

Main Idea

Vanilla sequence-to-sequence was degrading with sentence length.

Goal of attention:

- Do not force the network to catch long-distance dependencies.
- Use decoder state only for:
 - target sentence dependencies (=LM) and
 - a as query for the source word sentence