

Subject: Fwd: Your COLING 2020 Submission (Number 320)
From: Inigo <ijauregi@gmail.com>
Date: 30/09/2020, 10:47 pm
To: Massimo Piccardi <Massimo.Piccardi@uts.edu.au>, nazanin.esmail@gmail.com, Gholamreza.Haffari@monash.edu

Hi all,

The paper got accepted! Congratulations to everyone and thanks for a great collaboration!

All the best,
Inigo

----- Forwarded message -----
De: Chengqing Zong <czong@softconf.com>
Date: mié., 30 sept. 2020 a las 19:48
Subject: Your COLING 2020 Submission (Number 320)
To: <ijauregi@gmail.com>

Dear Inigo Jauregi Unanue:

On behalf of the COLING 2020 Program Committee, I am delighted to inform you that the following submission has been accepted to appear at the conference:

Leveraging Discourse Rewards for Document-Level Neural Machine Translation

The reviews with comments and suggestions for the final version of the paper are attached below.

The COLING2020 Program Committee, composed of 51 Area Chairs and 1561 Reviewers, worked very hard to thoroughly review over 1900 submitted papers. Please follow their suggestions when you revise your paper.

1. Final Submission Instructions

When you are finished, you can upload your final manuscript at the following site:

https://www.softconf.com/coling2020/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:

320X-G8B3G6B5E6

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/coling2020/papers/user/scmd.cgi?scmd=login&passcode=320X-G8B3G6B5E6

The reviews and comments are attached below. Again, try to follow their advice when you revise your paper.

2. Dual Submission Instructions

Please remember that if you have also submitted your paper to another conference or archival workshop, you will need to withdraw your paper from the other venues in order to present at COLING2020 and have your paper appear in the proceedings.

3. COLING2020 Conference

In response to the COVID-19 pandemic, we will be holding COLING2020 online on December, 8-13. The presentations at the conference will involve a combination of pre-recorded talks, live Q-A sessions, and interactive virtual poster rooms. We are still working on the final details of the virtual setup and will be sending you information about the details of your presentation in the next few weeks, along with details on registration and other practical aspects of the virtual conference. We very much appreciate your patience and understanding regarding the difficulties that the current COVID-19 pandemic is adding to the organization of a conference like ours.

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

Best Regards,
Program Chairs, COLING 2020
COLING 2020

COLING 2020 Reviews for Submission #320

Title: Leveraging Discourse Rewards for Document-Level Neural Machine Translation
Authors: Inigo Jauregi Unanue, Nazanin Esmaili, Gholamreza Haffari and Massimo Piccardi

REVIEWER #1
Reviewer's Scores
Relevance (1-5): 5
Readability/clarity (1-5): 5
Originality (1-5): 3
Technical correctness/soundness (1-5): 4
Reproducibility (1-5): 4
Substance (1-5): 3

Detailed Comments
This paper proposes to use a reinforcement learning approach that explicitly optimizes two established discourse metrics, lexical cohesion (LC) and coherence (COH). Extensive experiments on four different language pairs and three translation domains demonstrate the effectiveness of the proposed approach.

Strengths: 1. Overall, the paper is written well and easy to follow; 2. The experiments are sufficient and the results are comprehensive and convincing; 3. It shows that meta-learning can be used to effectively leverage training data from an auxiliary language for zero-shot and few-shot crosslingual transfer.
Weaknesses: The approach is straight and there is no much innovation about methodology. In particular, the authors should demonstrate the differences and relatedness with the papers (MIXER training proposed by Ranzato et al. 2016 and Minimum Risk Training for Neural Machine Translation proposed by Shiqi Shen et al. 2016)

Reviewer's Scores
Overall recommendation (1-5): 3
Confidence (1-5): 5
Presentation Type: Poster
Recommendation for Best Paper Award: No

REVIEWER #2
Reviewer's Scores
Relevance (1-5): 5
Readability/clarity (1-5): 5

Originality (1-5): 4
 Technical correctness/soundness (1-5): 4
 Reproducibility (1-5): 5
 Substance (1-5): 3

Detailed Comments

 This paper proposes doc-level NMT with minimum-risk training using coherence & cohesion objectives.

Strengths:

- The objectives are very well motivated based on classical NLP work.
- Plugging them as MRT objectives seems to slightly improve performance across multiple language pairs.
- The approach makes a lot of sense and I am surprised it has not been tried before.

Weaknesses

- The weighting parameter that trades off MRT with MLL should be a hyperparameter tuned on dev set, rather than "tuning" it on the test set as done in the tables.
- Missing related work.

Overall this is a nice paper with promising results. I believe it could use more work on thoroughly choosing the metrics, weighting, and other hyperparameters, but since the idea is encouraging I give a slight positive assessment.

References

- Using Context in Neural Machine Translation Training Objectives. Saunders, Stahlberg and Byrne, ACL 2020. -- similar MRT training for doc-level BLEU which is part of your objective.

Next ones are mostly for SMF but I believe much of the observations should still be very relevant:

- Document-Level Machine Translation: Ensuring Translational Consistency of Non-Local Phenomena. Eva Martinez Garcia, Ph.D. Thesis, 2019.
- The Trouble with Machine Translation Coherence: Karin Sim Smith, Wilker Aziz, Lucia Specia, EMW, 2016
- Coherence in Machine Translation. Karin Sim Smith, Ph.D. Thesis, 2017.

Thoughts

- While in general optimizing LC and COH makes a lot of sense, I am not sure about the choice of metrics selected here. How well does this pretrained topic model work qualitatively? Error propagation can be a major issue here.

- Deterministic learning with beam size=2 is a very rough estimate of the actual expected risk. I'm a bit surprised by the result in tdmuv bal that it outperforms sampling, and I suspect they might not be using any variance reduction strategies. I would suggest reconsidering sampling, possibly with the approach from Kool et al, Estimating Gradients for Discrete Random Variables by Sampling without Replacement, ICLR 2020.

Presentation:

page 2: double parantheses around Xiong et al

page 3 type: Texttilling, not Texttilling (single l)

page 4 and beyond. Why capitalize RLSA? it does not seem to be an acronym, right? The standard name for this is Expected Risk Minimization or Minimum Risk Training.

page 5 and results tables: do not write words in math mode (e.g. "a of cohesion devices", or "COH", or "BLEU-") latex interprets these as multiplication of several one-letter variables, as if you meant $c * o * h * e * s * i * o * n$... Instead use $\text{\texttt{\text{}}}$ or $\text{\texttt{\text{}}}$ depending of intent. In the results tables, you should have $\text{\texttt{\text{}}}$ and never $\text{\texttt{\text{}}}$.

page 6 inconsistent capitalization of Moses

Reviewer's Scores

 Overall recommendation (1-5): 4
 Confidence (1-5): 3
 Presentation Type: Oral
 Recommendation for Best Paper Award: No

 REVIEWER #3

Reviewer's Scores

 Relevance (1-5): 5
 Readability/clarity (1-5): 5
 Originality (1-5): 4
 Technical correctness/soundness (1-5): 4
 Reproducibility (1-5): 5
 Substance (1-5): 3

Detailed Comments

 The paper is proposing several linguistically motivated reward functions for text generation: lexical cohesion and coherence.

The authors illustrate the interest of such a reward signal for document-level translation on several language pairs with consistent improvements in BLEU score.

It could have been interesting to evaluate the approach on other tasks of generation like abstractive summarization.

Another interesting addition to the paper could have been to test on several seq2seq model architecture to verify consistent improvement in such settings.

Reviewer's Scores

 Overall recommendation (1-5): 4
 Confidence (1-5): 4
 Presentation Type: Poster
 Recommendation for Best Paper Award: No

 COLING 2020 - <https://www.softconf.com/coling2020/papers>

 ijauregi