Subject:
Fwd: Your NAACL-HLT 2019 Submission (Number 514)
From:
Inigo Jauregi <ijauregi@cmcrc.com>
Date:
23/02/2019, 9:11 am
To:
Massimo Piccardi <Massimo.Piccardi@uts.edu.au>, Ehsan Zare Borzeshi <ehsan.zareborzeshi@gmail.com>, Nazanin Esmaili <nazanin.esmaili@digitalhealthcrc.com>

Hi all,

The paper for NAACL has been accepted! Congratulations to all!!

Below the message from the organizers and the reviewers comments.

All the best,
Inigo

---------- Forwarded message ---------
From: Organizers, NAACL HLT 2019 <naacl2019@softconf.com>
Date: Sat, 23 Feb 2019 at 04:26
Subject: Your NAACL-HLT 2019 Submission (Number 514)
To: <ijauregi@cmcrc.com>

Dear Inigo Jauregi Unanue:

On behalf of the NAACL-HLT 2019 program committee, we are happy to inform you that the following submission has been accepted to appear at the conference:

ReWE: Regressing Word Embeddings for Regularization of Neural Machine Translation Systems

Our goal is to notify all authors within the next 2 weeks as to whether their paper will be presented as a talk or a poster. With this acceptance notice, we are confident that each set of authors can move ahead with their travel planning and paper revisions. The program committee worked very hard to thoroughly review all the submitted papers. Please repay their efforts by following their suggestions when you revise your paper, and remember that you have one extra page to address their comments.

Camera ready papers are due Monday, April 1st, 11:59pm UTC-12, and the final manuscript must be uploaded to the following site:

https://www.softconf.com/naacl2019/papers/

You will be prompted to login to your START account. If you do not see your submission, you can access it with the following passcode:

514X-C6H7H5E2C7

Alternatively, you can click on the following URL, which will take you directly to a form to submit your final paper (after logging into your account):

https://www.softconf.com/naacl2019/papers/user/scmd.cgi?scmd=aLogin&passcode=514X-C6H7H5E2C7

The reviews and comments are attached below. Remember that the selection process is not based solely on the scores you will get with your reviews, but also on detailed recommendations from the area chairs, discussions among reviewers, and a goal to assemble a varied and interesting program.

Congratulations on your fine work. If you have any additional questions, please feel free to get in touch.

Best Regards,
Christy Doran and Thamar Solorio - Program Co-Chairs NAACL HLT 2019
NAACL-HLT 2019

==
NAACL-HLT 2019 Reviews for Submission #514
==

Title: ReWE: Regressing Word Embeddings for Regularization of Neural Machine Translation Systems
Authors: Inigo Jauregi Unanue, Ehsan Zare Borzeshi, Nazanin Esmaili and Massimo Piccardi
==
REVIEWER #1
==

What is this paper about, and what contributions does it make?

Problem/Question:

This paper addresses the regularization problem of NMT systems. More specifically it tackles on the weakness of the MLE training objective to take into account the semantic similarity of different words in the target vocabulary.

Contributions (list at least two):

1. This paper proposes a new regularization term, Regressing Word Embedding Loss (ReWE loss), to be combined to the conventional MLE loss for NMT training.
2. This paper provides empirical results about the decent value of the combine coefficient between MLE and ReWE loss, and the performance comparison of MSE and CEL in term of distance metric used for the calculation of ReWE loss.
3. This paper proposes a method to leverage pre-train word embeddings into NMT training.

--

What strengths does this paper have?

--
Strengths (list at least two):

1. The proposed regularization technique applies for all seq2seq based NMT systems and is proven to have good results on strong settings.
2. The experimental approach is coherent: the targeted problem is well exposed, followed by a sound theoretical solution, then some conclusive experimental results.

--

What weaknesses does this paper have?

--
Weaknesses (list at least two):

1. Many aspects of this new approach need to be analyzed more thoroughly: for example, how does this approach perform on out-of-domain settings? How accurate is the learned target word embedding via this new ReWE loss? To what extent these learned target word embeddings can help the model to deal with OOV words during decoding?
2. No detail on the extra cost of using this ReWE loss for training is given, neither the information about how to pre-train the word level and BPE level embedding used for the calculation of ReWE loss.
3. Although the effectiveness of the new approach is proven over a pretty strong baseline, the training data scale remains too small for non low-resource settings (100k - 200k). Why not trying the approach on the state-of-the-art Transformer framework as the transfer of the new objective function is straight forward?
4. A related work is missing: Training Neural Machine Translation using Word Embedding-based Loss (https://arxiv.org/pdf/1807.11219.pdf)

--

--
Reviewer's Scores

--
Overall Score (1-6): 4
Readability (1-5): 4

===
===
REVIEWER #2
===
===

What is this paper about, and what contributions does it make?

--
Problem/Question:
The authors propose to include a regression module in an recurrent neural network based NMT model for the purpose of regularization. This module aims to directly predict the word embeddings of the correct target words from the decoder's hidden states. The module is trained either with minimum squared error or cosine similarity objectives. The method is evaluated on a number of language pairs in low-resource settings.

Contributions (list at least two):
1. A simple method for directly predicting word embeddings in the NMT framework is presented.
2. Two losses are empirically compared.
3. The approach is evaluated on three (relatively) low-resource translation tasks, showing improvements using the proposed technique (with cosine similarity as the loss).

--

What strengths does this paper have?

--
Strengths (list at least two):
- A new method for improving performance of neural machine translation systems is introduced. With the method, the authors achieve improvements over a number of existing state of the art MT systems.
- The method is evaluated on a number of language pairs, and statistical significance of the obtained results is reported.
- Optimizer instability was controlled in their experiments by repetition of training and testing.

--

What weaknesses does this paper have?

--
Weaknesses (list at least two):
- The authors restrict their evaluation to the low-resource setting, I would like to see this method evaluated on systems trained with larger data sets. Especially for the English-French language pair lots of data is available. Dropout, weight pruning and other methods are also useful for high-resource settings.
- It would be nice to see the exact formulation of the ReWE module, currently it's only described in running text.
- I'm missing a learning curve for the loss of the ReWE module, to estimate how the method performs. It would also be helpful to interpret the large weight assigned to the loss.
- The second loss that's reported on, MSE, seems to be deficient. If reported at all I would expect an analysis why it's deficient, e.g. with the help for the aforementioned learning curves.
- The comparison to the cited systems of Denkowski and Neubig (2017) seems arbitrary, the only interesting comparison in my opinion would be to the fully optimized system, and maybe with and without dropout.

--

--
Reviewer's Scores

--
                     Overall Score (1-6): 4
                       Readability (1-5): 4

Additional Comments (Optional)

--
Some comments by Section:

3 Model
- There should be a citation or a short explanation of the CEL and MSE losses in this context.

4 Experiments
- It's not really clear what the baseline of the authors exactly is (all features of Denkowski and Neubig's system, or just a subset?).
- I do not understand the sentence ending with "... for single models".
- Table 2: Lines 2-5 seems superfluous.
- "The values of our experiments are for blind runs ..." I also can't parse this sentence right.
- Is BPE used or not? I'm not sure why this is discussed here. There's some more discussion in the appendix, but I think it could be completely removed from the paper.

--

==
===
                            REVIEWER #3
==
===

What is this paper about, and what contributions does it make?

--
This paper described a method to increase the generalisation power of RNN-driven NMT based on joint learning and regularisation. The regularisation influences the translation model with contextual properties by simultaneously learning to predict the next word in the translation and regress toward its word embedding.
It’s a moderately original paper describing a regulation approaches to leverage the embeddings of the ground-truth tokens as targets.
The paper shows a novel way of using pre-trained word embeddings in NMT and show how the new method influences translation quality measured in terms of BLEU score. The paper also describes the training and hyper parameter tuning procedures, as well as bring informative translation examples.

--

What strengths does this paper have?

- Simple yet (seemingly) effective approach to increase the generalisation power of RNN-based NMT using regularization and pre-trained word embeddings
- Benchmark against the (Denkowski and Neubig, 2017) study which shows an advantage of the proposed method over various baselines
- Description of the training and hyper parameter tuning process (in Appendix)
- Informative translation examples (also in Appendix)

What weaknesses does this paper have?

- There is no benchmark against the state-of-the-art systems, like Transformer (Vaswani et al., 2017) presented in the paper
- Experimentation: both, selection of the language pairs and sizes of the corpora used in experimentation do not allow to reasonably estimate the impact of the proposed method
- The choice of BLEU score as the main evaluation metric without detailed error analysis apart from a small discussion in section 4.2

Reviewer's Scores

Overall Score (1-6): 4
Readability (1-5): 5

Additional Comments (Optional)

The paper was an interesting read and addresses one of the main machine learning problems in application to the task of MT. However, the paper would benefit from more convincing experimentation and better overview of existing regularisation methods for MT. For example, (Zhang et al., 2018) paper is not mentioned in the paper at all.

Zhang Zhirui, Shuangzhi Wu, Shujie Liu, Mu Li, Ming Zhou and Tong Xu. 2018. Regularizing Neural Machine Translation by Target-bidirectional Agreement. arXiv preprint arXiv:1808.04064.
