Overview of Machine Translation
Kenneth Heafield, University of Edinburgh
Seen at Golden Acre Shopping Centre
Airport Priority Queue
Edinburgh Translation Group 2/2

Marek Strelec
Research Assistant

Alham Aji
PhD student

Anna Currey
PhD student

Nikolay Bogoychev
Shared PhD student

Maxi Behnke
MSc+PhD student
Projects

English–isiXhosa for doctors with UCT

Potentially relevant other projects:
- Health information with UK National Health Service
- Harvesting translations from the web
- Low-resource languages
- Speech translation
- Speed
Projects

English–isiXhosa for doctors with UCT

Potentially relevant other projects:
- Health information with UK National Health Service
- Harvesting translations from the web
- Low-resource languages
- Speech translation
- Speed

Broader projects active in the group:
- Monitoring news in other languages
- Massively open online course translation
- Grammatical error correction
Open-Source Software

Marian
https://marian-nmt.github.io/
Neural networks in C++

Nematus
https://github.com/EdinburghNLP/nematus
Neural networks in Python and Theano

Moses
http://www.statmt.org/moses/
Phrase-based translation
Funding

EU Currently 6 projects.

UK Currently this project.

Industry Amazon, Booking.com, eBay, Facebook, Google, Intel, Microsoft, Mozilla, Samsung, WIPO
Funding

EU  Currently 6 projects.
    South Africa in Horizon 2020.

UK  Currently this project.
    South Africa eligible for “global challenges research fund” (10% of UK science)

Industry  Amazon, Booking.com, eBay, Facebook, Google, Intel, Microsoft, Mozilla, Samsung, WIPO
          Usually 1–1 arrangements.
1 Evaluating quality
2 Phrase-based models
3 Neural models
1 Evaluating quality
2 Phrase-based models
3 Neural models
Direct Assessment

Conference on Machine Translation: same annotators score competing systems
Winners in 2017 Conference on MT

Constrained news task, including ties:

<table>
<thead>
<tr>
<th>From English</th>
<th>To English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turkish</td>
<td>Edinburgh</td>
</tr>
<tr>
<td>Czech</td>
<td>Edinburgh</td>
</tr>
<tr>
<td>Chinese</td>
<td>Edinburgh</td>
</tr>
<tr>
<td>German</td>
<td>Munich</td>
</tr>
<tr>
<td>Russian</td>
<td>Edinburgh</td>
</tr>
<tr>
<td>Latvian</td>
<td>Tilde</td>
</tr>
<tr>
<td>Finnish</td>
<td>Helsinki</td>
</tr>
</tbody>
</table>

(Edinburgh did not participate in Finnish)
BLEU score

Human evaluation is expensive
⇒ Want automatic evaluation

BLEU: how much does the output match a human translator?
- 1-word matches
- 2-word matches
- 3-word matches
- 4-word matches
- Length

Typically expressed as a percentage: 0–100%
1% BLEU increase considered publishable
1 Evaluating quality
2 Phrase-based models
3 Neural models
Chambre

Bedroom
présidente de la Chambre des représentants

chairwoman of the Bedroom of Representatives
présidente de la Chambre des représentants

chairwoman of the House of Representatives
Phrase-based models

Extract translated phrases
Score phrases with translation probabilities
String together to form translations
Phrase-based models

Extract translated phrases
Score phrases with translation probabilities
String together to form translations

Neural is largely replacing phrase-based translation, except for low-resource
1 Evaluating quality
2 Phrase-based models
3 Neural models
Assign each word a unique row.
Recurrent Neural Network

\[
\begin{bmatrix}
1 \\
0 \\
0 \\
0
\end{bmatrix}
\]

\[
\begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

\[
p(<s>) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\]

\[
p(in) = \begin{bmatrix} 0.4 \\ 0.2 \\ 0.4 \end{bmatrix}
\]

\[
p(the) = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}
\]

\[
p(raw) = \begin{bmatrix} 0.4 \\ 0.4 \\ 0.4 \end{bmatrix}
\]

\[
\begin{bmatrix}
2.1 \\
-4 \\
0.3
\end{bmatrix}
\]
Recurrent Neural Network

\[
\begin{align*}
\text{Word} &:
\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\
\text{State} &:
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\end{align*}
\]

\[
\begin{align*}
p(<s>) &= \begin{bmatrix} 0 \end{bmatrix} \\
p(in) &= \begin{bmatrix} 0.4 \end{bmatrix} \\
p(the) &= \begin{bmatrix} 0.2 \end{bmatrix} \\
p(raw) &= \begin{bmatrix} 0.4 \end{bmatrix}
\end{align*}
\]

\[
\begin{align*}
\text{raw} &:
\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\
\end{align*}
\]

\[
\begin{bmatrix} 2.1 \\ -4 \\ 0.3 \end{bmatrix}
\]
Recurrent Neural Networks for Translation

Read source one word at a time (throw out predictions)
Then predict target words left to right.
Read source one word at a time (throw out predictions)
Then predict target words left to right.

Problem: sentences don’t fit in a 1024-dimensional vector
→ Look up source word ("attention") while producing target word
Transition to Neural Models

What types of models won pairs in the Conference on Machine Translation?

<table>
<thead>
<tr>
<th>Year</th>
<th>Neural</th>
<th>Phrase</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>1</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>2016</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>2017</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Neural is better at agreement

Source Byl to bratr, který bral věci takové, jaké jsou.
Reference He was the brother that went with the flow.
Phrase It was a brother who took things as they are.
Neural He was a brother who took things the way they are.

Decisions based on whole sentence, not just local context.
Source Seit Jahrzehnten fördert Langer den Nachwuchs.
Reference Langer has been encouraging up-and-coming talent for years.
Phrase For decades, Langer promotes the offspring.
Neural For decades, Langer has been promoting the young.

Larger context \(\Rightarrow\) generally better at fluency.
Rare words are hard

Source Jennifer Aniston: Ich werde immer in Schubladen gesteckt
Reference Jennifer Aniston: I’m always pigeonholed
Phrase Jennifer Aniston: I am always plugged in drawers
Neural Jennifer Aniston: I’ll always be put in drawers

⇒ Translations covering terminology are important.
Style

Source Erkek kardeşim, her duruma uyum sağlardi.
Reference He was the brother that went with the flow.
Phrase My brother, sağlardi fit in every situation.
Neural My brother was harmonised in every situation.

Neural fits style more closely:
Great if you add in-domain data
Awkward if you don’t. Subtitles swear a lot.