Unsupervised Pretraining for Neural Machine Translation Using Elastic Weight Consolidation

Dušan Variš, Ondřej Bojar

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



Introduction

The unsupervised pretraining using monolingual corpora is very important for the low-resource neural machine translation (NMT). However, due to the small amount of parallel data, the NMT network can still overfit during the MT finetuning phase so an additional model regularization during fine-tuning is necessary.

We show that Elastic Weight Consolidation (EWC) can be used during the fine-tuning phase to regularize parts of the network that were pretrained using monolingual data.

Methods

We initialize the NMT system using a pretrained source and target-side language models (LMs).

Then, we add the following source-side (and similarly target-side) regularization term to the NMT loss:

$$L_{ewc-src}(\theta) = \sum_{i,\theta_i \subset \theta_{src}} \frac{\lambda}{2} F_{src,i}(\theta_i - \theta_{src,i}^{\star})^2$$
(1)

We estimate Fisher information F_{src} using the source-side validation data.

Results

		SRC	TGT	ALL
Baseline	15.68		_	
Backtrans.	19.65		_	—
LM best		13.96	15.56	16.83
EWC best		10.77	15.91	14.10
LM ens.		15.16	16.60	17.14
EWC ens.		10.73	16.63	14.66

EWC is an useful **regularizer** if the **original and fine-tuning tasks** are **closely related**.



This work has been in part supported by project no. 19-26934X (NEUREM3) of the Czech Science Foundation, the by the Charles University SVV project number 260 453 and by the grant no. 1140218 of the Grant Agency of the Charles University.



EWC-regularized decoder fine-tuning converges much faster than the fine-tuning using the original LM objective as a regularizer.



EWC does not work well if the encoder LM pretraining (left-side-only context) differs from the NMT encoder (bidirectional context).



Code available in NeuralMonkey: Github Repository: **ufal/neuralmonkey** Branch: **ewc_aclsrw2019**