Towards Generative Commonsense Reasoning: A Concept Paper

Bill Yuchen Lin
yuchen.lin@usc.edu

Computer Science Department,
University of Southern California,
Los Angeles, California

Abstract

In this concept paper, we first review recent advances in machine commonsense reasoning and further investigate their potential connections with natural language generation. Current research efforts mainly focus on the learning of deterministic commonsense reasoners with multi-choice question answering datasets. We discuss the reasons why large pre-trained language encoders like BERT can easily achieve the state-of-the-art performance on these datasets. In a perspective of natural language generation, we advocate to evaluate the machine commonsense reasoning ability in a way of controlled language generation. We show related pilot works about this topic and discuss the future directions of the research in generative commonsense reasoning.

1 Commonsense Reasoning Challenges

There is a recent surge of novel large-scale datasets for testing machine commonsense with various focuses, such as situation prediction (SWAG) (Zellers et al., 2018), social behavior understanding (SocialIQA) (Sap et al., 2019a,b), visual scene comprehension (VCR) (Zellers et al., 2019), and general commonsense question answering (CommonsenseQA) (Talmor et al., 2019), which encourages the study of supervised learning methods for **deterministic commonsense reasoning**. Most of these current datasets for testing machine commonsense reasoning focus on multi-choice question answering tasks. These challenges expect machines perform much worse than general humans, due to that the commonsense is generally difficult to acquisition and utilize for reasoning in natural language understanding tasks.

However, fine-tuning large pre-trained language encoders (e.g. BERT, RoBERTa, XL-Net) has shown very promising results on such deterministic natural language understanding (NLU) challenges. This is mainly because the unsupervised pre-training objectives of such large language encoders are very similar to the design of the tasks. There are two main proxy tasks in BERT-like pre-trained models: masked language modeling (Masked-LM) and next-sentence classification (NextSent). The connections between these two tasks and the

- Masked-LM and CommonsenseQA: Before feeding word sequences into BERT, 15% of the words in each sequence are replaced with a “[MASK]” token. The model then attempts to predict the original tokens of the masked positions, based on the context provided by the other, non-masked, words in the sequence. The answers in CommonsenseQA are usually very short and can be filled as parts in the original questions. Thus, the supervised learning process for selecting a correct answer for a question is more like transforming the questions to masked sentences and conducting the masked-LM task.
- NextSent and SWAG: The correct answers are naturally next sentences of question content in SWAG. Thus, the NextSent objective perfectly aligns this task and data. Experimental results also show that BERT and other language models can almost perform as well as human performance.

Similarly, Trinh and Le (2018) find that the ensemble large pre-trained language models (including GPT-2) show promising results in WSC-like resolution task (Levesque, 2011).

2 Commonsense Knowledge and Language Modeling

Recent efforts also indicate that language models implicitly contain much commonsense knowledge and even factual knowledge. This is intuitive since a natural language sentence with high probability in language modeling is more likely to obey commonsense knowledge. Petroni et al. (2019) extensively study the possibility use language models as knowledge bases. They find that "(i) without fine-tuning, BERT contains relational knowledge competitive with traditional NLP methods that have some access to oracle knowledge, (ii) BERT also does remarkably well on open-domain question answering against a supervised baseline, and (iii) certain types of factual knowledge are learned much more readily than others by standard language model pre-training approaches". Feldman et al. (2019) also propose straightforward methods to effectively extract commonsense knowledge from pre-trained language models.

Applying commonsense knowledge in language generation is also explored recently for generating stories (Guan et al., 2018; Mao et al., 2019) and conversations (Zhou et al., 2018). While showing promising results, it is difficult to directly test machine commonsense reasoning in a language generation task. Thus, we would like to propose a new commonsense reasoning challenge in the setting of controlled language generation tasks. Specifically, given a set of common concepts, we would like to ask a language generation model to produce a sentence that contains them while being natural and make common sense.

3 Controlled Text Generation

Controlled text generation (Hu et al., 2017) studies how we can build a text generation model with prior constraints such as templates, sentiment, style, etc. To the best of our knowledge, few work has been done for constraining the target content having a set of pre-defined keywords. Different from the masked-LM objective, keywords-based sentence generation is to mask the majority of the context words and then try to generate a natural sentence to contain them. Also, this task eliminates the limitation of the word positions, such that the search space is further enlarged. Though the task seems to be overly open for the target outputs, we expect the sentences obey commonsense. Compared to other conditional text generation tasks like machine translation and abstractive summarization, our inputs are less structured (a set of words v.s. a meaningful sequence of words).

Potentially suitable methods for this problem is graph2sequence models (Song et al., 2018) for graph-based text generation tasks, such as AMR-to-text (Damonte and Cohen, 2019) and tripe-to-text generation (Zhu et al., 2019) challenges.

References

Marco Damonte and Shay B. Cohen. 2019. Structural neural encoders for amr-to-text generation. In *NAACL-HLT*.

Joshua Feldman, Joe Davison, and Alexander M. Rush. 2019. Commonsense knowledge mining from pre-trained models. *ArXiv*, abs/1909.00505.

Jian Guan, Yansen Wang, and Minlie Huang. 2018. Story ending generation with incremental encoding and commonsense knowledge. In *AAAI*.

Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P. Xing. 2017. Toward controlled generation of text. In *ICML*.

Hector J. Levesque. 2011. The winograd schema challenge. In *AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning*.

Huanru Henry Mao, Bodhisattwa Prasad Majumder, Julian J. McAuley, and Garrison W. Cottrell. 2019. Improving neural story generation by targeted common sense grounding. *ArXiv*, abs/1908.09451.

Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H. Miller, and Sebastian Riedel. 2019. Language models as knowledge bases? *ArXiv*, abs/1909.01066.

Maarten Sap, Ronan LeBras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019a. Atomic: An atlas of machine commonsense for if-then reasoning. In *Proc. of AAAI*.

Maarten Sap, Hannah Rashkin, Derek Chen, Ronan Le Bras, and Yejin Choi. 2019b. Socialiqa: Commonsense reasoning about social interactions. *CoRR*, abs/1904.09728.

Linfeng Song, Yue Zhang, Zhiguo Wang, and Daniel Gildea. 2018. A graph-to-sequence model for amr-to-text generation. In *ACL*.

Alon Talmor, Jonathan Herzig, Nicholas Lourie, and Jonathan Berant. 2019. Commonsenseqa: A question answering challenge targeting commonsense knowledge. In *Proc. of NAACL-HLT*.

Trieu H. Trinh and Quoc V. Le. 2018. A simple method for commonsense reasoning. *CoRR*, abs/1806.02847.

Rowan Zellers, Yonatan Bisk, Ali Farhadi, and Yejin Choi. 2019. From recognition to cognition: Visual commonsense reasoning. In *Proc. of CVPR*.

Rowan Zellers, Yonatan Bisk, Roy Schwartz, and Yejin Choi. 2018. Swag: A large-scale adversarial dataset for grounded commonsense inference. In *Proc. of EMNLP*.

Hao Zhou, Tom Young, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. 2018. Commonsense knowledge aware conversation generation with graph attention. In *IJCAI*.

Yaoming Zhu, Juncheng Wan, Zhiming Zhou, Liheng Chen, Lin Qiu, Weinan Zhang, Xin Jiang, and Yong Yu. 2019. Triple-to-text: Converting rdf triples into high-quality natural languages via optimizing an inverse kl divergence. In *SIGIR*.