

Sequence Length is a Domain: Length-based Overfitting in Transformer Models

Dušan Variš and Ondřej Bojar
varis,bojar@ufal.mff.cuni.cz

Charles University, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Prague, Czech Republic



Introduction

Transformers generalize poorly to longer AND shorter sequence editing examples. Similar trends can be observed on MT task.

Methods

Split data to buckets based on target-side length. Train a separate NMT system on each training bucket and evaluate it on the validation buckets.

Results

Strong implication of target-side-length overfitting in Transformers that use absolute position encoding.

- Higher train-test length difference → higher performance drop.
- Hypothesis length similar to that of training data.
- Length overfitting could be avoided with relative position embeddings (Neishi and Yoshinaga, 2019).

Transformers with absolute position encoding output sequences of length similar to sequences in training data.

Source (30-bucket)	The company does not collect its mail and it has closed its official headquarters in Žižkov more than six years ago.
Hyp1 (10-bucket)	Společnost nesbírá poštu a zavřel oficiální sídlo.
Hyp1 (gloss)	The company does not gather mail and closed official headquarters.
Hyp2 (30-bucket)	Společnost neshromažďuje poštu a již před více než šesti lety zavřela své oficiální sídlo v Žižkově .
Hyp2 (gloss)	The company does not collect mail and more than six years ago closed its official headquarters in Žižkov .
Hyp3 (60-bucket)	Společnost nevybírá poštu a uzavřela své oficiální sídlo v Žižkově více než šest let ago . v Žižkově. Společnost neshromažďuje poštu a uzavírá oficiální ústředí v Žižkově více než šest let agr. o .
Hyp3 (gloss)	The company does not pick up mail and closed up its official its official headquarters in Žižkov more than six years ago . in Žižkov. The company does not collect mail and closes up official headquarters in Žižkov more than six years agr. o .
Reference (30-bucket)	Nepřebírá poštu a oficiální sídlo na Žižkově zrušila před více než šesti lety.
Ref (gloss)	(The company) does not collect mail and official headquarters in Žižkov closed up more than six years ago.

Take a picture to
download the full paper

