Detecting Missing Information of Questions in Q&A Chatrooms

Anonymous Author(s)*

ABSTRACT

The appendix (Appendix A) of our main paper is outlined as follows. Section A.1 presents the performance of pre-trained models on reduced *validation set* of 1,000, 800, 600, 400, 200, respectively. The results indicate a certain level of robustness in detecting missing PS despite the reduction in *validation set* size.

CCS CONCEPTS

• **Software and its engineering → Software development process management**.

KEYWORDS

Chatroom, Lexico-syntactic Pattern, Pre-trained, LLM, CoT Prompt

ACM Reference Format:

Anonymous Author(s). 2018. Detecting Missing Information of Questions in Q&A Chatrooms. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 2 pages. https://doi.org/XXXXXXX.XXXXXXX

A ADDITIONAL STATISTICS AND ANALYSIS

A.1 Performance of pre-trained models on reduced validation set

In addition to the 10-fold cross-validation (10CV) experiments on 1,000 chatroom questions with pre-trained models presented in the main paper, we conducted another set of experiments to evaluate the performance of these models when reducing the size of the *validation set* from 1,000 to 200, decreasing by 200 each time. Given the limited size of the reduced *validation set*, we employed a 5-fold cross-validation (5CV) strategy, allocating 80%, 10%, and 10% of the data for training, parameter tuning, and testing per fold, respectively.

According to Figure 1, F1 score dramatically decreased when the *validation set* size was reduced to 400 for detecting missing ETD and 600 for detecting missing PS. However, the F1 score remained almost unchanged for detecting missing PS when the *validation set* size reduced from 1000 to 600, showing only a 10%-15% drop in F1 score between the best and worst cases. This indicates a certain level of robustness to the reduced *validation set* size for detecting missing PS. In contrast, the F1 score for detecting missing ETD exhibited a linear decrease before the dramatic drop at a *validation set* size of 400, with a 20%-40% decline in F1 score between the best and worst cases.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

REFERENCES

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

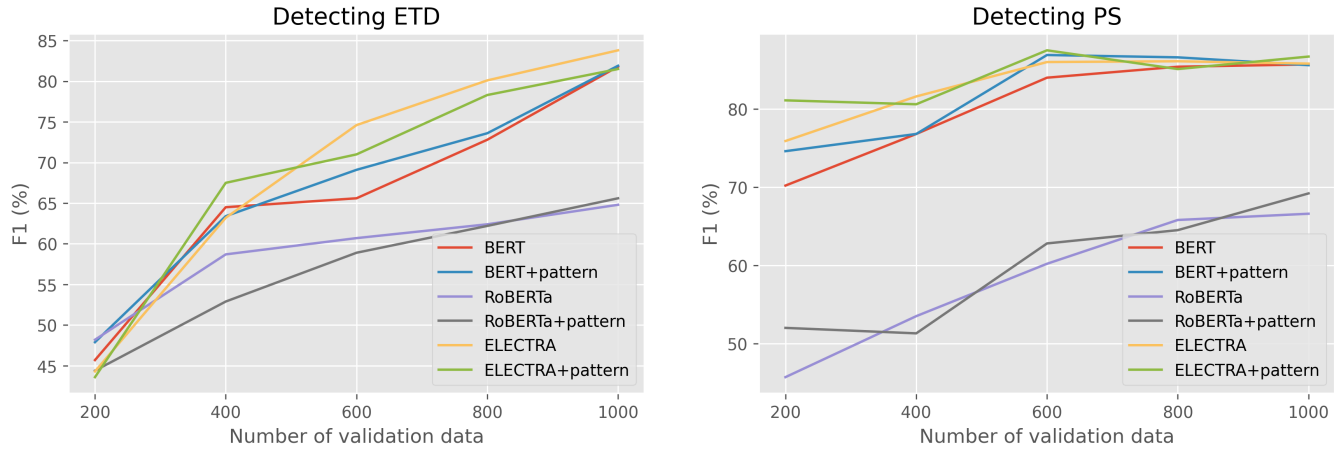

Figure 1: Performance of detecting missing ETD and PS for pre-trained models when reducing the *validation set* size.