Supplementary Material for ROCStories and the Story Cloze Test

1 Other Sources of Stories: Case of Weblog Stories

Personal stories from daily weblogs are good sources of commonsense causal information (Gordon and Swanson, 2009; Manshadi et al., 2008), but teasing out useful information from noisy blog entries is a problem of its own. Consider the following snippet from ICWSM 2011 Spinn3r Dataset of Weblog entries (Burton et al., 2009):

> "I had an interesting day in the studio today. It was so interesting that I took pictures along the way to describe it to you. Sometimes I like to read an autobiography/biography to discover how someone got from there to here.....how they started, how they traveled in mind and spirit, what made them who they are now. Well, today, my work was a little like that, as I had pieces, but didn't know how they were going to go together, or where they were going at all for that matter."

This text is rich, complex, and full of discourse complexities. A host of challenging language understanding tasks are required to get at the commonsense knowledge embedded within. What is needed is a simplified version of these narratives. This paper thus introduces a new corpus of 19,858 short commonsense stories.

2 Short Stories Corpus: Data Collection Methodology

We tested numerous pilots with varying prompts and instructions. We saw the best results when we let the workers write about anything they have in mind, as opposed to mandating a pre-specified topic. Instruction .1 is the final crowdsourcing prompt that we used in AMT to collect our corpus.

INSTRUCTION .1

Imagine that you want to **tell a five-sentence story to your friend**. It can be about something that happened, something you or someone else has experienced in the past, or simply any life story about someone or something. Your task is to write this five-sentence story. Your story should have all of the following three properties:

1. It should be entirely realistic, with no fairy tales or fictions.

2. It should read like a coherent story, with a specific beginning and ending, where something happens in between.

3. Each sentence in your story should be logically related to the next sentence and be about the characters of the story. Nothing should feel random, irrelevant or redundant.

The key in the above instructions is the second property: stories with a specific beginning and ending typically include causal and temporal links between events. We set a limit of 70 characters to the length of each sentence. This prevented long multi-part sentences which include unnecessary details. The workers were also asked to provide a title that best describes their story. Last but not least, we instructed the workers not to use quotations in their sentences and avoid using slang or informal language.

Quality Control. One issue with crowdsourcing is how to instruct non-expert workers. This task is a

type of creative writing, and is trickier than classification and tagging tasks. In order to ensure we get qualified workers, we designed a qualification test on AMT in which the workers had to judge whether or not a given story (total five stories) is an acceptable one. This not only eliminates any potential spammers on AMT, but also provides us with a pool of potential creative story writers. We had 767 workers take the qualification test, 417 of which were qualified and contributed to the data collection effort. Furthermore, we qualitatively browsed through the submissions and gave the workers detailed feedback before approving their submissions. We often bonused our top workers, encouraging them to write new stories on a daily basis.

3 Story Cloze Test: Data Collection Methodology

We sampled 2,500 stories from our Short Stories Corpus to get their first four sentence as the contexts. In order to get the two alternative endings, we crowdsourced the task on AMT. We achieved the most promising results by using the prompt in Instruction .2.

INSTRUCTION .2

You are given a sequence of four sentences which together form a coherent story. Your task is to write the ending fifth sentence in two ways:
(1) 'right ending': that naturally ends the story in a coherent and meaningful way.
(2) 'wrong ending': which is entirely impossible to be a correct/natural ending to the story. That is, if you add this fifth sentence to the four sentences it does not make sense as a meaningful story. Both your 'right ending' and 'wrong ending' should have the following properties:

1. The sentence should follow up the story by sharing at least one of the characters of the story.

2. The sentence should be entirely realistic, meaningful and sensible when read in isolation by itself. It should not include anything which is fictional or not likely to happen in the real world.

Quality Control. The accuracy of the Story Cloze Test can play a crucial role in directing the research community in the right trajectory. In order to ensure quality, we implemented the following two-step quality control:

1. Qualification Test: We designed a qualification test for this task, where the workers had to choose whether or not a given 'right ending' and 'wrong ending' satisfy our constraints. Out of 111 workers who took the qualification test, 34 were qualified who participated in the final data collection. At this stage we collected 2,500 cloze test cases.

2. Human Verification: In order to further validate the cloze test cases, we compiled the 2,500 Story Cloze Test cases into 5,000 full five-sentence stories. Then for each story we asked three crowd workers to verify whether or not the given sequence of five sentences makes sense as a meaningful and coherent story, rating within {-1, 0, 1}. Then we filtered cloze test cases which had 'right ending' with average rating higher than 0 and 'wrong ending' with average rating lower than 0. This process ensures that the 'right ending' is indeed a meaningful ending to the story and 'wrong ending' does not make sense as an ending to the story. This resulted in final 1,000 test cases, which hereinafter is called 'Story Cloze Challenge' set[1]. We also made sure to remove the original stories of the Challenge Story Cloze Test from our Short Stories Corpus.

References

K. Burton, A. Java, , and I. Soboroff. 2009. The icwsm 2009 spinn3r dataset. In In Proceedings of the Third Annual Conference on Weblogs and Social Media (ICWSM 2009), San Jose, CA.

Andrew S. Gordon and Reid Swanson. 2009. Identifying Personal Stories in Millions of Weblog Entries. In Third International Conference on Weblogs and Social Media, Data Challenge Workshop, San Jose, CA, May.

Mehdi Manshadi, Reid Swanson, and Andrew S. Gordon. 2008. Learning a Probabilistic Model of Event Sequences From Internet Weblog Stories. In 21st Conference of the Florida AI Society, Applied Natural Language Processing Track, Coconut Grove, FL, May.

[1]The current version of Story Cloze Challenge set can be found here: https://competitions.codalab.org/competitions/15333