A   Supplementary Material

A.1   Wikipedia Article Growth

Figure 3 shows how the monthly average of the article/edits length (in number of characters) varies over time for 4 different Wikipedia pages in two different languages: Donald Trump and World War II in the English Wikipedia, and Deutschland (Germany) and Zweiter Weltkrieg (World War II) in the German Wikipedia. It is possible to see that the average Wikipedia article size has been increasing dramatically over the years, and that this tendency seems to generalize across languages. In contrast, we also see that the average size of the edits applied remain relatively constant and that these are comparatively short.
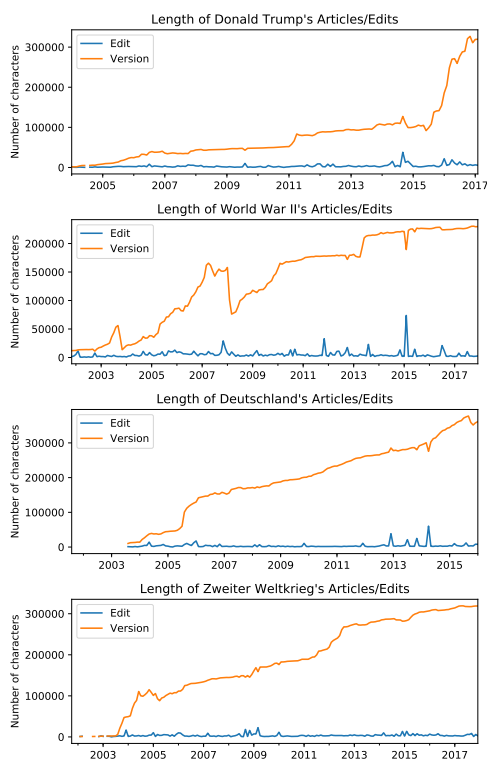

Figure 3: Monthly average of the article/edits length (in number of characters) for different Wikipedia articles.

A.2   Impact of document length on performance

We implement the approach by Dang and Ignat (2016) directly based on their code release. Their implementation uses the test documents when training the doc2vec model, which we consider inadequate. Instead we train only using the documents in the training split. The results for the original setting, although not reported here, are similar.

Table 4 shows how the performance the doc2vec-based approach on the test split of the *Wikiclass* dataset for different document lengths, in characters. This model is regarded as the state-of-the-art model on this dataset that does not require hand-crafted features.
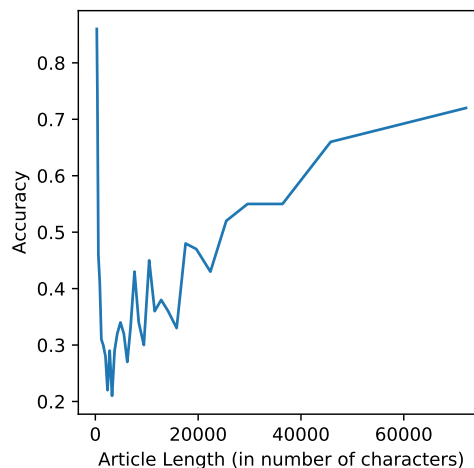

Figure 4: Performance of the doc2vec-based approach on the test split of the *Wikiclass* dataset, for different input lengths.