[External Mail]EMNLP21 - Author Response Period Starts

start@z.softconf.com 代表 Program Chairs, EMNLP 2021 <papers_emnlp2021@softconf.com>
周日 2021/7/11 2:51

收件人:Ming6 Chen 陈明 <chenming6@xiaomi.com>;

抄送:emnlp2021pcs@gmail.com <emnlp2021pcs@gmail.com>;

Dear ming chen,

The author response phase has begun and will be open until 17 July, 2021 (anywhere on Earth). You now have a chance to view the draft reviews of your submission, and, optionally, write a response:

Title: Emoji Knowledge Contained Pre-training for Sentiment Analysis

Number: 900

If you choose to respond to the reviews:

- Remember that reviewing is double-blind.
- Do not include any information in the response that can identify you or your co-authors.
- Do not include any URLs in your response.
- Use the response judiciously. If a reviewer has expressed uncertainty about an issue, or is making an incorrect assumption, or has misunderstood a point in the paper, please address these concerns in your author response.
- There is no need to respond to every minor question or suggestion for improvement.
- In interpreting your review, and the numerical scores, please refer to the review instructions at https://2021.emnlp.org/blog/2021-05-17-instructions-for-reviewers/.
- You are requested not to use the response form to debate the reviewers' subjective opinions regarding the merit of your work. Nor should you try to "correct" your paper in any way - either in terms of its basic technical arguments, or in the presentation of those arguments. Above all, the response facility should **not** be used to report on new results, obtained after the submission deadline.
- This process is intended to help produce more accurate reviews for your paper. It is not intended to strengthen the arguments in your paper.

If the paper is accepted, you will still have ample opportunity to make revisions, i.e., before submitting your camera-ready copy.

Responses must be **no more than 900 words in total**, counting the text boxes for responding to each reviewer, and the general response box. Comments to the chair do not count towards this word count.

Your reviews will appear at the end of this email. Each submission should have at least three reviews. If your paper has fewer than three reviews, please be assured that our senior program committee is working hard to chase the missing reviews. So please check START at a later point.

To make your response, go to the following URL:

https://www.softconf.com/emnlp2021/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:

900X-A7P6A9P5E2

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/emnlp2021/papers/user/scmd.cgi?scmd=aLogin&passcode=900X-A7P6A9P5E2

Best regards,

EMNLP 2021 Program Co-chairs

```
======================================================================
EMNLP 2021 Reviews for Submission #900
======================================================================

Title: Emoji Knowledge Contained Pre-training for Sentiment Analysis
Authors: ming chen, cheng zhou, like wang, xiaoshu chen and hucheng huang


======================================================================
                            REVIEWER #1
======================================================================

What is this paper about, what contributions does it make, and what are the main strengths and weaknesses?
----------------------------------------------------------------------
This paper proposes to integrate emoji into the pre-train model towards sentiment analysis. It considers not only the emoji itself but also their combinations by an unsupervised m

Strength: The paper studies an interesting problem. The proposed dataset is valuable for the community. The paper is basically clearly written.

Weakness: The downstream tasks are not sufficient to evaluate the effectiveness of the proposed PRE-TRAIN model. More downstream tasks such as aspect-level sentiment analysis, asp
----------------------------------------------------------------------


Reasons to accept
----------------------------------------------------------------------
A really interesting problem with reasonable technical contributions.
----------------------------------------------------------------------


Reasons to reject
----------------------------------------------------------------------
Insufficient references and unconvincing experimental results.
----------------------------------------------------------------------


Questions for the Author(s)
----------------------------------------------------------------------
Eq.(4), is the R(C) and AMI(C) have the same scale?
```

Why table 4 only contain the visualization of RoBERTa and EKCP? How about the other strong competitors?

Missing References

Recurrent Neural Network for Text Classification with Hierarchical Multiscale Dense Connections.
Beyond Polarity: Interpretable Financial Sentiment Analysis with Hierarchical Query-driven Attention.
Hierarchical Inter-Attention Network for Document Classification with Multi-Task Learning.

Typos, Grammar, Style, and Presentation Improvements

Line 356, no emoji -> non-emoji.
Line 281, W^{left} -> W^{right}

Reviewer's Scores

Reproducibility: 4
Overall Recommendation - Long Paper: 3.5

REVIEWER #2

What is this paper about, what contributions does it make, and what are the main strengths and weaknesses?

The authors propose a pre-training model Emoji Knowledge Contained Pre-training (EKCP) to learn the sentiment of Chinese sentences containing emojis. Unlike other BERT models, EKC

Reasons to accept

1) The use of the combination and sequence of emojis (for example multiple stars one after another as ratings).
2) Considering both Self Information and Mutual Information for building an Emoji Knowledge Dictionary.
3) Providing Attention Visualization for the final hidden vector to understand which words and emoji were emphasized by the model.
4) It shows a new direction on how to incorporate the sentiment of emojis to understand the sentiment of a given Chinese text containing emojis.

Reasons to reject

Annotating a dataset itself is considered a full paper in many top conferences. However, the authors claim to label a 50K dataset carefully for positive, neutral, and negative sen

Also, any pre-training language model can be trained on top of the existing model so it is not clear why the authors train their model with 100,000K samples that do not contain an

Questions for the Author(s)

Please refer to the "Reasons to reject" section

Reviewer's Scores

Reproducibility: 3
Overall Recommendation - Long Paper: 3

REVIEWER #3

What is this paper about, what contributions does it make, and what are the main strengths and weaknesses?

This paper proposes ways to embed emoji information into representations by BERT models for sentiment analysis where such information is just ignored in the canonical from of [UNK
mojis into consideration. The collected data itself is another contribution of this work since there are not many Chinese corpora in the community for sentiment analysis research

The strengths and weaknesses of this papers are described in the reasons to accept/reject.

Reasons to accept

1. Open-sourced dataset in Chinese, which can be another good resource for sentiment analysis research across multiple languages and should be introduced broadly in conferences li
2. Simple, but reasonable approaches introduced to take emoji combinations into consideration by automatically segmenting replicated or combined emojis in series

Reasons to reject

1. A little lack of analysis on the proposed ideas, more specifically Section 4.3 could have been expanded with more experiments to strengthen the claimed novelty

Questions for the Author(s)

- Can you elaborate a bit more how higher uncertainty of tokens in the left and right side of C means that C would be a more meaningful emoji combination? Likewise, how higher AMI
- How many emojis were considered? It seems this information isn't disclosed although it's critical. While Table 6 provides some perspective in this regard, the total number of si

Typos, Grammar, Style, and Presentation Improvements

- In Page 2 (left-column), emoji combination though emoji-focused --> emoji combination through emoji-focused
- Please be consistent with upper-/lower-cased forms of nouns like recurrent neural network or Convolutional Neural Networks. Style inconsistency like this would make your paper l
- The indexes of W^left and W^right are confusing. The indexes (j and a) of C refer to the location of tokens and these location indexes can be updated when they come to W^left an
- In Page 4 (right below Eq. 1), ER of W^left --> ER of W^right?
- In Page 6 (Section 4.1), 101,000k --> 101M & 100,000k --> 100M & 1,000k --> 1M
- In Page 6 (Section 4.1, too), with the their corresponding --> with their corresponding

Reviewer's Scores

Reproducibility: 3
Overall Recommendation - Long Paper: 3.5

--
EMNLP 2021 - https://www.softconf.com/emnlp2021/papers