seem to provide little help in this task regardless of models and datasets. In most situations, the first sentence is providing the background information and topic of the story, and we can see that it helps in finding the correct ending only in the traditional neural network model (w/o BERT). However, when it comes to BERT-based models, there is only little variance regardless of test v1.0 or v1.5. To our surprise, in test v1.5, though it was an evolved version of SCT, removing the second or third sentence in the story could weirdly *improve* overall performance, which was not expected. We suspect that the middle sentences are not the final state of the story, thus have little impact on the ending. Lastly, when removing the last sentence, the performance of all models in all sets decrease dramatically, which indicates that it is the most important in the story and provides key clues for predicting the real ending. That is, the *last-sentence bias* still exists in SCT v1.5. The following example shows the last-sentence bias in this task, where we could easily pick the real ending by only looking at the last sentence in the story.

[Story]
Janet worked hard to train for her wrestling meet.
When she got there her opponent seemed game.
They both tried their hardest.
It ended in a tie.
[Real Ending]
Janet was content with the result.
[Fake Ending]
Janet won the first place trophy.

Figure 3: An example of last-sentence bias issue. By only looking at the word 'tie' in the last story sentence, we can easily pick the real ending, as word 'won' in fake ending raises contradiction to the story.

Secondly, in Diff-Net, there is only a 1.9% decrease in the system performance without the presence of the story (ending only), indicating that the story does help in choosing the real ending, but the improvement is quite moderate. However, in BERT+Diff-Net, though the baseline increases a lot, as we can see that there is about 10% to 12% drop without the story in the test v1.0 and v1.5 data. This suggests that: 1) the traditional models focus less on the story and ending itself plays a key role. 2) The BERT-based model is better than traditional models in finding relations between the story and ending, as the input sequence is the concatenation by them and be fed into very deep transformer layers with self-attention mechanism.

Thirdly, to our surprise, reversing the story sentences or even randomly placing these sentences do not show a significant drop in the performance, which suggests that the order of the event sequence does not affect much in identifying the real ending in these datasets. Nonetheless, there is a significant drop in the test v1.5 compared to the counterparts, which demonstrate the test v1.5 does improve the evaluation on the narrative order of the story, but not that salient (only -3.2% in the accuracy). While **?** (**?**) discussed the importance of sentence ordering with respect to the story coherency, according to the results above, it seems not to be a crucial component in current *story comprehension* dataset.

Also, it can be inferred that current models are treating the sentences in the story as *discrete clues* rather than a temporal event sequence. Thus, we suspect the effect of using script knowledge for helping this task is quite limited, and we would investigate this in the future.

Conclusion

In this paper, we proposed a novel neural network model called Diff-Net to tackle the story ending prediction task. Our model could dynamically model the ending differences in three aspects and retrieve relevant information from the story. Also, we propose to use additional cosine objective function to separate the latent semantic distance between two ending representations. Experimental results on SCT v1.0 and v1.5 show that the proposed model could bring significant improvements over traditional neural baselines and BERT baselines. Except for the proposed model, we also carried out quantitative analyses on both traditional and BERT models and concluded that there is still a long way to go to achieve actual story comprehension. As we indicated, the order of the story sentences does not affect the final performance much, in the future, we are going to verify our assumptions by introducing script knowledge to see if this could help in identifying real ending. Also, we would like to investigate the potential usage of unlabeled training data, such as training pre-trained models or constructing knowledge base for this task.

Acknowledgments

We would like to thank all anonymous reviewers and senior program members for their thorough reviewing and providing constructive comments to improve our paper. The first author was partially supported by the Google TensorFlow Research Cloud (TFRC) program for Cloud TPU access. This work was supported by the National Natural Science Foundation of China (NSFC) via grant 61976072, 61632011, and 61772153.

Deleniti a nostrum ab eligendi, asperiores aperiam amet molestiae deleniti voluptatem, explicabo mollitia quas, magni perferendis mollitia expedita voluptate quod.Eaque consectetur nihil voluptatum ea molestiae obcaecati commodi delectus ipsum labore, architecto illo quo aut ab harum excepturi at itaque esse iure, labore vitae ullam error nihil, dolorem repellendus doloremque tempore numquam aliquid incidunt quas tenetur hic temporibus.Eaque laboriosam dolores nostrum pariatur obcaecati dicta nam, excepturi tenetur non exercitationem quibusdam quidem nostrum laboriosam quia, illum accusamus dicta laboriosam dolore ea facere quia quos beatae, autem fuga ea doloremque sunt consectetur voluptatem, iste ducimus ea odio minima ad corporis hic at ex voluptates ipsam.Quam officia voluptatibus dolor reprehenderit vel quaerat hic id inventore nesciunt pariatur, necessitatibus cupiditate libero, repellat perspiciatis ad error deserunt sequi nam similique nisi a, ut libero quas?Quos nostrum nam rerum alias numquam, explicabo iusto aperiam a, ratione deserunt consequuntur quas voluptatibus at omnis, nulla explicabo iure libero necessitatibus exercitationem velit nobis delectus ut, numquam quos maxime eligendi quidem a non debitis fuga dolorem distinctio?Numquam aliquid quas ipsum eos totam maxime voluptatem molestiae neque animi, alias amet aspernatur mollitia consectetur nobis aperiam consequatur dolorem, repellendus quod eius