



Templated Baseline

The ⟨team1⟩ (⟨wins1⟩-⟨losses1⟩) defeated the ⟨team2⟩ (⟨wins2⟩-⟨losses2⟩) ⟨pts1⟩-⟨pts2⟩.

(6×)

⟨player⟩ scored ⟨pts⟩ points (⟨fgm⟩- ⟨fga⟩ FG, ⟨tpm⟩-⟨tpa⟩ 3PT, ⟨ftm⟩- ⟨fta⟩ FT) to go with ⟨reb⟩ rebounds.

The ⟨team1⟩ next game will be at home against the Dallas Mavericks, while the ⟨team2⟩ will travel to play the Bulls.

		Development	
		PPL	BLEU
Beam	Model		
	Template	N/A	6.87
1	Joint Copy	7.46	10.41
	Joint Copy + Rec	7.25	10.00
	Joint Copy + Rec + TVD	7.22	12.78
	Conditional Copy	7.44	13.31
5	Joint Copy	7.46	10.23
	Joint Copy + Rec	7.25	10.85
	Joint Copy + Rec + TVD	7.22	12.04
	Conditional Copy	7.44	**14.46**

The Utah Jazz (38 - 26) defeated the Houston Rockets (38 - 26) 117 - 91 on Wednesday at Energy Solutions Arena in Salt Lake City . The Jazz got out to a quick start in this one , out - scoring the Rockets 31 - 15 in the first quarter alone . Along with the quick start , the Rockets were the superior shooters in this game , going 54 percent from the field and 43 percent from the three - point line , while the Jazz went 38 percent from the floor and a meager 19 percent from deep . The Rockets were able to out - rebound the Rockets 49 - 49 , giving them just enough of an advantage to secure the victory in front of their home crowd . The Jazz were led by the duo of Derrick Favors and James Harden . Favors went 2 - for - 6 from the field and 0 - for - 1 from the three - point line to score a game - high of 15 points , while also adding four rebounds and four assists ....

The Utah Jazz (38 - 26) defeated the Houston Rockets (38 - 26) 117 - 91 on Wednesday at Energy Solutions Arena in Salt Lake City . The Jazz got out to a quick start in this one , out - scoring the Rockets 31 - 15 in the first quarter alone . Along with the quick start , the Rockets were the superior shooters in this game , going 54 percent from the field and 43 percent from the three - point line , while the Jazz went 38 percent from the floor and a meager 19 percent from deep . The Rockets were able to out - rebound the Rockets 49 - 49 , giving them just enough of an advantage to secure the victory in front of their home crowd . The Jazz were led by the duo of Derrick Favors and James Harden . Favors went 2 - for - 6 from the field and 0 - for - 1 from the three - point line to score a game - high of 15 points , while also adding four rebounds and four assists ....

Generations are fluent and accurate...

- Along with the quick start , the Rockets were the superior shooters in this game , going 54 percent from the field and 43 percent from the three - point line

... but also have issues in the content and discourse

- The Rockets were able to out - rebound the Rockets (incorrect discourse)
- The Jazz were led by the duo of Derrick Favors and James Harden (wrong player selected)
- to score a game - high (non-factual) of 15 points

Question 2: How can we quantify the issues in generation?

Criteria:

1. Relation Generation: Referring expressions should be easy trace.
2. Content Selection: Relevant content should be generated.
3. Content Ordering: Discourse structure should be consistent.

Observation: NLU is currently a lot easier than NLG.

Extractive Evaluation

Use information extraction system for generations (details in paper)

Criteria:

1. Relation Generation: Referring expressions should be easy trace.
    - Precision and count of identified data points.
2. Content Selection: Relevant content should be generated.
    - F-score on generated data points.
3. Content Ordering: Discourse structure should be consistent.
    - Damerau-Levenshtein distance between ordered elements.

Higher-Level Properties

		RG		CS		CO
Beam	Model	P%	#	P%	R%	DLD%
	Template	**99.35**	49.7	**45.17**	24.85	**12.2**
B=1	Joint Copy	47.55	7.53	20.53	22.49	8.28
	Conditional Copy	68.94	9.09	25.15	22.94	9.00
B=5	Joint Copy	47.00	10.67	16.52	26.08	7.28
	Conditional Copy	71.07	12.61	21.90	27.27	8.70

Next Steps: Robust Generation

- Current systems target fluency, very hard to check for accuracy.
- Unlike NMT, NLG work needs high-precision outputs, very hard with black box models.
- Research focus?: learning *template-based* generation systems.

3) Uphill Battles in Generation

1. Challenges in Data-to-Document Generation
2. **Controllable Generation with Neural Templates**

(Work in Progress with Sam Wiseman)

Project Aim

Can we build a neural generation system that is:

1. Interpretable in its content selection.
2. Easily controllable in terms of style and form.

Tension between end-to-end neural approach and desired modularity.

Project Aim

Can we build a neural generation system that is:

1. Interpretable in its content selection.
2. Easily controllable in terms of style and form.

Tension between end-to-end neural approach and desired modularity.

Standard Copy Generation Approach

Step 1: Encode the Source

Fitzbillies,ty[coffee shop],pr[$<$ £20],food[Chinese],cust[3/5],area[city centre]

Step 2: Generate with Copy Decoder

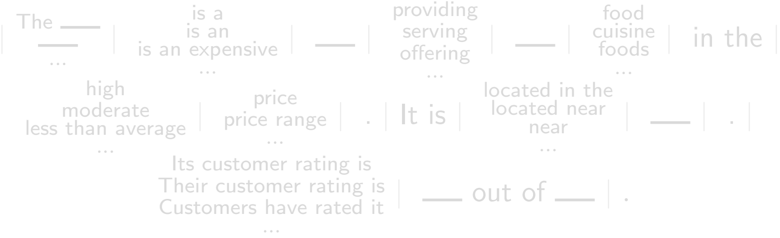
Fitzbillies is a coffee shop providing Chinese food in the moderate price range . It is located in the city centre . Its customer rating is 3 out of 5.

(Neural) Template Generation Approach

Step 1: Encode the Source

Fitzbillies,ty[coffee shop],pr[$<$ £20],food[Chinese],cust[3/5],area[city centre]

Step 2: Select a Template



| The ___
 ___
 … | is a
 is an
 is an expensive
 … | ___ | providing
 serving
 offering
 … | ___ | food
 cuisine
 foods
 … | in the |

high <br> moderate
 less than average
 … | price <br> price range
 … | . | It is | located in the
 located near
 near
 … | ___ | . |

Its customer rating is
 Their customer rating is
 Customers have rated it
 … | ___ out of ___ | .

Step 3: Fill-in Each Segment

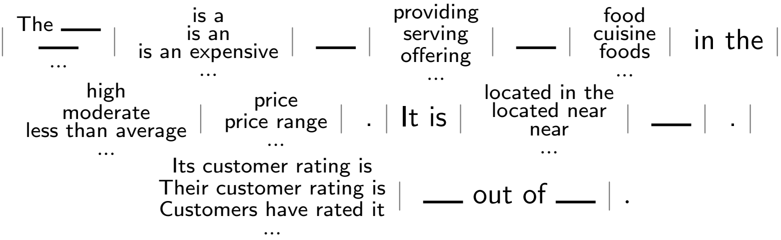
|| Fitzbillies || is a || coffee shop || providing || Chinese || food || in the || moderate || price range || . || It is || located in the || city centre || . ||

(Neural) Template Generation Approach

Step 1: Encode the Source

Fitzbillies,ty[coffee shop],pr[$<$ £20],food[Chinese],cust[3/5],area[city centre]

Step 2: Select a Template



| The ___ / ___ / … | is a / is an / is an expensive / … | ___ | providing / serving / offering / … | ___ | food / cuisine / foods / … | in the |

high / moderate / less than average / … | price / price range / … | . | It is | located in the / located near / near / … | ___ | . |

Its customer rating is / Their customer rating is / Customers have rated it / … | ___ out of ___ | .

Step 3: Fill-in Each Segment

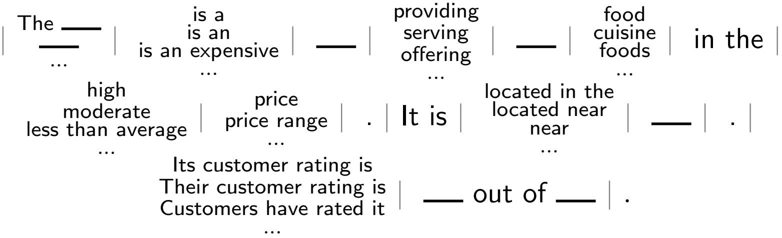
|| Fitzbillies || is a || coffee shop || providing || Chinese || food || in the || moderate || price range || . || It is || located in the || city centre || . ||

(Neural) Template Generation Approach

Step 1: Encode the Source

Fitzbillies,ty[coffee shop],pr[< £20],food[Chinese],cust[3/5],area[city centre]

Step 2: Select a Template



The ___	is a	___	providing	___	food	in the
___	is an		serving		cuisine	
…	is an expensive		offering		foods	
	…		…		…	

high | price | . | It is | located in the | ___ | . |
moderate | price range | | | located near | | |
less than average | … | | | near | | |
… | | | | … | | |

Its customer rating is | ___ out of ___ | .
Their customer rating is
Customers have rated it
…

Step 3: Fill-in Each Segment

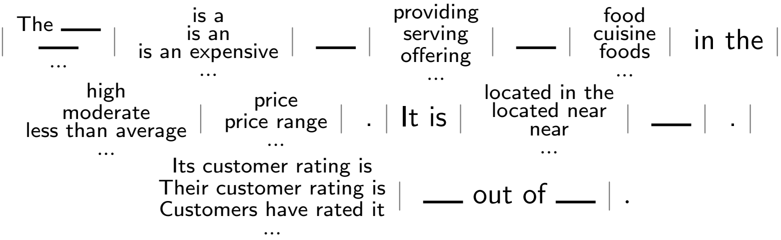
|| Fitzbillies || is a || coffee shop || providing || Chinese || food || in the || moderate || price range || . || It is || located in the || city centre || . ||

(Neural) Template Generation Approach

Step 1: Encode the Source

Fitzbillies,ty[coffee shop],pr[$<$ £20],food[Chinese],cust[3/5],area[city centre]

Step 2: Select a Template



| The ___
 ___
 … | is a
 is an
 is an expensive
 … | ___ | providing
 serving
 offering
 … | ___ | food
 cuisine
 foods
 … | in the |

| high
 moderate
 less than average
 … | price
 price range
 … | . | It is | located in the
 located near
 near
 … | ___ | . |

Its customer rating is
 Their customer rating is
 Customers have rated it
 … | ___ out of ___ | .

Step 3: Fill-in Each Segment

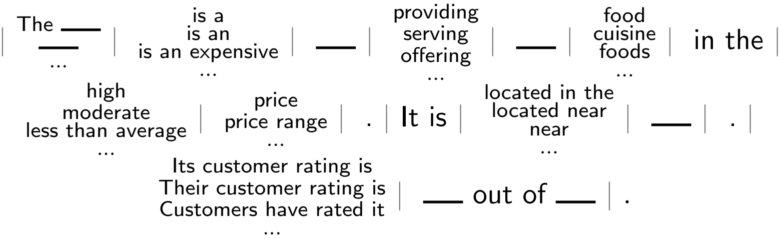
|| Fitzbillies || is a || coffee shop || providing || Chinese || food || in the || moderate || price range || . || It is || located in the || city centre || . ||

(Neural) Template Generation Approach

Step 1: Encode the Source

Fitzbillies,ty[coffee shop],pr[$<$ £20],food[Chinese],cust[3/5],area[city centre]

Step 2: Select a Template



| The ___
 ___
 … | is a
 is an
 is an expensive
 … | ___ | providing
 serving
 offering
 … | ___ | food
 cuisine
 foods
 … | in the |

high <br> moderate
 less than average
 … | price <br> price range
 … | . | It is | located in the
 located near
 near
 … | ___ | . |

Its customer rating is
 Their customer rating is
 Customers have rated it
 … | ___ out of ___ | .

Step 3: Fill-in Each Segment

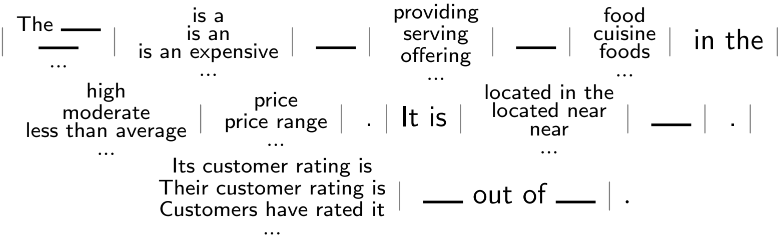
|| Fitzbillies || is a || coffee shop || providing || Chinese || food || in the || moderate || price range || . || It is || located in the || city centre || . ||

(Neural) Template Generation Approach

Step 1: Encode the Source

Fitzbillies,ty[coffee shop],pr[$<$ £20],food[Chinese],cust[3/5],area[city centre]

Step 2: Select a Template



| The ___
 ___
 … | is a
 is an
 is an expensive
 … | ___ | providing
 serving
 offering
 … | ___ | food
 cuisine
 foods
 … | in the |

high <br> moderate
 less than average
 … | price <br> price range
 … | . | It is | located in the
 located near
 near
 … | ___ | . |

Its customer rating is
 Their customer rating is
 Customers have rated it
 … | ___ out of ___ | .

Step 3: Fill-in Each Segment

‖ Fitzbillies ‖ is a ‖ coffee shop ‖ providing ‖ Chinese ‖ food ‖ in the ‖
moderate ‖ price range ‖ . ‖ It is ‖ located in the ‖ city centre ‖ . ‖

Criteria

1. Interpretable in its content selection.

   Decisions are localized to a segment of the template.

2. Easily controllable in terms of style and form.

   Alternative realizations through different templates.

However: templates feel much less “end-to-end”.

How can we learn them from data?

Criteria

1. Interpretable in its content selection.

   Decisions are localized to a segment of the template.

2. Easily controllable in terms of style and form.

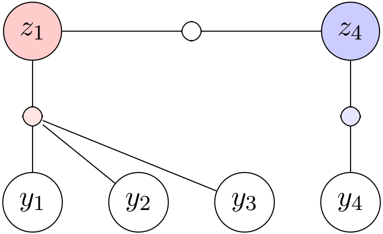
   Alternative realizations through different templates.

However: templates feel much less “end-to-end”.

How can we learn them from data?

Technical Methodology: Hidden Semi-Markov Model

- HMM: discrete latent states with single emissions (e.g. words).
- HSMM: discrete states produce multiple emissions (e.g. phrases).
- Parameterized with *transition*, *emission*, and *length* distributions.

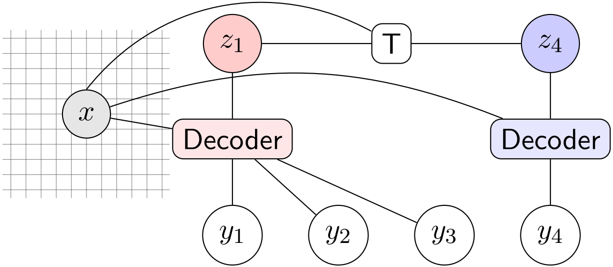


Technical Methodology: Neural Hidden Semi-Markov Model

- Employ HSMM as a conditional latent variable language model, $p(y_1, \ldots, y_T, z \mid x)$.

- Transition Distribution: NN between states.

- Emission Distribution: Seq2Seq+Copy-Attention, one per state k.



Technical Methodology: Learning Templates

- Fit model by maximizing log-marginal likelihood on training data.

$$\max_{\theta} \sum_{j} \ln \sum_{z} p(y^{(j)}, z \mid x^{(j)}; \theta)$$

Details: Pre-score segmentations, HSMM forward algorithm for sum, backprop with autograd, all inference is exact.

- Compute argmax segmentations to find common *templates*.

$$z^{(j)} = \arg\max_{z} p(y^{(j)}, z \mid x^{(j)}; \theta)$$

[The Wrestlers]$_{185}$ *[is a]*$_{29}$ *[coffee shop]*$_{164}$ *[that serves]*$_{188}$ *[English]*$_{139}$ *[food]*$_{18}$ *[in the]*$_{32}$ *[moderate]*$_{125}$ *[price range]*$_{180}$ *[.]*$_{90}$

Technical Methodology: Learning Templates

- Fit model by maximizing log-marginal likelihood on training data.

$$\max_{\theta} \sum_{j} \ln \sum_{z} p(y^{(j)}, z \mid x^{(j)}; \theta)$$

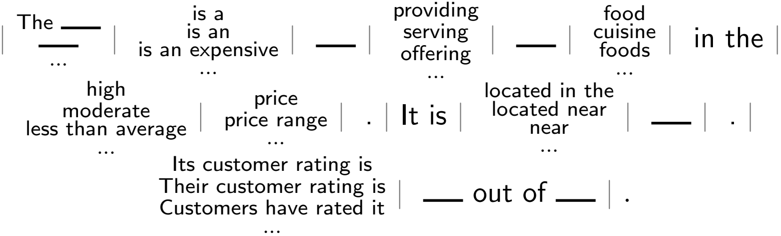
Details: Pre-score segmentations, HSMM forward algorithm for sum, backprop with autograd, all inference is exact.

- Compute argmax segmentations to find common *templates*.

$$z^{(j)} = \arg\max_{z} p(y^{(j)}, z \mid x^{(j)}; \theta)$$

[The Wrestlers]$_{185}$ *[is a]*$_{29}$ *[coffee shop]*$_{164}$ *[that serves]*$_{188}$ *[English]*$_{139}$ *[food]*$_{18}$ *[in the]*$_{32}$ *[moderate]*$_{125}$ *[price range]*$_{180}$ *[.]*$_{90}$

Neural Template

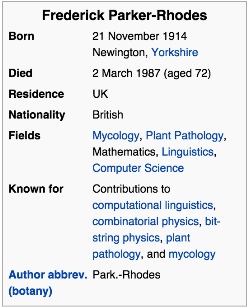


The ___ ___ ...	is a is an is an expensive ...	___	providing serving offering ...	___	food cuisine foods ...	in the
high moderate less than average ...	price price range ...	.	It is	located in the located near near ...	___	.
Its customer rating is Their customer rating is Customers have rated it ...	___ out of ___	.				

Experimental Setup

- Two datasets, E2E challenge and WikiBio
- Training with 35 and 65 state models, each 1x300 LSTMs.
- Extract 100 most common templates for each.
- Vocabulary limited to non-copy-able words.
- Generation with beam search with a pre-selected template.

WikiBio (500k)



Frederick Parker-Rhodes

Born	21 November 1914 Newington, Yorkshire
Died	2 March 1987 (aged 72)
Residence	UK
Nationality	British
Fields	Mycology, Plant Pathology, Mathematics, Linguistics, Computer Science
Known for	Contributions to computational linguistics, combinatorial physics, bit-string physics, plant pathology, and mycology
Author abbrev. (botany)	Park.-Rhodes

Frederick Parker-Rhodes (21 March 1914 - 21 November 1987) was an English linguist, plant pathologist, computer scientist, mathematician, mystic, and mycologist.

E2E Challenge

	BLEU	NIST
Val		
Substitution	43.71	6.72
Neural Template	66.06	7.93
Full Neural Model	69.25	8.48
Test		
Substitution	43.78	6.88
Neural Template	56.72	7.63
Full Neural Model	65.93	8.59

WikiBio

	BLEU	NIST	ROUGE-4
Conditional KN-LM	19.8	5.19	10.7
NNLM (field)	33.4	7.52	23.9
NNLM (field & word)	34.7	7.98	25.8
Neural Template	33.8	7.51	28.2

- Custom KN and NNLM Baselines from LeBret et al (2016)

Interpretability

kenny warren

name: kenny warren, **birth date:** 1 april 1946,

birth name: kenneth warren deutscher, **birth place:** brooklyn, new york,

occupation: ventriloquist, comedian, author,

notable work: book - the revival of ventriloquism in america

1. kenny warren deutscher (april 1, 1946) is an american ventriloquist.
2. kenny warren deutscher (april 1, 1946 , brooklyn,) is an american ventriloquist.
3. kenny warren deutscher (april 1, 1946) is an american ventriloquist, best known for his the revival of ventriloquism.
4. “kenny” warren is an american ventriloquist.
5. kenneth warren “kenny” warren (born april 1, 1946) is an american ventriloquist, and author.

Controllability

The Golden Palace

name[The Golden Palace], type[coffee shop], food[Chinese], priceRange[cheap] custRating[5 out of 5], area[city centre],

1. The Golden Palace is a coffee shop located in the city centre.
2. In the city centre is a cheap Chinese coffee shop called The Golden Palace.
3. The Golden Palace that serves Chinese food in the cheap price range. It is located in the city centre. Its customer rating is 5 out of 5.
4. The Golden Palace is a Chinese coffee shop.
5. The Golden Palace is a Chinese coffee shop with a customer rating of 5 out of 5.

Conclusion: Challenges in Text Generation

- End-to-end models are a remarkable step forward.
- Still significant challenges when we go beyond the sentence.
- Interpretability and controllability are non-trivial issues.
- Scaling probabilistic inference and structured prediction is getting easier, and presents an interesting step forward.