JoEMLS Open Peer Review Report

Rebuttal to the Comments

Open Point－the Open Peer Review System

Our journal has newly adopted the Open Point system, for illuminating precious records of dialogues and argumentations between authors and reviewers, and for encouraging them to express freely their discourses and opinions in comments and responses which both the parties agree to make public. In this well-intentioned mode of Open Peer Review, the presentation of dialogue contents enables more scholars to see viewpoints worth understanding and citing behind the published manuscripts. This mechanism helps our authors, reviewers and readers to receive authentic and substantial essence of academic communication.

Reviewed-Article Title	The Feasibility of Automated Topic Analysis: An Empirical Evaluation of Deep Learning Techniques Applied to Skew-Distributed Chinese Text Classification
Reviewer(s)	1. Anonymous reviewer A (only public the review comments) 2. Anonymous reviewer B (Name and review comments are not permitted to public)
Author(s)	Yuen-Hsien Tseng
Vol. & No.	Vol.57 No.1 (March 2020)
DOI：	https://doi.org/10.6120/JoEMLS.202003_57(1).0047.RS.CE_OPR

First Review Stage

Reviewer A: Anonymous Reviewer A

Review Comments：

1. This paper compared 5 different Chinese dataset against 8 different machine learning methods, including the most recent DL technologies such as CNN, RCNN, and BERT on document classification problems. It turns out that SVM and BERT perform better against others in some dataset. They evaluate in terms of micro and macro F1 scores. Tf-idf as well as word embedding are used as input features for the models and evaluated. Basically the evaluation experiments are rigorous and complete. However, the classification methods are popular and novelty is not as much.

- Author responded:
  - We thank the reviewer for recognizing our experiments.
  - This paper is not about proposing a novel or better approach for automated text classification (as stated in the 3rd and 4th paragraphs in the INTRODUCTION), but more about evaluating existing tools applied to real-world data which have some unique characteristics (skewdistributed) to reveal their practicality.
  - The possible contributions of this line of study are stated in the second paragraph of new section:“3. Research Questions”.

2. It is worthwhile to discuss more why BERT and SVM outperform others on some datasets. What properties of the dataset make the two classifiers perform better. It would be guideline for users to choose the classifier to the new dataset.

- Author responded:
  - We have made more discussions on how BERT and SVM outperform others. For BERT, please refer to:
    1) the existing 5th paragraph of Sub-Section “4.2 Machine Learning Models” on Page 11;
    2) the new paragraph beneath Table 4.

- For SVM, our experiment confirms the conclusions of past studies (as reviewed in Section 2) for the Chinese dataset. We also briefly discuss the strength of SVM, on Page 22, in a new paragraph starting "Based on Table 3, SVM is a competitive technique in TC tasks …".
- From our experiments, discussions, and past studies, we have added a guideline in the first paragraph of the CONCLUSION section for users to choose suitable classifiers for a new dataset.

3. Also word embedding and tf-idf input are compared, it is worthwhile to discuss the two encoding methods in comparison to the domain dataset properties.

- Author responded:
  - The comparison of tf-idf vectors and word embedding vectors have been discussed in the existing 2nd paragraph in Sub-Section "4.1 Feature Extraction".
  - Discussions of the two encoding methods in comparison to the domain dataset properties are difficult, because they are the initial step which indirectly affects the final effect of a TC task. We refrain these discussions from over speculations, because the evaluation of text classification for any domain dataset is typically conducted experimentally, rather than analytically, due to its subjective characteristics, as pointed out by Sebastiani (2002) shown in the new section "2. RELATED WORK".

Reviewer B: Anonymous Reviewer B

Review Comment：

(Reviewer B didn't public the reviewer's name and review comments)

Open Peer Review: How to cite this report?

APA

In-text Citation:

Anonymous Reviewer A and Anonymous Reviewer B (2020) ...

... (Anonymous Reviewer A & Anonymous Reviewer B, 2020).

Reference:

Anonymous Reviewer A & Anonymous Reviewer B. (2020). JoEMLS open peer review report: Rebuttal to the comments [Review of the article "The Feasibility of Automated Topic Analysis: An Empirical Evaluation of Deep Learning Techniques Applied to Skew-Distributed Chinese Text Classification," by Yuen-Hsien Tseng]. *Journal of Educational Media & Library Sciences*, *56*(3), D1-D4. https://doi.org/10.6120/JoEMLS.202003_57(1).0047.RS.CE_OPR

Chicago (Turabian)

Note

7. Yuen-Hsien Tseng, Peer commentary and rebuttal, January 2020, in "The Feasibility of Automated Topic Analysis: An Empirical Evaluation of Deep Learning Techniques Applied to Skew-Distributed Chinese Text Classification," *Journal of Educational Media & Library Sciences* 57, no. 1 (March 2020), D2, https://doi.org/10.6120/JoEMLS.202003_57(1).0047.RS.CE_OPR.

Bibliography

Tseng, Yuen-Hsien, Peer Commentary and Rebuttal, January 2020. In "The Feasibility of Automated Topic Analysis: An Empirical Evaluation of Deep Learning Techniques Applied to Skew-Distributed Chinese Text Classification." *Journal of Educational Media & Library Sciences* 57, no. 1 (March 2020): D1-D4. https://doi.org/10.6120/JoEMLS.202003_57(1).0047.RS.CE_OPR.