Visualizing Neural Machine Translation Attention and Confidence

Matīss Rikters[a], Mark Fishel[b], Ondřej Bojar[c]
[a]Faculty of Computing, University of Latvia
[b]Institute of Computer Science, University of Tartu
[c]Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics
E-mail: matiss@lielakeda.lv, fishel@ut.ee, bojar@ufal.mff.cuni.cz

Abstract

In this article, we describe a tool for visualizing the output and attention weights of neural machine translation systems and for estimating confidence about the output based on the attention.

Our aim is to help researchers and developers better understand the behaviour of their NMT systems without the need for any reference translations. Our tool includes command line and web-based interfaces that allow to systematically evaluate translation outputs from various engines and experiments. We also present a web demo of our tool with examples of good and bad translations: http://ej.uz/nmt-attention.

Confidence Scores

$$CDP = \frac{1}{J}\sum_j \log\left(1 + \left(\sum_i \propto_{ji}\right)^2\right)$$

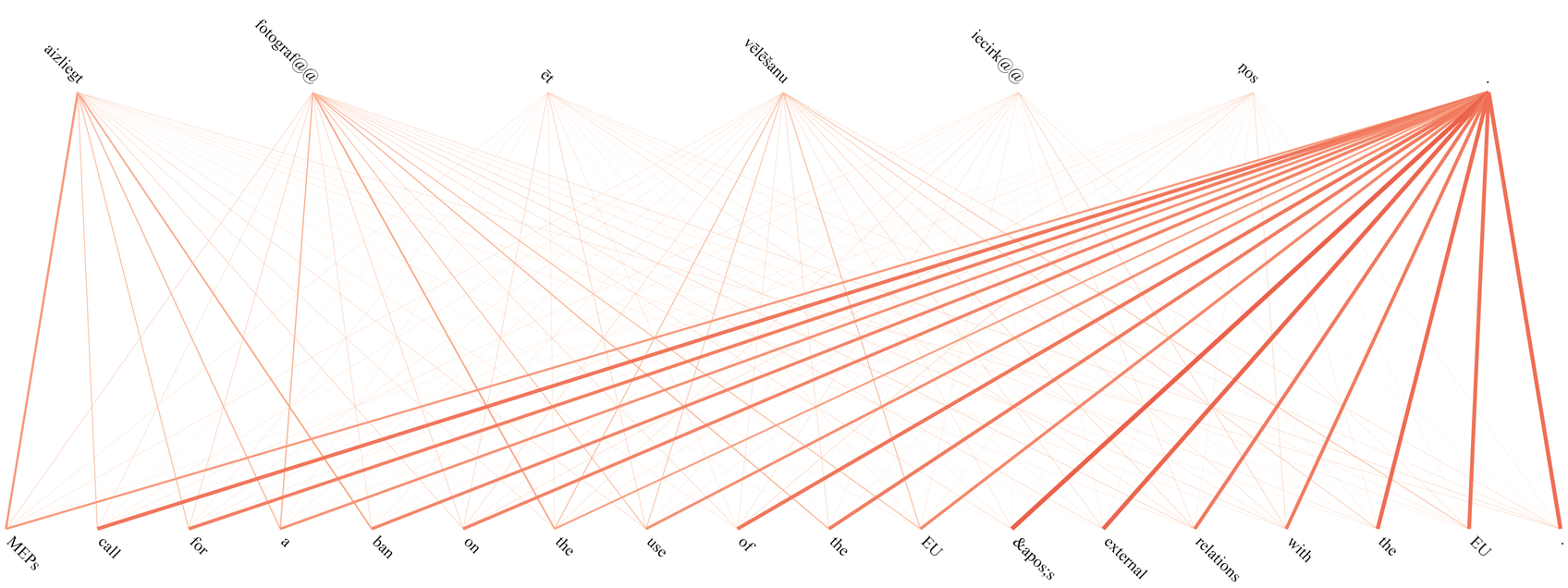
$$AP_{out} = -\frac{1}{I}\sum_i\sum_j \propto_{ji} \cdot log \propto_{ji}$$

$$confidence = CDP + AP_{out} + AP_{in}$$

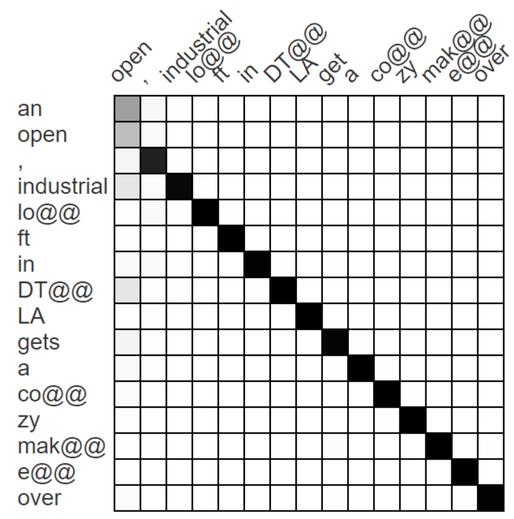
$$AP_{in} = -\frac{1}{I}\sum_j\sum_i \propto_{ij} \cdot log \propto_{ij}$$

$$\text{percentage} = e^{-C(X^2)}$$

Lacking Confidence



Excessive Confidence



Features

- **Works with attention alignment data from**
  - Nematus
  - Neural Monkey
  - AmuNMT
- **Visualise translations in**
  - Linux Terminal or Windows PowerShell
  - Web browser
    - Line form or matrix form
    - Save as PNG
    - Sort and navigate dataset by confidence scores

GitHub Link

http://ej.uz/nmt-attention

Poster Link

http://ej.uz/nmt-attention

Acknowledgements

This research was supported by the ICT COST Action IC1207 ParseME: Parsing and multi-word expressions - towards linguistic precision and computational efficiency in natural language processing, the grant H2020-ICT-2014-1-645442 (QT21) and Charles University Research Programme "Progres" Q18+Q48.