Appendix : A Span-based Dynamic Local Attention Model for Sequential Sentence Classification

Anonymous ACL-IJCNLP submission

A Detailed Description of Datasets

The two standard benchmark datasets we used are described as follows:

NICTA-PIBOSO (Kim et al., 2011) contains 1,000 biomedical abstracts with sentences classified into six categories (population, intervention, comparison, outcome, other, study design).

PubMed 20k RCT (Dernoncourt and Lee, 2017) consists of 200,000 PubMed abstracts with sentences annotated with five rhetorical labels (background, objective, methods, results, conclusions).

The statistics of datasets are shown in Table 1.

Dataset	$\|C\|$	$\|V\|$	Train	Validation	Test
NICTA-PIBOSO	6	17k	720 (7.7k)	80 (0.9k)	200 (2.2k)
PubMed 20k RCT	5	68k	15k (180k)	2.5k (30k)	2.5k (30k)

Table 1: The statistics of datasets. $|C|$ denotes the number of labels, $|V|$ denotes the vocabulary size. For the train, validation, and test sets, we indicate the number of abstracts followed by the number of sentences in parentheses.

References

Franck Dernoncourt and Ji Young Lee. 2017. Pubmed 200k rct: a dataset for sequential sentence classification in medical abstracts. *arXiv preprint arXiv:1710.06071*.

Su Nam Kim, David Martinez, Lawrence Cavedon, and Lars Yencken. 2011. Automatic classification of sentences to support evidence based medicine. In *BMC bioinformatics*, volume 12, pages 1–10. BioMed Central.